REVIEW ARTICLE

SPECIAL ISSUE

RSP Science Hub

International Research Journal on Advanced Science Hub
2582-4376
Vol. 05, Issue 05S
www.rspsciencehub.com

Check for updates

http://dx.doi.org/10.47392/irjash.2023.S017

International Conference on intelligent COMPUting TEchnologies and Research (i-COMPUTER) 2023

A Literature Review on Using Machine Learning Algorithm to Predict House Prices

Tanmoy Dhar[1], *Dr Manikandan P*[2]

[1]*Department of Data Sciences, Jain (Deemed-to-be-University), Bengaluru, India*
[2]*Professor, Department of Computer Science and Engineering, Jain (Deemed-to-be-University), Bengaluru, India*

Email: tanmoyd685@gmail.com.

Article History

Received: 27 February 2023
Accepted: 11 March 2023

Keywords:

House price Prediction;
Machine Learning;
Linear Regression;
Grid Search CV;
Lasso Regression;
Decision Tree;
Pickle File

Abstract

In this study, we use a variety of machine-learning methods to forecast the sale prices of residences. The size, location, building type, age, number of bedrooms, garages, and other characteristics of the property all affect how much it is worth when it is sold. Machine-learning algorithms are employed to develop the prediction model for houses in this article. Using machine learning methods, such as call trees, supply regression, support vector regression, and the Lasso Regression methodology, a prognostic model is developed in this case. Also, we have contrasted supported parameters for these algorithms such as MAE, MSE, RMSE, and accuracy. In this research, machine learning algorithms are used as a hunting tool to create models for predicting housing value.

1. Introduction

House price prediction is the process of estimating the future selling price of a residential property based on various factors such as location, size, age, features, and current market trends. It involves the use of data analysis and machine learning techniques to develop a predictive model that can accurately forecast the price of a house. The ability to accurately predict house prices can be beneficial for various stakeholders, including home buyers, sellers, real estate agents, and financial institutions. Home buyers can use the predictions to make knowledgeable purchasing decisions, while sellers and agents can use them to set appropriate prices for their properties. Financial institutions can also use the predictions to assess the risk associated with providing mortgages to home buyers. Overall, house price prediction is an important tool that can help individuals and organizations make informed decisions in the real estate market. As technology and data analysis continue to evolve, the accuracy and reliability of these predictions will likely improve, leading to more efficient and effective decision-making. A real estate appraisal is crucial to the process of purchasing a property. Professional appraisers with specialized training in real estate assessment typically do the appraisal. An automated technique for estimating prices can help buyers of real estate properties in estimating the prices of properties that are presently available for sale.

2. Theory of Background

House price prediction is a field that draws upon various theories and concepts from statistics, economics, and machine learning. Here are some of the key theoretical background concepts:

1. Regression analysis: In order to simulate the relationship between a dependent variable (in this case, the price of a house) and one or more inde-

pendent factors, regression analysis is a statistical approach (such as location, size, and, age of the property). Regression analysis helps to identify the key predictors of house prices and estimate their effects.

2. Hedonic pricing model: The hedonic pricing model is a type of regression analysis that focuses on the pricing of characteristics of a property, such as location, size, number of bedrooms, and amenities. The model assumes that the value of a property can be decomposed into the value of its individual characteristics. This model is widely used in house price prediction because it allows for the estimation of the marginal price effects of different property features.

3. Theory of supply and chain: The supply and demand theory of economics describes how the interplay between the quantity of a good that is supplied and the quantity of a good that is demanded in the market results in the price of a good. Because changes in housing supply and demand can have an impact on house values, this theory is essential to forecasting home prices.

4. Algorithms for the machine algorithm: Machine-learning algorithms are used to build predictive models that can learn from historical data and make accurate predictions about future house prices. These algorithms include linear regression, neural networks, decision trees, random forests, and. They are used to identify patterns and relationships between the variables that affect house prices, and then use these relationships to predict future prices.

5. Market Trends: Another theory that influences house price prediction is market trends. Market trends refer to the historical and current situation of the real estate market, such as the growth rate, inflation, interest rates, and overall economic performance of the region. Market trends can help predict the future value of a property by analyzing patterns and trends in the market.

6. Real Estate Cycles: The real estate market is well known for its cyclical character, with upswings and then declines. According to real estate cycle theory, past market trends have an impact on the current market situation and can be used to forecast future market trends.

