IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

Predicting The Housing Price using Artificial Intelligence/ Machine Learning

Aman Yadav
B.Tech Student, Department of Computer Science and Information Technology
Dronacharya College of Engineering, Gurugram, India

Abstract: *Data mining is now commonly used in the real estate market. The ability of data mining to extract relevant information from raw data makes it very useful to predict house prices, key housing features, and much more. Research has shown that fluctuations in housing prices often affect homeowners and the housing market. Literature research is done to analyze the relevant factors and the most effective models for predicting housing prices. The findings of this analysis confirmed the use of Artificial Neural Network, Support Vector Regression and XGBoost as the most efficient models compared to others. In addition, our findings also suggest that spatial and real estate agents are key factors in predicting house prices. This research will be of great benefit, especially to housing developers and researchers, to find the most important clues for determining housing prices and to identify the best machine learning model that will be used to conduct research in this field.*

Keywords: House Price Prediction, Machine Learning Model, Support Vector Regression, Artificial Neural Network, XGBoost

I. INTRODUCTION

A house is one of the most important necessities of life, as well as other basic necessities like food, water, and much more. Demand for housing has grown rapidly over the years as the standard of living of the people improves. Although there are people who make their homes as an investment and property, yet many people around the world buy a house as their living space or as a means of subsistence.

According to [1], the housing market has a positive impact on the national currency, which is an important measure of the national economy. Homeowners will buy furniture such as furniture and household items, and homeowners or contractors will buy raw materials to build houses to meet housing needs, which is an indication of the economic impact of the new housing market. Besides, buyers have the potential to invest heavily, and the construction industry is in good shape and can be seen by the country's high standard of housing.

According to [2], many international organizations and human rights groups have emphasized the importance of housing. The House focuses on the economic, financial, and political structure of each country. However, [3] reported that housing price volatility has always been a problem for homeowners, real estate and real estate, apart from that [4] he said that housing is not affordable as there has been significant inflation in several countries

housing sector. The quality of life of the citizens and the economy of the country depends on the possible increase in housing prices. Ultimately, the issue will affect investors who make their home as an investment.

The increase in housing demand occurs every year, resulting in an indirect increase in house prices every year. The problem arises when there are many variations such as location and the need for a building that may affect the price of a house, so most participants including buyers and developers, real estate developers and the real estate industry would like to know the exact factors or factors that influence a home. price to help investors make decisions and help homeowners set home prices.

Home price guessing can be done using multiple guessing models (Machine Learning Model) such as vector retrieval, neural artificial network, and more. There are many benefits to real estate buyers, real estate investors, and real estate agents from the real estate model. This model will provide a wealth of information and information to real estate agents, real estate investors and real estate agents, such as estimating real estate prices in the current market, which will help them to determine house prices. At the same time, this model can help potential buyers determine the features of the home they want in terms of their budget [5]. Previous research has focused on analyzing house price characteristics and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

predicting house prices based on a separate machine learning model. However, this article combines both the predictable house price and the properties together.

In this article, the literature review focuses on the pricing of home prices based on the machine learning model and analysis of the characteristics used primarily in previous research affecting house prices. This paper is organized as follows: the first phase summarizes the entire study. The second section describes the general characteristics used in real estate prices around the world. This was followed by a brief discussion of the machine learning model used in the previous study to predict house price. In the next section, the comprehensive results of the current model of house price forecast are considered. Finally, paragraph 5 and section 6 respectively provide the definition and conclusion of this comprehensive literature review..

II. ATTRIBUTES

Home price forecasts can be divided into two categories, first with a focus on housing features, and second with a focus on the model used in the home price forecast. Many researchers have developed a model for predicting house prices, including [1, 3, 6-8].

The study by [9] analyzes existing housing prices in Jakarta, Indonesia using a conceptual model and a questionnaire. Based on the results, the factors or factors affecting the price of the house vary with the construction of each house in Jakarta, so accepting the validity of this analysis as the main objective of this study is to distinguish the feature or attributes that affect the house price. Various considerations influence the price of a home. According to [10], factors that influence housing prices can be divided into three categories: location, status and neighborhood status.

A. Location

Location is considered the most important factor in determining the price of a house [6, 9-11]. [12] in his research he also recognized the importance of local attributes in determining the price of a house. Location area was subdivided into a fixed location attribute. All of these studies point to a close correlation between local features such as the distance from a nearby shopping mall, or an area that offers views of hills or the coast, and house prices.

