Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 12 , December - 2025

Home Price Prediction using ML

Arpit Bhandare, Ishita Ramteke, Shruti Atilawar , Shreyash Wankhade, Saniya Khobragade
Students, Department of Computer Science and Engineering
Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, Maharashtra, India

Prof. Vrushali Awale
Assistant Professor, Department of Computer Science Engineering
Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, Maharashtra, India

***Abstract* - Accurate prediction of house prices is essential for buyers, sellers, real estate developers, and financial institutions. With the rapid growth of urbanization and availability of large-scale housing data, machine learning (ML) techniques provide effective solutions for estimating property prices based on multiple influencing factors. This paper presents a machine learning–based home price prediction system that analyzes historical housing data and predicts prices using regression algorithms.**

The proposed system involves data collection, preprocessing, feature engineering, model training, and evaluation. Experimental results show that machine learning models can predict house prices with good accuracy and can support data-driven decision-making in the real estate sector.

Keywords - Home price prediction, machine learning, regression, real estate, data preprocessing.

INTRODUCTION

The real estate sector is one of the most important contributors to economic growth and urban development. Accurate estimation of house prices is crucial for buyers, sellers, investors, and financial institutions to make informed decisions. However, house price prediction is a complex task because property values depend on multiple factors such as location, size, number of rooms, availability of amenities, and prevailing market conditions.

Traditional methods of house price estimation rely on manual evaluation, expert opinions, and historical trends. These approaches are often time-consuming, subjective, and may not scale well with large datasets. With the rapid increase in the availability of digital housing data, machine learning techniques have emerged as a powerful alternative for analyzing complex patterns and predicting prices more accurately.

Machine learning models can learn relationships between input features and house prices from historical data and generate predictions for new properties. By applying supervised learning algorithms, it is possible to automate the prediction process and reduce human bias. This mini project focuses on developing a home price prediction system using machine learning techniques. The system includes data preprocessing, feature engineering, model training, and evaluation to predict house prices effectively. The proposed approach aims to provide a reliable and efficient solution for real-world real estate price estimation.

RELATED WORK

The real estate sector is one of the most important contributors to economic growth and urban development. Accurate estimation of house prices is crucial for buyers, sellers, investors, and financial institutions to make informed decisions. However, house price prediction is a complex task because property values depend on multiple factors such as location, size, number of rooms, availability of amenities, and prevailing market conditions.

Traditional methods of house price estimation rely on manual evaluation, expert opinions, and historical trends. These approaches are often time-consuming, subjective, and may not scale well with large datasets. With the rapid increase in the availability of digital housing data, machine learning techniques have emerged as a powerful alternative for analyzing complex patterns and predicting prices more accurately.

Machine learning models can learn relationships between input features and house prices from historical data and generate predictions for new properties. By applying supervised learning algorithms, it is possible to automate the prediction process and reduce human bias. This mini project focuses on developing a home price prediction system using machine learning techniques. The system includes data preprocessing, feature engineering, model training, and evaluation to predict house prices effectively. The proposed approach aims to provide a reliable and efficient solution for real-world real estate price estimation.

PROPOSED METHODOLOGY

The proposed methodology for the Home Price Prediction system follows a systematic and structured machine learning pipeline. Each stage plays a crucial role in building an accurate and reliable prediction model. The complete workflow is designed to transform raw housing data into meaningful insights and generate precise house price predictions.

1. Data Collection

The first step involves collecting housing data from publicly available datasets such as Kaggle or other open real estate data sources. The dataset typically includes attributes like location, area (in square feet), number of bedrooms (BHK), number of bathrooms, and price. These attributes serve as input features for the machine learning model.

2. Data Preprocessing

Raw housing data often contains missing values, inconsistencies, and noise. Therefore, data preprocessing is performed to improve data quality. Missing values are handled using techniques such as mean or median imputation. Duplicate records are removed to avoid bias. Outliers that can negatively affect model performance are detected and eliminated. Categorical features such as location are converted into numerical form using encoding techniques like one-hot encoding. Numerical features are normalized or scaled to ensure uniformity.

3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis is carried out to understand the underlying patterns and relationships within the dataset. Statistical summaries and visualizations are used to analyze feature distributions, correlations between variables, and price trends. EDA helps in identifying important features that significantly influence house prices.

