International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 03, ICCT - 2026

AI-POWERED REAL ESTATE MARKETPLACE: INTELLIGENT PRICE PREDICTION AND SMART BUYER–SELLER COMMUNICATION SYSTEM

J. PRINCESS BALA
Assistant Professor,
Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering,
Nazareth, India.

S. Vanaja
PG Scholar
Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering,
Nazareth, India.

***Abstract* - The real estate sector represents a significant contributor to national economic stability, with property valuation acting as a key determinant in financial and investment decisions. Traditional valuation approaches often depend upon human judgment, which can be subjective and inconsistent. This study introduces a data-driven framework for predicting property prices through the application of supervised machine-learning techniques. The methodology involves rigorous data preparation, including cleaning, feature engineering, encoding of categorical variables, and normalization, to ensure robust model performance. Several regression algorithms—Linear Regression, Decision Tree Regressor, Random Forest Regressor, and XGBoost—are examined and compared using evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Coefficient of Determination (R2). The most accurate model is deployed for final prediction. Experimental outcomes reveal that machine-learning approaches can successfully model the intricate relationships between property attributes and market prices, producing predictions that are more consistent and precise than traditional valuation techniques. The proposed solution demonstrates potential for integration within digital property platforms to support evidence-based decision-making and enhance transparency and efficiency in the housing market.**

Keywords* - *Artificial Intelligence, Machine Learning, Real Estate, Price Prediction, Deep Learning, Chatbot, Marketplace, Data Analytics

I. INTRODUCTION

Real estate plays an immensely influential role in shaping economies by impacting personal financial success across nations as well as fostering overall societal progress. Properties' values fluctuate due to various elements including geographical position, transportation networks, proximity to community services, prevailing socio-economic conditions, current business cycles, and consumer preferences for real estate transactions. Historically, property valuations have predominantly depended upon skilled professionals' judgment and physical inspection techniques, frequently requiring extensive effort, being influenced by personal perspectives, and susceptible to variability in results. With an expanding amount of detailed property information, there's rising demand for systems capable of making precise pricing forecasts through algorithmic analysis alone. Over the past few decades, ML has become an indispensable technique in forecasting activities throughout diverse sectors such as economics, medicine, and property management. Machine learning models can discern intricate connections among numerous data points, enabling highly accurate value predictions by considering various impacting variables. Utilizing archived real estate information enables advanced algorithms to analyze patterns in transactions and forecast future developments which aid more informed choices by purchasers, vendors, financiers, and regulatory bodies. The objective of this endeavor is to create and assess an artificial intelligence algorithm designed specifically for forecasting property values accurately. Key goals encompass: Gathering and preparing comprehensive sets of data related to residential properties featuring diverse characteristics. Employing feature selection methods for pinpointing key predictors. Utilizing various regression techniques alongside ensemble methods for forecasting real estate values. Analyzing model efficacy through statistical measures like MAE, MSE, and R2 index comparison. A new mechanism aims at boosting clarity and speed within property markets through precise, evidence-based valuation forecasts. This framework could seamlessly integrate within digital real estate networks or advisory systems for helping individuals assess home appraisals, refine financial planning approaches, and comprehend economic trends effectively. In conclusion, this research advances our understanding by integrating conventional appraisal methods with advanced machine learning algorithms for predicting the house price.

II. LITERATURE REVIEW

Decker D. C. (2011) - "Alternative housing data for Ames, Iowa". Dr. De Cock created the Ames housing data set, an extensive real-estate database extensively utilized in predictive analytics. A new research project substituted an old Boston data set for its replacement, offering insights through 79 factors related to domestic dwellings. Its status as an evaluation standard was recognized in assessing regression and machine learning algorithms.

