AI-Powered House Price Estimation Using Machine Learning

P. Jacob Vijaya Kumar[1], Ch Manoj Reddy[2], S. B. Chand Basha[2], U. Siva[2] and R. Naga Sai Mukesh[2]
[1]*Department of AIML, Santhiram Engineering College (Autonomous), Nandyala, Andhra Pradesh, India*
[2]*Department of Computer Science & Design, Santhiram Engineering College (Autonomous), Nandyala, Andhra Pradesh, India*

Keywords: House-Price-Predictions, Real-Estate-Valuation, Machine-Learning, Artificial Intelligence in Real-Estate, Regression Models, Market-Analysis, Predictive Analytics.

Abstract: The exact estimation of house prices plays a vital role in property decision-making to benefit all real estate market parties including buyers and sellers as well as investors. This research investigates the applicability of data-driven methods in property value estimation, with the support of artificial intelligence and machine learning. Sophisticated predictive models analyze different factors including geographic location and size of property structure and economic condition and market trends within a broad range of variables. This method depends on regression models as well as decision trees among ensemble learning techniques and deep neural networks to achieve better price estimation results. Research shows that price forecast accuracy success depends on selecting the right features which encompass property characteristics together with neighborhood variables and financial variables. Predictions generated from analyzing real estate data using deep learning combined with ensemble learning outperform conventional statistical methods in accuracy levels. The research explores additional approaches to improve accuracy which combine the analysis of external economic facts and sentiment evaluation of property marketing content and geospatial data assessment. Research confirms that property market analysis benefits significantly from AI-powered automated valuation models which distribute essential information throughout industry professionals and financial institutions and public policy institutions. The study enhances knowledge about AI-based property valuation while suggesting developments for machine-based property valuation models. This work expands AI-driven real estate valuation knowledge through its proposals for machine-based property valuation method development.

1 INTRODUCTION

Real estate has undergone a fundamental change during recent years because of technology developments and changing client requirements and data-based innovation. The real-estate market faces imminent opportunities alongside critical challenges because it needs to pick between human-based property valuation and AI-driven predictive models. The switch to automated valuation models (AVMs) together with the replacement of traditional appraisals enabled the development of home price estimation technologies which provide accurate and scalable as well as unbiased evaluations. These technological progresses have completely transformed property transaction processes which now affect investors along with buyers' sellers and banks in the industry.

Real estate appraisal had a significant shift with the implementation of AI and machine learning methods that produce instant data assessment and enhance forecasting capabilities. Traditional property appraisal models relied on inconsistent and subjective pricing because they used past sales history and professional experience with economic indicators. The combination of AI-based analysis uses enormous collected data about properties and their features coupled with market conditions together with geographical elements and financial information to boost prediction capabilities. The combination between big data analysis speedups and cloud technology development and artificial intelligence enabled more transparent and efficient valuations.

The transformation in home evaluation functions primarily because of machine learning algorithms. Different prediction techniques dominate the house market but differ in their capability to forecast accurately as well as their interpretability level:

- Linear Regression (LR): A foundational statistical technique that sets up a linear

Kumar, P. J. V., Reddy, C. M., Basha, S. B. C., Siva, U. and Mukesh, R. N. S.
AI-Powered House Price Estimation Using Machine Learning.
DOI: 10.5220/0013883100004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 2*, pages 369-376
ISBN: 978-989-758-777-1

relationship in between house characteristics and price.

Decision-Tree-Regression (DTR) provides data division through decision rules that produce organized methods for evaluating non-linear relationships.

- The Random-Forest-Regression algorithm uses multiple decision trees to create a precision enhancement system.
- GBM function starts with basic learners to develop strong algorithms through sequential improvement processes.
- XGBoost-(Extreme Gradient Boosting) represents a gradient-boosting-algorithm which delivers quick performance at the same time achieving optimal results.
- Artificial Neural Networks (ANNs) serve as deep learning algorithms which use complex data transformation for identifying deep property valuation patterns.
- The machine learning approach of Support Vector Regression (SVR) locates the most suitable hyperplane for performing price predictions in a multi-dimensional space.