These theories can be used in conjunction with statistical models, machine learning algorithms, and data analysis to develop accurate house price prediction models. Overall, the theoretical background of house price prediction involves a combination of statistical, economic, and machine-learning concepts. By applying these theories, analysts can develop accurate models to help real estate market stakeholders make informed decisions. Real estate theory encompasses the study of various aspects of real estate, including land use, property rights, property valuation, and the dynamics of real estate markets. Real estate is a significant sector of the economy, and understanding the principles that govern house price markets is essential for policymakers, investors, and property owners.

3. Literature Review

House price prediction has long been a subject of interest to scholars, most recently, the increasing availability of data and the creation of machine learning methods have enabled more accurate and sophisticated methods for predicting house prices. In the literature review, we will highlight some of the key research in this area.

The paper was conducted by Yashraj Garus and Himanshu in 2020, where the paper title is house price prediction using machine learning where they proposed a system including six-step of machine learning (Garud et al.). They mentioned in the paper ascending by cleaning the training dataset, feature engineering, test-train split, training the model's lasso and random forest algorithm will be used and then predict on the test set after the result will be visible and more accurate model will be picked. (Furia and Khandare)

In 2020, Ashutosh, Pranav, Deeksha, and Shreya completed a paper (Sharma et al.). Where they are using machine learning to build a system for forecasting and recommending House price prediction. Also, they are using linear regression as an algorithm to predict house prices. They are using ARIMA (Auto Regressive Integrated Moving Average Model) for forecasting models like sales prediction and for content-based recommendations have two approaches which are content-based filtering and collaborative filtering approaches and the accuracy of the model is 87%.

Neelam Shinde and Kiran Gawande published a paper in 2018 titled the valuation of house price using predictive techniques. (Guan, Zurada, and Levitan) In this paper, they are using four types of algorithms which are Logistic Regression, Support

vector regression, Lasso regression, and Decision tree. In the four algorithms, the highest accuracy is 84.64% in the Decision tree. Others are logistic regression, support vector regression, and lasso regression having 72.81%, 67.81%, and 60.32%, where they have some parameters which are Accuracy, R-squared value, RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MSE (Mean Square Error) for compare those models (Shinde and Gawande).

A case study was conducted in Malang, Indonesia by Adyan, Hilman, Ruth, and, Wayan in 2017. When regression analysis and particle swarm optimization are used to model a house price prediction (Alfiyatin et al.). The experimental procedure

investigates the Combination test for particles, iterations, and inertia weights, these three parameters employed in particle swarm optimization. Based on the NJOP data of 9 houses, the system modeled house price prediction into 7 models each of which represents one. The optimum parameters were 1800 particles, 700 iterations, and inertia weights of 0.4 and 0.8 which resulted in the lowest prediction error (RMSE) of IDR 14.186. Other prediction errors large compare to the previous.

A paper provides how to predict house costs utilizing different methods by using python libraries which is published by Manish, Himans, Neha, and Prenika in 2020. They used machine-learning techniques for implementation which are regression techniques, classification techniques, cross-validation techniques, and k-means to get the highest accuracy.

A review paper published by Harsha, Dr. Rajan, and Jagrati where predict house prices based on Fuzzy logic. Fuzzy logic is defined as a method for resembling human behavior and giving immediate values rather than 0 or 1, fuzzy logic can determine much flexibility and value. This study provides fuzzy logic as the greatest solution to predict house prices having a variety of demands of the client (Rajan, Jagrati, and Harsha).

The research was conducted in 2021 by PEI-YING, CHIAO-TING CHEN, and, three other authors. They build utilizing heterogeneous data analysis and a combined self-attention mechanism, a deep learning model for predicting home prices. They approach three experiments where the first experiment is the basis on a deep learning model and heterogeneous data compared to public facilities and satellite maps, the second one is for comparison with the outcomes of the joint-self attention mechanism, and multiple attention, mechanisms are used. The third experiment used the proposed model to analyze house prices in Taichang and Kaohsiw in Taiwan. To determine the essential characteristics of a home for buyers, joint self-attention weights are collected in the final trial (Wang et al.).

