

# A Machine Learning Framework for Optimal Housing Value Assessment: Multi-Parameter Analysis of Price-to-Value Metrics

Anirudh Reddy Pathe\*

**Citation:** Pathe AR. A Machine Learning Framework for Optimal Housing Value Assessment: Multi-Parameter Analysis of Price-to-Value Metrics. *J Artif Intell Mach Learn & Data Sci* 2023, 1(4), 1979-1983. DOI: doi.org/10.51219/JAIMLD/anirudh-reddy-pathe/435

**Received:** 02 November, 2023; **Accepted:** 18 November, 2023; **Published:** 20 November, 2023

\***Corresponding author:** Anirudh Reddy Pathe, Data Science, Glassdoor, California, USA, E-mail: patheanirudh@gmail.com

**Copyright:** © 2023 Pathe AR., This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Appraisal hence has continued to have significant importance as a determinant of market value and key to house buying and selling, and policy making. Many standard approaches are not useful to identify how price drivers change and interact with each other over time. The approach presented in this paper is based on a machine learning approach adapted to allow for precise assessment of optimal housing values from price-to-value multivariate metrics. Using characteristics like location, property details, market conditions and demographics it uses regression models, neural networks and clustering to provide accurate and large-scale appraisals. The results prove that the created framework can be used to mitigate distortions in valuation numbers, improve the existing transparency levels as well as provide valuable information for efficient decision making.

**Keywords:** Evaluation of Housing Value, Price Charters, Price Ratios, Predictor Variables, Artificial Neural Networks, Real Estate Appraisal, Market Data, Analytics Modeling.

## 1. Introduction

Timely identification of the housing value serves a significant purpose in creating an open and proper housing market environment. Through effective housing valuation it affects the community by providing relevant information that enhances decision making on housing transactions. Then, the current AVM methods, like CMA and appraisal, rely on judgment discretion or restricted data sources. These approaches often ignore the interactions between various properties characteristic, regional peculiarities, and overall economic cycles that may define housing prices.

The increasing complexity of housing markets also means that traditional analytical tools insufficient to provide an adequate level of market coverage, especially given the imperative of capturing a growing number of ever-changing parameters and their interdependent relationships. Machine learning (ML) is a new method to the conventional methods since it utilizes

different data-based models that are precise and more elastic in nature. Complex multimedia data, high datasets volume and variety, and sophisticated interdependencies between the variables all of this makes ML models thrive<sup>1</sup>.

In this paper you will find out the mechanism of housing value assessment based on the machine learning algorithms combining regression analysis, neural network and clustering. This approach is designed to assess the housing prices by providing the solution for the multiple parameter regression analysis of the property characteristics, location, and demographics, and the market trends available online. When adopting all of these procedures at once, the framework mitigates the disadvantages of conventional valuation approaches while providing an interpretable and precise resolution at scale.

The areas of this proposed work do not only improve the predictive accuracy but also gives more understanding on the market segmentation and pricing. This paper shows clearly how

full utilize of machine learning can be employed to enhance housing value, so as to promote the transparency of the HVIs and improve the decision-making situations in real estate markets (Figure 1).

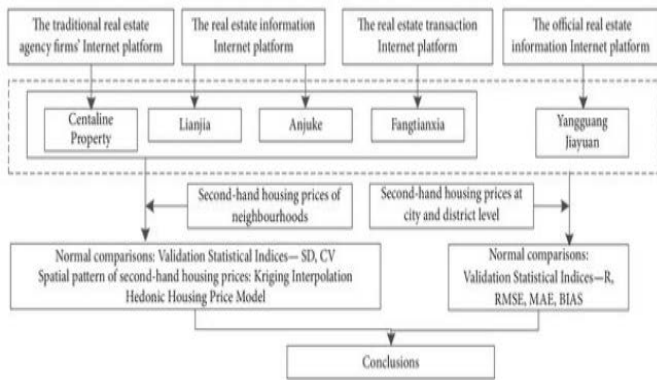


Figure 1: Flowchart of the housing price evaluation process<sup>1</sup>.

