KeAi
CHINESE ROOTS
GLOBAL IMPACT

International Journal of Innovation Studies

THE HOUSE PRICE PREDICTION SYSTEM USING AI FOR SUSTAINABLE URBAN DEVELOPMENT

Adithyan A
III BCA, Department of Computer Applications, Nehru Arts and Science College, Coimbatore, India
aadhi9685@gmail.com

Aswin R
III BCA, Department of Computer Applications, Nehru Arts and Science College, Coimbatore, India
aswinrameshm@gmail.com

Adithya M
III BCA, Department of Computer Applications, Nehru Arts and Science College, Coimbatore, India
madithya033@gmail.com

Dr. A. Kalaivani
Associate Professor, Department of Computer Applications, Nehru Arts and Science College, Coimbatore, India, nasckalaivani@nehrucolleges.com

Abstract
The rapid growth of urban populations and real estate markets has increased the demand for accurate housing price prediction systems. Traditional valuation methods are often inefficient and fail to capture complex relationships between property characteristics and market dynamics. This research presents an Artificial Intelligence (AI)–based house price prediction system that utilizes machine learning algorithms to analyze housing datasets and estimate property values accurately. The proposed system integrates predictive analytics with sustainability indicators such as environmental quality, proximity to public transportation, and green infrastructure to support sustainable urban development. Multiple machine learning algorithms including Linear Regression, Decision Tree, Random Forest, and Gradient Boosting are evaluated to identify the most effective predictive model. Experimental results demonstrate that ensemble learning methods outperform traditional models in terms of accuracy and reliability. The proposed system enhances transparency in real estate markets and contributes to data-driven decision-making in smart cities and sustainable housing planning.

Keywords— Artificial Intelligence, Machine Learning, House Price Prediction, Sustainable Development, Smart Cities, Real Estate Analytics, Urban Planning, Data Mining.

I. Introduction

The House Price Prediction System is a web-based application developed to estimate residential property prices using advanced Machine Learning techniques. The system is designed to provide accurate and intelligent property valuations based on real estate data analysis. It aims to simplify the complex process of price estimation by replacing traditional guesswork with data-driven predictions. By leveraging historical housing data, the system generates reliable estimates that assist users in making informed decisions. The application focuses on improving transparency and accuracy in property valuation. It

serves as a practical solution for modern real estate challenges. The system integrates data science concepts with user-friendly web technology.

In recent years, the Indian real estate market has witnessed rapid growth and continuous transformation. Urbanization, infrastructure development, and increasing population have significantly influenced property demand and pricing patterns. As cities expand and new residential projects emerge, property valuation has become more complex. Buyers and sellers often struggle to determine fair market prices due to fluctuating conditions. Accurate price estimation is therefore essential to avoid financial risk and uncertainty. A reliable prediction system can help stabilize pricing expectations. This highlights the importance of adopting intelligent technological solutions in the real estate sector.

Traditional property valuation methods largely depend on manual assessment, broker expertise, or comparisons with nearby properties. These approaches may not consider multiple influencing factors simultaneously. As a result, pricing decisions can sometimes be inconsistent or biased. Human error and subjective judgment may also reduce reliability. Additionally, static comparisons fail to capture dynamic market trends. With changing economic conditions, property prices can vary significantly over time. Therefore, a more systematic and analytical approach is required for better accuracy.

The proposed system introduces a Machine Learning-based solution to overcome these limitations. It analyzes historical datasets to identify patterns and relationships between property features and market prices. Important factors such as location, area in square feet, number of bedrooms, bathrooms, and amenities are considered in the prediction process. Location plays a crucial role, as connectivity, infrastructure, and nearby facilities directly impact property value. The system applies feature engineering techniques to preprocess and refine input data. Advanced regression algorithms are then used to generate precise price predictions.

The application is developed using Python and supported by powerful libraries such as Pandas, NumPy, Scikit-learn, and XGBoost. These tools enable efficient data processing, model training, and performance evaluation. The user interface is built using Streamlit to provide a simple and interactive experience. Users can enter property details and instantly receive predicted prices along with useful insights. The system also incorporates visualization techniques to display market trends clearly. This enhances user understanding and confidence in the results. Overall, the House Price Prediction System offers a modern, accurate, and user-friendly solution for real estate valuation and decision-making.

