



# INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

## HOUSE PRICE PREDICTION USING MACHINE LEARNING

MS.A.VIDHYAVANI<sup>1</sup>,O.BHARGAV SATHWIK<sup>2</sup>,HEMANTH.T<sup>3</sup>,

VISHNU VARDHAN YADAV.M<sup>4</sup>.

<sup>1</sup>Assistant Professor, Department of Computer Science and Engineering ,SRMISTRamapuram Chennai.

<sup>2</sup>Department of Computer Science and Engineering, SRMIST Ramapuram,Chennai.[UG Scholar]

<sup>3</sup>Department of Computer Science and Engineering, SRMIST Ramapuram,Chennai.[UG Scholar]

<sup>4</sup>Department of Computer Science and Engineering, SRMIST Ramapuram,Chennai.[UG Scholar]

### 1.ABSTRACT:

This project provides us an overview on how to predict house prices using various machine learning models with the help of different python libraries. This proposed model considers as the most accurate model used for calculating the house price and provides a most accurate prediction. This provides a brief introduction which will be needed to predict the house price. This project consists of what and how the house price model works with the assistance of machine learning technique using scikit-learn and which datasets we will be using in our proposed model.

Predicting the price of a house helps for determine the selling price of the house in a particular region and it help people to find the correct time to buy a home. In this task on House Price Prediction using machine learning, our task is to use data to create a machine learning model to predict house prices in the given region. We will implement a linear regression algorithm on our dataset. By using real world data entities, we are going to predict the price of the house in that area. For better results we require data pre-processing units to improve the efficiency of the model. for this project we are using supervised learning, which is a part of machine learning. We have to go through different attributes of the dataset.

### Keywords: -

Python, Machine learning, scikit-learn, python libraries, data pre-processing, Linear regression algorithm, Supervised learning

### 2.INTRODUCTION:

One of the basic requirements of livelihood in the recent world is to buy a house of your own. The price of the house may depend on various factors. Real estate agents and many who are involved in selling the house want a price tag on the house which would be the real worth of buying the house. The prediction of the price of the house is often very hard for the inexperienced.

House is a fundamental need to someone and their costs range from place to place primarily based totally at the facilities to be had like parking area, location, etc. The residence price is a factor that issues a lot of citizens either wealthy or white-collar magnificence as possible by no means decide or gauge the value of a residence primarily based totally on place or places of work accessible. Purchasing a residence is the best and unique desire of a own circle of relatives because this expends the whole thing of their funding budget and every now and then cover them under loan. It is a hard challenge for expecting the correct value of residence price. This proposed version could make viable who are expecting the precise costs of house. This model of Linear regression in machine learning takes the internal factor of house valuation dependencies like area, no of bedrooms, locality etc. and external factors like air pollution and crime rates. This Linear regression in machine learning gives the output of price of the house with more accuracy.

Here in this project, we are going to use linear regression algorithm (a supervised learning approach) in machine learning to build a predictive model for estimation of the house price for real estate customers. In this project we are going to build the Machine learning model by using python programming and other python tools like NumPy, pandas, matplotlib etc. We are also using scikit-learn library in our approach of this project. For going further into this project, we prepare dataset which consists of features like no of bedrooms, area of house, locality, etc. After preparing the dataset we will use 80 percentage of data for training the ml model and 20 percentage of data for testing the ml model.

Buying a house is one of the most important decisions for an individual. The price of a house is dependent on wide factors which are its features such as no of bedrooms, area of construction, locality of the house etc. and many external factors such as air pollution and crime rate. These all factors make the prediction of price of the house more complex. This house price prediction is beneficial for many real estates. So, an easier and accurate method is needed for the prediction of house price.

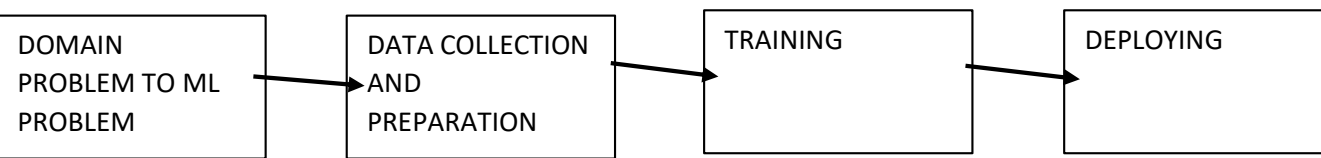
This project is proposed to predict the price of a house for real estate customers based on their preferences like no of rooms, no of bedrooms, area of the house, locality etc. This machine learning model to predict the house price will be very helpful for the customers, because it is being the major issue for the people, as they can never estimates the price of the house on the basis of locality or other features available.