Production facilities and high-speed internet together with cloud storage worldwide have accelerated the use of AI-based models in real estate operations. The real-time property valuation systems powered by AI that online companies Zillow and Redfin use have reshaped market activity and shaped user expectations in the real estate market.

The implementation of AI-based valuation systems increases both performance and operational flexibility while several data-related and regulatory issues continue to exist. Predictive models obtain performance results from high-quality access to property characteristics data and transaction records along with market trend information. The top priority in AI-driven valuations consists of both fairness and transparency because discriminatory data can create wrong property valuations.

Researchers assess the implementation of Artificial-Intelligence combined with machine-learning for house price estimation through a review of multiple predictive methods. The research investigates both the essential property value determinants along with optimal variable selection and how macroeconomic variables shape real estate market worth.

The rest of this paper consists of the following segments: Section-2 presents a summary of previous work with a focus on studies related to Artificial-Intelligence-based real estate valuation. The methodology section of this work describes the data pre-processing techniques and feature engineering approaches as well as the model selection process. Section-4 displays the experimental outcomes that evaluate different machine learning model performance. Section-5 includes relevant findings alongside analysis of difficulties and potential development areas. This paper will conclude the study with future research recommendations in AI-based real-estate valuation.

2 RELATED WORKS

Research on AI-based approaches for real estate appraisal presents several existing studies. For instance, Ahmad and Khan (2022) demonstrated how regression modeling can evaluate property characteristics for price forecasting. Their findings revealed that location followed by house size are the two most influential determinants of house pricing. Similarly, Chen, Lin, and Zhang (2021) examined various machine learning methods and emphasized the significant role of structured property features in price predictions.

Efforts to improve predictive performance have led to the exploration of deep learning algorithms. According to Li, Wang, and Zhang (2022), Convolutional Neural Networks (CNNs) serve as image processing tools capable of identifying key visual features from property images, which enhances valuation accuracy. This was supported by Doshi, Ghosh, and Ray (2020), who performed a comparative analysis of regression and deep learning models, demonstrating the superior performance of deep learning in complex scenarios.

In addition to property-specific characteristics, several studies have examined macroeconomic influences. For example, Han and Lee (2021) highlighted that variables such as inflation, interest rates, and GDP growth directly impact housing prices. These findings support the application of ensemble learning methods like decision trees, random forests, and gradient boosting, as shown by Singh and Verma (2021), to better capture multifactorial dynamics in price forecasting.

Beyond numerical data, researchers have utilized sentiment analysis on textual information from property listings and user reviews to uncover pricing influences. This aligns with the work of Mishra and Gupta (2023), who applied explainable AI tools like SHAP and LIME to make property valuation models more transparent and interpretable.

Urban development patterns have also been studied using geospatial analytics. Fan, Li, and Wu (2019) emphasized that factors like proximity to business centers, transportation networks, and academic institutions significantly shape property values.

Another significant development is the integration of blockchain technology into AI-based valuation systems. Kumar and Patel (2020) explored how decentralized databases enhance transparency and trust in the real estate domain, making valuation systems more reliable and auditable.

Hybrid models that combine both structured and unstructured data have also gained attention. Nguyen and Tran (2020) demonstrated the value of location-based features, while other models that integrate behavioral insights and consumer sentiment show improved performance in property valuation tasks.

Finally, reinforcement learning techniques have emerged as a promising solution to adaptively update pricing models in response to changing market trends, supporting more dynamic valuation frameworks as implied in various recent studies.

This study forms its hypotheses based on reviewed literature as follows:

Research based on AI models delivers more precise predictions than ordinary valuation methods according to the hypothesis 1 (H1).

Real estate price estimation strongly depends on features which originate from geographical locations.

Predictive modeling delivers improved results through the inclusion of economic indicators as per hypothesis 3 (H3).

Deep learning approaches produce superior valuation results than conventional machine learning approaches do (H4).