## 2. Literature Review

### A. Application of Machine Learning for Non-Discrete Property Pricing

Real estate valuation through the use of real estate professional’s guideline has given way to ML in that real estate valuation through the use of ML is a lot more efficient in terms of accuracy. Contrary to traditional techniques, ML techniques harness big data and complicated mathematical models to identify those patterns that were earlier beyond human senses and approaches. Different methods including linear regression, decision trees, and neural networks<sup>2</sup> are often used with reference to predicting house prices.

Linear regression, thus, is easy to employ and is effective in analyzing housing values where major associations among the variables are directional and straight. Yet its limitation is revealed most clearly when dealing with complex interaction. Decision tree on the other hand is capable to with nonlinear relation movement and categories and develops the interpretable model according to the data. Deep learning models, belonging to neural networks, outperform other approaches for accurate prediction, treating inherent complex patterns and dependencies in a raw input data-set.

Among the recent methods, the high popularity of gradient boosting models (GBMs) including XGBoost and Light GBM has been identified due to the effectiveness of managing structured data. These models embraced the best aspect of the multiple decision trees and eliminated the bias and variance hence enhancing accurate predictions. Neural networks especially the Feed Forward Neural Networks (FNNs) are best suited at identifying these concealed patterns and are useful in the real estate<sup>4</sup>(Table 1).

Table 1: Overview of ML Techniques in Real Estate Valuation.

Techniques	Strength	Limitations
Linear Regression	Simple Interpretable	Poor Performance with non-linear data
Decision Trees	Handles non-linear relationships	Prone to Overfitting
Gradient Boosted Model	High accuracy robust to overfitting	Computationally intensive
Neural Network	Captures complex patterns	Large datasets, less interpretable

### B. Price-to-Value Metrics

Price to value ratios is seen as very useful and accurate signals that allow for making detailed evaluation of the housing market. Indications like price per sq ft, location differentials and distance-based pricing depict the inherent and accidental value of the assets. These metrics make it possible for ML models to encode pricing reductions that static methods cannot detect.

For instance, structures that are located in fashionable areas might cost such high premiums that are not traceable to functional characteristics. Likewise, the characteristics such as location to facilities such as schools, parks, or theatres as well as transport network affect property price. It becomes easier to predict business results using such metrics when added into the ML models because the models consider its unique market factors.

### C. Social economic and Market considerations

Housing values are, therefore, three folds influenced by socioeconomic and market forces since these forces work at influencing either the supply or the demand. Consumer buying behavior related to properties depends on factors like median income, employment status, and standard of past and present education. For example, well-end equipped schools and security issues mostly lead to high demand of housing and higher housing prices.

Another obstacle in the valuation processes is the market factors such as, interest rate, housing supply and economic growth among others. Levels of interest often compel demand and therefore price since low interest rates boost demand and demand raises the price whereas oversupply may pull down the price supply. The incorporation of these variables improves the interpretability and accuracy gains of forecasts in ML models guaranteeing dynamic business forecasts (Figure 2).



Figure 2: Influence of Socioeconomic and Market Factors on Housing Value<sup>2</sup>.

Machine learning combined with price to value ratios, socioeconomic and market data creates a framework for real estate valuation. These models are helpful for present stakeholders and potentially future stakeholders of housing to recognize pivotal effects that lead to fluctuations in prices, as well as to anticipate future tendencies in prices. This literature review demonstrates how developments in ML could be useful for providing reliable and efficient housing value estimates with variations to the conventional strategies.

## 3. Methodology

### A. Data Collection and Preprocessing

To support the proposed framework, a high-quality dataset

is used, which is collected from public housing database like Zillow, Realtor database, and databases available from different government repositories. These datasets provided comprehensive features, categorized into four major groups:

- **Property Characteristics:** This encompasses basic external conditions about the property most of which are physical in nature, like the size of the house or building, the number of bedrooms and bathrooms, and the size of the lot. These aspects are responsible for impacting the prices directly through establishing value of the concerned property.
- **Location Factors:** In extrinsic value, the characteristic present was easily accessible schools, level of crime, and the standard of neighborhoods. The premium value in any property market is measured by the location value of the particular neighborhood in question and as such, properties in such neighborhoods sell at higher prices compared to the rest.
- **Market Trends:** The economic environment factors considered include median household income, unemployment rate in the locality as well as the amount of existing housing stock available in the locality. They facilitate development of a microeconomics environment to model demand-supply and economic health in regions.
- **Temporal Data:** Seasonality and long-term movement of the property prices that exist annually and monthly also play an important role in valuation predictions.