II. LITERATURE REVIEW

House price prediction has been widely studied using statistical models and machine learning techniques. Several researchers have developed predictive systems to improve the accuracy of real estate valuation and support decision-making in housing markets. This section reviews five important studies related to AI-based house price prediction.

1. Hedonic Pricing Model for Housing Valuation

One of the earliest approaches to housing price prediction was proposed by **David Harrison** and **Daniel L. Rubinfeld**, who developed the Hedonic Pricing Model. Their study analyzed housing prices based on individual property characteristics such as location, size, and neighborhood environment. The research demonstrated that environmental factors significantly influence property values. Although this

approach provided a theoretical framework for property valuation, it was limited by its inability to capture complex nonlinear relationships among variables.

2. Random Forest for Housing Price Prediction

The Random Forest algorithm introduced by **Leo Breiman** has been widely applied in housing price prediction systems. Random Forest is an ensemble learning technique that combines multiple decision trees to improve prediction accuracy and reduce overfitting. Studies have shown that Random Forest performs well when dealing with large datasets and complex relationships between housing features. This method has become one of the most reliable models for real estate price prediction.

3. Gradient Boosting and XGBoost Models

Boosting algorithms have also been widely used in predictive analytics. **Tianqi Chen** and **Carlos Guestrin** introduced the XGBoost algorithm, which is a scalable and efficient gradient boosting framework. Their research demonstrated that boosting models significantly improve predictive accuracy by sequentially correcting the errors of previous models. XGBoost has been successfully used in housing price prediction due to its ability to handle large datasets and complex feature interactions.

4. Artificial Neural Networks for Real Estate Prediction

Deep learning models have also been applied to real estate forecasting. Research by **Ian Goodfellow**, **Yoshua Bengio**, and **Aaron Courville** demonstrated the effectiveness of neural networks in capturing nonlinear relationships within large datasets. Artificial Neural Networks (ANNs) have shown promising results in predicting house prices by automatically learning complex patterns from housing data. However, deep learning models require large datasets and higher computational resources.

5. AI-Based Smart City Housing Analytics

Recent studies focus on integrating artificial intelligence with smart city infrastructure to improve housing market analysis. Research related to the **United Nations** Sustainable Development Goals highlights the importance of data-driven urban planning. AI-based housing prediction systems can support sustainable development by analyzing environmental factors such as access to public transportation, green spaces, and energy-efficient buildings. These systems help policymakers design sustainable housing policies and improve urban infrastructure planning.

III. PROJECT STATEMENT

The rapid growth of the real estate market has made property price estimation increasingly complex and uncertain. Traditional valuation methods mainly depend on manual assessment, broker experience, and comparison with nearby properties, which may not accurately reflect dynamic market conditions. These approaches often fail to consider multiple influencing factors such as location, area, amenities, and changing market trends simultaneously. As a result, property pricing may become inconsistent, biased, or unreliable, creating confusion for buyers, sellers, and investors.

To address these challenges, the House Price Prediction System aims to develop a machine learning-based web application that provides accurate and data-driven property price estimations. The system analyzes historical real estate data and key property features to predict market value efficiently. By automating the valuation process and incorporating advanced regression algorithms, the proposed solution reduces human error and enhances pricing transparency. The objective is to create a reliable, user-friendly platform that supports informed decision-making in the real estate sector.

IV. METHODOLOGY

The methodology of the House Price Prediction System begins with data collection from reliable real estate sources. Historical property data containing features such as location, area (square feet), number of bedrooms (BHK), bathrooms, amenities, and price is gathered. The collected dataset is carefull

examined to understand its structure, distribution, and relevance to the prediction objective. Proper data selection ensures that the model is trained on meaningful and high-quality information

The next step involves data preprocessing and cleaning. Missing values, duplicate records, and inconsistent entries are identified and handled appropriately. Outliers that may negatively affect model performance are detected and treated. Categorical variables such as location are encoded into numerical form, and numerical features are scaled if necessary. This step ensures that the dataset is structured and suitable for machine learning algorithms.

