Predictive Modeling of Housing Prices Using Machine Learning Techniques

[1]Kundeti Malakondala Rao, [2]S. Sreenivasulu
[1]M. Tech Student, Department of CSE, Prakasam Engineering College (Autonomous), Kandukur, India
[2]Associate Professor, Department of CSE, Prakasam Engineering College (Autonomous), Kandukur, India

ABSTRACT- ***Accurate house price estimation is important for buyers, sellers, investors, real-estate agencies, and policymakers. Conventional valuation methods often depend on manual appraisal, comparative market analysis, and statistical models that may not represent nonlinear relationships among property characteristics. This paper presents a machine-learning framework for predicting housing prices using a Multilayer Perceptron (MLP) Artificial Neural Network. The model considers property attributes such as location, area, number of bedrooms and bathrooms, property age, neighborhood quality, and proximity to amenities. The workflow includes data ingestion, data cleaning, missing-value treatment, categorical encoding, numerical normalization, outlier handling, exploratory data analysis, model training, and evaluation. The project also describes comparison with Linear Regression and evaluation using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared. The proposed system is intended to provide automated and consistent price estimates through a user-facing prediction and reporting interface. The supplied project report does not include a complete set of measured experimental scores; accordingly, this paper does not introduce unsupported numerical performance claims.***

KEYWORDS: - House Price Prediction, Artificial Neural Network, Multilayer Perceptron, Machine Learning, Real Estate, Regression.

INTRODUCTION

House price prediction is an important problem in the real-estate sector because property values influence household decisions, investment planning, lending, and urban development. A property's market value depends on several factors, including location, area, number of bedrooms and bathrooms, age, neighborhood infrastructure, access to transportation, nearby schools and hospitals, and prevailing economic conditions. These variables may interact in complex ways and can change across regions and time periods.

Traditional valuation relies on human expertise, comparative market analysis, and statistical methods. Although these approaches are useful, they can be time-consuming and subjective, and simple linear models may not capture nonlinear relationships among property features. The increasing availability of online listings, transaction histories, and demographic information has created opportunities for data-driven valuation.

This project uses a Multilayer Perceptron (MLP) Artificial Neural Network to learn relationships between housing attributes and observed prices. The workflow prepares historical data, trains the network, evaluates prediction error, and provides an interface

ISSN: 3049-0952
www.jsetms.com

through which users can obtain estimates from property details. The objective is to support informed decision-making rather than replace professional valuation or market due diligence.

RELATED WORK

House price prediction has been studied using econometric models, hedonic pricing, and machine-learning algorithms. Multiple Linear Regression is widely used because its coefficients are relatively easy to interpret, but its linear assumptions can limit its ability to represent complex interactions. Hedonic pricing models estimate the contribution of property attributes to observed value, although model performance may be affected by feature selection and changing market conditions.

Machine-learning approaches such as Decision Trees, Random Forests, Support Vector Machines, and Gradient Boosting have been used to model nonlinear relationships and interactions. Artificial Neural Networks extend this approach through layers of interconnected processing units. An MLP contains an input layer for property features, one or more hidden layers for nonlinear transformations, and an output layer for the predicted price. Data preprocessing, feature selection, and hyperparameter tuning are important to ensure stable training and reliable evaluation.

LITERATURE SURVEY

Research on predictive modeling has established the value of statistical learning and neural-network approaches for complex datasets. Breiman's Random Forest method provides an ensemble approach based on multiple decision trees, while Support Vector Networks introduced a margin-based classification framework that has also been adapted for regression tasks. General machine-learning texts describe model selection, regularization, feature representation, and validation as important elements of predictive system design.

Neural-network literature, including work on multilayer perceptrons and deep learning, explains how layered architectures can approximate nonlinear mappings between input features and target values. In housing applications, relevant input variables include property size, location, room count, property age, and neighborhood characteristics. The source project emphasizes MLP-based prediction and comparison with Linear Regression. It also identifies MSE, RMSE, MAE, and R-squared as evaluation metrics. The survey therefore motivates an ANN-based pipeline while retaining a conventional regression model as a baseline.

EXISTING METHOD

Existing house price estimation systems primarily use manual appraisal, comparative market analysis, Multiple Linear Regression, hedonic pricing, and basic econometric methods. These approaches generally consider selected attributes such as location, area, bedrooms, bathrooms, and property age. Linear Regression assumes that changes in the input variables have a linear relationship with the target price, whereas real-estate markets often contain nonlinear and interacting influences.

