



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 9      Issue: V      Month of publication: May 2021**

**DOI: <https://doi.org/10.22214/ijraset.2021.34804>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Review on House Price Prediction using Machine Learning

Monika Sahu<sup>1</sup>, Mrs. Awantika Singh<sup>2</sup>

<sup>1</sup>M.tech Scholar, <sup>2</sup>Assistant Professor, Computer Science & Engineering Department, Kalinga University, Raipur (C.G.), India

**Abstract:** *House costs increment consistently, so there is a need for a framework to anticipate house costs later on. House cost expectation can assist the designer with deciding the selling cost of a house and can assist the client with orchestrating the perfect chance to buy a house. There are three factors that impact the cost of a house which includes physical condition, idea and area. House price forecasting is a crucial topic of land. Machine learning techniques are applied to research historical property transactions in World to get useful models for house buyers and sellers. In the current paper we examine about the expectation of future lodging costs that is produced by AI calculation. For the choice of forecast strategies we look at and investigate different expectation techniques. For the choice of prediction methods we compare and explore various prediction methods.*

*We utilize regression as our model due to its adaptable and probabilistic methodology on model selection. Our result exhibit that our approach of the difficulty got to achieve success, and has the power to process predictions that might be comparative with other house cost prediction models. This study utilizes machine learning algorithms as a search method that develops housing price prediction models.*

*We create a housing cost prediction model in sight of machine learning algorithm models. House price prediction on a knowledge set has been done by using all the above mentioned techniques to seek out the simplest among them. The motive of this paper is to assist the vendor to estimate the selling cost of a house perfectly and to assist people to predict the precise time slap to accumulate a house. A number of the related factors that impact the value were also taken into considerations like physical conditions, concept and site etc. Here we aim to form our evaluation supported every basic parameter that's considered while determining the worth. We use linear regression techniques during this pathway, and our result aren't sole determination of one technique rather it's the weighted mean of varied techniques to offer most accurate results. The results proved that this approach yields minimum error and maximum accuracy than individual algorithm applied.*

**Keywords:** *House price prediction, Machine Learning, AI, Model, Linear Regression, Algorithm.*

## I. INTRODUCTION

Thousands of homes are sold everyday. And houses price increases every year. So there is a need for system to predict house prices in the future. During this project, a machine learning model is proposed to predict a house price supported data associated with the house (its size, the year it had been inbuilt, etc.). Prediction house prices are expected to assist people that decide to buy a house in order that they can know the worth home in the longer term, then they will plan their finance well. Additionally, house price predictions also are beneficial for property investors to understand the trend of housing prices during a certain location. House price prediction can help the developer determine the asking price of a house and may help the customer to rearrange the proper time to get a house.

There are three factors that influence the worth of a house which include physical conditions, concept and site. Machine learning may be a subfield of AI that works with algorithms and technologies to extract useful information from data. Machine learning methods are appropriate in big data since attempting to manually process vast volumes of knowledge would be impossible without the support of machines.

Machine learning in computing attempts to unravel problems algorithmically instead of purely mathematically. Therefore, it's supported creating algorithms that let the machine to find out. There are two type in machine learning which are supervised and unsupervised.

Supervised is where the program gets trained on pre-determined set to be ready to predict when a replacement data is given. Unsupervised is where the program tries to seek out the connection and therefore the hidden pattern between the info Several Machine Learning algorithms are wont to solve problems within the world today. Thus, this paper attempts to use Linear Regression algorithms to match their performance when it involves predicting values of a given dataset.