4. Feature Engineering and Selection

Feature engineering involves creating new meaningful features or transforming existing ones to enhance model performance. Irrelevant or highly correlated features are removed to reduce dimensionality and prevent overfitting. Feature selection techniques ensure that only the most influential attributes are used for training the model.

5. Dataset Splitting

After preprocessing and feature selection, the dataset is divided into training and testing sets. Typically, 70–80% of the data is used for training the model, while the remaining 20–30% is reserved for testing and validation. This helps in evaluating the model’s generalization capability.

6. Model Selection and Training

Supervised machine learning regression algorithms are used to train the model. Algorithms such as Linear Regression, Decision Tree Regression, and Random Forest Regression are applied to the training dataset. The models learn the relationship between input features and house prices. Hyperparameter tuning is performed to optimize model performance.

7. Model Evaluation

The trained models are evaluated using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE). These metrics measure the difference between actual and predicted house prices. The model with the lowest error is selected as the final prediction model.

8. Prediction and Output Generation

Once the best-performing model is selected, it is used to predict house prices for new input data provided by users. The system outputs the estimated house price based on the given features. This prediction can be displayed through a user interface or a web application.

9. Deployment (Optional)

For real-world usability, the trained model can be deployed using a web framework such as Flask. This allows users to input property details and obtain real-time price predictions through a web-based interface.

EXPERIMENTAL RESULTS AND DISCUSSION

The experimental evaluation of the proposed Home Price Prediction system was conducted using a real-world housing dataset. The dataset was divided into training and testing subsets to analyze the prediction capability of the machine learning models. The experiments were implemented using Python and standard machine learning libraries.

The dataset contained important housing attributes such as area, number of bedrooms, number of bathrooms, and location. After preprocessing and feature selection, 80% of the data was used for training the models, while the remaining 20% was used for testing. Multiple regression models were trained and evaluated to compare their performance.

To measure the accuracy of the models, standard regression performance metrics were used. Mean Absolute Error (MAE) was used to calculate the average absolute difference between the actual and predicted house prices. Mean Squared Error (MSE) measured the average squared error, while Root Mean Square Error (RMSE) provided an overall measure of prediction accuracy by penalizing larger errors.

The experimental results indicate that Linear Regression provided a reasonable baseline performance but showed higher error values due to its assumption of a linear relationship between features and house prices. Decision Tree Regression improved prediction accuracy by capturing non-linear patterns in the data. Among all the evaluated models, Random Forest Regression achieved the best performance with the lowest error values.

The improved accuracy of the Random Forest model can be attributed to its ensemble learning approach, which combines multiple decision trees and reduces overfitting. The results demonstrate that proper data preprocessing, feature engineering, and the selection of suitable machine learning algorithms significantly enhance prediction performance.

Overall, the experimental findings validate the effectiveness of the proposed system and confirm that machine learning–based approaches can be successfully applied for accurate and reliable home price prediction in real-world scenarios.

PRACTICAL CONSIDERATIONS AND APPLICATIONS

Practical Considerations

1. **Data Quality and Availability**
   The accuracy of home price prediction largely depends on the quality of the dataset. Missing values, outdated records, or incorrect entries can significantly affect model performance, making proper data cleaning essential.

2. **Location dependency**
   House prices vary greatly across different regions. A model trained on one city or locality may not generalize well to another, requiring location-specific training or frequent model updates.

3. **Feature Selection**
   Choosing relevant features such as area, number of rooms, and location is critical. Irrelevant or redundant features can reduce prediction accuracy and increase computational complexity.

4. **Model Selection and Complexity**
   Simple models like Linear Regression are easy to interpret but may underperform on complex datasets, while advanced models such as Random Forest offer better accuracy at the cost of higher computational resources.

5. **Scalability and Performance**
   In real-world applications, the system should efficiently handle large datasets and multiple prediction requests without performance degradation.

6. **Model Maintenance**
   Real estate markets change over time. Periodic retraining of the model with updated data is necessary to maintain accurate and reliable predictions.

