

Housing Price Prediction using Machine Learning

Shivangi Chaurasiya¹, Hetvi Bhagat², Kartika Musle³, Shubham Gowda⁴

¹Member, Thakur Polytechnic, Mumbai, Maharashtra – 400101, India

²Member, Thakur Polytechnic, Mumbai, Maharashtra – 400101, India

³Member, Thakur Polytechnic, Mumbai, Maharashtra – 400101, India

⁴Member, Thakur Polytechnic, Mumbai, Maharashtra – 400101, India

Abstract - The House Price Index (HPI) is commonly used to estimate changes in house prices. Since house prices are closely related to other factors, such as location, region, and population, information other than a house price index is needed to predict single-family house prices. There have been far too many research papers that use traditional machine learning approaches to accurately predict house prices, but they rarely care about the performance of individual models and overlook the less common but complex ones. Therefore, to explore the effects of different features on prediction methods, this article will apply both traditional and advanced machine learning approaches to study the difference between different advanced models. This article will also comprehensively validate several techniques for implementing regression models and provide an optimistic result for house price forecasting.

Key Words: Machine learning algorithm, decision tree regression, housing price prediction of Boston, USA.

1. INTRODUCTION

The real estate market is an area where machine learning can be applied to improve and predict prices with high accuracy. Housing pricing is a basic customer decision model in which a number of criteria can be taken into account to predict the price of a desired home. Participants in the process are unfamiliar with the various analytical techniques available to estimate the price of real estate taking into account various characteristics related to the environment, the environment, other structures, etc. The design will help users invest in a property without going through an agent. It also reduces the risks associated with the transaction. Lasso regression was used as a model due to its transferable and probabilistic methodology in model selection. The result shows that the approach to the problem must be successful and have the ability to make predictions comparable to other house price prediction models. One of these methods used to predict house prices are based on multiple factors. By subway cities like Mumbai, consider the potential buyer of a home different factors such as location, size, proximity to parks, schools, hospitals, power generation plants and, above all, the house price.

1.1 LITERATURE SURVEY

Studies on Housing Price calculation approaches have been carried out in previous years. But few give prices based on these attributes. In this section, we will examine the old system and its approaches to find its strengths and limitations.

In 2018, Rohan Bafna, Anirudh Dhole, Ankit Jagtap, Asif Kazi, Arbaz Kazi specified that the price fluctuation rate should have a method for its traceability. All over the world, forms such as hedonic price method, multiple linear regression analysis and decision tree regression, travel cost method, fuzzy logic system, etc.

In 2014, Azme Bin Khamis, Nur Khalidah Khalilah Binti Kamarudin said that the accuracy of the house price estimation model is measured by several criteria. The values of R² and MSE were compared with the selected model.

In 2017 Calhoun CA said the proposed asset estimation models and house price indices represent the first attempt to implement standard hedonic methods to analyze housing values in Thailand and develop general regional HPIs.

In 2013, Frew J. Jud GD applied hedonic modeling techniques to estimate the value of a sample of apartments sold between 1996 and 1999 in the Portland, Oregon metropolitan area. The results provide a model for applying hedonic models to apartment valuation. The values decrease with increasing economic activity in the neighborhood and increase with residence income, but the effects are not statistically significant at reasonable levels of probability.

In 2016, Kanojia Anita specified hedonic pricing the method is relatively simple and non-controversial to apply, since it is based on real market prices and easily measurable data. If the data is readily available, can be relatively inexpensive to apply. If the data needs to be collected and compiled, the cost of an application can increase significantly.

In 2003, Limsombunchai, Christopher Gan Minsoo Lee empirically the power to predict the hedonic price model by comparing it to the real estate price prediction model of the artificial nerve network. The results of the hedonic

value models support the above conclusions. Although R2 hedonic pricing models are high (over 75%) for forecast in the sample, hedonic pricing models are no larger than network models. Models are high (over 75%) to predict, for example, hedonic value models are not superior to neural network models.

In 2018, Neelam Shinde, Kiran Gawande. Procedure described step by step to analyze the data set. and find the correlation between the parameters. He calculated the performance of each model using different performance matrix and compared them

In 2014 Azme Bin Khamis, Nur Khalidah Khalilah Binti Kamarudin explained the accuracy of the model in the house price forecast has been measured by a number of Criteria. Using the artificial neural network, the accuracy was approximately 26.47% more magnification than linear multiRegression. They conclude that the neural network model is preferred to predict the price of real estate compared to Multiple Linear regression model and can be used as another way to estimate the price of real estate in the future.