Feature engineering is then applied to enhance model performance. Important attributes influencing property prices are selected based on correlation analysis and domain knowledge. New features, such as location-based scoring or price per square foot, may be derived to improve predictive capability.

Irrelevant or redundant features are removed to reduce noise and improve efficiency. This process helps the model better understand relationships between variables.

After preprocessing and feature selection, the dataset is divided into training and testing sets. The training set is used to train regression models such as Linear Regression, Random Forest, and XGBoost. These algorithms learn patterns and relationships between input features and property prices. Model parameters are tuned using techniques such as cross-validation to improve prediction accuracy and prevent overfitting.

Model evaluation is performed using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) score. These metrics help assess how well the model predicts property prices compared to actual values. The best-performing model is selected based on accuracy and generalization capability. This ensures reliable predictions for unseen data.

Finally, the trained model is integrated into a web application developed using Streamlit. A user- friendly interface allows users to input property details and instantly receive predicted prices. The system processes inputs through the trained model and displays results along with visual insights. This deployment step makes the prediction system accessible, interactive, and practical for real-world real estate decision-making.

V. General Process of the Proposed System

The general process of the proposed House Price Prediction System begins with user interaction through a web-based interface. Users enter essential property details such as location, area in square feet, number of bedrooms (BHK), number of bathrooms, and available amenities. The system is designed to provide a simple and intuitive form so that users can easily input the required information without technical knowledge. Once the details are submitted, the data is prepared for processing.

In the second stage, the system performs input validation and preprocessing. The entered values are checked for correctness and completeness to ensure reliable predictions. Categorical inputs such as location are converted into numerical representations using encoding techniques. Numerical inputs are scaled or transformed if necessary. This preprocessing ensures that the data format matches the structure used during model training.

The third stage involves feature engineering and data transformation. The processed inputs are passed through a predefined feature pipeline that selects and organizes relevant attributes. Additional derived features, such as location-based scores or price per square foot indicators, may be generated to enhance prediction accuracy. This step ensures that the model receives optimized and meaningful input features.

In the fourth stage, the trained machine learning model analyzes the processed features. Regression algorithms such as XGBoost or other optimized models evaluate the relationships between input parameters and property prices. Based on learned patterns from historical data, the model generates an estimated price for the given property details. The system ensures that predictions are computed quickly and efficiently.

Finally, the predicted property price is displayed to the user through the web interface. The system may also present additional insights such as confidence levels, comparative pricing, and visual trend charts. This helps users better understand the estimated value and market positioning of the property. Overall, the proposed system provides a smooth, accurate, and user-friendly process for real estate price estimation.

Methodology

The proposed House Price Prediction System using Artificial Intelligence follows several steps such as data collection, preprocessing, feature engineering, model training, prediction, and deployment. These steps help the system analyze housing data and predict property prices accurately.

1. Data Collection

The first step is collecting housing data from reliable sources such as online real estate websites, real estate agencies, and government property records. The dataset includes important information about properties. Common features include location, area in square feet, number of bedrooms (BHK), number of bathrooms, available amenities, and the property price. A larger and more diverse dataset helps improve the accuracy of the prediction model.

2. Data Preprocessing

The collected data may contain missing values, duplicate records, or incorrect information. Data preprocessing is used to clean the dataset. Missing values are filled or removed, duplicate records are deleted, and outliers are identified. Categorical data such as location is converted into numerical values so that machine learning algorithms can process it. Numerical features are also scaled to maintain consistency in the data.

3. Feature Engineering

Feature engineering helps improve the prediction model by creating new useful features. For example, the system can calculate price per square foot or assign a score to locations based on their popularity. Important features are selected based on their relationship with the property price. This helps the model focus on the most relevant information.

4. Model Selection and Training

In this step, machine learning models are selected and trained using the prepared dataset. Common regression models used for house price prediction include Linear Regression, Random Forest, and XGBoost. The dataset is divided into training and testing sets so that the model can learn from past data and then be tested on new data. Cross-validation techniques are also used to improve the model performance and prevent overfitting.