Scientists employed this information set in order to assess models such as Linear Regression, Decisions Trees, and Random Forests concerning their ability to forecast house prices. Park, S. , and Bae, J. In 2015, researchers developed an artificial neural network model for predicting house prices. Park and Bae utilized artificial neural networks (ANNs) for forecasting house values using past information. The results indicated that artificial neural network (ANN) models excel over conventional regression methods in identifying intricate connections among input variables and real estate values

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 03, ICCT - 2026

due to their ability to handle non-linear patterns. Research highlighted the viability of employing deep learning techniques within property analysis, particularly advantageous when dealing with extensive data sets.

Aydin, S. et al. In 2019, researchers focused on developing models to predict real estate prices by analyzing spatial and temporal data patterns. Aydin and his team integrated geographical information and chronological elements within their pricing forecast algorithms. The study revealed that incorporating geographical locations along with temporal economic indicators enhanced predictive model precision. Research indicates that housing values aren't solely determined by physical attributes; they're affected significantly by geographical closeness and market trends as well. This underscores the relevance of utilizing data-driven methodologies tailored specifically for geographic analysis in predicting these fluctuations effectively.

III. METHODOLOGY

The problem of automated detection of coronary artery lesions from invasive Coronary Angiography (ICA) images is addressed by this work. The purpose of the proposed method is to develop a reliable and easily implementable deep learning framework that can accurately categorize lesion and non-lesion regions while also delivering clinically interpretable results. Preprocessing, feature extraction, classification, and visualization stages are part of the patch-based CNN approach that the system follows.

A. Existing System

Existing systems for real-estate valuation primarily depend on manual or rule-based assessments. In such frameworks, values inspect properties and reference recent sales to 5 determine market value. While professional judgment is valuable, these systems are inconsistent due to variations in experience, access to data, and subjective interpretation. Recent research efforts have explored statistical models and basic ML techniques. The Hedonic Pricing Model and regression analysis provide insight into how different features affect price but cannot capture complex, non-linear interactions. Datasets such as Ames Housing have enabled evaluation of models like Decision Trees and Random Forests, offering improved accuracy but limited scalability. Moreover, most available systems lack integration with user-friendly platforms that could allow buyers and sellers to interact directly based on model-predicted prices. Thus, there remains a need for a unified, automated solution that leverages robust ML algorithms while providing an accessible interface for stakeholders.

B. Proposed Method

The proposed system adopts a structured machine learning pipeline to predict real estate prices accurately. Initially, a comprehensive dataset is collected containing property attributes such as location, size, number of rooms, amenities, and pricing history. The data is then preprocessed by handling missing values, removing outliers, encoding categorical variables, and applying feature scaling techniques. Feature engineering is performed to extract meaningful insights, including location-based importance and derived attributes. The processed data is split into training and testing sets, and multiple machine learning models such as Linear Regression, Random Forest, and XGBoost are trained to learn patterns and relationships between features and property prices. The best-performing model is selected based on evaluation metrics like RMSE and R^2 score.

In addition to price prediction, the system integrates a smart buyer–seller communication module powered by an AI chatbot. This module enables users to interact in real time, ask property-related queries, and receive instant automated responses, thereby improving user engagement and reducing dependency on manual agents. The entire system is deployed through a user-friendly web interface where users can input property details and receive predicted prices along with recommendations. Continuous learning is incorporated by updating the model with new market data to maintain accuracy over time. This integrated approach ensures a scalable, efficient, and intelligent real estate marketplace.

1) Data collection

This module is responsible for gathering historical real estate data from various sources such as online property listings, public datasets, or real estate APIs. The data typically includes features like location, property size, number of bedrooms and bathrooms, age of the property, proximity to facilities, and price. Accurate and comprehensive data collection is crucial as it directly affects the performance of the prediction model.

2) Data Preprocessing

The collected raw data may contain missing values, duplicates, or inconsistencies. This module handles data cleaning by imputing missing values, removing duplicates, and correcting inconsistencies. It also involves encoding categorical features (like location, property type) into numerical format, normalizing numerical features, and scaling the data to prepare it for machine learning models.

3) Future Engineering & Selection

This module identifies the most relevant features that impact property prices. It may create new features such as price per square foot or proximity to key landmarks. Feature selection techniques like correlation analysis or feature importance from models are applied to remove irrelevant or redundant features, improving model efficiency and accuracy.