Model	Accuracy
Decision Tree	84.64
lasso	60.32
Logistic Regression	72.81%
SVR	67.81%

IN 2018 Sharmila Muralidharan, Katrina Phiri, Sonal

K. Sinha nal K. Sinha MSD, MAD and MAPE calculated for

residential and non-residential datasets, and analyzed in the case of the residential and non-residential models, the linear regression model showed the lowest error rate. However, it should be noted that due to to the fact that linear regression cannot deal with categorical variables, this is not entirely accurate image. Between decision trees and the neural network models, neural networks performed better.

In 2015, Hromada describes software that can be used for real estate price appraisals and real estate analysis real estate ads posted on the web in Czech Republic. The software systematically collects, analyzes and evaluates the data set on changes in real estate Marlet. For each semester, the software collects more than 65,000 quotes for the sale or rental of apartments, houses and building lots. all real estate

ads are permanently stored in the software database and thoroughly analyzed for

1.2 Attribute

The house price forecast can be divided into two categories, the first focusing on the characteristics of the house and the second on the model used in the house price forecast. Many researchers have produced a house price forecasting model, which includes. A survey by analyzing the price of existing housing in Jakarta, Indonesia, using the conceptual model and questionnaires. Based on the results, the attributes or factors that affect housing price differ for each housing construction in Jakarta, thus accepting the validity of this analysis, since the main objective of this research is to rank the factor or attributes that affect the price of housing. Several considerations influence the price of a home. According to the factors that influence housing prices, they can be classified into three categories: location, structural condition and neighborhood.

- **Location^[5]:**

Location is considered the most important characteristic in determining house prices. In his study, he also observed the importance of location attributes in deciding house price. The location of the property has been classified in a fixed location attribute. All of these studies point to a strong association between location attributes, such as distance to the nearest shopping center, or position offering a view of the hills or the coast, and variations in housing prices. Structural Another important feature that influences the price of housing is the structural structure or some researches have included it as physical attributes.

- **Structural characteristic^[5]:**

The structural characteristic is a characteristic that people can identify, whether it is the number of bedrooms and bathrooms, the floor area, the garage and the patio. These structural assets, often offered by home builders or real estate developers to attract potential buyers, satisfy the desires of potential buyers. In their previous study, structural attributes would be the primary consideration for home hunters when determining what to buy, as these attributes represent their market value. In their previous study, they stated that all of these attributes have a positive relationship with rising house prices.

- **Neighborhood^[5]:**

The qualities of the neighborhood can be taken into account when determining the price of the house. According to the effectiveness of public education, the social status of the community and the proximity to shopping centers often improve the value of a property. There is a substantial increase in housing prices from the fifth-class suburban community to the affluent neighborhood, as expected. However, the study found that these qualities tend to be culturally based, as they are not equally relevant across cultures.

2. Machine Learning Model

According to the housing demand assessment paradigm, it can be classified into two classes which are the traditional method and the advanced assessment method. The traditional evaluation scheme, including the multiple regression method and decision tree regression. The selection of models to use to predict house prices is quite critical as there are a variety of models available. One of the most widely used models in this area of research is regression analysis, which is used in many studies, including.

☺ **Multiple Linear Regression:** Regression analysis is a model used to determine the relationship between variables. To assess the correlation of the variables, the correlation coefficient or the regression equation can be used. Multiple regression models can determine which characteristics are most important in explaining the dependent variable. Multiple regression analysis also enables certain price predictions by capturing data from independent and dependent variables. The power of the multiple regression model can be seen when the strength of the relationship between dependent and independent variables is measured. They use a multiple regression model to describe improvements of an independent variable with a dependent variable. This model can be done by using housing price projection as separate and dependent variables, such as housing price, house size, property type, number of bedrooms, and ownership, among others. Therefore, housing price is defined as a target or dependency variable while other attributes are defined as independent variables to determine the main variables by identifying the correlation coefficient of each attribute.

☺ **Decision Tree Regression:** One of the most widely used and effective methods of authority education is the decision tree. Can be used to solve back-and-forth tasks, the latter is widely used in practice.

It is an organized group of trees with three types of nodes. The root node is the core node that represents the whole pattern and can be further subdivided into nodes. Internal nodes represent the data set assets and branch decision rules. Finally, a leaf node shows the result. This algorithm is very helpful in solving decision-making problems.

