

MACHINE LEARNING FOR Home Value PREDICTION

Mayank Kumar Singh
Department of Computer Science
and Engineering
Galgotias University
Greater Noida

Arvind Kumar Singh
Department of computer Science
and Engineering
Galgotias University
Greater Noida

Priyanshu Singh
Department of Computer Science
and Engineering
Galgotias University
Greater Noida

Abstract: Our ecosystem's least transparent industry is real estate. Housing prices fluctuate on a daily basis and are sometimes hyped rather than based on valuation. Our research project's main focus is on predicting housing prices using real-world factors. Our goal is to make evaluations on each and every basic parameter that is properly considered when determining the price. We used multiple linear regression to estimate house prices based on square footage and the number of bedrooms in this paper. The relationship between the mean value of one variable and the values of other variables is measured by regression. Regression analysis is a collection of statistical for calculating the association between multivariate in statistical modelling. Multiple variables Explain the association between one uninterrupted Based on variable (y) and two or more individualistic variables. linear regression (x1, x2, x3...etc). Three modules were used to implement this: The data entry module is used to provide the project with the information when it requires. The Analysis module is used to examine and forecast house prices based on the needs of the customer. The Front-end module is used to create the project's required GUI screens.

Index Terms– machine learning, supervised learning, prediction, multiple linear regression, parameters.

I.INTRODUCTION

In this era of globalization, investment is a business activity that most people are interested in. Gold, stocks, and real estate are just a few examples of commonly used investment objects. Property investment, in particular, has increased significantly. Housing price trends are not only a source of concern for buyers and sellers, but they also reveal the state of the economy. House cost are impact by a variety of point, plus the number of sleeping chamber and shower room. The nearby location, which has excellent access It also leads to roads, superhighway, schools, marketplace, and local duty, opportunities. the increase in house price. Manual house prediction becomes difficult, so many systems for house price prediction have been developed.

Accurately predicting house prices has always attracted the interest in both purchaser and vendor, as well as broker dealer capitalist, croupier, house teller, treasurer, usurer, banking executive. Many scholars have already attempted to decipher the enigmas of house price forecasting. House price prediction is a classic problem that has been extensively studied using machine learning tools and techniques. Many theories have emerged as a result of the contributions of various researchers from around the world. Some of these theories suggest that the geographical position and culture of a given region decide how home prices will rise

or fall, whereas others believe that socioeconomic factors are a major factor in increasing home prices. We all recognize that house worth may be a variety from some outlined assortment, thus clearly the process of forecast the price of a house is a regression task. An individual can forecast the cost of a house by using the following formula will typically look for similar properties in his or her neighborhood and then attempt to predict the price based on the information gathered. All of this suggests that The field of regression research into home price forecasting is a pretty modern one that necessitates machine learning expertise. This has sparked my interest in working in this field.

Researchers are faced with two major challenges. The most difficult task is determining the area of regression research into. Home price forecasting is a relatively new concept. Housing price growth is determined by various residential construction sectors. Kahn's model illustrates the impact that house prices can have an essentially Home wealth grows faster than wages, giving it a trendy look for a long time before collapsing and experiencing a long period of decline. The researchers' second major challenge is to determine which In terms of accurately forecasting house prices, machine learning would be the most popular tool. Ng and Deisenro develop a smart phone framework for Gaussian process regression. To develop precise mathematical models for forecasting housing prices. Hu et al use the maximum information coefficient (MIC). However, we now recommend using multi variate regression because it is more accurate and highlights some key attributes.

LITERATURE SURVEY

Property investment, like any other investment, generally requires the decision to set money put something away for the future Real estate differs from other paper investments in that it needs a significant quantity money, which the investor

would not lose by failing to follow proper decision-making procedures. The developers, for example, have the option of either producing or buying real estate. The developers have the option of selecting the market for commercial property is marked by low demand and high supply, or low supply and high supply. In the real estate business, these are the facts. Improper analysis may lead to overdevelopment or a surplus in the market. A thorough analysis should accompany any decision to invest or develop. To improve the reliability of decision-making, a modern methodology should be used to conduct the research be used. Real estate investment trusts are a newer way to invest in real estate. Direct property investment, as opposed to 'paper investment,' entails the Owning rights to a plot of land, typically with a house.

Multiple Regression Analysis (MRA) was used by Aminah, et al. to describe price difference for a limited number of homes Each attribute that has been established as a theoretical price determinant is priced within, and the expected contribution of each is seen clearly. Their research paper showed how statistical analysis can be used to assess property investment by taking into account a number of variables. 'Location, location, location' is an old adage when it comes to determining price and rental terms. Nonetheless, the old price model has changed as time and technology have advanced. As we all know, all decision makers make decisions based on their convictions about all of the preceding elements of a decision. The synthesizing stage is part of the decision-making procedure. Developers or owners must conduct a synthesis of nominee, growth, and investment alternatives. A decision situation's collection of alternative candidates (a, b, c...) is crucial. Synthesis is described as the process of identifying and describing possible alternatives. The following stage entails a review of all behaviors' that could be used to predict outcomes.

Multiple Regression Analysis is one of the most commonly used explanation and forecasting methods for property results (Chaplin, 1998). 'The' is an abbreviation for 'the'. The method is then extended to Hedonic Regression and implemented in a Simultaneous Equation System (Thompson and Tsolacos, 2000). A type of regression analysis based on multiple regression is multiple MRA (Multiple Regression Analysis). Analysis and multiple regression analysis can be used on small samples. The MRA's strength, according to Steven (1992), is that it can be used to determine the influence of individualistic variables on the based-on variable is ranked in order of significance (the dependent variable is the phenomenon to be explained, and individualistic variables are the factors used to explain the based-on variable). Only by recognizing the fundamental principles of MRA can the advantages of MRA be fully understood. MRA is used for prediction in its traditional form. In recent conversation, however, MRA is usually used to clarify the topic under investigation. MRA is used as a causal model by Cohen and Cohen (1983) to describe shifts in independent variables and to determine the relative value of each individualistic variable. Each regression coefficient calculates the amount of change in the based on variable when the individualistic variables in the equation change by a unit.

