International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

# House Price Prediction Using Linear Regression

Prof. Swati Shamkuwar
Asst. Prof Dept of Information Technology
G.H.Raisoni College of Engineering, Nagpur.

Kanchan Kaushik, Lenyandri Aswale, Krishna Tidke

## *Abstract:-*

***Due to the ever-growing population and people moving to other cities for employment opportunities, housing demand is rising continuously. For individuals who are planning to remain for a significant period of time although not perpetually, in addition to individuals that aren't interested in taking risks while building is underway, it is important to predict the house price on a long-term temporary basis. The price of a home is closely related to other elements like location, region, and population. The system would forecast the price of the home using the information on the province, length, bedroom count, accommodations, and other elements. This would give a ballpark figure for the property's expected market value. It could also be used to make comparisons of different properties so that buyers could indeed make good decisions. To estimate the price, previous market dynamics, variation in prices, and future shifts will all be taken into consideration. It's necessary to have a system to predict house prices because house prices rise every year.***

***Keyword:- Machine learning, Linear Regression***

## I) Introduction:-

The mission's aim is to forecast effective home pricing for real estate clients taking into account their priorities and budgets. We compare and investigate various prediction methods in order to choose the best one.

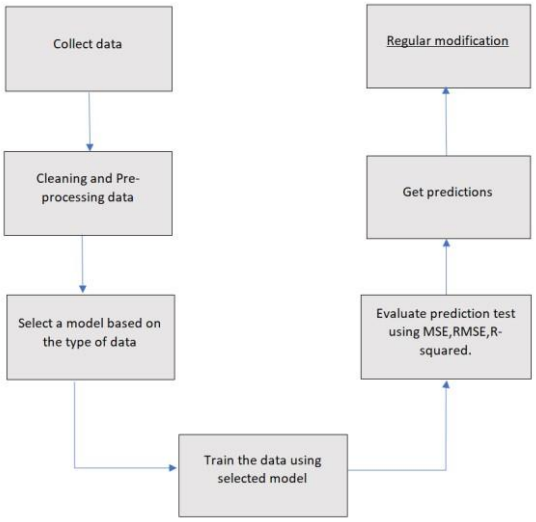
Other theories highlight the socio-economic conditions that primarily play a role in such house price rises, whereas others emphasize how the geographic location and social features of a specific area determine how the house price will increase or decrease.

Several character selection methods, like variance interaction aspect, effectively executing, and principal component evaluation, in addition to the conversion of data mechanisms, including dealing with anomalies and values that are not present are employed in this work. The four parameters of accuracy, precision, specificity, and sensitivity are used to evaluate the effectiveness of machine learning techniques.

Support vector regression, models of neural networks, and other prediction models may all be applied to forecast home prices. The House-price model offers financiers of real estate,

buyers of houses, and constructors of houses a number of advantages. This model will offer a wealth of data and provide information to home buyers, investors in real estate, and builders that will help them estimate house prices, such as the pricing of home values in the marketplace. Using Linear Regression & Random Forest Regression the model is trained, then evaluated

to measure the performance of the model.




## II) Literature survey:-

Three major categories can be used to categorize the literature review on this subject: hybrid models, machine learning-based models, and traditional statistical models. Conventional Statistical Models: Conventional analytical techniques including multiple linear regression (MLR), time series analysis, and autoregressive integrated moving averages (ARIMA) have been utilized to forecast housing prices in the past. To create predictions, these models rely on a set of specified features like location, square footage, the number of rooms, etc. These models can make predictions that are reasonably accurate, according to studies, but how well they work is greatly influenced by the caliber of the input data and the features chosen.

Machine learning-based models: Due to their high prediction accuracy and capacity to automatically learn features from data, machine learning-based models like decision trees, random forests, support vector machines (SVMs), and neural networks have grown in popularity in recent years. The non-linear correlations between the input features and the output variable can be captured by these models, which can also manage complicated data structures. Moreover, the input data in these models can accommodate missing values and outliers. Several research has demonstrated that when it comes to predicting property prices, machine learning-based algorithms perform better than conventional statistical models.

Hybrid Models: To produce better forecasts, hybrid models integrate the advantages of both conventional statistical models and machine learning-based models. A typical instance of an integrated model is one that uses data mining to determine the connection between the components of the entries and the outcome parameter before putting the learnt model's equations into a more traditional analytical framework, such as MLR. Several studies have demonstrated that when it concerns forecasting house prices, hybrid models can outperform both conventional statistical models and systems based on machine learning.

In conclusion, the literature review on house price prediction demonstrates that hybrid models, which combine the positive characteristics of both machine learning and classic statistical models, perform even better than traditional statistical models.

## III)Proposed methodology:-

**Collection of data:** Obtain past information on home pricing and relevant factors including dimensions, locality, bedrooms, and restrooms. You can get this information from a number of places, including real estate websites, federal databases, and private businesses.

**Data Cleaning:** To minimize any outliers, missing numbers, or discrepancies, the obtained data must always be cleaned and pre-processed.

This phase is essential to guaranteeing the model's correctness and dependability.

**Model choice:** Depending on the data type and the issue at hand, select the best model. Systems such as neural networks, decision trees, random forests, and linear regression are often used techniques for evaluating house prices.

**Model Training:** Training the selected algorithm via tidy, automated inputs. To achieve this, sets for training and verification of information must be created, and the framework's parameter values needs to be modified.

**Model Simulation:** Utilize numerous metrics, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared, to analyze the performance of the trained model. This process aids in determining the model's advantages and disadvantages.

**Model Implementation:** After the framework has been trained and evaluated, it can be used to make estimations on newly acquired information. A web application, API, or command-line interface can all be used to do this.

