



## Real Estate Price Prediction Using ML: a Survey-Based Study

---

Yash Chothani, Jay Lathiya, Param Jadav and Darshak Dhangan

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

March 18, 2024

# REAL ESTATE PRICE PREDICTION USING ML: A Survey-Based Study

1<sup>st</sup> CHOTHANI YASH

*Department of Computer Science and Engineering  
Parul University  
Vadodara, India  
200303105011@paruluniversity.ac.in*

2<sup>nd</sup> JAY LATHIYA

*Department of Computer Science and Engineering  
Parul University  
Vadodara, India  
200303105030@paruluniversity.ac.in*

3<sup>rd</sup> PARAMSINGH JADHAV

*Department of Computer Science and Engineering  
Parul University  
Vadodara, India  
200303105020@paruluniversity.ac.in*

4<sup>th</sup> DHANGAN DARSHAK

*Department of Computer Science and Engineering  
Parul University  
Vadodara, India  
200303105013@paruluniversity.ac.in*

**Abstract**—This study proposes a performance comparison between machine learning regression algorithms and Artificial Neural Network (ANN). The regression algorithms used in this study are Multiple linear, Least Absolute Selection Operator (Lasso), Ridge, Random Forest. Moreover, this study attempts to analyse of the correlation between variables to determine the most important factors that affect real estate prices and house prices. There are two datasets used in this study which called public and local which contain real estate prices . The accuracy of the prediction is evaluated by checking the root square and root mean square error scores of the training model. The test is performed after applying the required pre-processing methods and splitting the data into two parts. However, one part will be used in the training and the other in the test phase. We have also presented a binning strategy that improved the accuracy of the models.

**Index Terms**—Multiple linear regression, Lasso Regression, Ridge Regression, Random Forest Regression, Artificial Neural Network, Machine Learning, Real Estate Price Prediction

## I. INTRODUCTION

The real estate market stands as a pivotal sector in the global economy, reflecting the complexities of various interconnected factors such as economic conditions, demographic trends, and geopolitical influences. The ability to accurately predict real estate prices is of paramount importance for investors, homeowners, and policymakers alike, as it empowers informed decision-making, risk management, and strategic planning. In recent years, the integration of advanced computational techniques, particularly in the realms of machine learning and data science, has reshaped the landscape of real estate price prediction. This survey aims to provide a comprehensive exploration of the methodologies, challenges, and advancements in the field of real estate price prediction. By synthesizing existing literature and surveying the diverse range of models employed, we seek to unravel the intricate tapestry that characterizes the intersection of real estate and predictive analytics. The significance of understanding real estate price prediction methodologies cannot be overstated, given their potential

impact on various stakeholders. Investors can optimize their portfolios, homeowners can make informed decisions about their properties, and policymakers can devise strategies to address housing market dynamics. This survey delves into the historical evolution of real estate price prediction, reviews the array of methodologies employed, examines the challenges faced, and contemplates the future directions of this dynamic field.

## II. LITERATURE REVIEW

### A. Historical Perspective of Real Estate Price Prediction

The roots of real estate price prediction can be traced back to traditional statistical methods such as regression analysis, which laid the foundation for understanding the relationship between various predictors and property values. Early studies primarily focused on macroeconomic indicators and demographic variables, striving to establish causal links with real estate prices. Over time, this paradigm evolved with the integration of more sophisticated statistical techniques, leading to a nuanced understanding of the multifaceted factors influencing property values.

### B. Ensemble Learning Methods

Ensemble learning methods, such as random forests, gradient boosting machines (GBM), and XGBoost, have gained popularity in real estate price prediction due to their ability to capture nonlinear relationships and handle large datasets effectively. Random forests, in particular, have been shown to perform well in this domain, leveraging the power of decision trees to make accurate predictions.