4) Machine Learning Model

This module implements multiple machine learning algorithms such as Linear Regression,

Decision Trees, Random Forest, and Gradient Boosting to train on the preprocessed data. Each model learns patterns and relationships between the property

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 03, ICCT - 2026

features and their prices. Hyperparameter tuning is also performed in this module to optimize model performance.

5) Model Evaluation

After training, each model is evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score. This module helps determine which model predicts property prices most accurately and reliably.

6) Prediction & Deployment

The selected best-performing model is deployed to predict the prices of new properties. This module provides a user interface or API for users to input property details and receive predicted prices. It may also include visualizations to show predicted vs actual prices for better interpretability.

C. Algorithm Used

The proposed system primarily utilizes supervised machine learning algorithms such as Linear Regression, Random Forest, and XGBoost for real estate price prediction. Linear Regression is used as a baseline model to establish a linear relationship between input features (e.g., location, area, number of rooms) and the target variable (price). However, since real estate data often contains complex and nonlinear relationships, ensemble learning methods like Random Forest are employed. Random Forest constructs multiple decision trees during training and outputs the average prediction, which improves accuracy and reduces overfitting. XGBoost, a gradient boosting algorithm, further enhances performance by sequentially building trees that correct the errors of previous models, resulting in high predictive accuracy and efficiency.

The algorithm workflow begins with feeding preprocessed data into the selected models for training. Each model learns patterns and correlations between property features and their prices. After training, the models are evaluated using metrics such as Root Mean Square Error (RMSE) and R-squared (R^2) to determine their performance. The best-performing model is then deployed for real-time predictions. When a user inputs property details, the trained model processes the input and generates a predicted price instantly. Additionally, the system can continuously improve by retraining the model with new data, ensuring adaptability to changing market trends and maintaining prediction accuracy over time.

D. Architecture Diagram

The system architecture for the Real Estate Price Prediction model is structured into four main layers: data input, data preprocessing, machine learning model processing, and output prediction. In the first layer, real estate data containing property details such as location, size, number of rooms, age, and market price is collected from public datasets and websites. This raw data is then passed to the preprocessing layer, where missing values are handled, irrelevant attributes are removed, categorical data is encoded, and numerical values are normalized to ensure consistency.

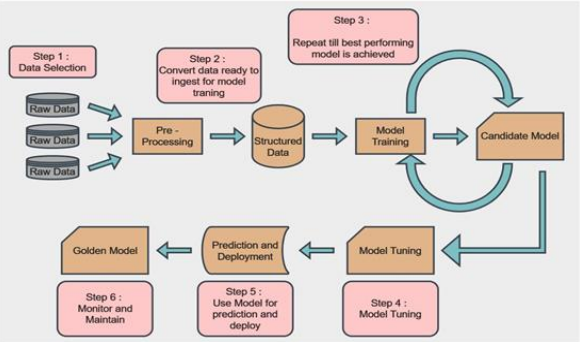

Figure 1.Architecture Diagram

In the machine learning layer, algorithms such as Linear Regression, Decision Tree, Random Forest, and XGBoost are trained using the processed data to learn patterns and relationships between features and property prices. Finally, in the output layer, the trained model predicts the estimated price of a property based on new input features, and evaluation metrics such as MAE, MSE, RMSE, and R^2 Score are used to measure the accuracy and reliability of the model. prediction. In the first layer, real estate data containing property details such as location, size, number of rooms, age, and market price is collected from public datasets and websites. This raw data is then passed to the preprocessing layer, where missing values are handled, irrelevant attributes are removed, categorical data is encoded, and numerical values are normalized to ensure consistency. In the machine learning layer, algorithms such as Linear Regression, Decision Tree, Random Forest, and XGBoost are trained using the processed data to learn patterns and relationships between features and property prices. Finally, in the output layer, the trained model predicts the estimated price of a property based on new input features, and evaluation metrics such as MAE, MSE, RMSE, and R^2 Score are used to measure the accuracy and reliability of the model.