## II. LITERATURE SURVEY

Throughout the most recent twenty years there have been an enormous number of exact examinations dissecting land costs. Kilpatrick showed the helpfulness of time-arrangement relapse model which utilized financial information to give figure of Central Business District (CBD) land cost in moving business sector. Wilson et al contemplated the private property market represents a significant extent of UK monetary movement. Valuers gauge property estimations dependent on current bid costs. In this paper, the public lodging exchange information was prepared utilizing Artificial Neural Organizations (ANN), which estimates future pattern of the real estate market. Imprint and John fostered a relapse model with empty land deals. The model disclosed up to 93% of the market esteems. Wang and Tian utilized the wavelet Neural Network (NN) to conjecture the land value file. This sort of wavelet NN incorporated the value of the wavelet examination and the custom NN. It additionally contrasted the determining result and smoothing strategy and the NN figure. Zhangming estimated the land value list by utilizing the Back Propagation (BP) NN. The BPN utilized the sigmoid capacity. Tinghao utilized the Auto Regressive Integrated Moving Average (ARIMA) model and conveyed the illustrative investigation on year information from 1998 to 2006. He utilized the set up model to make the gauge to the land value record of 2007. An indulgent relapse on the cost of land recommended that true arrangement contrasts between political purviews have significantly affected land costs between 1970 and 1980. Steven and Albert utilized 46,467 private properties crossing 1999 - 2005 and exhibited that utilizing coordinated with sets that comparative with straight libertine estimating models, ANN produce lower dollar evaluating mistakes, had more noteworthy valuing exactness out-of-test, and extrapolate better from more unpredictable evaluating conditions. ANN is more qualified to indulgent models that use huge quantities of factors. Sampath kumar and Santhi examined the land value pattern of Sowcarpet which is the focal part. They created measurable model utilizing financial factors and anticipated that the yearly ascent in land cost would be of 17%. Urmila revealed that the previous patterns were dissected to determine the pace of development or decrease and the patterns are utilized in determining. Monetary boundaries may be acquainted with plan more practical relationship. A portion of different procedures they Mansural Bhuiyan and Mohammad Al Hasan 2016 use is relapse, profound figuring out how to take in the idea of models from the past outcomes (the property/land which were auctions off beforehand which are utilized as preparing data). There are various models utilized, for example, straight model information utilizing just one element, multivariate model, utilizing a few highlights as its info and polynomial model utilizing the information as cubed or squared and henceforth determined the root mean squared mistake (RMS an incentive) for the model.

## III. RESEARCH GAP

In all over world, there are various land ordered sites where properties are recorded for sell/purchase/lease purposes. Be that as it may, in every one of those sites we will see part of irregularities as far as valuing of a condo and there are a few situations when comparative condos are valued distinctively and hence there's part of in-straightforwardness. Now and then the shoppers may feel the valuing isn't supported for a particular recorded condo however there no way to affirm that by the same token. Legitimate and defended costs of properties can introduce huge loads of straightforwardness and trust back to the significant domain industry, which is critical with respect to numerous purchasers particularly in India the exchange costs are very high and resolving this issue will help both the buyers and along these lines the land industry inside the day's end . We propose to utilize AI and counterfeit. Knowledge strategies to foster a calculation which will anticipate lodging costs dependent on certain info highlights. The business utilization of this calculation is that arranged sites can straightforwardly utilize this calculation to anticipate costs of most recent properties that are having the opportunity to be recorded by taking some information factors and foreseeing the privilege and defended cost for example try not to take value contributions from clients and hence not letting any blunder crawling inside the framework. This concentrate on proactive valuing of homes inside the Indian setting has never been accounted for prior inside the writing to the least complex of our insight.

## IV. METHODOLOGY

- 1) *Dataset*: The dataset utilized during this work is obtained from land Agents within the US to gauge the prices of homes. This dataset was intended to help them find the foremost appropriate cost of the house within the in respective locations within the US. For predicting the prices of the homes, the following attributes are identified and included, in conjunction with the acronym used for his or her representation within the dataset snippet as shown. Base Price (price), Size of Plot, House age, Number of Bedrooms, Number of rooms, Area population, Address, of the house. These attributes, or predictor variables, are the factors which are majorly considered during house purchase and thus influence the pricing of homes .The dataset utilized in this work is obtained from land Agents within the US to judge the costs of homes.



The data-set we have used is a group of houses of USA .The size of the dataset is of 5,000 houses which are divided into training data and testing data. The dataset contains 7 columns and 5000 rows with CSV extension. The data contains the following columns:

- a) 'Avg. Area Income' – Avg. The income of the householder of the city house is located.
  - b) 'Avg. Area House Age' – Avg. Age of Houses in the same city.
  - c) 'Avg. Area Number of Rooms' – Avg. Number of Rooms for Houses in the same city.
  - d) 'Avg. Area Number of Bedrooms' – Avg. Number of Bedrooms for Houses in the same city.
  - e) 'Area Population' – Population of the city.
  - f) 'Price' – Price that the house sold at.
  - g) 'Address' – Address of the houses.
- 2) *Fetch the Data:* Set by the help of the pandas in python platform and analyze the data set.
  - 3) *Data Preprocessing:* Data Preprocessing: Is a process to convert raw data into meaningful data using different techniques. Data in real word are incomplete, noisy, duplicate or inconsistent, so we can convert this data into accuracy, completeness, consistency by various data preprocessing techniques.