**Model ongoing maintenance:** Frequently add fresh data to the model and assess its performance. This step makes that the scheme has been accurate and current throughout time. Data collection, cleaning, feature engineering, feature selection, training, evaluation, deployment, and maintenance are all part of the approach for predicting housing prices. These procedures can be used to create an accurate and trustworthy model for forecasting home values.

## IV) Future scope:-

The capability to forecast home prices is a field that is continually developing as new methods and information become accessible. In order to achieve higher and better accuracy, the algorithms can be augmented with some alternative solutions.

As consumer demand for housing in large cities rises, there's been an obvious spike in the number of private builders that add additional features to real estate in order to draw potential buyers.

There are additional different models available for prediction. This kinds of system's data supply must be appropriate for the user running the workflow and the device that is getting used.

Furthermore, extra data sets can be used to boost the model's accuracy.

## V) System Design:-

Our proposed system framework includes the following components:

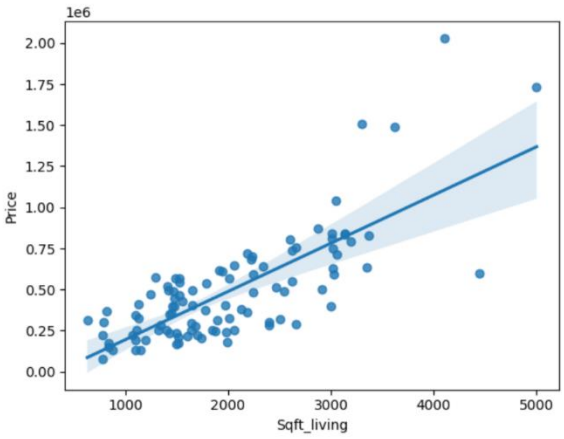
**A) The user interface:-** The House Price Prediction and Evaluation System's user interface is a web application. There are two different user kinds. One is a buyer looking to invest in real estate. Also included is the Seller. Whereas for Buyer, the registration process is optional, Seller must be required to register. Both users must submit the information for the property they want before the system can forecast its value accurately.

**B) Dataset used:-** For house prediction, we are using the "House Prices" dataset.

This dataset was obtained using a web-based methodology. This collection comprises housing attributes and zipcodes from several cities. Ex: square feet, bedrooms, floors, etc.

| | ID | Price | Bedrooms | Bathrooms | Sqft_living | Sqft_lot | Floors | Waterfront | View |
|---|---|---|---|---|---|---|---|---|---|
| count | 100.000000 | 1.000000e+02 | 100.000000 | 100.000000 | 100.000000 | 100.000000 | 100.000000 | 100.0 | 100.000000 |
| mean | 50.500000 | 4.877884e+05 | 3.320000 | 2.032500 | 1999.910000 | 9823.750000 | 1.465000 | 0.0 | 0.260000 |
| std | 29.011492 | 3.278720e+05 | 1.062492 | 0.815456 | 851.148839 | 7113.578422 | 0.472769 | 0.0 | 0.760383 |
| min | 1.000000 | 7.800000e+04 | 1.000000 | 1.000000 | 630.000000 | 1148.000000 | 1.000000 | 0.0 | 0.000000 |
| 25% | 25.750000 | 2.530000e+05 | 3.000000 | 1.000000 | 1430.000000 | 5292.500000 | 1.000000 | 0.0 | 0.000000 |
| 50% | 50.500000 | 4.045000e+05 | 3.000000 | 2.000000 | 1820.000000 | 7985.500000 | 1.500000 | 0.0 | 0.000000 |
| 75% | 75.250000 | 6.126545e+05 | 4.000000 | 2.500000 | 2555.000000 | 10450.500000 | 2.000000 | 0.0 | 0.000000 |
| max | 100.000000 | 2.030000e+06 | 6.000000 | 4.500000 | 5000.000000 | 43000.000000 | 2.000000 | 0.0 | 3.000000 |

**C) Implementation :-**




## VI) Conclusion:

Based on the features and characteristics of a given property, house price prediction techniques periodically provide a foretold value. The estimated value is frequently calculated through prior sales records as well as other pertinent components like the property's location, size, age, and characteristics, as well as its current market and financial circumstances.

It's essential to remember that while house price prediction models can be a useful tool, their accuracy cannot be relied upon to anticipate a property's true value in the market. The true sales value of a property can be impacted by a variety of other elements, including market swings, inventory and demand, special features, and house conditions.

The market for real estate offers buyers, sellers, and investors a number of useful tools, which include house price prediction. It can offer insightful information about trends in the market and assist in guiding decisions on real estate purchases.

## VII) References:

A) Adyan Nur Alfiyatin, Hilman Taufiq "Modelling house price prediction using regression analysis" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017

B) Farah Khanum , Nisarga .P .M , Navitha Pawar "Real Estate House Price Prediction using Machine Learning", IJESC, Volume 11, Issue No.07, 2021

C) Abigail Bola Adetunjia, Oluwatobi Noah Akande, Funmilola Alaba Ajala , "House Price Prediction using Random Forest Machine Learning technique"

D) https://github.com/YBI-Foundation/Dataset/blob/main/House%20Prices.csv

E) Alisha Kuvalekar, Shivani Manchewar, Sidhika Mahadi, Shila Jawale "House Price Forecasting Using Machine Learning" Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST) 2020

F) Anand g. Rawool, Dattatray v. rogye, Dr. Vinayak bharadi "House Price Prediction Using Machine Learning" IRE Journals | Volume 4 Issue 11

G) "House Price Prediction using Machine Learning Algorithms: A Comparative Study" by Hoda Tahmasebi, Sajjad Hamidian, and Narges Sahraei, in the International Journal of Advanced Computer Science and Applications (2018): https://thesai.org/Publications/IJACSA