



Systematic study of Real estate valuation using a Machine Learning Model

Guide: Rushikesh Chaudhari

Kamma Udaya Lakshmi (190304105203)

Govvala Bhavana (190304105006)

Devipriya Pulivarthy (190304105057)

Dasari Tarani (190303105378)

Abstract

Machine learning is widely used in real estate valuation to predict the price of house with high accuracy. Real estate valuation is a decision model in which number of attributes such as structural attributes and locational attributes are taken to predict the house price. The fluctuation of prices in real estate market are always a concern for both the land buyers and seller. So, literature survey has been done to analyse the most efficient attributes and models to predict the accurate house price. In the analysing process there are many types of models to predict the house price like hedonic price model, multiple liner regression, artificial neural network, support vector regression, XG boost etc... This study will help the future researches and housing agents to select the most significant machine learning model for accurate prediction of house pricing.

KEYWORDS: Hedonic price model, Multiple liner regression, Artificial neural network, Support vector regression, XG boost.

1. INTRODUCTION

Everyone has a dream to build or buy their own house. In order to achieve this dream, they face certain obstacles one of it is not knowing the house price or land. There are many points of interests to determine the price of house such as distance of the house from railway stations, shopping malls, schools, markets, hospitals etc.... the houses which are near to these points of interest will have higher price compared to the houses which are far away. Hence these points of interests determine the price of the house. These points of interests vary from buyer to buyer on their own preferences. In the research of survey papers there is no methos mentioned to sort the preferences of the buyer.

From time to time the development of area takes place like train stations etc.... Due to this the house demand may increase in that area. This may become problem for the real estate dealers and buyers to fix the price considering the accurate key factor.

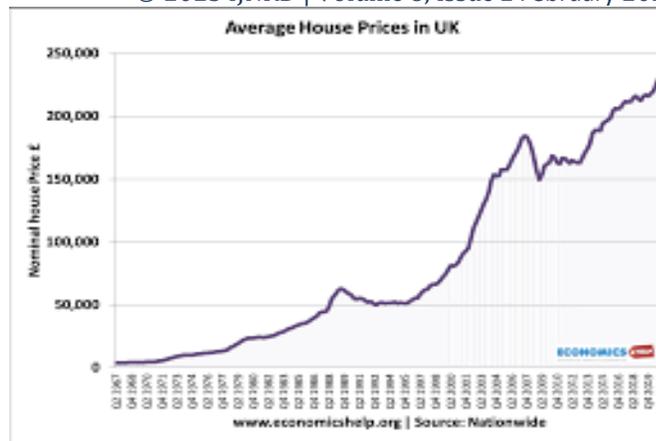


Fig 1. House price fluctuation in UK from 1967 to 2021

In this article, literature survey focuses on house price prediction based on which model is more efficient than the other models. Every model has its own pros and cons. So, the model which gives the accurate house price will be the best.

2. LIETRETURE SURVEY

In house price prediction a machine learning model is required to get the accurate house price. Machine learning is actually a type of artificial network which used to get accurate predictions by training the machine using the training data. Model or the algorithm which we use is to fit the data into training data and using that we evaluate the house price is accurate or not. So, in house price predication we need to classify the training data and need a model to get the house price. The house price will be predicted by them. The inputs will be the attributes or the key interest of the buyer which are used to get the house price. There are many types of attributes in house price predication.

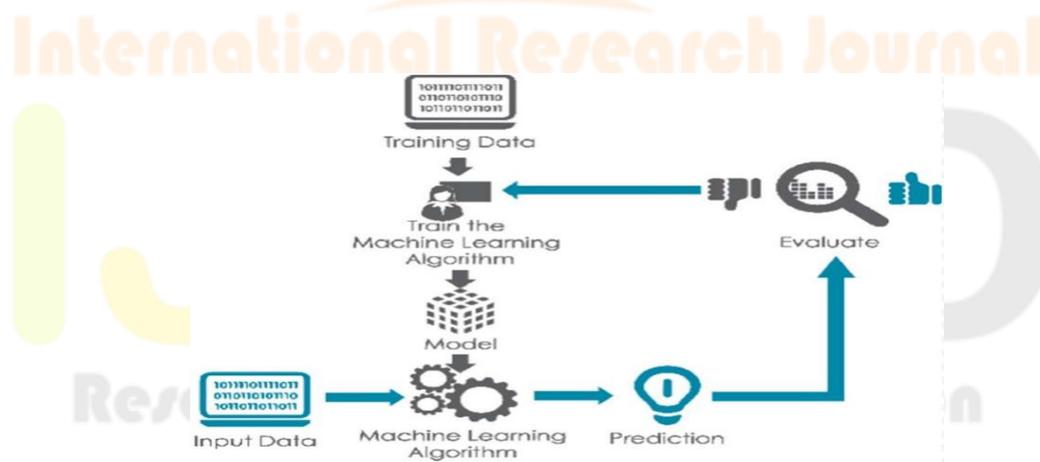


Fig 2. Steps involved in house price prediction using machine learning.

