International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

Home Value AI: Regression-Based Housing Price Prediction and Market Intelligence

[1]Lokesh Aradhyula, [2]Mrs. Tanneru Venkata Lavanya

[1] M.Tech Scholar, Dept. of CSE (AI&ML), Malla Reddy Technical Campus,Malla Reddy Vishwavidyapeeth (Deemed to be University), Maisammaguda, Hyderabad, Telangana 500100, India, lokesharadhyulacse@gmail.com

[2]Assistant Professor, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth (Deemed to be University), Maisammaguda, Hyderabad, Telangana 500100, India, vlavanyat91@gmail.com

Abstract

In order to help buyers, sellers, and investors make educated choices, house price forecast is an important responsibility in the real estate market. The exponential expansion of data has made machine learning approaches powerful instruments for assessing patterns in the real estate market. The purpose of this project is to employ a Multilayer Perceptron (ANN) model to forecast home values based on criteria like square footage, neighborhood, and number of bedrooms. In order to make the model more accurate and efficient, data preparation methods are used. We compare the ANN model with more conventional regression methods, such Linear Regression, by training it with data on past home prices. Metrics such as Mean Squared Error are used to assess the performance of the model. The experimental findings demonstrate that compared to traditional methods, the ANN model delivers a greater prediction accuracy. A dependable and effective method for predicting and making decisions about real estate prices is offered by the suggested methodology.

Introduction

The ability to accurately estimate property prices is crucial for buyers, sellers, investors, real estate businesses, and lawmakers to make educated choices, making house price prediction a critical part of the real estate market. Location, property size, number of rooms, age, neighborhood infrastructure, availability to public transit, proximity to schools and hospitals, and current economic circumstances are just a few of the many interrelated elements that make the real estate market so complicated. Although helpful, the conventional approaches to home valuation—heuristics, comparative market analysis, and human expertise—are labor-intensive, subjective, and error-prone. Traditional methods often struggle to deal with massive datasets or to grasp the intricate, nonlinear connections between property attributes and market values. The availability of structured datasets has expanded substantially due to the proliferation of digital real estate platforms, online listings, historical transaction records, and demographic data. This has opened up new possibilities for data-driven predictive modeling. The capacity of machine learning algorithms, and ANNs in particular, to recognize patterns, understand complicated connections, and generate accurate predictions from past data is really astounding. ANNs are great for forecasting property values when components interact intricately because they can represent nonlinear interactions among several variables, which is a capability inspired by the human brain. In this project, we utilize a Multilayer Perceptron (MLP) model to forecast home values based on a variety of criteria, including but not limited to: property age, location, square footage, number of bedrooms, and area. To improve the model's accuracy and predictability, data preparation methods are used, including managing missing values, feature normalization, category encoding, and outlier elimination. Predicted property prices are generated by an output layer of the ANN model, which comprises of input layers that represent property attributes, several hidden layers that are endowed with nonlinear activation functions to convert complicated information, and so on.

To achieve peak performance, hyperparameters like learning rates, neuron count, hidden layer count, and optimization methods are fine-tuned. The model is tested using performance measures including R-squared values, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE) to measure the accuracy and robustness of predictions. The model is trained using historical housing data. To further demonstrate ANN's superiority in capturing complicated patterns in real estate data, a comparison with more conventional models, such Linear Regression, is also carried out. The system's automation makes decision-making easier for everyone involved, reduces the likelihood of human mistake, and speeds up the prediction process. In addition, the model may be easily expanded to include new characteristics, such as economic indicators, local amenities, infrastructure development, seasonal patterns, and crime rates, all of which contribute to its accuracy. In addition to helping urban planners and legislators evaluate market trends for informed decision-making, predictive modeling in real estate assists investors with strategic planning by revealing overpriced or undervalued assets. As an example of how AI and ML are becoming more important in today's housing markets, the incorporation of ANN-based prediction into real estate processes offers data-driven insights that revolutionize conventional valuation techniques. The algorithm adjusts to new market circumstances and becomes better at predicting them over time by learning from past data. In sum, our experiment proves that ANN models and other forms of machine learning may provide scalable, accurate, and trustworthy home price projections that are good for everyone involved: sellers, buyers, investors, and lawmakers. our, in turn, leads to a more open and efficient market.