Martin Hesliand and Louis Johner published the study in 2022 (Hoesli and Johner). The author uses simulation and property-level data for the United States to diversify its portfolio among gateway and non-gateway real estate areas. They contrast risk-adjusted performance shown to be comparable to different market types with performance indicators for portfolios with varying splits between the gateway and non-gateway markets. The non-gateway market provides output with greater income returns while the paper gateway market has a better appreciation and overall returns. Although full draw-down and recovery times are often shorter for gateway markets than for non-gateway markets, downside risk does appear to be slightly higher for gateway markets. This post also offers tips on how to define gateway markets in the most effective way.

4. House Price Prediction Techniques

There are a few approaches to predicting house prices. Most commonly two techniques are used commonly which are machine learning and neural network model. Both machine learning and neural network models have been widely used in house price predicting actions, as they are capable to learn complex relationships between input features and the output variable.

5. Machine learning model

Machine learning has become increasingly popular in house price prediction due to its ability to handle complex relationships between input features and the output variable. Here are some of the commonly used machine learning models for house price prediction:

A. Linear Regression: Linear regression is a simple and widely used machine learning model that can predict house prices based on various input features. In this model, the relationship between the input features and the output variable is regarded as linear. However, when there is a nonlinear relation-

ship between the input characteristics and the output variables, linear regression models may not work well.

B. Decision Tree: Decision trees are another popular machine learning model used for house price prediction. This model can handle both numerical and categorical input features and can capture nonlinear relationships between input features and the output variable. Decision trees can also handle missing data, which is a common problem in real-world datasets.

C. Support Vector Machine: Support Vector Machines are a popular machine learning model used for regression and classification tasks. In house price prediction, SVMs can be used to predict the price of a house based on various input features. SVMs can handle both linear and nonlinear relationships between input features and output variables and can handle large datasets.

D. Classifier with Random Forest: Random forests are an ensemble learning technique that uses multiple decision trees to predict house prices. Random forests can handle both categorical and numerical data and are less prone to over-fitting than decision trees. They also have the ability to handle missing data and outliers.

6. Neural Network:

Neural networks, on the other hand, have shown great potential in predicting house prices due to their ability to learn nonlinear relationships between input features and the output variable. Feed-forward neural networks, recurrent neural networks, convolutional neural networks, and auto-encoders are some of the commonly used neural network architectures in house price prediction tasks. Neural networks are suitable for large datasets, as they require more computational power than traditional machine learning models.

A. Feed-Forward Neural Networks: Feed-forward neural networks are the most widely used kind of neural network used in house price prediction. In this architecture, the input features are fed into an input layer, and then they are processed through one or more hidden layers before reaching the output layer. The output layer provides the predicted house price. These networks can learn nonlinear relationships between input features and output variables, but they require a large amount of training data to overcome over-fitting.

B. Convolutional Neural Networks: CNNs (Convolutional neural networks) are widely used in image-processing tasks. In the context of house price prediction, CNNs can be used to learn features from images of houses or neighborhoods. For example, CNN can learn to recognize the architectural style of houses, the presence of trees, or the quality of roads. These features can be used as input to a feed-forward neural network to predict house prices.

C. Auto Encoders: Auto-encoders are a type of neural network used for unsupervised learning tasks. Auto-encoders can be used to learn compressed representations of input characteristics in the context of predicting home prices. These representations can be used as input to a feed-forward neural network to predict house prices. Autoencoders can help reduce the dimensionality of the input features, improving the neural network's performance.

In general, the choice between using a machine learning or neural network model in house price prediction depends on the complexity of the problem and the amount of data available. For smaller datasets or less complex relationships, traditional machine-learning models may be sufficient. For larger datasets or more complex relationships, neural networks may provide better accuracy and generalization.

7. Comparative Study of the Existing Method

A comparison of the current methods for predicting home prices has been done by a number of scholars. Here are some of the key findings from these studies. A study conducted by K. Mehta and S. P. Singh found that linear regression performed well when predicting house prices based on location, bedrooms available, and the quantity of bathrooms, achieving an accuracy of around 85%. However, it did not perform as well when predicting house prices based on more complex features such as crime rates and school ratings. G. Manogaran et al. found that decision trees achieved an accuracy of around 77% when predicting house prices. Gradient boosting achieved an accuracy of around 93% and was also able to capture nonlinear relationships between input features and house prices. Neural networks achieved an accuracy of around 94% when predicting house prices based on features such as location, bedrooms available, and square footage. Authors try to pre-

