

## Machine Learning For Home Value Prediction

Mayank Kumar Singh<sup>1</sup>, Arvind Kumar Singh<sup>2</sup>, Priyanshu Singh<sup>3</sup>

### Abstract

The real estate industry is the minimal pellucid in our surroundings. Home prices vary on a regular basis and are sometimes blown up rather than based on evaluation. Our study project's major focus is on predicting house values using real-world factors. Our goal is to base our assessments on each and every key parameter that is taken into account when establishing the pricing. We used multiple linear regression technique to determine property prices based on square footage and the number of rooms in this paper. The relationship between the average value of one variable and the values of other variables is calculated through regression. Regression is a type of statistical methods for estimating the relationships between variables in statistical modelling. The relationship between one dependent variable ( $y$ ) and two or more independent variables ( $x_1, x_2$ , etc.) is explained by multiple linear regression Three modules were used to implement this: The data entry module is used to give the project with the information it requires. The Analysis module is used to examine and forecast housing prices based on the demands of the customer. The Front-end component has been used to construct the blueprint required GUI screens.

**Keywords:** *Machine learning, supervised learning, forecast, multiple linear regression, and factors*

### Introduction

In current era of globalisation, investing is a commercial activity that most customers are engaged in. Gold, equities, and real estate are just a few examples of commonly used investment commodities. Real estate investment, in especially, has risen dramatically. Home

---

\* Correspondence Author

<sup>1</sup> Department of Computer Science and Engineering, Galgotias University, Greater Noida  
mayank.bns2000@gmail.com

<sup>2</sup> Department of Computer Science and Engineering, Galgotias University, Greater Noida  
arvi.96480699@gmail.com

<sup>3</sup> Department of Computer Science and Engineering, Galgotias University, Greater Noida  
singhpriyanshu555@gmail.com

values patterns are not only a source of anxiety for traders, but they also reveal the state of the economy. There are numerous options. The neighbouring area, which has easy accessibility to roads, highways, schools, shopping malls, and nearby job opportunities, also led to the high in home value. Traditional home prediction becomes challenging, hence various techniques for home value forecast have been developed. Accurately predicting home values has long piqued the interest of purchaser, trader and real estate agents.

Home value forecasting is a basic subject which has been thoroughly examined using machine learning methods and techniques. Many hypotheses have emerged as a result of the contributions of numerous investigators. A few of these theories contend that a region's geographical location and culture influence that how home is designed. We clearly understand that a home value is a value from a set of numbers, so predicting home values is clearly a regression issue. To predict house price, an individual will often look at similar properties in his or her region and then attempt to predict the price based on the information gathered. All of this suggests that home value forecasting is a developing field of research.

Investigators are confronted with two significant obstacles. The most difficult task is determining the optimal amount of variables that will aid in faultless predicting the path of housing values. As per Kahn, productivity growth in several home construction sectors has an impact on property price rise. The model that Kahn used illustrates how home values might affect people's lives. The investigators' second significant difficulty is to determine which machine learning methodology would be the most successful in properly forecasting property prices. Ng and Deisenro create a mobile phone application for regression via Gaussian processes. Hu et al employ the information gain coefficient (MIC) to create precise mathematical model that can predict hoarding. However, we now recommend using multi variate regression because it is more precise and highlights some key qualities.

### **Literature Survey**

Real estate investment, like any other investment, necessitates the option to set money available for future gains. The primary difference between real estate as well as other paper investments is that real estate investment requires a large amount of startup investment, that the investor would not jeopardise by not following good decision-making methods. Builders, for example, are in charge of deciding whether or not to develop or invest in real estate. Builders are in charge of determining which constructions will be profitable and acceptable at a certain specific property prices. Real estate investment also has a market that is defined by low inventory and excessive supply, or vice versa. These are facts of life in the real estate industry. Improper assessment might result in economic overbuilding or a glut. A thorough analysis should accompany each choice to invest or expand. To increase the precision of judgement, an analysis with the aid of a modern methodology should be used. Real estate investment trusts (REITs) are a newer way to invest in property. Direct real estate investment, as opposed to 'paper investment,' entails the ownership rights to a value of homes, usually with a house.