There are many types of models which can be used among them hedonic price model, multiple liner regression, artificial neural network, support vector regression, XGboost are the mostly used methods or models in our research.

i) Hedonic Price Model

Hedonic price model is generally used in the housing industries to determine the price of house based on the environment and non-environment characteristics (like internal and external factors).

The Hedonic price function $p.f(S, N, Z)$:

S= Structural characteristics (like type of construction, house size, no of rooms)

N= Neighbourhood characteristics (like accessibility to work, crime rate, distance to the market, school, hospitals)

Z= Characteristics of the environment

An example to understand the Hedonic price model Area A: Already established building, 2 bed rooms, nearby railway, school, and 5 mins from the highway. Area B: New building, 2 bedrooms, 20 mins from the railway station. In the above case, Area A house cost is high than compared to Area B house because the railway station distance will impact the price of both houses. The buyer will prefer Area A compared to Area B because transport is easily available and provide flexibility. Since Area B has more disadvantages compared to Area A so the buyer is ready to spend more money since it saves his time and money because it is near to railway station to come and go to the office [1,4,7,11,12].

ii) Multiple Linear Regression

Multiple linear regression is a model to output the dependent variables but taking the independent variables as input. Here independent variables are variables like access to hospitals, shopping malls, schools, restaurants, railway station, public transportation, house age, number of bedrooms, number of bathrooms, floor area, landscape, place of worship, crime rates, noise, local government, income, cost of material. These all-independent variables are used to predict the dependent variable that is house price. Using these variables will accurately give the house price [1,2,6,15,16,17,18,19].

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon$$

y_i = dependent variable

x_i = explanatory variables

β_0 = y-intercept (constant term)

β_p = slope coefficients for each explanatory variable

ϵ = The model's error term

iii) Gradient Boosting

Gradient boosting is a method of regression and classification tasks in the gradient boosting the performance stage is low compared to the XG boost. XG boost Algorithm stands for extreme gradient boosting. It helps us for the accurate predictions. This model is used to work with multiple datasets. In the house predictions we have taken large number of datasets. so, we can use the XG boost method. By using the python in XG boosting it gives a large number of inner parameters. The algorithm is most effective in producing a more stable prediction. The cost of the model is less. XG boost is it perform well for the structured data and fits for a smaller amount of data. We can say that XG boosting uses more regularized forms comparing to the gradient boosting. The gradient boosting is used to minimise the bias error of the model and it has a lot of flexibility [2,18].

iv) Artificial Neural Network

Normally in our neuron system when we have signals to the neuron then a neuron will process and transfer. The same thing happens in artificial neural network when we give attributes as input it will process and give the result which is our house price. In multi-layer neural network there are three layers. First layer takes the inputs, in our model they are attributes. Second layer that is hidden layer, the inputs will be processed to get the output and transfer the result to the result layer. In every layer there are number of circles called as nodes. Third layer is the result layer, it gathers the result transferred from the hidden layer and present it as the final output that is house price [2,7,10,13].

v) Support Vector Regression

Support vector regression is the learning algorithm that is used to predict the discrete values. In the house price prediction, we can't predict the accurate price of the house. The sellers should balance the house price and maintain the requirements of the buyers. So, we take the different requirements from the buyer and we analyse the data by using the support vector regression method. We make a hyperplane by the values. We take the best possible one which analyse and make the possible requirements for the buyers. So that both the sellers and buyers can reach an agreement [2,14].

3. VARIABLES

Variables are the main key points which are used in the house price prediction to get the accurate house price these variables are the requirements which buyers or the customers requires

The key points used are classified as given below:

a) Locational Variables

They are the basic requirements for a person to live comfortably. These variables or key points are the requirements or need for a person to live like access to shopping malls, office, school, convenience store, local market etc.... These factors also play a viral role in house price decision. Mostly the people in cities or metropolitan cities needs or require these attributes as only the cities have these attributes.

b) Structural variables

The structural variables or interest points are the basic information about condition of land or the house such as land size, number of bedrooms, number of bathrooms, housing age, floor size etc...

c) Neighbourhood variables

These variables explain about the surroundings of the house like landscape, place of worship, crime rates, noise, local government, pollution etc....

d) Economic variables

These variables are like cost of material used in building house, real income etc....

