iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** II **Month of publication:** February 2022

DOI: https://doi.org/10.22214/ijraset.2022.40272

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue II Feb 2022- Available at www.ijraset.com

House Price Prediction Using Regression

Meha Ajay Kumar Shukla
[1]*Information Technology, Gujarat Technological University*

Abstract: ***The housing sector is the second largest employment provider after agriculture sector in India and is estimated to grow at 30% over the next decade. Housing is one of the major sectors of real estate and is well complemented by the growth of the urban and semi-urban accommodations. Ambiguity among the prices of houses makes it difficult for the buyer to select their dream house. The interest of both buyers and sellers should be satisfied so that they do not overestimate or underestimate price. Our system provides a decisive housing price prediction model to benefit a buyer and seller or a real estate agent to make a better-informed decision system on multiple features. To achieve this, various features are selected as input from feature set and various approaches can be taken such as Regression Models or ANN.***

Keywords: Machine Learning, Artificial Intelligence, Supervised Machine Learning, Regression, Artificial Neural Network (ANN)

I. INTRODUCTION

Food, clothing, and housing (shelter) are the primary requirements of life. The availability of these requirements increases the physical efficiency and productivity of the people. So housing is a factor of prime importance in human resource development. At one point in life, everybody has to deal with the housing dilemma. For many people housing is one of the major investment of their life, people pay their fortune to buy the Dreamhouse.

The gap between the demand and the unsold residential inventory is increasing day by day. Most of the time the seller/buyer has no idea about the house price, and they may overstate or understate the price. This works as an advantage for the middleman to manipulate the price to maximize their profits.

In my current project, I am trying to develop a better, efficient, and accurate house price prediction system. It is a supervised learning method in which we are following a sequence of steps: First step is Data Collection, the second step is Developing a model for the dataset, the third step is finding out the relationship among the various features with its respective price and last is to predict the price of a particular house.

II. ALGORITHMS USED

A. Multiple Regression

The most common form of linear regression is Multiple linear regression. According to predictive analysis, multiple linear regression explains the relationship between one continuous dependent variable and two or more independent variables. The independent variables can be continuous or categorical (dummy coded as appropriate).

A central task of the multiple linear regression analysis is to fit a single line through a scatter plot. Accurately the multiple linear regression fits a line through a multi-dimensional space of data points. The simplest form has one dependent and two independent variables. The dependent variable may also be referred to as the outcome variable or regressors. The independent variables may also be referred to as the predictor variables or regressors.

There are 3 major uses for multiple linear regression analysis. First, it might be used to identify the strength of the effect that the independent variables have on a dependent variable.

Second, it can be used to forecast effects or impacts of changes. That is, multiple linear regression analysis helps us to understand how much will the dependent variable change when we change the independent variables. For instance, a multiple linear regression can tell you how much GPA is expected to increase (or decrease) for every one point increase (or decrease) in IQ.

Third, multiple linear regression analysis predicts trends and future values. The multiple linear regression analysis can be used to get point estimates. An example question may be “what will the price of gold be 6 month from now?”

When selecting the model for the multiple linear regression analysis, another important consideration is the model fit. Adding independent variables to a multiple linear regression model will always increase the amount of explained variance in the dependent variable (typically expressed as R^2). Therefore, adding too many independent variables without any theoretical justification may result in an over-fit model.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue II Feb 2022- Available at www.ijraset.com

B. Artificial Neural Network

Artificial neural networks are one of the main tools used in machine learning. As the "neural" part of their name suggests, they are brain-inspired systems which are intended to replicate the way that we humans learn. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that transform the input into something that the output layer can use. They are excellent tools for finding patterns which are far too complex or numerous for a human programmer to extract and teach the machine to recognize.

While neural networks (also called "perceptron") have been around since the 1940s, it is only in the last several decades where they have become a major part of artificial intelligence. This is due to the arrival of a technique called "backpropagation," which allows networks to adjust their hidden layers of neurons in situations where the outcome doesn't match what the creator is hoping for — like a network designed to recognize dogs, which misidentifies a cat.