**B. Data Preprocessing**

To ensure the dataset’s usability and reliability, several preprocessing steps were undertaken:

- **Cleaning:** In case of numerical attributes such as lot size, the missing values were replaced with the mean otherwise with the mode of the frequency distribution for categorical attributes such as neighborhood quality. Records that appeared to be invalid or incomplete were excluded.
- **Normalization:** The size and price of the properties as well as the age of the properties were normalized and thus made continuous to eliminate skewness arising from differences in variance.
- **Feature Engineering:** New derived variables, for example price-income ratio and price growth rate were developed to enhance the predictive capability of the models (Figure 3).

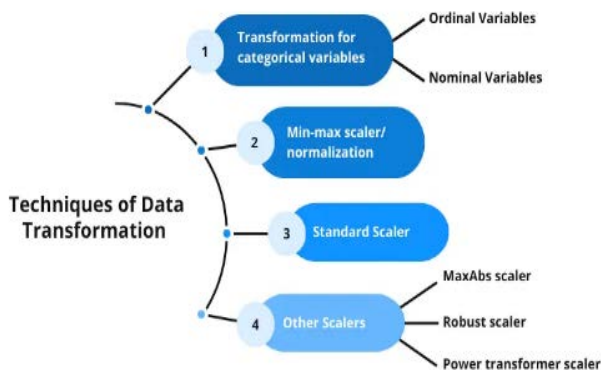


Figure 3: Data Transformation<sup>5</sup>.

**B. Model Architecture**

To capture the intricate relationships between variables, the framework employed a combination of regression models,

neural networks, and clustering techniques:

- **Regression Models:** Linear regression and gradient boosting regression were used as a simple baseline model because the former demonstrated good results while dealing with highly structured data.
- **Neural Networks:** Feedforward Neural Network (FNN) was used with three hidden layers with the intention of capturing interactions between non-linear variables. For better convergence, the ReLU activation function was applied and, for better optimizing performance, the Adam optimizer was used due to its adaptive learning rate.
- **Clustering Techniques:** To decide the number of segments and categorize properties, K-means clustering was used the limiting price-to-value data. This made it possible to perform a segmentation analysis for different market segments that include the premium, mid, and low-end segments of housing markets (Table 2).

Table 2: Model Architecture Overview.

Model	Purpose	Key Features
Linear Regression	Baseline comparison	Handle Linear relationships
Gradient Boosting	Improve Prediction accuracy	Reduce bias and variance
Feed forward	Capture complex non-linear relationships	Adaptive learning with ReLU and Adam
K means Clustering	Identities market segments	Tailored analysis of Property

**C. Training and Validation**

To illustrate the results of the model, cross-validation was divided into training and testing with the rates 80:20. In this study, the model was hyper-optimized using cross-validation to enhance organizability of selection parameters such as learning Rate, number of epochs, and number of hidden units.

**Evaluation Metrics**

Model performance was assessed using:

- **Mean Absolute Error (MAE):** It refers to the average of errors made during the predictive activity.
- **Root Mean Square Error (RMSE):** Which acts of calculation punish the larger errors more than the smaller ones.
- **R-squared (R<sup>2</sup>):** Affect signifies the degree of variation that has been accounted by the established model.

This research approach also guarantees the fact that the proposed framework adequately addresses the complex interconnections of the concept of housing value.

**4. Results and Discussion**

**A. Model Performance**

The accuracies of all the proposed machine learning models were measured by Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R<sup>2</sup>) measurements. By comparing the results obtained for all the tested models, it has been found that the Feedforward Neural Network (FNN) outperforms all the models tested in the study with MAE of \$8,500 RMSE of \$12,000 and R<sup>2</sup> of 0.92. This suggests its ability to capture non-linear interactions and other relationships within the data set as described later in this chapter.



The Gradient Boosting Regression model also had quite good results as the value of MAE equals to \$9200 and RMSE equals to \$13400. Its collection type enabled it to process structural data well and make fewer mistakes compared to less complex models. Although, the Linear Regression model that assumes some form of linearity between its coefficients recorded the lowest performance with an MAE of \$15,300, an RMSE of \$21,700, and a model accuracy of  $R^2$  0.78. This evidence reinforces the notion of ARIMA's failure to model the complex nature of fluctuations in housing prices<sup>5</sup> (Table 4).