Manual data collection and valuation require time and may produce inconsistent estimates. Hedonic models can be rigid when market conditions change, and conventional workflows may struggle with large or high-dimensional datasets. Limited automation also increases the effort required for cleaning, feature selection, and repeated analysis. These limitations motivate a predictive framework that can learn complex patterns from historical housing records.

ISSN: 3049-0952
www.jsetms.com

PROPOSED METHOD

The proposed system uses a Multilayer Perceptron Artificial Neural Network to estimate house prices from historical property data. The input feature set may include location, property area, number of bedrooms and bathrooms, property age, proximity to amenities, neighborhood quality, and other available attributes. Data ingestion collects housing records from sources such as online listings, historical transactions, government databases, and local property registries, as described in the project report.

Data preprocessing handles missing values, removes duplicate or inconsistent records, treats outliers, scales numerical attributes, and encodes categorical variables using suitable encoding methods. Feature engineering may derive additional attributes from the available property information. Exploratory Data Analysis uses descriptive statistics and visualizations, including histograms, scatter plots, and correlation matrices, to understand feature distributions and relationships.

The MLP architecture consists of an input layer corresponding to the selected property features, one or more hidden layers with nonlinear activation functions, and an output layer that produces the predicted house price. Training adjusts weights and biases through backpropagation and an optimization algorithm. Hyperparameters such as learning rate, number of hidden layers, number of neurons, and activation functions are tuned to improve model performance.

The proposed workflow evaluates the ANN using MSE, RMSE, MAE, and R-squared and compares its results with a Linear Regression baseline. A prediction and reporting module accepts property details and displays the estimated price. Visual reports can compare actual and predicted prices and provide model-performance summaries. In comparison with the existing method, the proposed system automates preprocessing and prediction, learns nonlinear feature relationships, supports repeatable evaluation, and reduces reliance on manual estimation.

SYSTEM ARCHITECTURE

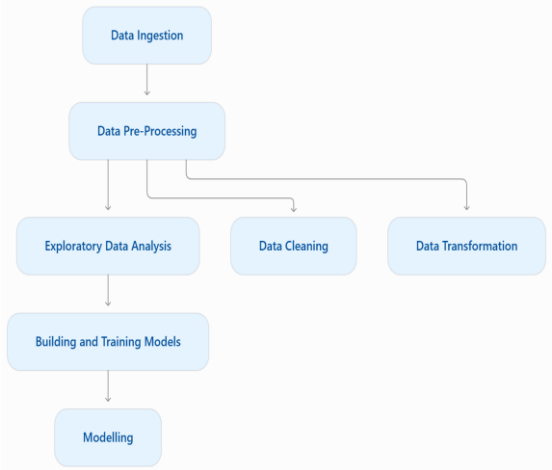

Fig. 1. Housing price prediction architecture adopted from the source project.

The architecture begins with data ingestion, followed by preprocessing and exploratory analysis. The processed dataset is supplied to the model-building and training stage. The trained model is then used for prediction and reporting.

METHODOLOGY DESCRIPTION

Data Ingestion Module: Collects housing datasets from online real-estate listings, historical transaction records, government databases, and local property registries. The records contain features such as location, area, bedrooms, bathrooms, property age, and nearby amenities.

Data Preprocessing Module: Cleans missing, duplicate, inconsistent, and noisy records. Numerical variables are normalized or standardized, categorical features are encoded, and feature engineering is used to construct meaningful attributes.

Exploratory Data Analysis Module: Examines distributions, correlations, outliers, and relationships through statistical

ISSN: 3049-0952
www.jsetms.com

summaries, histograms, scatter plots, and correlation matrices. The analysis supports feature selection and model preparation.

Model Building and Training Module: Constructs the MLP network using an input layer, hidden layers, and an output layer. The model learns from historical data through backpropagation and optimization of weights and biases. Hyperparameter tuning is applied to select a suitable network configuration.

Model Evaluation Module: Measures prediction performance using MSE, RMSE, MAE, and R-squared. Linear Regression is used as a comparison baseline, and evaluation should be conducted on data not used for model fitting.

Prediction and Reporting Module: Accepts property details, applies the same preprocessing operations used during training, and generates a predicted price. The system can present the result alongside visual comparisons and performance reports.

Deployment and User Interface Module: Provides a web-based or standalone interface for entering property information and viewing predictions. The modular structure supports maintenance and future integration of additional economic, demographic, and image-based features.

