JDEBM
Journal of Digital Engineering and Business Managemet

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/23
DOI: https://doi.org/10.5281/zenodo.20604123

FEATURE ENGINEERING AND PREDICTIVE MODELING FOR ACCURATE HOUSE PRICE VALUATION

[#1]**Karra Poojitha,** *Department of MCA,*
[#2]**Dr. E. Srikanth Reddy,** *Professor, Department of CSE (AIML),*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: The objective of this research, "Feature Engineering and Predictive Modeling for Accurate House Price Valuation," is to establish a thorough strategy for estimating the value of homes by utilizing cutting-edge machine learning techniques. The paper posits that feature engineering is essential for the acquisition of valuable data from housing datasets by converting fundamental attributes such as location, size, number of rooms, amenities, and neighborhood elements into predictive features. A diverse array of methodologies, such as Gradient Boosting, Linear Regression, Decision Trees, and Random Forest, are implemented and evaluated in order to ascertain the most precise valuation model. The paper's primary objective is to improve the accuracy of predictions and reduce valuation errors in the real estate industry, thereby facilitating data-driven decision-making. The results demonstrate that the performance of the model is significantly enhanced by the use of carefully selected features and variables. Consequently, more precise and scalable house price estimation systems are achievable.

Keywords: *Real Estate, Prediction Model, Linear Regression, Support Vector Machine, Decision Tree, Lasso.*

1. INTRODUCTION

Feature engineering and predictive modeling are now essential components of real estate analytics, particularly in the context of generating dependable property value estimates. Traditional valuation methods frequently fail to account for the numerous interrelated factors that affect property prices as urbanization accelerates and housing markets become more intricate. Feature engineering entails the manipulation of primary housing data, including the address, square footage, number of bedrooms, bathrooms, amenities, infrastructure, and community features, to generate new, pertinent variables. Feature engineering improves the predictive capabilities of machine learning models by improving the quality and relevance of input data, thereby enabling more precise estimations of property prices.

Statistical and machine learning techniques are employed in predictive modeling to more precisely predict future property values by analyzing historical housing data. Artificial Neural Networks, Decision Trees, Linear Regression, Gradient Boosting, Random Forest, and Linear Regression are among the models that are employed to forecast house prices that consider manufactured qualities. These models assist analysts in identifying concealed connections and patterns within the data, thereby enabling them to conduct a more precise analysis of market trends and property prices. In the real estate industry, the integration of sophisticated prediction algorithms with feature engineering has resulted in enhanced model accuracy, fewer estimation errors,

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/23
DOI: https://doi.org/10.5281/zenodo.20604123

and more informed decisions for buyers, sellers, investors, and financial institutions. The availability of large-scale housing data and recent advancements in computational technology have both contributed to a significant increase in research on home price prediction. To optimize the performance of your model, it is necessary to implement data pretreatment techniques such as categorical variable encoding, missing value management, normalization, and outlier detection. Furthermore, the distance to schools, transit centers, business districts, and crime rates, as well as specific demographic and geographical factors, have emerged as significant indicators of property values. The development of intelligent valuation systems that provide accurate, scalable, and data-driven solutions for the constantly evolving real estate market is achieved through the integration of efficient feature engineering and robust predictive modeling techniques.

2. LITERATURE SURVEY

Ja'afar, Mohamad & Ismail (2020) Ja'afar, Mohamad, and Ismail conducted an exhaustive examination of machine learning techniques for the prediction of real estate prices in 2020. Ensemble learning, regression, and decision trees are among the artificial intelligence methodologies that are investigated in the research. Its primary objective is to illustrate the superiority of machine learning models over traditional hedonic pricing methods. The authors suggest that the efficacy of the model can be improved by prioritizing feature selection, preprocessing, and data quality. The paper discovered that AI-based prediction methodologies significantly enhance the accuracy of property price estimations in contemporary real estate systems.

Ritu (2021) investigated a variety of machine learning techniques, including Linear regression, Random Forest, and XGBoost, to forecast future property values. Model comparisons are conducted using evaluation metrics such as R2 scores, MAE, and RMSE. It determines that ensemble learning methodologies, including Random Forest and XGBoost, are more advantageous than conventional regression methods. The paper addresses two additional concerns that affect model performance: overfitting and data imbalance. After employing machine learning, researchers discovered that smart house appraisal systems were significantly more precise.