Major steps in data preprocessing:

- a) Data Cleaning
  - b) Data Integration
  - c) Data Reduction
  - d) Data Transformation
  - e) Data Discretization
- 4) *Train Test Split Evaluation:* It is a technique for evaluating the performance of a machine learning algorithm. We have segregated dependent and independent variable in two different vectors respectively x and y, we handled our by data preprocessing. We have 500 records inside of our Data set and we split the data into training set and test set. Vector y is dependent or target variables. Training part of data is there that will pass to our machine learning model for the training purpose and one machine learning model is trained, the rest remaining data that is test data will use to validate machine learning model. We split a data in percentage between 0 and 1 for either the train or test datasets. Here we take test set with the size of 0.40 (40 %) means that the remainder percentage 0.60 (60 %) is assigned to the training set.

#### A. Linear Regression

Linear regression is a statistical model that attempts to point out the connection between two variables with the equation. It is the one of the easiest algorithm in machine learning. It is calculated by using formula,

$$y = mx + b$$

y = Dependent variable

m = Coefficient rate and slope of line

x = Independent variable

b = Where line crosses the y-axis

From the above formula we try to find the value of x and y that every value of x has a corresponding value of y in it if it is continuous. The reason for this is linear regression is always continuous. The output of the linear regression is the value of the variable. The accuracy or the great fit is calculated by using the r squared method.

Why we select linear Regression

Regression model predicts a continuous variables and linear Regression predicted value is continuous and it also remove the outliers there for our model will perform in a better way.

Linear regression are easily comprehensive and transparent. They can be understood very easily because it is represented by simple mathematical notation.

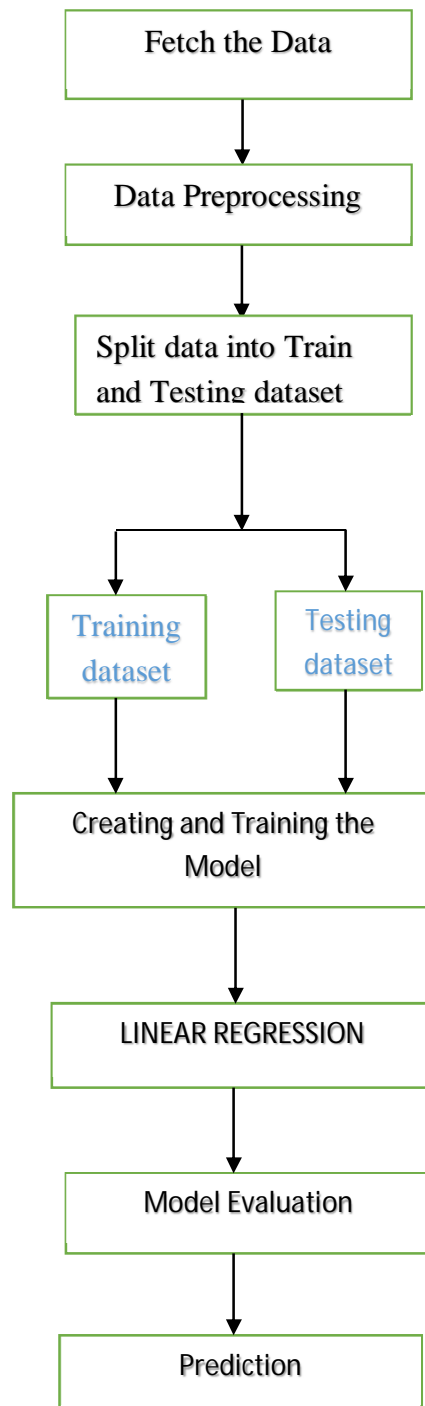


Fig. Data flow diagram

## V. CONCLUSION

Linear regression means we will predict a variable from an independent one, so whenever we'd like to understand from the start whenever we add information. The regression curve is vital because it makes the estimation of a variable more accurate and it allows the estimation of a response variable for people with values of the carrier variable not included within the data. We also inferred there are two methods of predicting a variable either from within the range of values of experimental variable of the sample given (interpolation) or outside this range (extrapolation). The house price and the linear regression is the most effective model for our dataset.