### C. Evolution of Methodologies and Techniques

The advent of machine learning marked a paradigm shift in real estate price prediction. Traditional linear regression models gave way to more flexible and powerful algorithms. Decision trees and random forests emerged as popular choices

due to their ability to capture nonlinear relationships. Furthermore, the rise of deep learning, particularly neural networks, introduced a new era of predictive modeling, enabling the extraction of intricate patterns from vast datasets. This section will delve into the key milestones and innovations that have shaped the evolution of methodologies in real estate price prediction.

#### D. Deep Learning Approaches

In recent years, deep learning techniques, especially neural networks, have emerged as powerful tools for real estate price prediction. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants have been applied to learn intricate patterns from raw real estate data, such as images, text descriptions, and time-series data. These models often outperform traditional methods, but they require large amounts of data and computational resources for training.

#### E. Key Challenges in Real Estate Price Prediction

While significant strides have been made, real estate price prediction is not without its challenges. This subsection will explore the common obstacles faced by researchers and practitioners, including issues related to data quality, model interpretability, and the dynamic nature of real estate markets. Understanding these challenges is essential for developing robust and reliable prediction models and will provide insights into areas where further research is needed.

### III. METHODOLOGY

The Methodology represents a description about the framework that is undertaken. It comprises various milestones that should be accomplished in order to satisfy the objective. We have undertaken different data mining and machine learning concepts. The accompanying figure, Fig.1 represents step-wise undertakings that should be finished.

#### A. Data Collection

To collect data for real estate price prediction, various factors must be considered. This includes gathering property features such as square footage, number of bedrooms, and amenities, along with location details such as neighborhood, proximity to amenities, and crime rates. Market trends, economic indicators, environmental factors, and external influences like government policies also play a role. Data sources range from online platforms like real estate websites and government databases to surveys and historical records of property sales. Ensuring data accuracy, reliability, and relevance is crucial, as is addressing ethical considerations such as privacy and biases in the data.

#### B. Data Visualization

Visualizations are important for understanding data, as they provide a visual summary of information and make it easier to identify patterns and trends. Charts and graphs make communicating data findings easier and help us gain valuable insights.

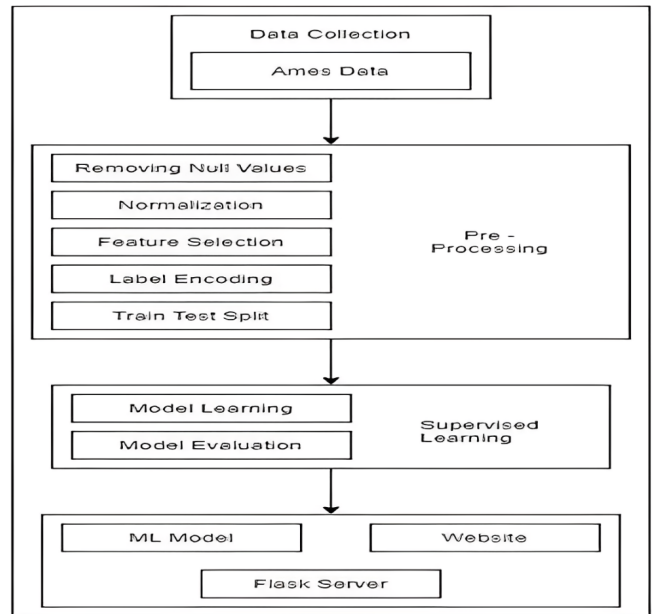


Fig. 1. Methodology

#### C. Data Pre-processing

Handling missing values is important for machine learning algorithms, as it can lead to inaccurate results. A bar plot is constructed to show the count of missing values in every attribute. After removing the null values, we need to check the distribution of the target variable, Sale Price. The target variable is right-skewed, meaning that the majority of the data has low-priced houses and a limited number of records for high-priced houses. To convert the labels into the machine-readable form, all categorical variables were converted into numeric form using the Label Encoder from Sklearn library.