To tackle this problem, we are going to design a ML model with the assistance of a linear regression algorithm (supervised learning method).

In order to predict the house price some regression techniques such as multiple regression, ridge and lasso regression, support vector regressions, and boosting algorithm like extreme gradient boost regression, are used. This technique is used for building a predictive model to predict the house price. Among the Machine learning models prepared by using these methods the best model is picked by doing a comparative analysis on these models.

Comparative analysis is a statistical method that is used to find the errors between the machine learning models and the model with less error is picked as the best model for prediction.

### 3.PROJECT WORKFLOW:



#### Domain problem to ML problem: -

Domain Problem:

In this project we are going to predict the price of a house for real estate customers based on their preferences like no of bedrooms, area of the house, locality etc.

ML problem:

This problem can be solved using supervised learning in ML. It is treated as a Regression problem as the target is continuously valued.

#### Data collection and preparation: -

This phase consists of three parts they are: -

- 1.Data collection
- 2.Data pre-processing
- 3.EDA and feature Engineering

Data collection:-In order to proceed further into this project, first of all we should collect the data, after collecting a dataset, then we should pre-process the data and then we do the exploratory data analysis on the given data set.

Data pre-processing:- This pre-processing of data consists of Handling missing values which means some of values in the dataset may be missing. We should handle those missing values.

Handling duplicates which mean some of the values in the dataset may be their multiple times we should handle those duplicate values that means we should delete those duplicates.

Handling outliers, it consists of identifying the outliers and hence removing those outliers.

EDA and feature Engineering:-

By using matplotlib package in python we will do EDA on a given dataset. Normally we will do EDA on a given data set to establish a relation between given feature of the data set.

#### Training: -

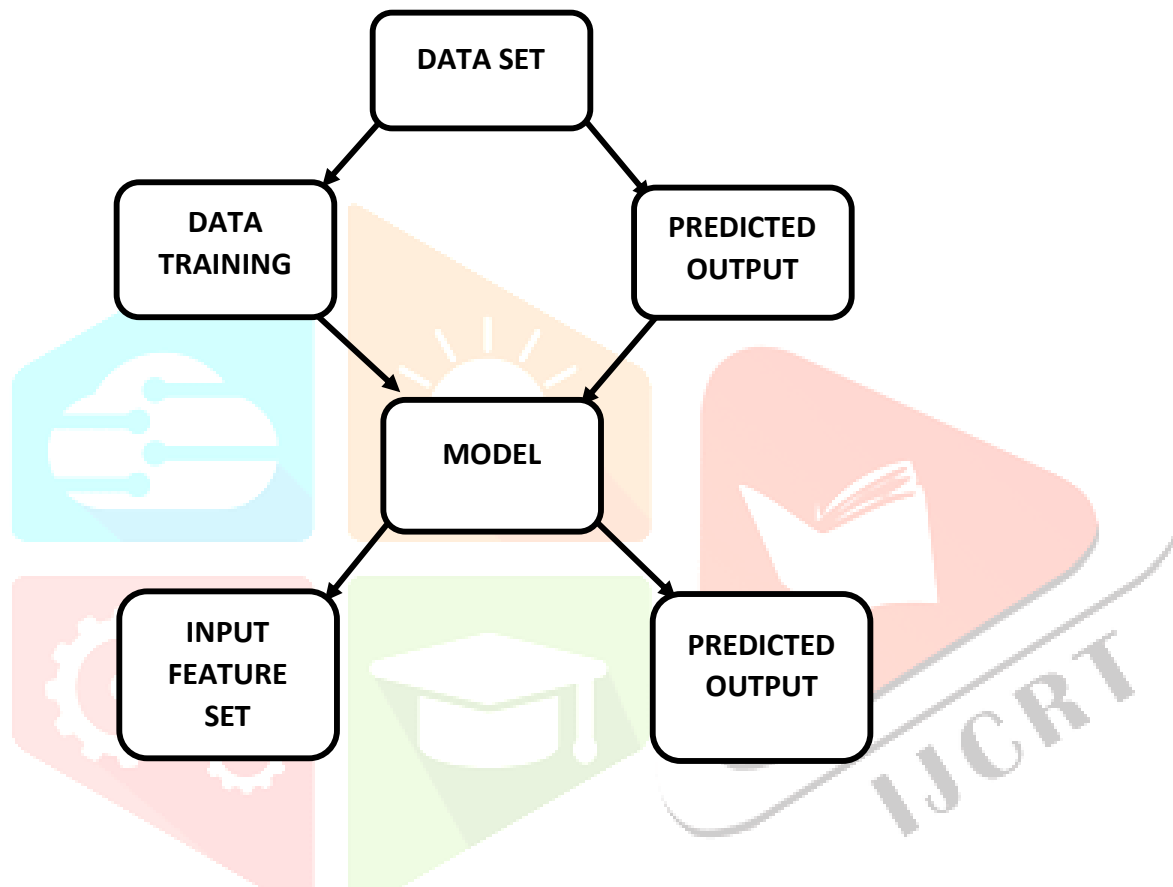
This phase consists of 4 parts they are:-