4. DISCUSSION

Hedonic price model

In this model It identifies attributes into two categories environmental and non-environmental attributes from the attributes of locational, structural, neighbourhood and economic [2]. This data makes it possible to assert that increase or a decrease the aesthetic quality of the landscape of 10 points [4,11,12]. This model prefers non environmental or external factors and predict the house price based on them. In some research papers hedonic model is used for house price prediction it is mentioned that it is costly and depends on only external factors like access to shopping malls, railway station, schools etc.... to predict the house price [12]. This may affect the house price by preferring the non-environmental characterises. It instantaneously increases and reduces the house price based on external factors. As Hedonic price model gives more importance to the external factors upon internal factors. The price varies in huge amount due to the change in the external factors.

Multiple linear regression

In this model all the structural, locational, neighbourhood and economic attributes are named as dependent variables and are used to find the dependent variable that is house price [1,16]. As it gives equal importance to all the attributes or variables it can predict the house price more efficiently and has high accuracy. It is a best model for large number of data set compared to other models. It is sensitive to outliers as it gives equal importance to the variables sometimes the house price varies because of them.

Gradient boosting

In this model it collects structural attributes information related to house or land like number of bedrooms, area of land, number of bathrooms etc.... This model is mostly used for its performance, consistency and interpretability. However, this model also faces challenges with regard to the balance between accuracy and performance [2]. To overcome the challenge, it needs to follow certain steps and optimize the attributes which results in it can work only for smaller data set.

Artificial neural networks

In this model the local and structural attributes are collected. As we know that hidden layer is the second layer of artificial neural network, there can be many hidden layers as per the user requirements but the problem is a greater number of layers then more complicated it will be. It accepts non-linear attributes which are more in number for a house price prediction this result in complex of the model as it has limited performance [2].

Support vector regression

In this model it used to collect neighbourhood, structural and locational attributes. It is easily adaptable and also works on non-linear attributes [2]. As the model is trained by the training data set samples, if the number of features for each data point exceeds the number of training data samples the model will underperform. It best for smaller data sets but not for large data sets as becomes complex. Even if the data set has a greater number of overlapping then it may not perform well.

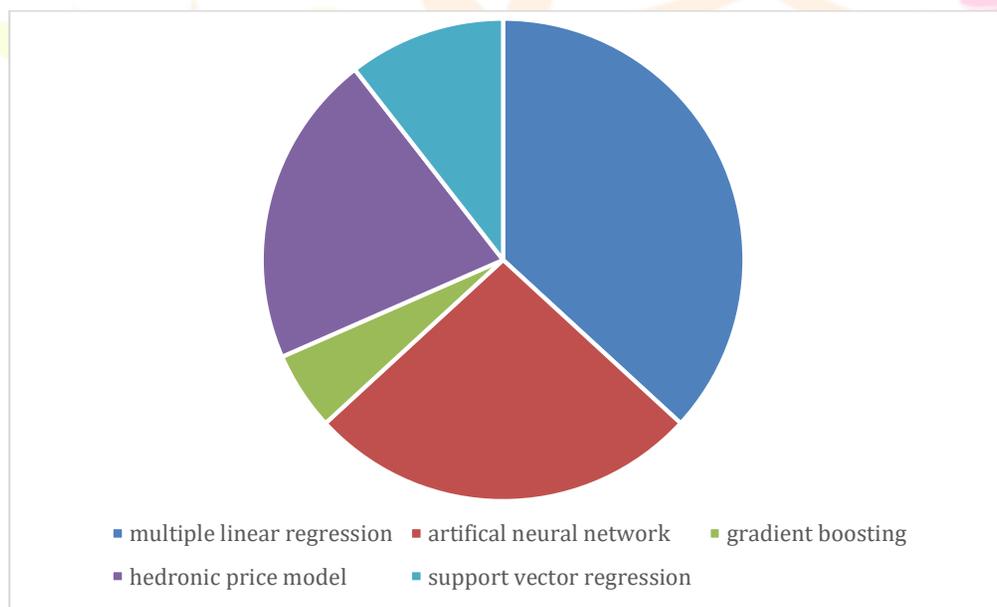


Fig 3. Pie chart for the number of articles used these models.

5. ANALYSIS

This paper analysed the current research work on the house price prediction using machine learning with the help of attributes and which model or algorithms gives accurate house price. By taking the attributes into consideration the attributes like access to convenience store, access to shopping malls, plot size are most commonly required attributes requested by the buyer. After the analysis of predictive models, the most commonly used models are multiple linear regression, hedonic price model, gradient boosting, artificial neural network and support vector regression. From the

reference, multiple linear regression and artificial neural network has more potential compared to other models in house price prediction. These models take the most significant attributes and give the accurate house price.