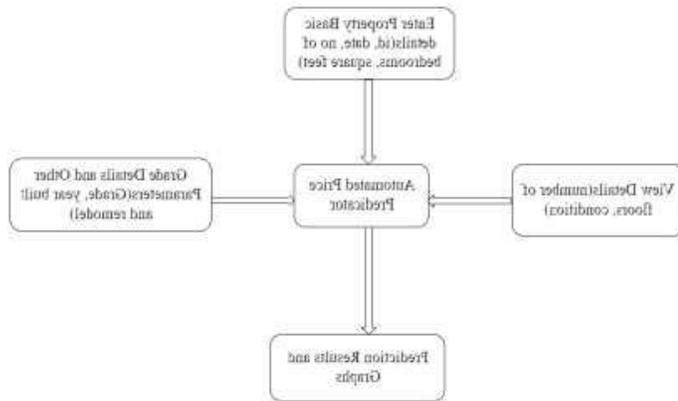
PROPOSED SYSTEM

We use a machine learning algorithm called multi variate regression to predict the house price in this proposed scheme. We proposed the system "House price prediction using machine learning," and we used multiple linear regression to estimate the house price. We were able to train the computer using different attributes of data points from the past in this proposed method to make an educated guess about the future. To train the model, we used data from previous housing prices. To solve the

problem, we primarily used two machine-learning libraries. The first was pandas, which was used to clean and manipulate data before placing it into a format that could be analyzed. The other was sklearn, which was used for real-time analysis and prediction and came with a number of built-in features. We used data from previous years' housing and land prices, which we received from an online public database. Some of the data was used to train the computer, while the rest was used to test the data. The supervised learning model's basic approach is to learn patterns and relationships in data from the training set and then replicate them for the test data. For data processing, we used the Python Panda library, which merged several datasets into a single data frame. Because of the raw data, we must prepare it for feature recognition. Id, Date, Price, Bedrooms, Sqft living, condition, grade, and yd built were the attributes. It uses linear regression and multiple linear regression for the house to predict the price based on the given attributes and data. We also calculated the accuracy by comparing the test set predictions to the actual values. The Predicted price is given by the proposed scheme.

SYSTEM ARCHITECTURE

The first phase in housing price prediction is gathering raw data from different sources and entering it into a database. Any historical data from the company can be included in the dataset. We may extract the attributes that are used for estimation and evaluate the dependent and independent variables from the raw data. Using multiple linear regression to measure the intercepts and coefficients, we can get the expected price as an output.



**ALGORITHM:
MULTIPLE LINEAR REGRESSION**

For data processing, we used the Python Panda library, which merged several datasets into a single data frame. Because of the raw data, we must prepare it for feature recognition. Date, available, low, big, and closing price for a given day were the attributes. We used all of these features to train the computer to predict the object variable, which is the price for a given day, using a Decision Tree model. We also calculated the accuracy by comparing the test set predictions to the actual values.

- 1.Importing Data manipulation library files.
//Pandas and NumPy
- 2.Initialize attributes. #col_names=[
'id','date','price','bedrooms','bathrooms','sqft_living',
'condition','grade','yr_built']
- 3.Load dataset and read using read_csv.
- 4.Caluate total housing price using intercepts and coefficients.
- 5.Calculate using Multiple regression Formula $Y = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$
- 6.dot file generation:
- 7.from sklearn import linear_model
- 8.Put dataset feature columns in X.
//X=eqts[feature_cols]

- 9.Set Target value//y
- 10.Fit values to the model//regr=regr.fit(X,y)
- 11.Built linear model using linear regression and Multiple linear regression.
- 12.Predict House price

MODULES

I.Data Collection

To begin, Dataset can be gathered from a variety of sources within any organization. The correct dataset aids in prediction and can be manipulated to suit our needs. Our data is primarily comprised of house prices from the previous year. The information can be gathered from the company based on its venue, square footage, and number of rooms. By accumulating these, it is possible to make accurate predictions.

II. Data Processing

The housing information are fetched from the database and saved in a Comma Separated Value (CSV) file after the data is collected. The sqlite3 module is used to retrieve data from the database, and the csv module is used to create and write house information records to a csv file.The dataset is represented by this csv file. The dataset's attributes include id, date, number of spaces, and sqft living. All of the attributes' values were chosen 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 because it does not require any domain knowledge to create graphs using linear regression, which we can do by using the linear model.The

sklearn library has a function called `linearregression()`. We use the algorithm to train the model using the attribute we considered in the dataset. Models are fine-tuned and fitted using training sets.

4. Sifei Lu, Zengxiang Li, Zheng Qin ,Xulei Yang , Rick SiowMong Goh - “A hybrid regression technique for house prices prediction” 2017,IEEE.

IV. Deploying the Model

The multiple linear regression algorithm generates the multi variate analysis rules. The data can be tested using the qualified data. It helps in the development of a performance or an effective house price forecast by using this model.

CONCLUSION

This paper, titled "House price prediction using machine learning," describes a method for predicting house prices using multi variate regression analysis on a given collection of data. The paper is especially helpful to potential real estate investors because it offers timely advice. It is also helpful to home owners who are trying to sell their home because it estimates the price. Finally, the efficiency of transactions in the real estate industry improves.

REFERENCES

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