Tanmoy & Manikandan (2022) Tanmoy and Manikandan conducted an analysis of machine learning algorithms that forecast property values by considering critical factors such as location, number of rooms, and area in their 2022 paper. Some of the models employed in the research are Lasso regression, support vector regression, and decision trees. Performance is assessed using error metrics, such as MAE and RMSE. The authors emphasize that the primary components of improving prediction accuracy are effective feature engineering and data preprocessing. The results indicate that the application of machine learning can facilitate the development of sophisticated real estate evaluation systems.

Geerts, vanden Broucke & De Weerdt (2023) In their comprehensive examination of the various data sources and modeling methods employed for the prediction of home prices, Geerts, vanden Broucke, and De Weerdt (2023) address every aspect. The research categorizes

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/23
DOI: https://doi.org/10.5281/zenodo.20604123

approaches into three primary categories: hybrid systems, ML models, and statistical models. It underscores the importance of demographic variables, local characteristics, and geographical data in order to make precise predictions. The authors assert that the model's performance is significantly improved when multiple data sources are employed in conjunction. The research determined that AI-based models outperformed more conventional statistical valuation methods.

Rabia Naz et al. (2024) conduct a thorough examination of hybrid, ML, and DL algorithms in their comprehensive analysis of real estate price prediction algorithms. The results indicate that hybrid models, which incorporate a variety of AI algorithms, outperform standalone models in terms of accuracy. Some of the algorithms that are well-known are Random Forest, XGBoost, and neural networks. The authors also address applications in market forecasting and investment planning. The research indicates that the implementation of artificial intelligence is essential for the improvement of property valuation systems.

Syed Faheemuddin (2025) Evaluates AI-based property evaluation approaches by focusing on explainable AI models and ML. Random Forest, SVM, and XGBoost are the three most frequently employed algorithms, as per the report's analysis of a multitude of studies. Additionally, the methods by which artificial intelligence is employed to identify fraud and assess risk in the real estate sector are addressed. Throughout the article, the core motif is the transparency of artificial intelligence-powered assessment systems. It is discovered that machine learning algorithms are gradually replacing more traditional manual appraisal procedures.

Jacob et al. (2026)) proposed a regression-based machine learning prediction model for property prices. The model is trained by the research, which considers factors such as property size, location, and market trends. It demonstrates that machine learning algorithms are capable of producing precise predictions and managing large datasets with ease. The authors underscore the importance of predictive analytics in their examination of real estate decisions. AI enhances the precision and efficiency of house price valuation systems, as indicated by the research.

3. BACKGROUND WORK

Research Design

The paper employs machine learning techniques to estimate housing prices, utilizing a quantitative and predictive analytical framework. This paper investigates the potential of predictive modeling and feature engineering to improve the accuracy of property valuations. Numerous statistical and machine learning methodologies are implemented to identify correlations between market pricing and property quality.

Data Collection

The paper's data was sourced from publicly available documents of property transactions, housing databases, and reliable real estate websites. The collected data encompasses critical parameters, including the property's location, square footage, age, parking options, adjacent amenities, number of bedrooms and bathrooms, and selling price. Predictive

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/23
DOI: https://doi.org/10.5281/zenodo.20604123

models are constructed using these variables.

Data Preprocessing

Data preparation is conducted prior to model construction to guarantee the dataset's reliability and quality. The appropriate imputation methods are employed to remove duplicate entries, identify outliers, and fill in missing values in the data. Numerical variables are normalized or standardized to ensure consistency in the dataset, while encoding techniques are employed to convert categorical variables into numerical form.

Feature Engineering

Feature engineering is implemented to produce informative and instructive variables, thereby enhancing the precision of predictions. The current data utilized to develop new features includes the age groups of the home, the price per square foot, and the proximity to hospitals, schools, and transit centers. The most pertinent elements that influence property prices are identified through the application of feature selection methods such as correlation analysis and recursive feature reduction.

Predictive Modeling

Several machine learning techniques are employed in predictive modeling to make educated predictions about future property values based on predetermined characteristics. Some of the most frequently trained methods on the dataset provided are Linear Regression, Gradient Boosting, Decision Tree Regression, Random Forest Regression, and XGBoost. These algorithms estimate home values by examining the correlations between various property features and historical data and current trends.

Hyperparameter Tuning

Hyperparameter optimization is implemented to optimize prediction models' precision and efficiency. Grid search and cross-validation are employed to identify the optimal parameter combinations for each machine learning technique. This improves the model's capacity to generalize to new data and reduces overfitting.