#### D. Machine Learning Approaches

- In recent years, deep learning has reshaped the landscape of predictive modeling, with neural networks leading the way.
- Deep learning models, characterized by their ability to automatically learn hierarchical representations, have demonstrated considerable success in capturing intricate patterns within real estate data.
- This subsection will explore the application of neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) in real estate price prediction, shedding light on their strengths and challenges.

#### E. Ensemble Methods and Hybrid Models

- Ensemble methods, such as bagging and boosting, have gained popularity in real estate price prediction for their capacity to enhance predictive performance.
- This subsection will delve into the principles behind ensemble methods, emphasizing their role in improving model accuracy and robustness.

- Additionally, the exploration of hybrid models, combining elements of traditional methods, machine learning, and deep learning, will be addressed, showcasing the potential synergy of diverse methodologies. This section aims to provide a comprehensive overview of the methodologies employed in real estate price prediction.
- By examining traditional methods, machine learning approaches, deep learning techniques, and hybrid models, this survey aims to elucidate the diverse toolkit available to researchers and practitioners in the pursuit of accurate and reliable real estate price predictions.

#### IV. CHALLENGES AND LIMITATIONS

##### A. Data Quality and Availability

- Limited availability of high-quality and comprehensive datasets, leading to challenges in building accurate prediction models.
- Inconsistencies, errors, and missing data in real estate records, necessitating extensive preprocessing and cleaning.

##### B. Spatial and Temporal Dynamics

- Spatial variation in property prices and market conditions, requiring models to account for geographical factors and neighborhood characteristics.
- Temporal dynamics, including seasonality, cyclical trends, and short-term fluctuations, pose challenges in capturing and forecasting price changes over time.

##### C. Model Complexity and Interpretability

- Balancing model complexity with interpretability, as more complex models may offer higher predictive performance but are harder to interpret and explain.
- Trade-off between accuracy and transparency, with complex models often lacking transparency in decision-making processes.

##### D. Overfitting and Generalization:

- Overfitting to training data, resulting in models that perform well on historical data but fail to generalize to unseen data.
- Challenges in achieving robust generalization, especially in dynamic real estate markets with evolving trends and market conditions.

Some other challenges for real estate price prediction are Feature Engineering and Selection, Market Uncertainty and External Factors, Ethical and Regulatory Considerations, Computational Complexity and Scalability, Market Dynamics and Behavioral Factors, etc

#### V. APPLICATION AND CASE STUDIES

##### A. World Applications of Real Estate Price Prediction Models

- The integration of predictive models into real-world scenarios has yielded significant benefits across various sectors.

- In the realm of real estate, these models find applications in diverse areas. One notable application is in investment decision-making, where predictive analytics aids investors in identifying lucrative opportunities and optimizing their portfolios.
- Additionally, real estate professionals leverage predictive models to assist clients in pricing their properties competitively, leading to more informed listing strategies.

Paper Name	Author	Method	Result
Real Estate Price Prediction Using Regression Techniques	Andrea del Carmen Salazar Zozaya, Sharon Ramírez-Lechuga, Eri Luna López	Ridge regression, Random forest, Lasso regression, XGBoost	Random forest gives highest $R^2$ score 0.8500.
REAL ESTATE PRICE PREDICTION	Abdul Salam Ansar, Faisal Ansari, Rani Mario	Ridge regression, Random forest, Lasso regression, XGBoost, etc	Accuracy is 86.45%.
Residential Asset Pricing Prediction using Machine Learning	Yiyang Luo	Random Forest and support vector machine	$R^2$ for both methods turned out above 0.9.
House Resale Price Prediction Using Classification Algorithms	P. Durganjali, M. Vani Pujitha	Logistic regression, Decision tree, Naive Bayes, and Random forest	Adaboost gave the best accuracy of 96%.
Comprehensive Analysis of Housing Price Prediction	Rushab Sawant, Yashwant Jangid, Tushar Tiwari, Saurabh Jain, Ankita Gupta	Random Forest and Decision Tree	Random forest provides the best $R^2$ score 0.9996.
House Price Prediction Using Regression Techniques	CH. Raga Madhuri, Anuradha G. M.Vani Pujitha	Multiple linear, Ridge, LASSO, Elastic Net, Gradient boosting, and Ada Boost Regression	Gradient boosting techniques have the highest accuracy compared to other methods.