3. Related Work

A total of 14 items were reviewed and scored to capture all the attributes that influence home prices. Correlation is a term used to represent a statistical measure of the direct relationship between two variables. It can also be described as a measure of dependence between two different variables. If there are multiple variables and the goal is to find the connection between all these variables and store them using the appropriate data structure, a

matrix data structure will be used. Such a matrix is called a related matrix.

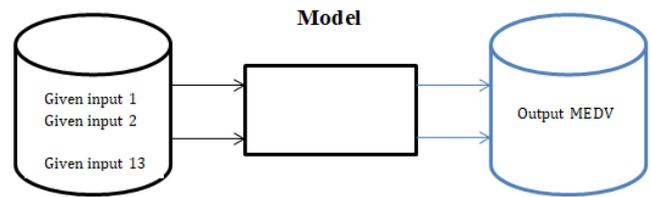


Fig 1: A schematic of model with thirteen input and one output.

- If the value is 1, it is said to be positive correlation between two variables. This means that when one variable increases, the other variable also increases [6].
- If the value is -1, it is said to be -ve correlation between two variables. This means that when one variable increases, the other variable decreases [6].
- If the value is 0, there is no correlation between two variables.[6].

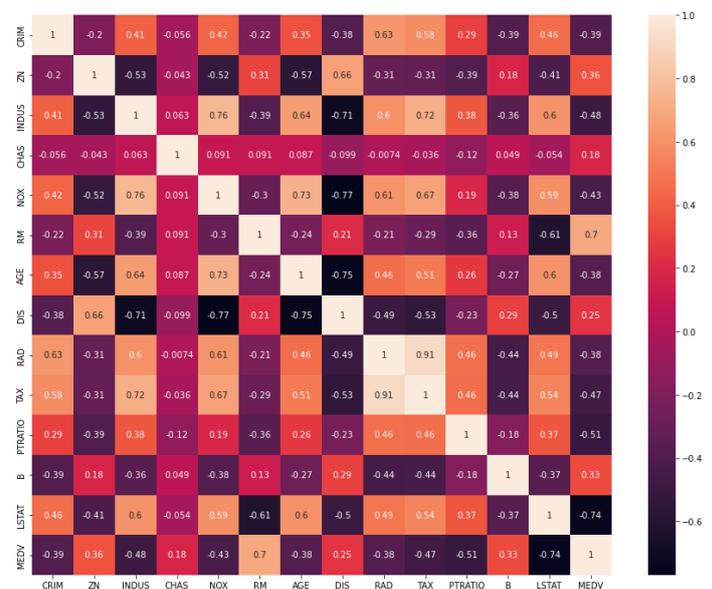


Fig 2: Attributes and correlation

According to figure 2 our attributes RM (Average number of rooms per dwelling) is highly correlated which is 0.7 and highly negative correlated is LSTAT (Percentage of lower status of the population). This LSTAT is -0.74 negative correlated.

3. CONCLUSIONS

This article reviews and discusses current research on significant real estate price attributes and discusses data mining techniques used to predict real estate prices. Technically, houses with a strategic location, such as

accessibility to the mall or other facilities, tend to be more expensive than houses in rural areas with a limited number of facilities. The accurate prediction model would allow home investors or buyers to determine the realistic price of a house, as well as real estate developers to decide the affordable price of the house. This article discussed the attributes used by previous researchers to predict the price of a house using various prediction models. These models were developed based on various input attributes and perform significantly positively with real estate prices. In conclusion, the impact of this research was to aid and assist other researchers in developing a real-world model that can easily and accurately predict house prices. It is necessary to continue working on a real model with the use of our results to confirm them.



Shubham Gowda a member of group studying in Thakur Polytechnic worked on the accuracy and other implementations of the project.

REFERENCES

1. D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," *Science*, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.
2. M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
3. R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.
4. K. Elissa, "Title of paper if known," unpublished.
5. https://www.researchgate.net/publication/347584803_House_Price_Prediction_using_a_Machine_Learning_Model_A_Survey_of_Literature
6. <https://vitalflux.com/correlation-heatmap-with-seaborn-pandas/>

BIOGRAPHIES



Shivangi Chaurasiya a member of group studying in Thakur Polytechnic executed the program of the project.



Hetvi Bhagat a member of group studying in Thakur Polytechnic worked on the base of the project.



Kartika Musle a member of group studying in Thakur Polytechnic worked on the criteria of the project.