5. User Input Interface

The system provides a user interface where users can enter property details such as location, area, number of bedrooms, number of bathrooms, and amenities. The system checks and processes this input data so that it matches the format used by the trained model.

6. Prediction

After receiving the input details, the trained machine learning model predicts the estimated price of the property. The prediction is based on patterns learned from historical housing data. The system may also provide a confidence score to show how reliable the prediction is.

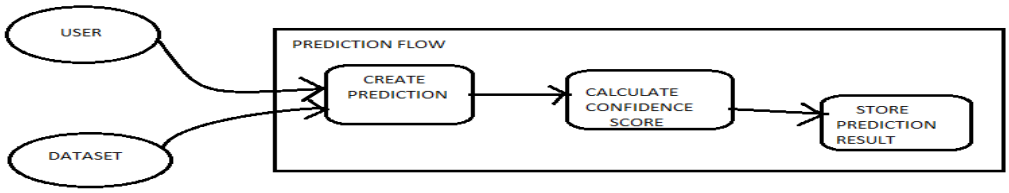
7. Visualization and Insights

The predicted results are displayed using graphs and charts to make them easier to understand. The system may show price comparisons, market trends, and factors affecting property prices. These insights help users make better real estate decisions.

8. Deployment

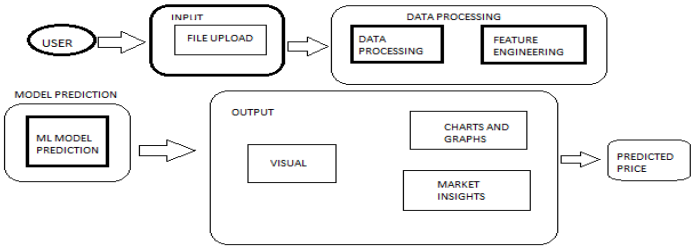
Finally, the system is deployed as a web application using tools such as Streamlit. This allows users to access the system through a web browser. Users can enter property details, view predicted prices, and analyze market trends easily.

VI.Data Flow Diagram (DFD)



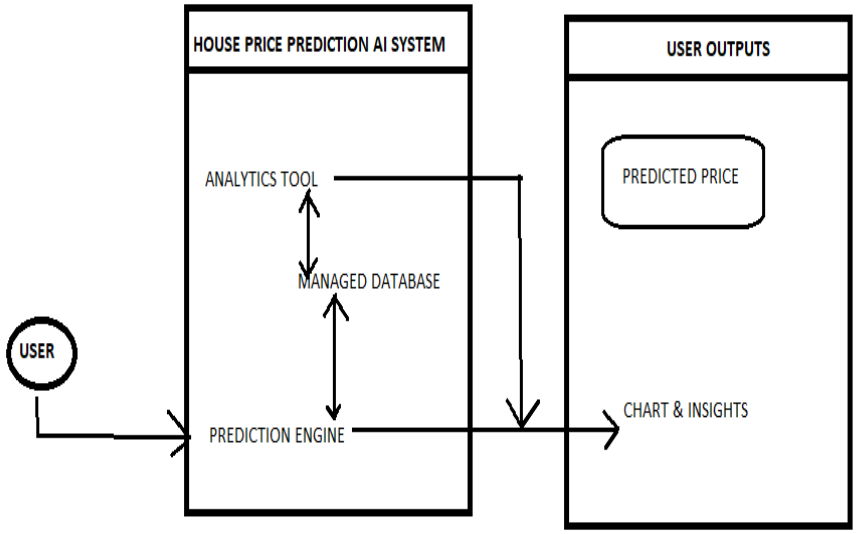
DATA FLOW DIAGRAM

Level 0 DFD (Context Diagram)