TABLE I  
FINDING AND GAPS

##### B. Case Studies Highlighting Successful Implementations

- Examining specific case studies provides valuable insights into the practical implications and successes of real estate price prediction models.
- This subsection will delve into noteworthy case studies where predictive models have played a pivotal role.
- Examples may include instances where machine learning algorithms accurately predicted market trends, leading to profitable investments, or situations where deep learning models effectively captured subtle nuances in property valuations.

##### C. Lessons Learned from Practical Experiences

- While successes are noteworthy, it is equally crucial to explore the lessons learned from practical implementa-

tions of real estate price prediction models. This section will address challenges faced during the application of these models, including issues related to data quality, model interpretability, and market dynamics.

- Understanding these lessons is paramount for refining methodologies and improving the practical utility of predictive models in real-world scenarios.
- By examining both applications and case studies, this section aims to showcase the diverse and impactful ways in which real estate price prediction models are integrated into decision-making processes.
- Highlighting successes and lessons learned provides a balanced perspective on the real-world implications of predictive modeling in the dynamic and complex domain of real estate.

## VI. CONCLUSION AND FUTURE SCOPE

### *Conclusion*

Real estate price prediction is a complex task that is influenced by a variety of factors such as location, economic conditions, and housing supply and demand. There are various techniques that can be used to predict real estate prices with reasonable accuracy, such as machine learning, statistical modeling, time series analysis, and neural networks. However, predictions are only estimates and can be influenced by unforeseen factors, so it is important to consider multiple sources of data and expertise when making important real estate investment decisions. Overall, real estate price prediction represents a promising area for innovation and research, with significant implications for stakeholders across the real estate ecosystem. By addressing the challenges outlined in this survey and embracing emerging trends, researchers, practitioners, and policymakers can unlock new opportunities to enhance decision making and drive sustainable growth in the real estate market.

### *Future Scope*

The scope of Real estate price prediction is continually expanding as technology advances, data sources grow, and the importance of informed decision-making in the real estate market becomes increasingly evident. However, users of housing price predictions should always exercise caution, consider expert advice, and understand that predictions are subject to uncertainties and market fluctuations.

## REFERENCES

- [1] Andrea del Carmen Salazar Zozaya, Sharon Ram´irez-Lechuga, Eri Luna L´opez A, Real Estate Price Prediction Using Regression Techniques, 2021
- [2] Abdul Salam Ansar, Faisal Ansari, Rani Mario, REAL ESTATE PRICE PREDICTION A Case Study (2020).
- [3] Residential Asset Pricing Prediction using Machine Learning ,Yiyang Luo. ,2020.
- [4] Teuben, B., Bothra H., Real Estate Market Size 2019—Annual Update on the Size of the Professionally Managed Global Real Estate Investment Market; Technical Report; MSCI, Inc.: New York, NY, USA,2019 .
- [5] Thuraiya Mohd, Suraya Masrom, Noraini Johari, "Machine Learning Housing Price Prediction ", International Journal of Recent Technology and Engineering (IJRTE), Volume-8, Issue-2S11, 2019.
- [6] CH. Raga Madhuri, Anuradha G, M.Vani Pujitha, "House Price Prediction Using Regression Techniques", 2019.
- [7] S.B. Jha, R.F. Babiceanu, V. Pandey, R.K. Jha, Housing Market Prediction Problem using Different Machine Learning Algorithms: A Case Study, ArXiv Prepr. ArXiv2006.10092. (2020)