RESULTS AND DISCUSSION

The source project describes the use of an MLP model for housing price prediction and identifies Linear Regression as a comparison model. It specifies MSE, RMSE, MAE, and R-squared as the principal evaluation measures. These metrics quantify absolute and squared prediction errors and the degree to which the model explains variation in observed house prices.

Figures 2 to 5 show the implemented HouseAI web application, from the home page and user login to the property input form and the prediction result.

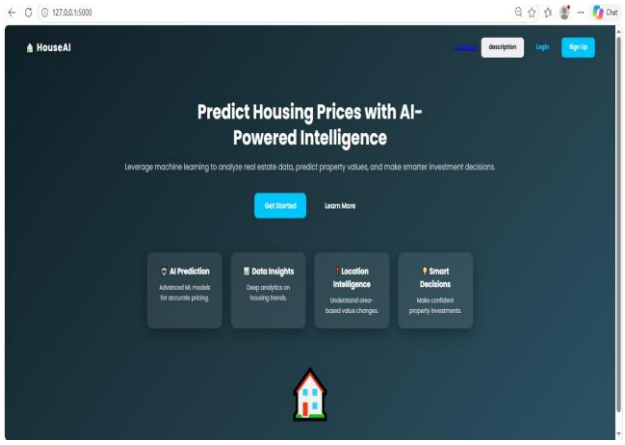

Fig. 2. Home page of the HouseAI application.

Landing page introducing AI prediction, data insights, location intelligence and smart decisions, with Login and Sign Up options.

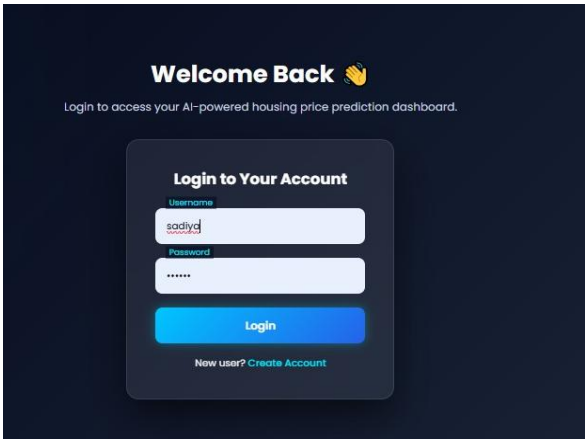

Fig. 3. User login page.

Username and password form that gives registered users access to the prediction dashboard.

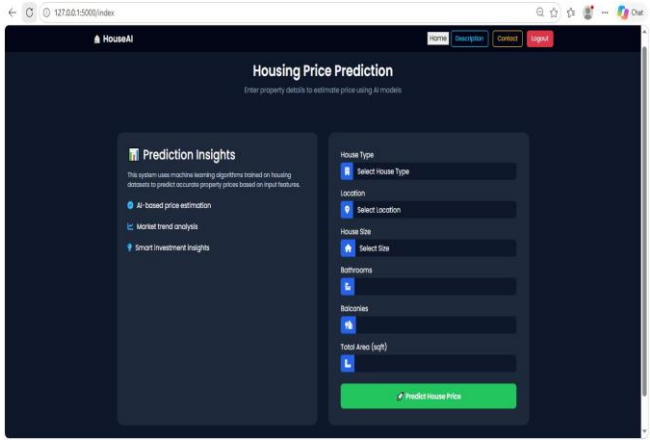

Fig. 4. Property details input form.

Form that collects house type, location, size, bathrooms, balconies and total area (sqft) before prediction.

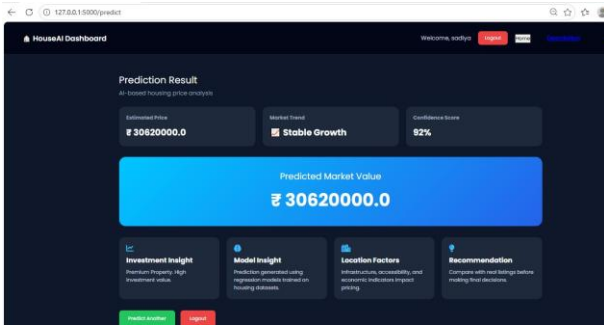

Fig. 5. Prediction result dashboard.

Predicted market value of ₹30,620,000 is shown with a 92% confidence score and a "Stable Growth" market trend.