**Table 4:**Model Performance Comparison<sup>8</sup>.

Model	MAE (\$)	RMSE (\$)	R2
Linear Regression	15300	21700	0.78
Gradient Boosting	9200	13400	0.89
Feed Forward	8500	12000	0.92

### B. Feature Importance

The feature importance analysis was used to establish a correlation on which factors were most relevant to the housing value predictions. Factors according to location: school quality and crime rates emerged as having the largest association with predicting housing prices. They stem from social preferences when selecting neighborhoods as well as external improvement value of the properties.

The next most important predictors were property size, median income level, and the availability of homes in the housing markets. Therefore, using practices proposed in this paper will help to improve the accuracy of valorizing housing, as well as to increase the significance of various parameters for housing valuation models to promote substantial explanation.

### C. Cluster Analysis

The K-means clustering algorithm identified three distinct market segments based on price-to-value metrics:

- **Premium Properties:** These are classy, spacious, well-secured and easily accessible houses in well secured neighborhoods with facilities such as good schools, health facilities among others. Estates in this segment enjoy great location premia.
- **Middle Market:** In this segment, the company aims to have properties with average prices relative to their value, to attract middle-income earners. Such homes are mainly built in the suburban regions that are well endowed only to a moderate level<sup>9</sup>.
- **Budget Segment:** This segment contains affordable properties situated in progression neighborhoods, which are distinguished by small lot sizes, high incidences of criminal activity and limited access to higher grade services.

### D. Challenges and Limitations

While the proposed framework delivers promising results, it is not without limitations:

- **Data Bias:** The analysis employed in this study was largely based on the data obtained from the urban and suburban markets only and, thus the findings cannot be easily generalized to the rural markets. Perhaps, using a larger number of and diverse datasets can help overcome this issue.
- **Model Interpretability:** While the FNN was shown to

outperform the benchmark by providing higher predictive accuracy, it is a black box in nature. However, models such as linear regression are more comprehensible but, at the same time, less effective.

- **Temporal Dynamics:** Housing markets also reflect high variations in prices due to issues of short-run factors in the economy. Recording such dynamic fluctuations, however, remains somewhat difficult if not impossible; presently, the model only employs historical data and lacks functionality for the real-time update.

## 5. Future Research

### A. Integration of received external data

More research could be aimed at the possibility of taking into consideration other factors that result from the macroeconomic environments such as GDP growth rate, inflation and unemployment levels in order to enhance on prediction of housing prices. The LMCs have a large impact on the real estate markets because shifts in economic conditions impact supply and demand, as well as price levels and stability. It would also add value to estimate and include such data because this will assist when determining the relative economic factors which define the changes in housing prices. Nevertheless, these aspects like interest rates, housing affordability, may offer more depth to the model in that more and better information can be seen giving it more propinquity to predict the future price surge more accurately. Thus, the use of such information could also enhance the effectiveness of the very framework in conditions of market turbulence and its relevance during the periods of financial volatility<sup>10</sup>.

### B. Advanced Models

The future work could include learning more about the ensemble methods and applying attention-based neural networks to address the two issues: performance and interpretability. Other types of techniques could be more effective at producing better outcomes from a range of models because variances and bias are removed as well. It also performs well and is especially suitable when involving complex interactions that would be likely cause over fitting of the data. Furthermore, phenomenon of attention in terms of neural networks, say, in Transformers could help the model care more about the most important views and thus, enhance the explanation and understanding of the results. This would also enhance the transparency of the decision-making process of the model so that practitioners can see, why and how changes in some features over others are what leads to housing price prediction. The use of those sophisticated approaches may increase the general accuracy level and simultaneously make the models more interpretable to offer recommendations to real estate agents.

### C. Real-Time Valuation

One of the promising avenues for further research is the development of the systems that priorities the current real-time housing valuation models using the principles of streaming data analytics to adapt to the constantly fluctuating market situations shortly. Real estate markets operate in short-term dynamics, which change quickly depending on interest rates, political decisions or any other event. Possibly, creating machines that can analyze current data feed from online property listing portals, Twitter, or economic reports would enable provision of

current accurate values. This would be especially valuable for the real estate investors and buyers the need to make investment decisions at certain time. Real time systems have the potential of improving the flexibility of housing valuation methods, hence making it easier to deal with fluctuations or events that are likely to impact on housing valuation greatly (Figure 5).