Animah, et al. demonstrated how Multiple Regression Model was used to describe price variation for a group of residences in. Each feature that has been established as a price decider in theory is priced, and the recognize significance of one is stated clearly. Their study article revealed how statistical method can be used to analyse real estate investment when

multiple determinants are taken into account. Area is an old adage when it comes to determining price and leasing terms. However, as time and technology have progressed, the traditional pricing model has evolved. As we all know, all policy makers make decisions based on their beliefs about all of the preceding elements of a decision. The synthesising step is part of the consideration process. Builders or owners must conduct a synthesize of candidate, development, or investment alternatives. A decision situation is probably set of alternative candidates (a, b, c...) is a critical component. Integration is the process of identifying and describing alternative approaches. A next stage entails a review of all actions that could be utilised to estimate results.

Multiple Regression is one of the most extensively utilised methodologies for explaining and forecasting asset management (Chaplin, 1998). After that, the method is expanded to Hedonic Regression and used in Concurrent Equation Systems (Thompson and Tsolacos, 2000).

Analysis of Multiple Regressions. The correlation analysis is the foundation for multiple regression analysis. On tiny amounts, analysis and multiple regression analysis may usually be employed to undertake multivariate analysis. According to Steven (1992), the MRA's strength resides primarily in its application as a tool for determining the relative relevance of individualistic variables on the relativist variable. Just by grasping the fundamental concepts of MRA can the benefits of MRA be fully comprehended. MRA is used for prediction in its traditional form. In modern practise, though, MRA is typically employed to explain the topic under inquiry. MRA is used as a causal model by Cohen and Cohen (1983) to explain changes in independent variables and to evaluate the possible significance within each variable. For a unit change in the variables in the equation, each regression coefficient measures the amount of change in the relativist variable.

### **Proposed System**

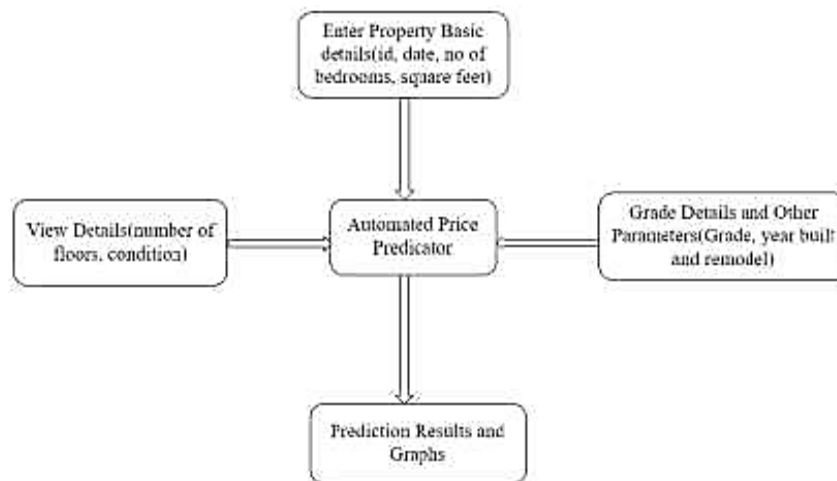
They employ a machine learning approach called multi variate regression to forecast the home value in this proposed model. We suggested the system "Home value forecasting utilising multi variate analysis," and we used multiple linear model to predict the home value. They were able to train the computer using different aspects of data points as from past to produce a future forecasting in our proposed system. To train the model, we used data from prior house prices. To overcome the challenge, we primarily used two machine-learning libraries. The first one was pandas, that was used to filter and alter data before putting everything into a format that could be analysed. The other was sklearn, that was used for real-world estimation and forecasting and had a number of built-in functions. We utilized data from past years' home and land prices, which we obtained from such an online public database. A few of the data was utilised to train the machine, while the remaining was utilised to analyse the results.