6. CONCLUSION

In the conclusion, the model which gives accurate house price and most commonly used was multiple linear regression. It is basically used for its flexibility, objective approach and sticking to core. In a project like real estate evaluation model there will be number of attributes which need to be considered. With the help of machine learning and this multiple linear regression the data will be classified into training data and testing data so that it can fit into multiple linear regression and gives the accurate house price. This way the buyer and seller can have a fair deal on house price.

7. REFERENCE

1. Adyan Nur Alfiyatin , Hilman Taufiq ,2017,Brawijaya University, Malang, Indonesia Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization.
- 2.Nor Hamizah Zulkifley.,Shuzlina Abdul Rahman,Nor Hasbiah Ubaidullah,Ismail Ibrahim,15 July 2020,House Price Prediction using a Machine Learning Model: A Survey of Literature, 08 December 2020.
3. Hilary I. Okagbue¹, Nkolika J. Peter², Adedotun O. Akinola³, Chukwuemeka O. Iroham², Akunnaya P. Opoko³ ,01/2020, Smart Review of the Application of Genetic Algorithm in Construction and Housing,02/2020.
- 4.Alexandre Tangerini and Nils Soguel, —Bringing the Hedonic Price Method into Fashion when Valuing Landscape Quality||, December 2006, Working paper de l'IDHEAP, pp 1-3, 10.
5. Oyedeji, Joseph Oyewale, 9 Oct 2020, Estate Management Department Bells University of Technology, Ota Impact of Land Administration on Private Housing Delivery in Lagos, Nigeria.
6. Adyan Nur,2017, Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization.
7. 1Visit Limsombunchai, 2Christopher Gan and 2&3Minsoo Lee,2004 , Hedonic Price Model vs. Artificial Neural Network House Price Prediction.
8. Sarkar Snigdha Sarathi Das,Mohammed Eunus Ali,Yuan-Fang Li,Yong Bin Kang, September 2020, Boosting House Price Predictions using Geo-Spatial Network Embedding.
9. Eunah Jung ¹ and Heeyeun Yoon, 12 September 2018 , A Mahalanobis-Metric Matching Approach to the Housing Market in Gyeonggi, South Korea ,1 November 2018.
10. Nkolika J. Peter¹, Hilary I. Okagbue² *, Emmanuela C.M. Obasi³, Adedotun O. Akinola⁴ ,2020, Review on the Application of Artificial Neural Networks.
11. 1 Anita Kanojia, 2Dr. R. B. Magar, 3Umesh Jadhav, 2016, Valuation of Residential Properties by Hedonic Pricing Method (HPM).
12. Nana Cui ,Hengyu Gu ,Tiyen Shen ,* and Changchun Feng, 19 September 2018, addressed: <https://doi.org/10.3390/su10124343> , The Impact of Micro-Level Influencing Factors on Home Value: A Housing Price-Rent Comparison.
13. Olalekan Oshodi¹, Oyewale Oyedeji¹ , and Clinton Aigbovboa¹,2020, does proximity to tourist site have an impact on the rental value of a residential property? ,addressed :<https://doi.org/10.1051/mateconf/202031204003>.
14. 1Sefa Awaworyi Churchill, 2Abebe Hailemariam, 3,4Michael Odei Erdiaw-Kwasie, April 2020, examine relationship between urbanization and real house prices for a panel of OECD countries using a long historical data that covers the period 1870 to 2016.
15. Chenhao Zhou ,2021,House price prediction using polynomial regression with Particle Swarm Optimization, New York University, USA

16. Antony Christopher, Jan 13 2021,Applying Multiple Linear Regression in house price prediction.
17. R Manjula, Shubham Jain, Sharad Srivastava and Pranav Rajiv Kher , 14th 2017, Real estate value prediction using multivariate regression models.
18. Zhou, Yichen , 2020,UNIVERSITY OF CALIFORNIA Los Angeles, Housing Sale Price Prediction Using Machine Learning Algorithms.
19. Lishun Yuan, Spring 2019 , A REGRESSION MODEL OF SINGLE HOUSE PRICE IN LA CONSTRUCTING A PREDICTED MODEL FOR HOUSE PRICES.
20. Steven C. Bourassaa, Martin Hoeslib, Donato Scognamiglioc and Philippe Sormani, Constant-Quality House Price Indexes for Switzerland.