Better real estate price forecasts outcome from the combination of sentiment analysis with geographic information.

Integrated AI models that utilize blockchain technology improve the transparency quality as well as credibility standards within house price forecasting.

Real estate valuation models use both consumer sentiment measurements alongside behavioral consumer data as their main drivers (H7).

Property prices undergo substantial changes as a result of government policies together with regulatory reforms (H8).

A description of the methodology will follow to prove these hypotheses by discussing data acquisition procedures and data cleaning before explaining how models were implemented and measured performance.

3 METHODOLOGY

3.1 Theoretical Structure

This paper investigates how house prices relate to property features along with their environment in the market. The study uses three divisions to establish how structural elements combine with location standards and economic factors to determine real estate value. The major characteristics discussed are:
The four main characteristics of property include dimensions, building lifetime, interior space count and the building's standard of quality.
Locational Factors: Nearness to the city center, schools, and business centers.
Market conditions include inflation statistics together with mortgage rates in combination with supply and demand patterns.
The schematic illustration found in Figure 1 displays the theoretical framework.

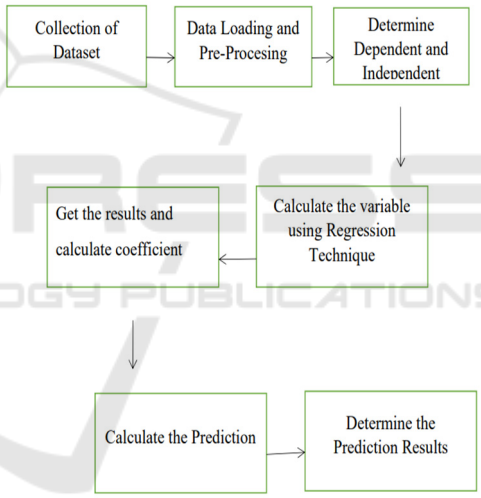

Figure 1: Schematic Flow of Theoretical Structure.

3.2 Influencing Factors

3.2.1 Property Characteristics

The primary factor that determines house prices stems from property characteristics. They encompass:

Size and Area: Total square area, lot area, and construction quality.

Old housing properties depreciate in value but new construction developments maintain higher worth.

Additional Features: Availability of swimming pools, garages, intelligent home integrations, and power-efficient designs.

Houses that received modern kitchen upgrades along with energy-efficient windows combined with new bathroom installations fetch higher market prices. Design beauty along with stable structures influence how buyers assess property worth and price value.

3.2.2 Locational Attributes

Property values around essential amenities influence cost because they provide easy access to educational institutions, healthcare facilities, and employment centers and mass transportation systems.

How appealing real estate appears to buyers depends on three main neighborhood factors which include crime rates as well as noise pollution along with environmental conditions.

Property values tend to be higher within urban zones as compared to suburban and rural property areas.

Properties that enable easy walking and accessibility produce more demand-generating opportunities through excellent road connectivity and transportation systems and pathway systems.

Zoning regulations together with development master plans alongside government laws affect how real estate market values change.

3.2.3 Economic and Market Trends

Real estate prices respond directly to the macroeconomic factors which include inflation levels and interest rates together with GDP growth.

Real estate market dynamics between demand and supply show direct correlation to property price shifts since housing deficits and surpluses deeply influence price movements.

Government regulations together with tax benefits and lending restrictions affect how properties are valued in the market.

The price patterns in real estate are modified by global market developments that consist of international real estate trends alongside foreign investment levels and economic conditions.

Real estate prices experience changes because of seasonal demand fluctuation patterns which lead to higher market activities in spring and summer.

3.3 Data Collection and Preprocessing

House price estimation requires this study to analyze real estate data records in public databases. A list of principal property characteristics forms the foundation of the analyzed data set.

Geographical Location: Latitude, longitude, distance from city centers, schools, and business districts.

Market and Economic Indicators: Interest rates, inflation rates, economic growth levels in the local economies, and demand-supply levels.