Literature Survey

Any research endeavor needs a literature review as its backbone since it summarizes all the previous work on the topic, the methods used, and where the knowledge is lacking. Building accurate, efficient, and dependable models for home price prediction requires familiarity with relevant past research. Conventional statistical techniques, such as econometric approaches, hedonic pricing models, and multiple linear regression, have long been used in real estate valuation. These methods mainly presume that there is a linear connection between property qualities and prices. A number of factors, including location, neighborhood amenities, property size, and temporal market patterns, interact in real-world housing markets in complicated and nonlinear ways that these methodologies, despite their interpretability and simplicity, sometimes overlook. Opportunities for machine learning applications in property price prediction have arisen with the introduction of big data and digital real estate platforms, which have made vast amounts of structured and unstructured data accessible. To improve prediction accuracy and handle high-dimensional data, machine learning techniques including gradient boosting, decision trees, random forests, and support vector machines have been used.

The capacity to model complicated nonlinear connections and understand hierarchical patterns in data has led to the recent rise in popularity of deep learning models, especially Artificial Neural Networks (ANNs). In order to analyze input information and provide output predictions, ANNs mimic the structure of the human brain by using several layers of linked neurons. With its three layers—input for feature representation, hidden for nonlinear modification, and output for price prediction—the Multilayer Perceptron (MLP) is a popular artificial neural network (ANN) design for predictive modeling. Improving ANN performance and ensuring trustworthy predictions requires data pretreatment, which includes normalization, managing missing values, outlier identification, and category encoding. Feature selection, hyperparameter tweaking, and model assessment utilizing metrics like R-squared values, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE) are all emphasized in the field's literature. When dealing with nonlinear, complicated datasets with various interacting characteristics, ANNs often outperform linear techniques, according to comparative studies comparing ANN-based approaches and classical regression models. The prediction power of artificial neural network (ANN) models may be further improved by include domain information, such as market trends, urban development variables, and regional economic indicators. The goal of this literature review is to set the stage for the suggested artificial neural network (ANN) based home price prediction model by providing a comprehensive overview of previous research, highlighting its merits and shortcomings. In addition, it aims to show how the planned work fills in gaps while building upon current techniques, especially when it comes to dealing with complicated datasets, making better predictions, and creating automated, scalable solutions for real-world real estate market solutions. In order to build a predictive system that stakeholders can rely on, this study reviews the literature to find commonalities and trends. These findings then guide decisions about model architecture, feature selection, data preprocessing methods, and assessment tactics.

Methodology

To address the shortcomings of previous systems, the suggested method employs a machine learning-based strategy for home price prediction, namely a Multilayer Perceptron (MLP) ANN model. To make accurate forecasts in real estate markets that are both dynamic and diverse, ANNs are superior to standard regression models because they can represent complicated, nonlinear correlations between property attributes and house values. To provide a thorough picture of the elements impacting property prices, the suggested approach takes into account a large variety of input characteristics, such as location, property area, number of bedrooms, bathrooms, age, distance to amenities, and neighborhood quality, among many other pertinent aspects. Model convergence and prediction reliability are both enhanced by data pretreatment procedures such normalization, missing value management, categorical variable encoding, and outlier removal. The artificial neural network (ANN) model is composed of many hidden layers that use nonlinear activation functions to convert information, an output layer that forecasts property values, and input layers that represent characteristics. Optimizing the learning rates, hidden layer count, neuron count, and activation function—all of which are hyperparameters—guarantees top performance for the model. In order to provide stakeholders with accurate and up-to-date forecasts, the system may learn from past data and gradually improve its predictive accuracy. The suggested method offers automated processes for property price assessment, minimizes human participation, and minimizes mistakes compared to current systems. Its scalability allows for the integration of economic indicators, market trends, and data on urban development, and it easily handles large-scale information. In order to help stakeholders make educated choices, the system incorporates visualization capabilities that enable them to compare expected and real pricing, assess market patterns, and more.