E. Performance Evaluation

The Real Estate Price Prediction system involves using Python as the primary programming language, implemented in a Jupyter Notebook or Google Colab environment, with essential libraries such as pandas, numpy, scikit-learn, matplotlib, and seaborn for data processing, model training, and visualization. The dataset used consists of property-related attributes including location, number of rooms, total area, age of the building, and actual market price, collected from publicly available real estate databases. The data is preprocessed by handling missing values, removing outliers, encoding categorical variables, and normalizing numerical features to ensure uniformity. Various machine learning models such as Linear Regression, Decision Tree, Random Forest, and XGBoost are trained and tested, with the dataset split into 80% for training and 20% for testing. The performance of each model is evaluated using metrics

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 03, ICCT - 2026

like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score to identify the most accurate and reliable prediction model.

IV. CONCLUSION

This project has demonstrated that machine-learning techniques can significantly enhance the accuracy and efficiency of real-estate price estimation. Through the implementation of multiple regression models and comprehensive data preprocessing, the system achieved high predictive performance, with the XGBoost model proving most effective. The integrated web-based platform further enables users to access real-time valuations and interact directly with one another, creating a more transparent and cost-efficient property market.

By reducing human bias and manual effort, this approach offers a reliable alternative to traditional valuation methods. The framework can be expanded for use by financial institutions, real-estate developers, and policy analysts to assess market trends and investment potential.

V. FUTURE ENHANCEMENT

Future work may include the incorporation of deep learning models such as Convolutional Neural Networks (CNNs) to analyze property images and extract visual features that influence pricing. Additionally, integrating geo-spatial data and real-time market trends using APIs can provide more dynamic and location-aware predictions. The system can also be extended with blockchain technology to ensure secure and transparent property transactions. Furthermore, enhancing the chatbot with natural language processing (NLP) capabilities and voice-based interaction can make communication more intuitive. Continuous model retraining with updated datasets and the inclusion of recommendation systems for buyers will make the platform more intelligent, personalized, and scalable in real-world applications.

REFERENCES

[1] Kumar, S., & Bhatia, R. (2020). Real Estate Price Prediction Using Machine Learning Algorithms. International Journal of Innovative Research in Computer and Communication Engineering, 8(6), 1217–1224.

[2] Sirmans, G. S., & Macpherson, D. A. (2003). The Structure of Hedonic Pricing Models. Journal of Real Estate Literature, 11(1), 3–43.

[3] Yacim, J. A., & Boshoff, D. G. (2018). Comparative Evaluation of Machine Learning Algorithms for Predicting Residential Property Values. International Journal of Housing Markets and Analysis, 11(4), 678–696.

[4] Bolar, K. (2021). Real Estate Valuation Using Advanced Regression and Ensemble Learning Techniques. International Journal of Data Science, 6(2), 105–120.

[5] Wang, C., & Li, J. (2022). Using XGBoost and LSTM for Real Estate Price Forecasting. Neural Computing and Applications, 34, 13241–13252.

[6] Chau, K. W., & Chin, T. L. (2003). A Critical Review of Literature on the Hedonic Price Model. International Journal for Housing Science and Its Applications, 27(2), 145–165.

[7] Miller, N. G., & Sklarz, M. A. (1987). A Note on Leading Indicators of Housing Market Activity. Journal of Real Estate Research, 2(3), 195–203.

[8] Rahman, M., & Islam, M. (2020). Machine Learning Approaches for Housing Price Prediction: A Case Study. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 6(3), 258–266.

[9] Li, W., & Luo, Q. (2021). Real Estate Price Forecasting Using Random Forest and Gradient Boosting. Journal of Big Data Analytics in Real Estate, 3(1), 45–58.

[10] Kaggle. (2024). House Prices – Advanced Regression Techniques Dataset. Retrieved from https://www.kaggle.com/c/house-prices-advanced-regression-techniques.