- 1.Choosing a machine learning algorithm
- 2.Training
- 3.Evaluation
- 4.Hyperparameter tuning

**Deploying: -**

Deploying a machine learning model involves:-

1. Interpreting the results
2. Packaging and shipping
3. Maintenance and monitoring

**4. ARCHITECTURAL DIAGRAM:**

In this project a linear regression is used for predicting the price of the house using a machine learning approach.

**5. LINEAR REGRESSION ALGORITHM:**

Linear regression is a supervised learning algorithm in machine learning field.

It is a statistical method which aims for developing a linear relation between a dependent variable or target variable and one or more independent variables by developing a linear equation to the given data.

Formulae:

$$Y = b_0 + b_1x + E$$

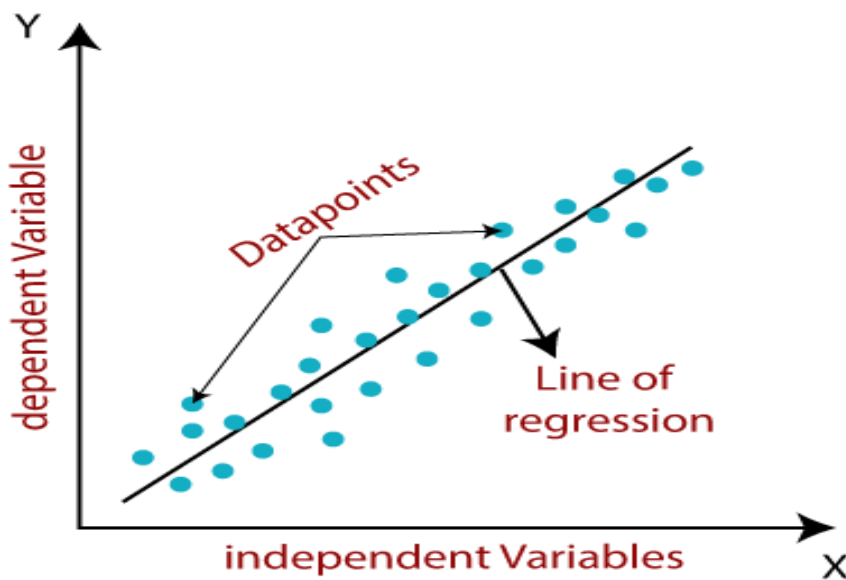
Where, Y=target variable or dependent variable

X=independent variable

E=error

b1=linear regression coefficient

b0=intercept



Actually, Linear regression are of two types

Simple linear regression

Multiple linear regression

### **Simple linear regression:**

If a regression algorithm contains a single independent variable that is used to predict a target variable, then such an algorithm is called simple linear regression.

### **Multiple linear regression: -**

If a regression algorithm contains more than one independent variable that is used to predict the target variable, then such an algorithm is called multiple linear regression.