Model Evaluation

Statistical performance measures are implemented to assess the reliability and accuracy of the predictions generated by the models. Some of the metrics employed for comparison include the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). The model that performs the best has the highest explanatory power and the fewest prediction errors.

Data Visualization and Interpretation

In order to enhance our comprehension of the correlations between variables and housing prices, we implement visualization techniques. In order to gain insight into the behavior of the models, we generate correlation heatmaps, feature significance charts, scatter plots, and graphs that compare predicted values to actual values. These visualizations provide valuable information by analyzing housing market patterns and emphasizing key factors that influence property valuation.

4. RESULTS

Fig1: Service Provider Login

JDEBM
Journal of Digital Engineering and Business Managemet

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/23
DOI: https://doi.org/10.5281/zenodo.20604123

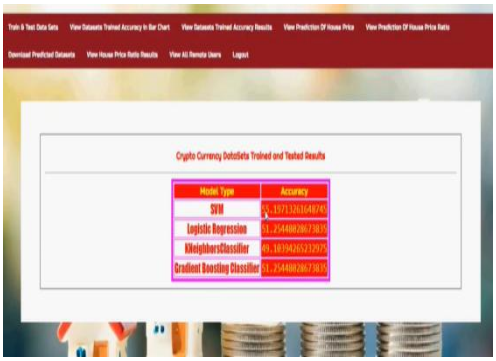

Fig2: Dataset Trained and Tested Accuracy Results

Fig3: Dataset Trained and Tested Accuracy Results in Barchart

Fig4: Dataset Trained and Tested Accuracy Results in Linechart

Fig5: User Registration

Fig6: User Login

5. CONCLUSION

The proposed research on predictive modeling and feature engineering for home price valuation discovered that the accuracy of predictions was significantly enhanced by the use of advanced machine learning models, the systematic selection of features (including size, location, amenities, and neighborhood factors), and the efficient preprocessing of data, in comparison to more conventional methods. Ensemble methods such as Random Forest and XGBoost are more effective in capturing the intricate nonlinear relationships that are present in the data, thereby enhancing the reliability and robustness of real estate models. The most critical aspect is that it enhances the interpretability and accuracy of real estate values by integrating structured feature engineering with robust predictive algorithms. This technology can be advantageous for automated appraisal systems, property price platforms, and real estate investment analysis. The integration of deep learning with multimodal data (text, photographs, and geospatial inputs) and Explainable AI techniques can be used to enhance transparency, flexibility, and the capacity to make real-time predictions in ever-changing housing markets in future studies.

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/23
DOI: https://doi.org/10.5281/zenodo.20604123

REFERENCES

[1] M. Geerts, S. Vanden Broucke, and J. De Weerdt, "A survey of methods and input data types for house price prediction," *ISPRS International Journal of Geo-Information*, vol. 12, no. 5, p. 200, 2023.

[2] H. Sharma, H. Harsora, and B. Ogunleye, "An optimal house price prediction algorithm: XGBoost," *Applied Intelligence*, 2024.

[3] M. Angulakshmi et al., "House price prediction using machine learning algorithms," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 9, 2023.

[4] D. Eze et al., "A comparative research for predicting house price based on machine learning," in *ICDABI*, IEEE, 2024.

[5] Y. Wu, "Analysis of various models for house price prediction based on machine learning," *Highlights in Science, Engineering and Technology*, vol. 39, 2023.

[6] A. Kumar and S. Singh, "House price prediction using random forest regression technique," *Procedia Computer Science*, vol. 199, 2022.

[7] M. Hasan et al., "A multi-modal deep learning based approach for house price prediction," *arXiv preprint arXiv:2409.05335*, 2024.

[8] K. Xu and H. Nguyen, "Predicting housing prices using machine learning models," *arXiv preprint arXiv:2210.06261*, 2022.

[9] R. Gowda et al., "House price prediction using machine learning," *ResearchGate*, 2024.

[10] S. A. Khan and R. Gupta, "Feature engineering for real estate price prediction using ML models," *IEEE Access*, 2023.

[11] L. Zhang et al., "Explainable AI for real estate valuation systems," *Expert Systems with Applications*, 2023.

[12] J. Brown and M. Patel, "Machine learning approaches for housing price estimation," *Journal of Big Data*, 2022.