TABLE 1. Summarizing Recent Techniques

Technique	Contribution	Drawbacks
Artificial Neural Networks (ANN)	Able to capture complex patterns in data and predict house prices with high accuracy	Require a large amount of data for training, and the training process can be time-consuming. The model's complexity can lead to over-fitting, and the interpretability of the model is limited.
Random Forest	Able to handle non-linearity and interactions between variables and provide accurate predictions.	The model's accuracy can decrease when dealing with high-dimensional data. The model's interpretability is limited, and it may suffer from the over-fitting problem.
Gradient Boosting	Able to learn from the previous model's errors and improve the accuracy of the predictions.	The model's complexity can lead to over-fitting, and the training process can be time-consuming. It also requires careful hyper-parameter tuning to achieve optimal performance.
Support Vector Regression (SVR)	Able to handle non-linearity and high-dimensional data and provide accurate predictions.	The model's training can be computationally expensive, and the model's interpretability is limited. It also requires careful hyper-parameter tuning to achieve optimal performance.

dict house prices more efficiently basis on these approaches. It appears that neural networks and gradient boosting tend to perform the best in accuracy, followed closely by random forests and SVMs. Here is a table summarizing some recent techniques used for house price prediction, their contributions, and drawbacks

8. Observation

This survey paper presents an overview of recent updates on House Price Prediction with these different machine-learning algorithms. Neural networks and gradient boosting tend to perform the best in terms of accuracy, followed closely by random forests and SVMs. The selection of features or input variables that are used to predict house prices is so crucial for the accuracy of the model. Researchers have found that the most important features for predicting house prices include location, square footage, the number of bedrooms and bathrooms, and property age. House prices do not always have a linear relationship with input features. For example, the relationship between the price of a house and its square footage may not be linear. As a result, machine-learning techniques can capture nonlinear relationships. Overall, house price prediction is a complex task that requires careful consideration of input features, data quality and quantity, and the choice of machine-learning technique. While there are many approaches that can be used, the choice of technique ultimately depends on the specific needs and constraints of the problem at hand.

9. Conclusion

The suggested approach forecasts the price of real estate in Bangalore based on a number of characteristics. To find the optimal model, we will test various machine-learning techniques. The Linear Regression Algorithm achieved the lowest loss and the highest R-squared when compared to all other methods. The website was made with Flask. Ultimately, a house price prediction's conclusion would depend on a number of variables and would need to be carefully analyzed based on the prediction's particular context.

References

Furia, Palak and Anand Khandare. "Real Estate Price Prediction Using Machine Learning Algorithms". *Data Science and Intelligent Computing E-Conference* (2022): 19–32. 10 . 1002 / 9781119792437.ch2.

Garud, Y, et al. "Home Loan Prediction Using Machine Learning Models". *Elementary Education Online* 20.3 (2021). 10 . 17051 / ilkonline . 2021.03.359.

Guan, Jian, Jozef Zurada, and Alan S Levitan. "An Adaptive Neuro-Fuzzy Inference System Based

Approach to Real Estate Property Assessment". *Journal of Real Estate Research* 30.4 (2008): 395–422. 10.1080/10835547.2008.12091225.

Hoesli, Martin and Louis Johner. "Portfolio Diversification Across U.S. Gateway and Non-Gateway Real Estate Markets". *Journal of Real Estate Research* 44.4 (2022): 523–552. 10 . 1080 / 08965803.2022.2038902.

Rajan, Dr, Jagrati, and Harsha. "Predicting Home Price Based on Fuzzy Logic". (2022).

Sharma, A, et al. "Home Price Prediction Forecasting and Recommendation System Using Machine Learning". (2020).

Shinde, N and K Gawande. "Utilizing Predictive Methods to Value Home Prices". (2018).

Wang, T. -Y, et al. "Integrated Self-Attention Mechanism and Heterogeneous Data Analysis in a Deep Learning Model for Home Price Prediction". (2022).

Embargo period: The article has no embargo period.

To cite this Article: Dhar, Tanmoy, and Dr Manikandan P. "**A Literature Review on Using Machine Learning Algorithm to Predict House Prices**." International Research Journal on Advanced Science Hub 05.05S (2023): 132–137. http://dx.doi.org/10.47392/irjash.2023.S017