B. Structural

Another important factor influencing the price of a home is the design of a building or a particular study listed as factors (10, 13]. A building element is an element that people can identify, whether the number of bedrooms and bathrooms, or the basement, or garage and balcony. These building properties, which are often given to home builders or engineers to attract potential buyers, therefore meet the wishes of potential buyers. [14] In his previous research, building properties would be a major consideration for homeowners in deciding what to buy as such features represent their market value. In their previous research, [15] he stated that all these attributes are positively related to rising house prices [16].

C. Neighbor

Neighbor attributes can be included in determining the price of the house. According to [13], the efficiency of public education, the social environment and the proximity to shopping malls often enhance local value. There is a significant increase in housing prices from the lower fifth category to the wealthy as predicted [16]. However, [13] research has found that these traits are often rooted in culture, as they do not apply in the same way to all cultures.

Previous Study		[6]	[3]	[4]	[17]	[18]	[19]	[5]	[10]	[9]	Data Source
Locational Attributes	Access to shopping mall	✓	✓	✓	✓	✓		✓	✓		7
	Access to schools	✓			✓			✓	✓	✓	5
	Access to hospitals	✓	✓	✓	✓			✓	✓	✓	7
	Restaurants				✓			✓	✓	✓	4
	Public Transportation	✓	✓	✓				✓	✓	✓	6
Structural Attributes	No of bedroom	✓	✓	✓	✓			✓			5
	No of bathroom		✓			✓	✓				3
	Floor area	✓			✓	✓					3

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

	Garage and patio				✓	✓	✓				3
	Property age housing		✓	✓	✓	✓					4
	Lot size		✓	✓	✓	✓	✓	✓	✓		7
Neighborhood Attributes	Socio-economic variables				✓	✓				✓	3
	Local government										0
	Crime rates										0
	Place of worship							✓	✓		2
	Pleasant landscape		✓					✓			2
	Quiet atmosphere		✓								1
Economic Attributes	Income									✓	1
	Cost of material									✓	1

Table 1. Summarization of Attributes used in Previous Study

Based on Table 1, it is clarified that spatial factors including store access, school access, hospital access, restaurants and public transport are the most used indicators for neighborly pricing and economy.

III. MACHINE LEARNING MODEL

According to [20], the paradigm of assessing the need for housing can be divided into two categories which is the standard method and the advanced measurement method. A standard measurement scheme, which includes a recurring regression method and a gradual retrospective procedure, while a hedonic pricing tool, a synthetic neural network (ANN) and a spatial analysis framework is a pre-determined measurement method. The choice of model to be used to predict the price of the house is very important as there are a variety of models available. One of the most widely used models in this field of research is Regression Analysis which is used in many studies, including [3, 10, 21]. Another common model of real estate prices is Support Vector Regression (SVR) [7, 22, 23].

A. Regression Analysis i. Hedonic Price Model

The housing market is a little different from the common good use. According to [13], the housing market is unique in that it reflects the characteristics of durability, flexibility and spatial harmony. Therefore, the hedonic method is chosen to accurately predict market differences. [24] conceived a hedonic model back in 1939, but this study became famous in the early 1960s with the full use of Zvi Griliches and Rosen [24]. In the early 1930's, the Court used this model to analyze the number of vehicles in terms of quality and quality. [25] hedonic has defined it as "the ambiguous prices of attributes and disclosures to economic actors based on the perceived value of different products and the specific number of its characteristics." After years of progress, Rosen used the method in the study of real estate prices and was often included in the real estate industry research [3].

Rosen's philosophy or model consists of two distinct categories. The reversal of the product price in its attributes is done in the first phase to calculate the total price of the component. The value of the asset will be determined in the first phase, but the function of the opposite need cannot be reproduced in this phase. Therefore, the second phase of the measurement is required to identify the function of the cross-reference that can be found in the first phase of the transparent price function.. In an earlier study, a study compared three commonly used house price measurement methods which are a simple standard method, a hedonic model, and a simulation method. The result found that when they were adopted in the housing market, the two most common methods and the matching method proved to be biased. Therefore, the hedonic model provides the highest results associated with those two versions that are often encountered [3].