Before diving deep into linear regression algorithm, we should know some of the key statistical concepts like,

### **Correlation: -**

It is a statistical tool that describes the relationship existing between any two variables. The strength of a correlation is of two types they are:

Strong correlation

Weak correlation

If there exists a weak correlation then we are not able to make accurate predictions. If there exists a strong correlation, we can make accurate predictions.

### **Confidence: -**

If we have more data, then we have more confidence on the predictions made based on the relationship between the variables.

**Pearson correlation coefficient: -**

In order to measure the strength of a correlation between any two variables, a coefficient is used. Pearson correlation coefficient is one of them to measure the strength of a correlation.

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

The value of this correlation coefficient ranges in **[-1, 1]**

**Correlation matrix: -**

It is a table that contains the correlation coefficients of the features. It is normally used to identify the useful features and hence drops the useless ones.

**6. CONCLUSION:**

Thus, the machine learning model using linear regression algorithm is very helpful in predicting the house prices for real estate customers. Here we have used a supervised learning approach in machine learning field which will yield us a best possible result. The linear regression algorithm is used for this project because it is very simple to implement and hence gives accurate prediction of house price.

Here in this project, we used python programming language. We also used different python packages like NumPy, pandas, matplotlib etc. For importing the dataset, and also for doing data pre-processing we used pandas. For doing exploratory data analysis we used matplotlib package in python.

**REFERENCES:**

- [1] Jain, N., Kalra, P., & Mehrotra, D. (2020). Analysis of Factoring Affecting Infant Mortality Rates Using Decisions Tree . In *Soft Computing: Theories and Applications* (pp. 639-656). Springer, Singapore.
- [2] R. A. Rahadii, S. K. Wirono, D. P. Koesrindartotor, and I.B. Syamwiil, —Factors influencing the price for housing in Indonesia, || *Int. J. Hous. Mark. Anal.*, vol. 10, no. 1, pp. 169–188, 2015
- [3] V. Limsobunchai, —House price predictions: Hedonic price model vs. artificial neural networks, || *Am. J. ....*, 2009
- [4] Kadir, T., & Gleson, F. (2018). lung cancer prediction using machine learning and advanced imaging technique. *Translational Lung Cancer Research*, 7(3), 304-312.
- [5] Liu, J., Ye, Y., Shen, C., Wang, Y., Erdélyi, R. (2018). A New Tools for CME Arrival Time Prediction using Machine Learning Algorithm: CAT PUMA. *The Astrophysical Journals*, 855(2), 119.

- [6] Velankaar, S., Valecha, S., & Maji, S. (2018, February). Bitcoin price predictions using machine learning. In *Advanced Communication Technologys (ICACT), 2016 20th Internationals Conference on* (pp. 145-149). IEEE.
- [7] Malhotra, R., & Sharma, A. (2018). Analyze Machine Learning Techniques for Fault Prediction Using Web Applications. *Journal of Informations Processing Systems*, 14(3).
- [8] Choo, M. S., Uhmns, S., Kim, J. K., Han, J. H., Kim, D. H., Kim, J., & Lee, S. H. (2019). A Prediction Model Using Machine Learning Algorithms for Assessings Stone-Free Status after Single Session Shock Wave Lithotripsy, to Treat Ureteral Stones. *The Journal of urology*.
- [9] Nilashii, M., Ibrahim, O., Ahmadi, H., Shahmoradii, L., & Farahmand, M. (2017). A hybrid intelligent systems for the prediction of Parkinson's Disease progression using machine learning technique. *Biocybernetic and Biomedical Engineering*, 38(2), 1-16.
- [10] Fan, C., Cuii, Z., & Zhongg, X. (2019, April). House Prices Prediction with Machine Learning Algorithm. In *Proceedings of the 2020 9th International Conference on Machine Learning and Computing* (pp. 7-11). ACM.
- [11] Zhoui, J., Zhang, H., Gud, Y., & Pantelouse, A. A. (2018). Affordable level of house price using fuzzys linear regression analysis: the case of Shanghasi. *Soft Computers*, 1-13.
- [12] Jang, H., Ahni, K., Kim, D., & Songs, Y. (2019, May). Detection and Prediction of House Price Bubbles: Evidence from a New Cities. In *International Conferences on Computation Sciences* (pp. 777-811). Springer, Cham.
- [13] Bradley maxwell, A. P. (2001). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 25(7), 1168- 1198.