## REFERENCES

- [1] Akash Dagar and Shreya Kapoor, "A Comparative Study on House Price Prediction", International Journal for Modern Trends in Science and Technology, 6(12): 103-107, 2020.
- [2] Puneet Tiwari<sup>1</sup>, Varun Singh Thakur 'Review on house price prediction through Regression technique' International Journal of Scientific Progress and Research (IJSPR), Issue 173, volume 73.
- [3] R Manjula, Shubham Jain, Sharad Srivastava and Pranav Rajiv Kher 'Real estate value prediction using multivariate regression models' et al 2017 IOP Conf. Ser.: Mater. Sci. Eng. **263** 042098
- [4] Housing Price Prediction using Machine Learning Yashraj Garud , Hemanshu Vispute , Nayan Bisai and Prof. Madhu Nashipudimath<sup>4</sup> Volume: 07 Issue: 05 | May 2020 (IRJET)
- [5] G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu "house price prediction using machine learning" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019.
- [6] Darshil Shah, Harshad Rajput, Jay Chheda "house price prediction using machine learning and RPA"International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 03 | Mar 2020
- [7] "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization "(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017
- [8] Eli Beracha, Ben T Gilbert, Tyler Kjorstad, Kiplan womack, "On the Relation between Local Amenities and House Price Dynamics", Journal of Real estate Economics, Aug. 2016.
- [9] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest?" In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7376 LNAI, 2012, pp. 154–168, ISBN: 9783642315367. DOI: 10 . 1007 / 978 - 3 - 642 -31537-4\ 13.
- [10] S. Ray, "CatBoost: A machine learning library to handle categorical (CAT) data automatically," CatBoost: Analytics Vidhya, 14-Aug-2017.
- [11] R. J. Shiller, "Understanding recent trends in house prices and home ownership," National Bureau of Economic Research, Working Paper 13553, Oct. 2007. DOI: 10.3386/w13553.
- [12] S. C. Bourassa, E. Cantoni, and M. Hoesli, "Predicting house prices with spatial dependence: a comparison of alternative methods," Journal of Real Estate Research, vol. 32, no. 2, pp.139–160, 2010.
- [13] Li, Li, and Kai-Hsuan Chu. "Prediction of real estate price variation based on economic parameters." Applied System Innovation (ICASI), 2017 International Conference on.IEEE, 2017.
- [14] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of machine learning research 12.Oct (2011): 2825-2830.
- [15] Byeonghwa Park , Jae Kwon Bae (2015). Using machine learning algorithms for housing price prediction , Volume 42, Pages 2928-2934.
- [16] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, 2015. Introduction to Linear Regression Analysis.
- [17] A. Azadeh, B. Ziaei, and M. Moghaddam, —A hybrid fuzzy regression-fuzzy cognitive map algorithm for forecasting and optimization of housing market fluctuations, I Expert Syst. Appl., vol. 39, no. 1, pp. 298–315, 2012.
- [18] F. S. Gharehchopogh, T. H. Bonab, and S. R. Khaze, —A Linear Regression Approach to Prediction of Stock Market Trading Volume: A Case Study, I Int. J. Manag. Value Supply Chain., vol. 4, no. 3, pp. 25–31, 2013.
- [19] H.-I. Hsieh, T.-P. Lee, and T.-S. Lee, —A Hybrid Particle Swarm Optimization and Support Vector Regression Model for Financial Time Series Forecasting, I Int. J. Bus. Adm., vol. 2, no. 2, pp. 48–56, 2011
- [20] F. Marini and B. Walczak, —Particle swarm optimization (PSO). A tutorial, I Chemom. Intell. Lab. Syst., vol. 149, pp. 153–165, 2015.
- [21] A. Hayder M. Albehadili Abdurrahman and N. . Islam, —An Algorithm for Time Series Prediction Using, I Int. J. Sci. Knowl. Comput. Inf. Technol., vol. 4, no. 6, pp. 26–33, 2014.
- [22] Y. P. Anggodo and W. F. Mahmudy, —Automatic Clustering and Optimized Fuzzy Logical Relationship for Minimum Living Needs Forecasting, I J. Environ. Eng. Sustain. Technol., vol. 4, no. 1, pp. 1–7, 2017.



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)