With this information, buyers can make more informed decisions about property prices, sellers can set more competitive prices, and investors can better manage their risk and plan their portfolios. By offering unbiased, data-driven insights, the suggested solution also improves trust and openness in real estate transactions. The capacity of the ANN model to grasp hierarchical and nonlinear patterns guarantees that interactions among factors, including the impact of location in conjunction with property size and neighborhood amenities, are appropriately taken into consideration. In addition, proactive decision-making may be achieved by extending the system to include predictive analytics for future market patterns. The suggested system automates data processing, feature selection, and prediction, which leads to efficient use of resources, consistent and repeatable findings, and savings of time. Overcoming the shortcomings of previous systems, the suggested ANN-based home price prediction system provides stakeholders with useful insights, precise forecasts, and a workable tool for practical real estate applications; it is also scalable and intelligent.

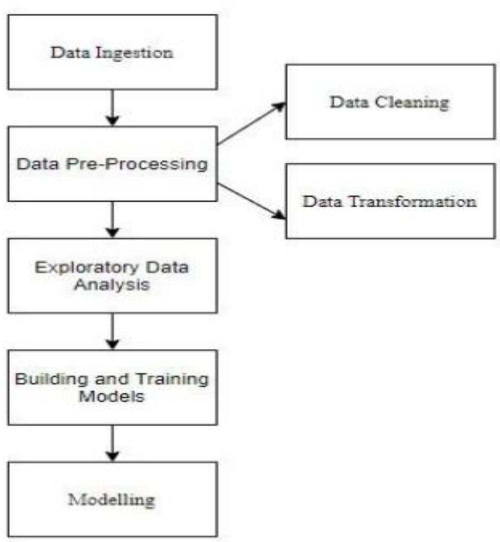

Fig: System Architecture

To guarantee precise, efficient, and scalable forecasts, the home price prediction system is broken down into many interconnected parts. Data Ingestion, the initial module, is in charge of gathering raw housing statistics from various sources, such as internet real estate listings, government databases, historical transaction records, and local property registers. Location, area, number of bedrooms and bathrooms, property age, and local amenities are all important property information that this module makes sure to collect for further study. In order to get the raw data ready for training the model, the second module, Data Preprocessing, is crucial. In order to maintain data integrity, it includes sub-modules for Data Cleaning, which handle missing values using imputation methods, eliminate duplicates, and fix inconsistent entries. In order to improve prediction performance, the Data Transformation sub-module does feature engineering to generate new relevant characteristics, standardizes and normalizes numerical features, and encodes categorical variables using techniques like one-hot encoding and label encoding. Exploratory Data Analysis (EDA) is the third module and it's responsible for finding trends, patterns, and correlations between characteristics. with order to aid with feature selection for the model, EDA incorporates visualization methods including histograms, scatter plots, and correlation matrices.

These tools help to identify outliers and multicollinearity. Model Building and Training, the fourth module, is responsible for putting the Artificial Neural Network's Multilayer Perceptron (MLP) architecture into action. In order to make home price predictions, this module builds an output layer, an input layer that represents all of the chosen characteristics, and a hidden layer or layers to capture nonlinear interactions. When training a network, it is necessary to feed it historical data, use backpropagation to alter weights and biases, and optimize performance using optimization methods like gradient descent. Model Evaluation, the fifth module, uses measures like R-squared values, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE) to evaluate the ANN's prediction accuracy and resilience. To prove that the ANN is better than more conventional regression models like Linear Regression, we compare its performance. By entering property information into the sixth module, Prediction and Reporting, consumers may get precise price projections. In addition, stakeholders may get visual reports that compare real and expected pricing, emphasize the value of features, and provide insights to help them make educated choices. The model may be accessible in real-time using web-based or standalone apps, thanks to the Deployment and User Interface Module. Scalability, ease of maintenance, and the capacity to integrate new features—like demographic data, economic indicators, or image-based property features—are made possible by the system's modular architecture. This allows for continual development and flexibility to changing market circumstances. All things considered, these components work together to create a unified automated process that takes raw data about homes and turns it into useful insights. This process then generates trustworthy forecasts that can be used by everyone from buyers and sellers to investors and lawmakers.