The supplied report does not provide a complete numerical results table containing measured values for the ANN and Linear Regression models. Therefore, a numerical accuracy improvement or a quantified superiority claim cannot be established from the available material. For a final experimental paper, the authors should insert the actual test-set metrics, describe the dataset size and train-test split, and include an actual-versus-predicted plot. Results should also discuss regional representation, missing data, outliers, temporal changes in market prices, and the possibility of prediction bias.

CONCLUSION

This paper presents an MLP-based machine-learning framework for predicting housing prices. The proposed workflow combines data ingestion, preprocessing, exploratory analysis, neural-network training, model evaluation, and prediction reporting. By learning nonlinear relationships among property attributes, the ANN provides an approach that can address limitations of conventional linear valuation models. The framework includes comparison with Linear Regression and evaluation using MSE, RMSE, MAE, and R-squared. The system is intended to support buyers, sellers, investors, and real-estate stakeholders with consistent data-driven estimates, while professional assessment and local market knowledge remain important.

FUTURE SCOPE

Future development may incorporate updated transaction feeds, regional economic indicators, neighborhood-level geospatial information, and richer property attributes. Ensemble learning and advanced neural architectures may be evaluated alongside the MLP. Explainable AI can help users understand the contribution of individual features to a prediction. A real-time web or mobile interface, continuous model retraining, anomaly detection, and fairness assessment across regions and property categories may further improve reliability and practical value.

REFERENCES

[1] Y. Kütük, "Multidimensional house price prediction with SOTA RNNs," *International Journal of Strategic Property Management*, vol. 28, no. 6, pp. 411–423, 2024. DOI: 10.3846/ijspm.2024.22661.

[2] R. Cellmer and K. Kobylińska, "Housing price prediction—Machine learning and geostatistical methods," *Real Estate Management and Valuation*, vol. 33, no. 1, pp. 1–10, 2025. DOI: 10.2478/remav-2025-0001.

[3] Wang *et al.*, "Forecasting second-hand house prices in China using the GA-PSO-BP neural network model," *PLOS ONE*, 2025. DOI: 10.1371/journal.pone.0322821.

[4] N. Chuhan, "House price prediction based on different models of machine learning," *Applied and Computational Engineering*, vol. 49, pp. 47–57, 2024. DOI: 10.54254/2755-2721/49/20241058.

[5] R. Naz, B. Jamil, and H. Ijaz, "Machine learning, deep learning, and hybrid approaches in real estate price prediction: A comprehensive systematic literature review," *Proceedings of the Pakistan*

Academy of Sciences: A. Physical and Computational Sciences, vol. 61, no. 2, pp. 129–144, 2024. DOI: 10.53560/PPASA(61-2)863.

[6] C. Li, "Review of housing price forecasting methods based on machine learning and deep learning," *Applied and Computational Engineering*, vol. 118, pp. 145–150, 2025. DOI: 10.54254/2755-2721/2025.20931.

[7] I. Moreno-Foronda, M. T. Sánchez-Martínez, and M. Pareja-Eastaway, "Comparative analysis of advanced models for predicting housing prices: A review," *Urban Science*, vol. 9, no. 2, Art. no. 32, 2025. DOI: 10.3390/urbansci9020032.

[8] H. Zhang, "Residential real estate price prediction based on adaptive loss function and feature embedding optimization," *Humanities and Social Sciences Communications*, vol. 12, Art. no. 832, 2025. DOI: 10.1057/s41599-025-05217-9.

[9] F. Wu and T. Kee, "Smart urban forecasting: AdvancedNet's AI-driven approach to housing price prediction in prefecture-level cities," *Urban, Planning and Transport Research*, vol. 13, 2025. DOI: 10.1080/21650020.2025.2598533.

[10] N. Singh and K. Shanmugam, "Exploring predictive models to improve the accuracy of Housing Price Index forecasts in India's real estate sector," *PLOS ONE*, vol. 21, no. 1, 2026. DOI: 10.1371/journal.pone.0341026.

[11] Q. Guo, S. Li, T. Xia, and D. Chen, "IALA-BPNN: Superior house price prediction through multi-strategy optimized Artificial Lemming Algorithm and BP neural network," *Scientific Reports*, vol. 16, 2026. DOI: 10.1038/s41598-026-52995-6.

[12] A. D. Gbadebo and A. O. Ayodele, "Enhancing house price prediction using ensemble machine learning models: A comparative study with U.S. housing data," *Mundo FESC Journal*, vol. 16, no. 34, pp. 501–522, 2026. DOI: 10.61799/2216-0388.1885.