Hedonic Price Model is a mathematical model that believes that value for space is the sum of all its features based on a hedonic market perspective..

B. Multiple Linear Regression

Deviation analysis is a model used to determine the relationship between variables. To assess the relative affinity, a coefficient of integration or regression equation can be used [26]. Many retrospective models can determine which factors are most important for the different dependent definitions. Multiple retrospective analysis also allows for certain price

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

forecasts by capturing separate independent and dependent data. In [27], the dynamics of the multiple regression model can be seen when the value of the relationship between dependent and independent variables is measured. [28] use a regression model frequently to describe the development of independent variables with dependent variability.
This model can be achieved using house price as varied and dependent such as house prices, house size, building type, number of bedrooms, and much more. Therefore, house price is set as targeted variation or dependency, while other attributes are set as independent variables to determine the main variable by determining the relative coefficient of each attribute.

Support Vector Regression

The back of the support vet is a predictable model based on SVM, a neural network that usually has a three-layer, powerful way of studying surveillance. The model is based on a subset of training data. The advantages of vector regression are that it can process indirect results, provide a single potential solution, and is able to overcome small sample learning problems [23].
The ability to generate market forecasts in a number of markets, including real estate, shows that this model can overcome non-linear retrieval problems and smaller sample learning problems. Furthermore, since this model was not dependent on distribution considerations, as well as the ability to map input feature, either linear or indirect, this model was often used in house price comparison [22]. Vector rotation offers significant advantages in many respects as this model can avoid over-equilibrium problems, while ensuring one good solution by minimizing structural hazards and potential risks [29].
In this field of research, retrospect vector support is used to collect location data, properties and spatial attributes...

Artificial neural Network

In 1958, [30] formed a synthetic neural network known as ANN. Walter Pitts and Warren McCulloch published a paper entitled "A Logical Calculus of Imaginary Ideas in Nervous Activity" in 1943 that noted that the neural network could be created automatically, based on the role and structure of the biological neural network. In some studies, as this model often promotes learning, artificial neural networks are said to be artificial brain diagrams [31, 32].
The artificial neural network model is always selected when indirect attribute is involved. Real estate valuation should also use this model as real estate considerations are also not linear. Therefore, as in [32 - 35], their research produces a positive result, and therefore promises to provide an accurate predictive model using a synthetic neural network algorithm. This program, however, has very limited functionality. ANN can model complex non-linear relationships as house price forecasts include many non-line variables.

Gradient Boost

Gradient boost was created [36] in 1999 and is a widely used machine learning algorithm due to its functionality, consistency and interpretation. Gradient boosting brings advanced features to a variety of machine learning tasks, such as multiple classification, click-through guess and level. With the advent of big data in recent years, gradient expansion is facing new challenges, especially in terms of balance between accuracy and performance [37]. There are a few parameters for increasing the gradient. To ensure a consistent balance between competence and familiarity, the following steps can be taken to select parameters: (1) Setting familiarity parameters (lambda, alpha), (2) lowering the level of learning and determining those correct parameters and [19].

IV. RELATED WORK

A total of 14 articles were reviewed and evaluated to take into account all factors affecting the price of the house. [3] In his article he stated that the square footage of a house unit is the most important difference in predicting the price of a house, followed by the number of bathrooms and the number of bedrooms. In addition, research shows that the value of a house increases by 2.6% when the floor area of a house is raised by 100 square meters. They also conclude that there building's operating year is 1 year lower, the value increase by 0.3 percent. In addition, the price of a house would increase by 10.4 or 13.7 percent, with one more bedroom or one more bathroom.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