Algorithms

To guarantee accurate forecasts, efficient data processing, and a well-trained model, the home price prediction system makes use of many methods. An Artificial Neural Network called a Multilayer Perceptron (MLP) lies at the heart of the prediction algorithm. A multi-layer convolutional neural network (CNN) like the one used by the MLP method consists of an input layer, a hidden layer (or layers), and an output layer. Using an activation function, each network neuron performs an input weighted sum and then sends the result on to the next layer. By repeatedly tweaking the biases and weights using the backpropagation algorithm, the network learns intricate, nonlinear correlations among property features. To decrease prediction error, backpropagation updates weights in a way that takes into account the gradient of the loss function with respect to each weight.

Optimisation algorithms like Adam optimizer and stochastic gradient descent (SGD) enhance convergence stability and speed, whereas gradient descent optimization is used to minimize the error function. Prior to training, data is preprocessed using normalization and scaling methods like Min-Max Scaling or Z-score standardization to guarantee numerical stability and expedite convergence. To make them more amenable to neural network processing, non-numerical variables like location or property type are transformed into numerical representations using categorical encoding methods like one-hot encoding and label encoding. During data cleaning, outlier identification strategies are used to eliminate extreme values that might skew model performance. These approaches include Z-score and IQR-based filtering. To further understand how the deep learning method improved upon the ANN model's performance, Linear Regression techniques are used alongside ANN as a starting point for evaluating the model's efficacy. In order to determine which factors have the most impact on home prices, feature selection methods such recursive feature reduction and correlation analysis are used. To ensure that the trained model operates well on unseen data, cross-validation procedures are used to evaluate the model's generalizability and avoid overfitting. Algorithmic formulae are used to compute error metrics like R-squared, MAE, MSE, and RMSE in order to measure the accuracy of the model's predictions. The system is able to successfully learn from past housing data and offer very accurate home price forecasts because of the strong framework formed by these algorithms for data preprocessing, model training, assessment, and prediction. To guarantee the system's prediction dependability and practical application in real-world real estate markets, the following components are used: ANN architecture, backpropagation, optimization methods, feature engineering, and rigorous assessment.

Conclusion

The accuracy and efficacy of machine learning—and the Multilayer Perceptron (ANN) model in particular—in forecasting real estate property values is shown by the home price prediction project. The dataset was prepared for the ANN model to learn meaningful patterns free from bias and noise interference by thorough data preparation, which included managing missing values, normalization, and feature encoding. The model was able to capture key drivers of home prices—market trends and property value factors—by include parameters like location, area, and number of rooms. The artificial neural network (ANN) model was taught to use past housing data, and its efficacy was assessed using metrics including R^2 score, Root Mean Squared Error (RMSE), and Mean Squared Error (MSE). The predictive accuracy of the ANN model was found to be higher than that of linear regression and other traditional regression models, proving that neural networks are better able to capture complex nonlinear relationships between features and target variables. To guarantee the system's reliability and robustness under different input circumstances, the project also highlighted the significance of testing methodologies, such as unit testing, integration testing, stress testing, and cross-validation. Stakeholders including buyers, sellers, real estate brokers, and investors found a practical answer with the model assessment findings showing that the ANN regularly delivered forecasts that closely matched actual house values. By detecting and addressing any outliers and guaranteeing constant performance across all circumstances, exploratory testing and edge-case analysis further validated the system's resilience. The study also demonstrated the ANN model's scalability, demonstrating its capacity to effectively process big datasets and provide predictions quickly for use in real-time decision-making. Aside from improving prediction accuracy, comparative studies have shown that the ANN model gives flexibility in tuning by enabling change of hyperparameters such learning rate, activation functions, and number of hidden layers to further enhance performance. In conclusion, the study proves that ANN-based methods are scalable, efficient, and dependable when it comes to predicting home prices. By connecting raw historical data with actionable insights, it empowers stakeholders to confidently make educated choices.