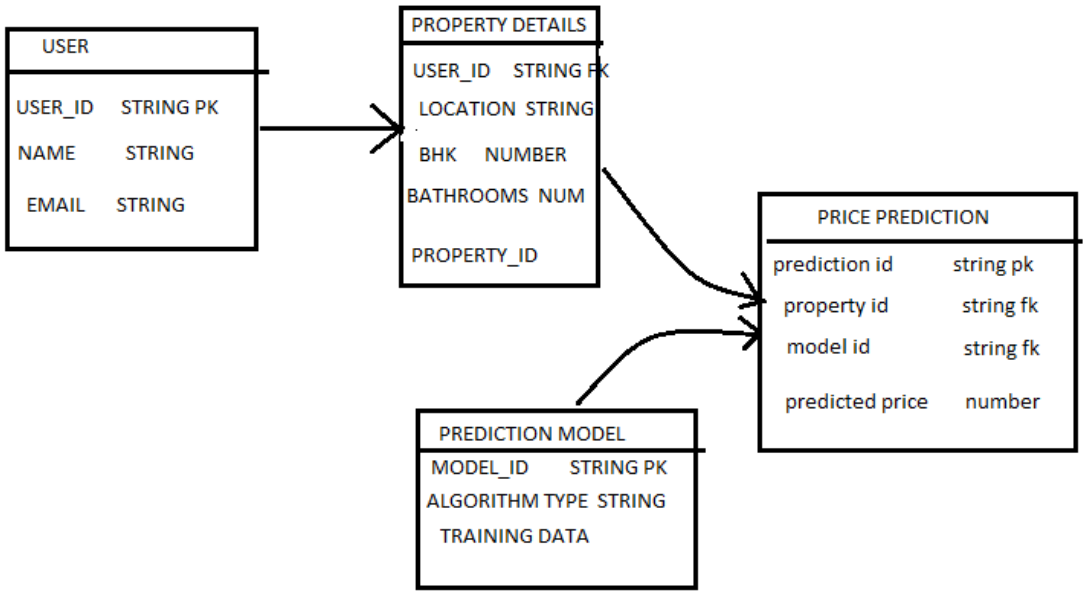
LEVEL 0 DATA FLOW DIAGRAM

Level 1 DFD



LEVEL 1 DATA FLOW DIAGRAM

Prediction model



UML DIAGRAM

VII. Sustainable Development Impact

The proposed system contributes to sustainable development in several ways:

1. Supports **smart city planning**

2. Encourages **efficient land use**
3. Promotes **green infrastructure development**
4. Improves **housing affordability analysis**
5. Enables **data-driven urban policy decisions**

Housing prediction systems can guide policymakers in building **sustainable and resilient cities**.

IIX. Conclusion

The House Price Prediction System provides an intelligent, data-driven solution for estimating residential property prices with high accuracy and reliability. By leveraging historical real estate data and machine learning algorithms such as XGBoost, Random Forest, and Linear Regression, the system effectively captures complex relationships between property features and market values. The integration of location, area, number of bedrooms and bathrooms, amenities, and market trends ensures comprehensive analysis, reducing the dependency on traditional manual valuation methods.

Thorough testing—including unit testing, input validation, model accuracy testing, integration testing, and system testing—ensures that the system is robust, reliable, and user-friendly. The deployment through a Streamlit web interface allows users to interact with the system easily, enter property details, and receive instant price predictions along with visual insights. Overall, the House Price Prediction System serves as a modern, efficient, and practical tool for buyers, sellers, and investors, enabling informed decision-making in the real estate market and setting a foundation for future enhancements such as real-time data integration and predictive analytics.

IX. Future Work

Future improvements include:

- Integration with GIS-based mapping systems
- Real-time housing market prediction
- Deep learning models such as Neural Networks
- Mobile and web-based applications
- Integration of environmental sustainability metrics

References

[1] D. Harrison and D. L. Rubinfeld, “Hedonic housing prices and the demand for clean air,” *Journal of Environmental Economics and Management*, vol. 5, no. 1, pp. 81–102, 1978.

[2] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[3] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, 2016, pp. 785–794.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[5] S. Mullainathan and J. Spiess, “Machine learning: An applied econometric approach,” *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 87–106, 2017.

[6] J. B. Duarte, F. Lourenço, and P. M. Ferreira, “Machine learning approaches for real estate price prediction,” *Expert Systems with Applications*, vol. 42, no. 21, pp. 7539–7547, 2015.
[7] United Nations, “Transforming our world: The 2030 agenda for sustainable development,” United Nations, New York, NY, USA, 2015.
[8] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA, USA: O’Reilly Media, 2019.