Another important advance has been the arrival of deep learning neural networks, in which different layers of a multilayer network extract different features until it can recognize what it is looking for.

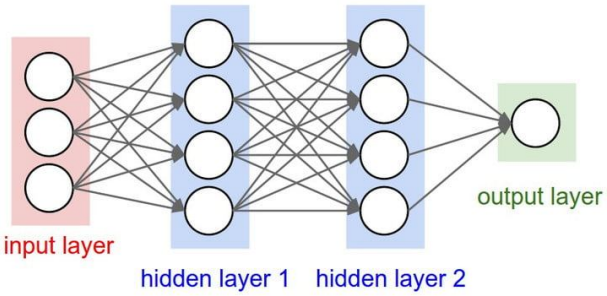


Fig. 1 Neural Network

III. STEPS TAKEN

A. Data Gathering and Handling

The first and foremost step is to gather good quality data with integrity. Various preprocessing steps must be done on this data to ensure that the data is good to be fed to the model designed. The data may contain noise, wrong data, or missing data. Those all must be removed unless the model will perform poorly. Also, the data is split into training and test data.

B. Model Development

The model is created here. The coding takes place in this step. All the parameters are defined, and all the attributes are taken into consideration

C. Training, Testing and Optimizing the model

After the model is developed, it is trained on the training dataset. The model will form its logics according to this dataset. After the model is trained, it is tested on the test dataset which it has never seen. This way we can check the accuracy/precision of the model. If the parameters of the model can be tweaked and adjusted to obtain a better accuracy/precision, then it is done to optimize the model.

IV. METRICS CONSIDERED

Various metrics could be considered to check the precision of the model. We have taken in consideration the following metrics.

A. Variance Score

The variance is a numerical value used to indicate how widely individuals in a group vary. If individual observations vary greatly from the group mean, the variance is big; and vice versa.

It is important to distinguish between the variance of a population and the variance of a sample. They have different notation, and they are computed differently. The variance of a population is denoted by σ2; and the variance of a sample, by s2.

The variance of a population is defined by the following formula:

$$\sigma^2 = \Sigma (X_i - \bar{X})^2 / N$$

Fig. 2 Variance of population

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

where σ^2 is the population variance, $\bar{X}$ is the population mean, X_i is the *i* element from the population, and N is the number of elements in the population.
The variance of a sample is defined by slightly different formula

$$s^2 = \Sigma (x_i - \bar{x})^2 / (n - 1)$$

Fig. 3 Variance of a sample

where s^2 is the sample variance, $\bar{x}$ is the sample mean, x_i is the *i*th element from the sample, and n is the number of elements in the sample. Using this formula, we conclude that the variance of the sample is an unbiased estimate of the variance of the population. And finally, the variance is equal to the square of the standard deviation.

B. Absolute Error

Absolute Error is the amount of error in your measurements. It is the difference between the measured value and "true" value.
The formula for the absolute error (Δx) is:

$$(\Delta x) = |x_i - x|,$$

Fig. 4 Absolute error

where x_i is the measurement, x is the true value.

C. Mean Absolute Error

The Mean Absolute Error (MAE) is the average of all absolute errors. The formula is:

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} | x_i - x |$$

Fig. 5 Mean absolute error

Where n is the number of observations.