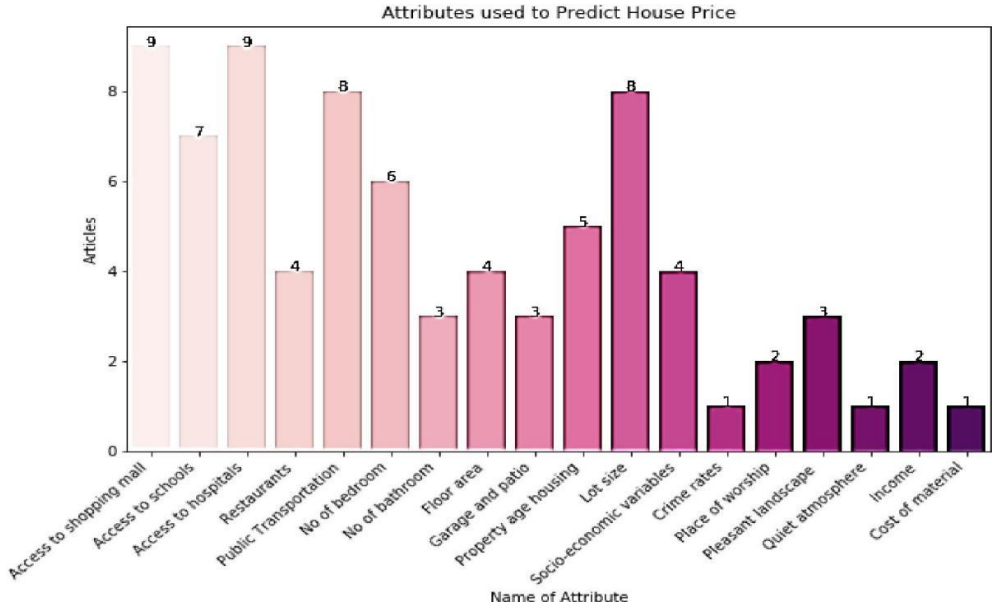

Fig.1. Types of Attributes used in Previous Study

Since the previous reviewed study, 19 attributors are said to have been widely used by other researchers to assess house prices. Characteristics from 12 articles, shown in Fig. 1, both collected and presented in a bar graph. It is therefore clear what factors researchers have used most in determining housing prices. The number in the bar bar above represents folding pages that use attributes as a prediction.

The diagram above shows that the shopping mall, the hospital entrance and the size of the house are the main sources of housing. Recent research has dominated controversies over local features, including access to shopping malls and hospitals, as well as building features, including the number of bedrooms and the size of the house.

In fact, [16] defined the area as an important house price forecast. Local contribution to lower housing prices as predicted from first-class residential districts to fifth-class residential districts. [5] pointed out that the four most affected housing prices are hospitals, schools, campuses, and recreational parks, which can be incorporated into local features.

In comparison, 8 out of 14 studies used structural factors to determine housing prices, including the number of bedrooms, the number of bathrooms and the size of the house. [38] specified that the three main factors influencing house sales prices are the total of square feet, the overall efficiency and total number of bath units. These findings are consistent with [7] finding that the number of bedrooms and bathrooms and the price of a house have a significant relationship. In the same vein, [17] he emphasized that the additional floor, bedroom and bathroom add 13 percent, 16 percent and 2 percent of the total household value, respectively.

In addition to geographical and structural properties, many researchers use local attributes to determine house prices. This can be seen in [12], where he says that the influence of neighbors has affected the price of a house, because residents may prefer a better place today. Neighborhood factors include low crime rates, a pleasant environment and a peaceful atmosphere. These factors will determine whether the price of the house is high or low.

Although only a few researchers have chosen an economic indicator, which includes individual income and housing costs, as a factor in determining the price of a home, we acknowledge that economic analysts have a significant impact on house prices. [9] stated in his study that the price of a house can be determined based on the income of an individual because the government plays a role in determining the price of a house based on individual financial circumstances. [12] support this study by saying that the relationship between house price and income is important to define affordability. This is one of the reasons why every person can afford to rent or rent. An examination of the key factors affecting the price of the house is important and is related to the initial research question of this study.

After examining the key factors affecting house price determination, the mining data model (within the context of this study is a forecast model) can be used to estimate house price.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

To predict house price, predictable models such as vector reduction support or artificial neural network were used. Predictable modeling uses data mining to predict what it saw during the research phase. Figure 2 shows the types of the predictive model used by researchers in a previous study. However, the four most popular speculation models, or more commonly known as dividers, are used by researchers to construct this predictive model which is multiple linear regression, support vector regression, artificial neural network and classifier gradient booster..