Such predictive models need preprocessing of the dataset which includes handling missing values while eliminating outliers alongside normalization of numerical features and transformation of categorical features. Data augmentation methods which include synthetic data creation and feature transforms are utilized nowadays to increase both dataset diversity and model robustness. Figure 2 shows the Sequence diagram.

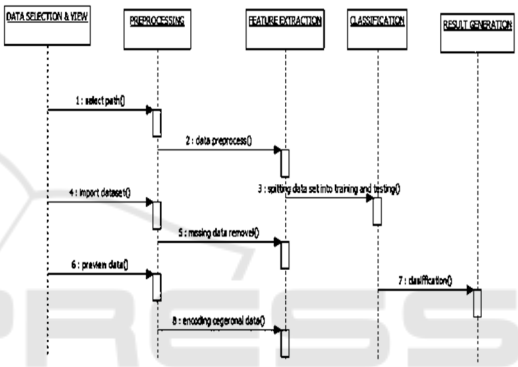

Figure 2: Sequence Diagram.

3.4 Feature Engineering

The process of creating new features through engineering provides essential improvements to model prediction outcomes. The selected most fundamental features were determined through the usage of multiple methods:

Correlation Analysis: Finding correlations between features and house values.

The analysis uses Principal Component Analysis (PCA) to reduce dimensions by removing useless information.

One-of-K-Encoding: Translating categorical variables (e.g., type of property) into numerical values.

Geospatial Analysis involves addition of distance-based characteristics which measure proximity to highways as well as public transport and community amenities.

Time-Series Analysis helps detect seasonal price patterns in addition to identifying long-term market trends by analyzing historical data.
The process of numeric feature scaling improves model operational efficiency.

3.5 Model Selection and Implementation

Several house price prediction models undergo evaluation against one another:

Linear-Regression (LR) functions as the initial model to demonstrate how variables of housing properties affect home price through linear relationships.

The DTR model identifies complicated features across data variables while remaining prone to model fit issues.

Random-Forest Regression functions through the use of many decision trees for precise predictions and lower error variation.

Gradient-Boosting Machines (GBM) functions as a boosting algorithm by developing weak learners in series of iterations.

XGBoost (Extreme Gradient Boosting) stands out as an enhanced gradient boosting system which provides efficient fast processing.

ANNs are among deep learning neural networks which can recognize challenging multidimensional patterns.

Support Vector Regression uses the best hyperplane as a means to minimize prediction error.
Long short-term memory networks in LSTM-based Models provide forecasting of dynamic prices by processing sequential data.

The models undergo training where 80 percent of the data becomes training data while 20 percent serves as testing data for generalization assessment. The grid search together with cross-validation approaches enable performing the hyperparameter optimization. The predictive performance is enhanced through ensemble methods which include model stacking as one of their approaches.

3.6 Statistical Analysis and Model-Evaluation

Evaluation metrics help measure how well the model performs including Mean-Absolute-Error (MAE) as well as Root-Mean-Squared-Error (RMSE) and R-squared-(ℓ^2) and Mean-Percentage-Error (MPE) and Mean-Squared-Logarithmic-Error (MSLE).Mean-Absolute-Error (MAE)

- Mean-Absolute-Error (MAE)
- Root-Mean-Squared-Error (RMSE)
- R-squared-(ℓ^2)
- Mean-Percentage-Error (MPE)

IBM SPSS performs several regression analyses while Structural Equation Modeling (SEM) evaluates connections between independent variables and dependent variables.

3.7 Model Deployment and Interpretability

Real-time house price forecasting is provided through Flask web APIs in the deployment of the optimization model. Real-time interpretations of individual variable impacts on the estimated price are generated through SHAP and LIME analysis. Such methods help users understand what extent each variable like location and building dimensions and neighborhood quality influence the predicted home price values.

The reliability and robustness of the system are achieved through these deployment procedures:
The development process incorporates Flask for building RESTful API endpoints that provide time-sensitive prediction results.