PRACTICAL APPLICATIONS

1. **Real Estate Price Estimation**
   The system helps buyers and sellers estimate fair market prices of properties based on current data, reducing dependency on manual valuation methods.
2. **Decision Support for Property Investment**
   Investors can use the predicted prices to analyze profitability and make informed decisions regarding property purchases or sales.
3. **Banking and Loan Approval Systems**
   Financial institutions can integrate the model to assess property value during home loan approval, helping in risk assessment and credit evaluation.
4. **Real Estate Agency Automation**
   Real estate agencies can automate property valuation processes, improve operational efficiency, and provide quick price estimates to clients.
5. **Urban Planning and Policy Making**
   Government authorities and urban planners can use housing price predictions to support infrastructure planning, taxation, and smart city development.
6. **Web and Mobile Applications**
   The model can be deployed in web or mobile applications, allowing users to input property details and obtain instant price prediction

CONCLUSION AND FUTURE WORK

CONCLUSION

This research paper presented a machine learning–based approach for predicting home prices using historical housing data and regression techniques. The proposed system focused on analyzing key property features such as area, number of bedrooms, number of bathrooms, and location to estimate house prices accurately. A systematic methodology involving data collection, preprocessing, feature engineering, model training, and evaluation was implemented to ensure reliable predictions.

Experimental results demonstrated that machine learning models are effective in capturing complex relationships between housing attributes and property prices. Comparative analysis showed that advanced regression models, particularly ensemble-based techniques, achieved better prediction accuracy compared to traditional linear models. The use of appropriate performance metrics further validated the effectiveness of the proposed approach.

The study also highlighted important practical considerations, including data quality, feature selection, and model generalization. By addressing these factors, the proposed system can be applied effectively in real-world scenarios. The developed model has potential applications in real estate valuation, investment analysis, banking systems, and urban planning.

In conclusion, the proposed home price prediction system demonstrates that machine learning provides a reliable and scalable solution for real estate price estimation. With further enhancements such as the integration of real-time market data, advanced learning models, and deployment as a web-based application, the system can be extended to support intelligent and data-driven decision-making in the real estate domain.

FUTURE WORK

1. **Integration of Real-Time Data**
   Future versions of the system can incorporate real-time housing market data to improve prediction accuracy and reflect current market trends.

2. **Use of Advanced Machine Learning Models**
   Deep learning techniques such as Artificial Neural Networks (ANNs) and Gradient Boosting models can be explored to capture more complex relationships in housing data.

3. **Location-Specific Models**
   Developing region-wise or city-specific models can enhance prediction performance by accounting for local market variations.

4. **Feature Expansion**
   Additional features such as proximity to schools, hospitals, transportation facilities, and neighborhood quality can be included to improve predictions.

5. **Web and Mobile Application Deployment**
   The system can be deployed as a full-scale web or mobile application to provide real-time predictions to end users.

6. **Explainable AI Integration**
   Incorporating explainable AI techniques can help users understand how different features influence house price predictions, increasing transparency and trust.

REFERENCES

1. Research Papers and Articles

a. Y. K. Jain and S. Gupta, “House Price Prediction Using Machine Learning Techniques,” *International Journal of Computer Applications*, vol. 182, no. 25, pp. 1–6, 2020.

b. A. A. Oluwaseun, M. A. Adeyemi, and R. O. Ibrahim, “Real Estate Price Prediction Using Regression and Machine Learning Algorithms,” *International Journal of Advanced Research in Computer Science*, vol. 11, no. 4, pp. 45–50, 2020**.**

2. Organizations and Websites

1. Kaggle, “House Price Prediction Datasets,” Available: https://www.kaggle.com
2. Scikit-learn Developers, “Scikit-learn: Machine Learning in Python,” Available: https://scikit-learn.org
3. Python Software Foundation, “Python Programming Language,” Available: https://www.python.org
4. UCI Machine Learning Repository, “Housing Datasets,” University of California, Irvine. Available: https://archive.ics.uci.edu

TOOLS AND MODELS

1. **Python Programming Language**
   Python was used as the primary programming language due to its simplicity, flexibility, and extensive support for machine learning and data analysis.

2. **Scikit-learn Library**
   Scikit-learn was used for implementing machine learning algorithms such as Linear Regression, Decision Tree Regression, and Random Forest Regression, along with model evaluation metrics.

3. **Pandas and NumPy**
   Pandas was used for data handling and preprocessing, while NumPy supported numerical computations and array operations.

4. **Matplotlib and Seaborn**
   These libraries were used for data visualization and exploratory data analysis to understand patterns and relationships within the housing dataset.

5. **Machine Learning Models**
   Supervised regression models including Linear Regression, Decision Tree Regression, and Random Forest Regression were employed to predict home prices based on historical data.

6. **Jupyter Notebook / IDE Environment**
   Jupyter Notebook and Python IDEs were used for model development, experimentation, and result analysis.