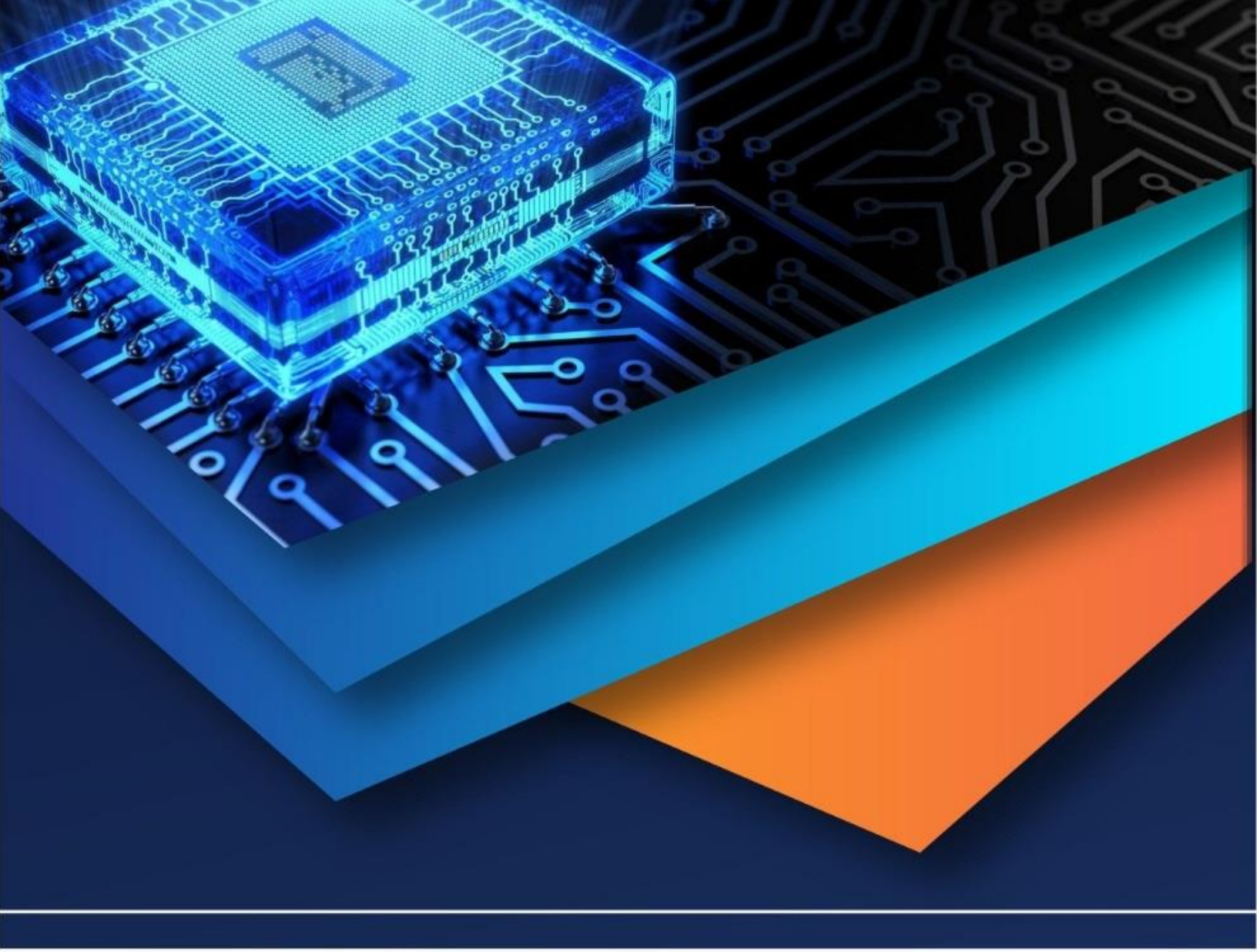
V. CONCLUSIONS

The system developed can provide a prediction accuracy of 87% for a housing price. It uses Multi Linear Regression on defined variable optimally. The use of Neural networks has increased the efficiency of the algorithm. The accurate predictive model will help user invest in valuable property. In future additional features benefiting the customer can be added to the system without disturbing core functionality. Training the model with larger set of data can increase the efficiency and benefit the model to get higher accuracy with more accurate dependent variables.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] K. Elissa, "Title of paper if known," unpublished.
[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)