The supervised machine learning model's basic approach is to learn patterns and relationships in data from the training dataset and then replicate them for the testing data. For data analysis, we used the Python Panda package, which aggregated many datasets into a single data framework. Because of the original data, we must prepare it for feature identification. Id, date, value, rooms, sqft living, condition, quality, and year built were the qualities. It uses linear regression and multiple linear regression for the home to forecast the value depending

on the provided qualities and data. We also calculated the efficiency by comparing the test set forecasts to the actual values. The Predicted pricing is provided by the suggested mechanism.

### System Architecture

The basic stage in home values forecasting is gathering raw data from various sources and entering it into a system. Any actual data from the organisation can be included in the dataset. You can retrieve the qualities that are needed for forecasting and derive the relativist and individualistic variables from the original data. Calculate the intercepts and coefficients using multiple linear regression and it gives the predicted price as output to us.



### Algorithm:

#### MULTIPLE LINEAR REGRESSION

For data analysis, we used the Python Panda module, which aggregated many datasets into a single data frame. Because of the original data, we must prepare it for feature recognition. Date, opening, lowest, highest, and close price for a given day were the qualities. We used all of these factors to train the machine to forecast the target variables, which seems to be the value for a specific day, using a Decision Tree classifier. We also calculated the efficiency by comparing the test set forecasts to the original values.

1. Importing Pandas and Numpy data processing library files
2. Set up the qualities. Column names=['id','date','value','rooms','bathrooms','sqft','condition','year built']
3. Use read csv to load the dataset and read it.
- 4 Applying intercepts and coefficients, estimate the total home value.
5. Estimate using the multiple regression formula  $Y = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$
- 6.dot file generation:
- 7.import linear model from sklearn
- 8.Put dataset feature columns in X.//X=eqts[feature\_cols]
- 9.Determine the goal value//y
- 10.Fit values to the model//regr=regr.fit(X,y)
- 11.Linear regression and multiple linear regression models were used to create a built-in linear model.
- 12.Estimate the value of a home

## MODULES

### *I. Data Collection*

*To begin, Set of data can be gathered from a variety of streams inside any company. The correct dataset aids in forecasting and can be adjusted to meet our needs. Our data is primarily comprised on home values from the preceding year. The information can be gathered from the company vary by location, square footage, and count of bedrooms. By accumulating these, it is possible to create precise forecasts.*

### *II. Data Processing*

After the collection of data, the housing details are fetched from the database and stored in Comma Separated Value (CSV) file. We are using sqlite3 module to fetch data from the database and csv module for creating and writing house details records into the csv file. This csv file acts as the dataset. The attributes of the dataset are id, date, number of rooms and sqft\_living. All the values of the attributes selected were continuous numeric values.

### *III. Training the Data*

Following the preparation and transformation of the data, the linear model was built using multiple linear regression. The regression methodology was chosen since it does not need any industry knowledge to generate graphs using linear regression, which we can do by utilising the linear model. The function called linearregression(). We have considered the characteristic when it is used.

### *IV. Deploying the Model*

The multiple linear regression technique generates the multi variate analysis rules. The data can be tested using the training data. It aids in the production of an outcome or an efficient house price prediction when employing this model.

## **Conclusion**

The paper "MACHINE LEARNING FOR HOME VALUE PREDICTION" describes a method for predicting home values using multivariate regression analysis on a set of data. The report is particularly beneficial to new property investors because it provides timely advice.

## **References**

- Ayush Varma, Abhijit Sarma, Sagar Doshi, Rohini Nair - "Housing Price Prediction Using Machine Learning and Neural Networks" 2018, IEEE.
- G.Naga Satish, Ch.V.Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu "House Price Prediction Using Machine Learning". IJITEE,2019.
- CH. Raga Madhuri, G. Anuradha, M. Vani Pujitha -" House Price Prediction Using Regression Techniques: A Comparative Study" 2019 in (ICSSS),IEEE.