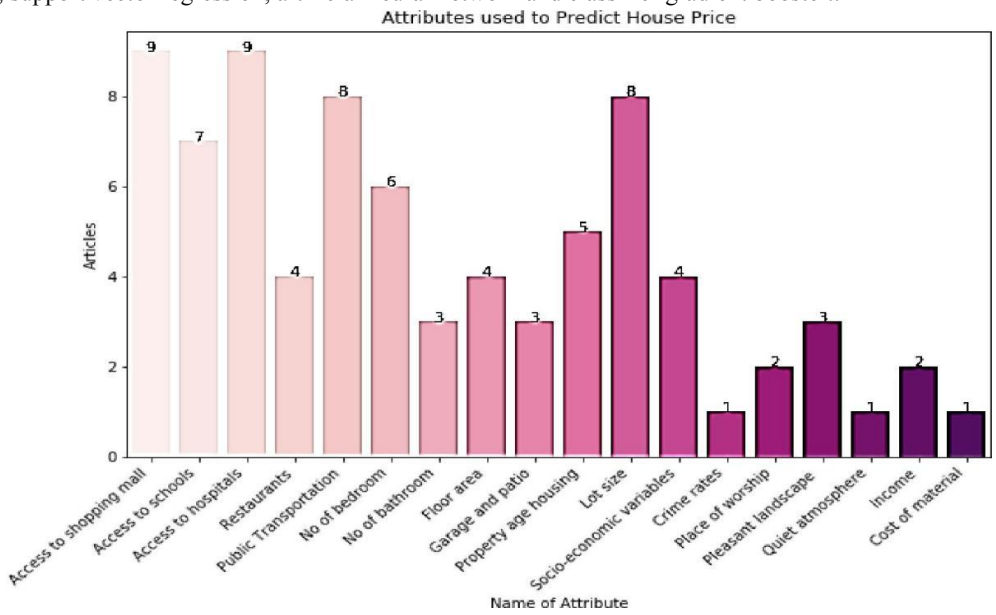

Fig 2. Types of Predictive Model used in Previous Study

Previous study	[38]	[3]	[7]	[39]	[34]	[32]	[40]	[22]	[10]	[19]	[41]	[11]	[23]	Data Source
Multiple Regression	✓	✓		✓			✓		✓		✓	✓		7
Support Vector			✓					✓					✓	3
Artificial Neural			✓		✓	✓	✓						✓	5
XGBoost	✓									✓				2

Table 2. Summarization of Predictive Model used in Previous Study

Table 2 shows that the most popular model used to predict house prices is to use multiple retrospective analysis. The hedonic price model was commonly used to distinguish key variables with another regression model, such as vector regression, multiple retrospective analysis and other models. At the time, the study conducted [38] selected XGBoost as the best model as it offers a lower RMSE value compared to other models in his study. Such an analysis is related to the second research questionnaire of this study.

VI. CONCLUSION

This paper examined and analyzed current research on key aspects of real estate prices and analyzed data mining techniques used to predict house prices. Technically, affordable housing such as access to shopping malls or other facilities is often more expensive than rural homes with limited resources.

An accurate forecasting model will allow investors or real estate buyers to determine the real estate price and real estate developers to determine the affordable housing price. This paper discusses the characteristics used by previous researchers to predict house prices using various speculative models. Taken together, the survey results showed the power of SVR, ANN and XGBoost in predicting house prices. These models are developed based on a few installation features and work very well for house price. In conclusion, the impact of this study was intended to assist and assist other researchers in developing a real-life model that could easily and accurately predict house prices. More work on the real model should be done using what we have found to validate it.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

REFERENCES

[1]. A. S. Temür, M. Akgün, and G. Temür, "Predicting Housing Sales in Turkey Using Arima, Lstm and Hybrid Models," J. Bus. Econ. Manag., vol. 20, no. 5, pp. 920–938, 2019, doi: 10.3846/jbem.2019.10190.

[2]. A. Ebekozien, A. R. Abdul-Aziz, and M. Jaafar, "Housing finance inaccessibility for low-income earners in Malaysia: Factors and solutions," Habitat Int., vol. 87, no. April, pp. 27–35, 2019, doi: 10.1016/j.habitatint.2019.03.009.

[3]. A. Jafari and R. Akhavian, "Driving forces for the US residential housing price: a predictive analysis," Built Environ. Proj. Asset Manag., vol. 9, no. 4, pp. 515–529, 2019, doi: 10.1108/BEPAM-07-2018-0100.

[4]. Choong Wei Cheng, "Statistical Analysis of Housing Prices in Petaling," Universiti Tunku Abdul Rahman, 2018.