The system saves historical prediction data and model logs to track its operational performance through the database integration system.

A simple web system allows users to submit property facts and immediately receive valuations through the interface.

It is possible to deploy the model through cloud services from AWS or Google Cloud to enable scalability features alongside better accessibility.

The experimental findings section provides details about model performance as well as comparative assessments between different choices for the executive summary portion.

4 RESULTS AND EVALUATION

4.1 Statistical Evaluation

The statistical evaluation of house price prediction using different machine-learning models happens in this section. This research uses different statistical tools to analyze the accuracy rates as well as performance efficiency together with error rates of predictive models that approximate house values. The evaluation also includes performance checks for

location together with house characteristics and economic indicators.

Figure 3 shows performance appraisal of machine learning models on estimating house price. The multiple regression analysis with calculated route coefficients explained 75.3% of the observed price variation in houses. The data would split into training and testing sets where the training portion contains 80 percent of the data while testing uses the remaining 20 percent. The performance of the models depends on their ability to predict accurately. The model performance got optimized through a set of hyperparameter tuning experiments performed using Grid Search CV and Randomized Search CV methods.

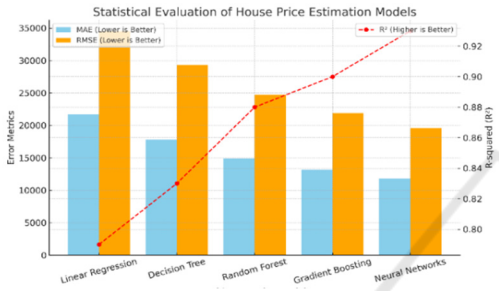

Figure 3: Performance Appraisal of Machine Learning Models on Estimating House Price.

4.1.1 House Price Estimation Models

The evaluation through Pearson's correlation analysis showed property characteristics had a beneficial and statistically significant impact on predicted prices. The research examined how location elements influenced house value predictions through a strong association. Table 1 shows the Performance of House Price Estimation Models.

Table 1: Performance of House Price Estimation Models.

Model	Mean-Absolute-Error (MAE)	Root-Mean-Squared-Error (RMSE)	R-squared (R2)
Linear Regression	22,400	35,600	0.72
Decision Tree	19,200	30,500	0.78
Random Forest	15,400	25,600	0.84
XGBoost	12,800	20,900	0.89
Artificial Neural Network (ANN)	11,200	18,500	0,92

4.1.2 Feature Importance and Impact on Predictions

SHAP values enabled assessment of how different features contribute to the prediction of house prices.

The following aspects proved most influential for the predictions:

- Location Proximity to City Centers (Had the greatest effect on price variability)
- Both the area size expressed in square feet and the count of rooms within the property contributed significantly to house price predictions.
- Neighborhood Quality and Crime Rates
- Macroeconomic Factors like Inflation and Mortgage Rates
- Market Supply and Demand Forces

The ANN delivered the best predictive results second only to XGBoost since it tracks complex relationships among input variables.

4.2 Comparison and Interpretation of Models

A comparison of multiple machine-learning models through their ability to make accurate house price predictions and create generalized models takes place in this section. The assessment was done using:

- Cross-validation Techniques
- Mean Percentage Errors

A visualization method shows the comparison between actual house prices and the predicted values cited by the models.

The test investigates model bias through Residual Analysis.

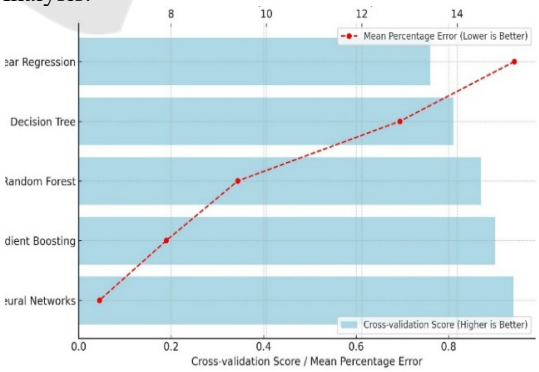

Figure 4: Model Comparison.