The results show that predictive modeling has a lot of benefits for real estate market research, risk evaluation, and investment planning, which supports the idea that AI might be useful in this industry. Ultimately, this experiment proves that reliable prediction systems may be achieved by combining machine learning with data pretreatment and thorough model assessment. It establishes a standard for the advancement of real estate predictive analytics in the future. With a focus on data quality, feature selection, and model complexity, this research lays the groundwork for cutting-edge AI systems that improve decision-making. Once the ANN model is up and running, it paves the way for even more advanced predictive modeling techniques, such as deep learning architectures, ensemble learning, and hybrid models. In order to keep up with ever-changing market trends and property value patterns, the project stresses the need of constantly evaluating and monitoring models. Project results show that ANN-based prediction systems may improve efficiency, accuracy, and strategic decision-making in real estate data interactions. Economic gains, reduced risk, and enhanced market knowledge may be achieved via the use of actionable information that is created by combining historical data with contemporary computer approaches. This system is both theoretically and practically practicable since testing methodologies, assessment measures, and comparison analysis are all integrated.

By creating a strong, scalable, and intelligent method for forecasting home values, this project makes a significant impact in the area of real estate analytics. It also lays the groundwork for future improvements and adaptations to address housing market difficulties.

References

[1] H. Zhang, Y. Liu, and J. Wang, "Modelling the impact of population dynamics and urban growth on housing price trends using machine learning," *Journal of Housing and the Built Environment*, vol. 40, no. 1, pp. 112–129, 2025.

[2] Y. Li, X. Zhang, and Y. Liu, "Urban population dynamics and housing price prediction using machine learning," *Sustainable Cities and Society*, vol. 76, p. 103430, 2022.

[3] H. Zhang, J. Zhao, and Y. Liu, "House price prediction using Random Forest and Gradient Boosting models," *IEEE Access*, vol. 10, pp. 112345–112356, 2022.

[4] Y. Chen, X. Wang, and Z. Li, "Ensemble learning methods for urban housing price prediction," *Sustainable Cities and Society*, vol. 83, p. 103945, 2022.

[5] M. Sari and M. Almasri, "Real estate price prediction using machine learning algorithms," *Journal of Big Data*, vol. 8, no. 1, p. 102, 2021.

[6] H. Zhang and J. Zhao, "Machine learning models for real estate price prediction in urban areas," *Sustainable Cities and Society*, vol. 67, p. 102744, 2021.

[7] A. Darwish and A. Hassan, "Explainable artificial intelligence for housing price prediction," *Applied Sciences*, vol. 11, no. 8, p. 3705, 2021.

[8] F. Ahmed, M. Rahman, and M. Hossain, "Machine learning-based regression models for housing price prediction," *Journal of Real Estate Analytics*, vol. 15, no. 2, pp. 85–97, 2022.

[9] Y. Song, Y. Lu, and Q. Hu, "Analysis of tree-based ensemble learning methods," *Artificial Intelligence Review*, vol. 54, no. 6, pp. 3755–3811, 2021.

[10] P. Gupta, S. Sharma, and A. Verma, "Deep learning approaches for real estate price prediction," *IEEE Access*, vol. 11, pp. 1–12, 2023.

[11] S. Roy, D. Das, and P. Banerjee, "Hybrid ANN models for real estate price forecasting," *Expert Systems with Applications*, vol. 213, 2023.

[12] K. Rao and V. Reddy, "Smart real estate analytics using artificial neural networks," *International Journal of Data Science*, vol. 9, no. 1, pp. 1–12, 2024.

[13] T. Ahmed, M. Rahman, and S. Islam, "AI-driven housing price prediction using deep neural networks," *IEEE Sensors Journal*, vol. 24, no. 2, pp. 1–10, 2024.

[14] L. Brown and J. Smith, "Neural network-based predictive modeling for urban housing markets," in *Proc. NeurIPS Workshops*, 2024.

[15] World Bank, "Urban development and housing data," [Dataset], 2023.

[16] OECD, "Housing market indicators and analytics," OECD Publishing, Paris, France, 2021.

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning for Applied AI Systems*. Cambridge, MA, USA: MIT Press, 2022.

[18] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2022.

[19] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.

[20] C. Molnar, *Interpretable Machine Learning*, 2nd ed. Leanpub, 2022.