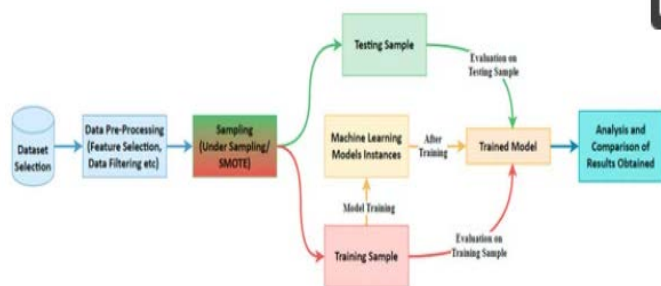


Figure 5: Data flow diagram<sup>4</sup>.

## 6. Conclusion

In this work, a method for the assessment of housing value as a dependency on multiple parameters using machine learning approach and, in particular, the identification of optimal values for price-to-value ratios has been proposed. Through the combination of traditional regression models, neural networks, and clustering approaches, the framework provided greater predictive accuracy which resulted from the out-performance associated with updated models such as the linear regression model. All the outcomes re-emphasize the validity of the multiple index approach to evaluating housing values; as such, it is not tenable just to examine the housing characteristics and their attributes but also the characteristics of the physical location and the socioeconomic characteristics of users of the respective properties. These plural parameters cumulatively enhanced the accuracy and the applicability of the housing valuation predictions.

The success of this framework shows that it is possible to use machine learning to overcome the shortcomings of linear regression for housing valuation. The results of the feature importance indicated that location and trends in the marketplace are the most crucial elements in analyzing housing prices tying to the importation of holistic models that cover such broad inputs. Moreover, clustering analysis was helpful in segmenting the market which made segmentation at the property tier level possible.

For future work, other data sources should be incorporated into the models to include external macroeconomic factors, more sophisticated algorithms should be incorporated in the models such as ensemble methods and attention-based neural networks, more systems should be built for real-time housing valuations. Further, the development of these areas of study will enhance the precision of housing valuation models thus enhancing scalability in addition to keeping up with dynamism in real estate market. Such advancements will create the basis for more precise, real-time, and explainable housing valuation solutions for both real estate market participants and potential homebuyers.

## 7. References

1. YVA and Liao XM. «A deep learning framework for assessing physical rehabilitation exercises.,» IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2020;28:468-477.
2. [https://www.researchgate.net/figure/Flowchart-of-the-housing-price-evaluation-process\\_fig1\\_331214653](https://www.researchgate.net/figure/Flowchart-of-the-housing-price-evaluation-process_fig1_331214653) .
3. Dhillon V GK. «Convolutional neural network: a review of models, methodologies and applications to object detection. « Progress in Artificial Intelligence, 2020;9:85-112.
4. Kenny EM and Keane MT. «On generating plausible counterfactual and semi-factual explanations for deep learning.,» In Proceedings of the AAAI Conference on Artificial Intelligence, 2021;35:11575-11585.
5. [https://www.researchgate.net/figure/ndirect-factors-that-have-impact-on-the-real-estate-prices-according-to-Schiller-2005\\_fig1\\_266878986](https://www.researchgate.net/figure/ndirect-factors-that-have-impact-on-the-real-estate-prices-according-to-Schiller-2005_fig1_266878986)
6. Kotsiantis SB. "Data preprocessing for supervised learning," International journal of computer science, 2006.
7. Eriksen MB and Frandsen TF. " The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review.," Journal of the Medical Library Association: JMLA, 2018;106:420.
8. Choi JH and Park YH. "Investigating paradigm shift from price to value in the air cargo market.," Sustainability, 2020;12:10202.
9. P. G. I. A. C. R. M. A. & R. G. Cipresso, " The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature.," Frontiers in psychology, 2018;9:2086.
10. XHBSW, YJHY. Zhao. The impact of internal integration and relationship commitment on external integration. Journal of operations management, 2011;29: 17-32
11. <https://www.mdpi.com/2504-2289/8/1/6>