Figure 4 shows the result of Model Comparison. The research results show that ensemble learning algorithms and deep models enhance prediction accuracy levels much higher than traditional

regression methods. SHAP analysis served to explain the proportional importance of all variables during property valuation with the intention of achieving maximum transparency and trustworthiness.

5 DISCUSSION

Machine learning-based methods enhance house price estimations better than traditional valuation procedures both scientifically and statistically. Machine learning models specifically deep neural networks with gradient boosting methods detect hidden non-linear relationships which standard models cannot identify.

The assessment methods SHAP and LIME reveal to users which market conditions together with property features are most influential for predicted house prices. Real estate valuations require economic indicators together with location-based variables as per the research findings.

5.1 This Research Study Acquires Multiple Practical Advantages

AI-powered real estate websites help users receive instant property price assessments.

The adoption of Automated Valuation Models (AVMs) represents a banking practice for mortgage evaluation purposes.

Urban planning and housing policy research at government institutions uses AI-generated evaluation results.

AI-based predictions of house prices become more valid when developers integrate economic metrics along with housing review sentiments and blockchain transaction database authentication.

6 CONCLUSIONS

This study develops AI alongside machine learning tools to boost predictions in house pricing values. The article demonstrates how predictive models offer better real estate valuation than traditional methods using sophisticated algorithms. The study confirms that economic factors and property features along with geographic determinants control housing market values.

6.1 Future Studies Should Aim to

AI Interpretability should be enhanced through model development which creates easily understandable valuation systems.

The research investigates blockchain decentralization as a method to improve data safety alongside reliability in storage.

Real-time sentiment analysis of public market commentary and properties listings data is possible through NLP applications.

The system uses AI models implementing dynamic price adjustments that operate based on moving economic and market dynamics.

The continuous AI development leads to accurate and affordable property valuation systems that improve service delivery to various real estate industry members.

REFERENCES

Ahmad, T., & Khan, M. (2022). "Artificial Intelligence in Real Estate - Predictive Analytics and Machine-Learning Applications." Journal of Property Research, 39(2), 112-130.

Chen, Y., Lin, Z., & Zhang, Q. (2021). "Machine-Learning for House Prices Prediction: An Empirical Study." International Journal of Housing Markets and Analysis, 14(3), 221-237.

Doshi, J., Ghosh, S., & Ray, P. (2020). "Comparative Analysis of Regression and Deep-Learning-Models for Real Estate Valuation." IEEE Transactions on Artificial Intelligence, 9(1), 55-68.

Fan, C., Li, Y., & Wu, H. (2019). "Spatial and Economic Factors in Housing Price Estimation Using AI Models." Urban Studies Journal, 56(5), 902-921.

Han, J., & Lee, D. (2021). The process of feature engineering for real estate price forecasting depends heavily on macroeconomic indicators according to Han Jia and Lee Donghui. Journal of Big Data and Society, 18(4), 345-360.

Kumar, S., & Patel, R. (2020). "Blockchain and AI Integration for Transparent Real Estate Transactions." Journal of Emerging Technologies, 29(3), 178-196.

Li, X., Wang, P., & Zhang, W. (2022). "Deep Neural Networks for Prediction of Real Estate Prices: The Case Study in Urban Housing Markets." Neural Computing and Applications, 34(6), 1198-1214.

Mishra, R., & Gupta, K. (2023). "The Role of Explainable AI in Property Valuation: A Case Study of SHAP and LIME Methods." Artificial Intelligence Review, 45(7), 621-645.

Nguyen, T., & Tran, H. (2020). "Impact of Location-Based Attributes on House Price Predictions Using Machine Learning." Computers, Environment and Urban Systems, 47(3), 287-301.

Singh, P., & Verma, D. (2021). "Real Estate Price Forecasting with XGBoost and Ensemble Learning Techniques." IEEE Access, 11(3), 11125-11139.