HOUSE PRICE PREDICTION USING MULTI VARIATE ANALYSIS

[1]Arshiya Shaikh, [2] R.Vinayaki, [3] G. Siddhanth, [4] Y. Phanindra Varma

[1]Assistant Professor, [2]Student, [3]Student, [4]Student
[1]Department of Information Technology
[1]JBIET, Hyderabad, India

Abstract: Real estate is the least transparent industry in our ecosystem. Housing prices keep changing day in and day out and sometimes are hyped rather than being based on valuation. Predicting housing prices with real factors is the main crux of our research project. Here we aim to make our evaluations based on every basic parameter that is considered while determining the price. In this paper, we performed multiple linear regression for estimating house price based on area in square feet and number of bed rooms. Regression is a measure of the relation between the mean value of one variable and corresponding values of other variables. In statistical modelling, regression analysis is a set of statistical processes for estimating the relationships among variables. The multiple linear regression explains the relationship between one continuous dependent variable (y) and two or more independent variables (x1, x2, x3...etc). Here we implemented through three modules: Data entry module, is used to provide the needed data to the project. The Analysis module is used to analyse and predict the house prices, based on the customer needs. The Front-end module is used to create the needed GUI screens for the project.

IndexTerms– machine learning, supervised learning, prediction, multiple linear regression, parameters.

I. INTRODUCTION

Investment is a business activity on which most people are interested in this globalization era. There are several objects that are often used for investment, for example, gold, stocks and property. In particular, property investment has increased significantly. Housing price trends are not only the concern of buyers and sellers, but it also indicates the current economic situation. There are many factors which has impact on house prices, such as numbers of bedrooms and bathrooms. Even the nearby location, a location with a great accessibility to highways, expressways, schools, shopping malls and local employment opportunities contributes to the rise in house price. Manual house predication becomes difficult, hence there are many systems developed for house price prediction. Accurate prediction of house prices has been always a fascination for the buyers, sellers and for the bankers also. Many researchers have already worked to unravel the mysteries of the prediction of the house prices.

House price prediction is a classic problem which has been analyzed extensively using tools and techniques of machine learning. There are many theories that have been given birth as a consequence of the research work contributed by the various researchers all over the world. Some of these theories believe that the geographical location and culture of a particular area determine how the home prices will increase or decrease whereas there are other schools of thought who emphasize the socio-economic conditions that largely play behind these house price rises. We all know that house price is a number from some defined assortment, so obviously prediction of prices of houses is a regression task. To forecast house price one person usually tries to locate similar properties at his or her neighbourhood and based on collected data that person will try to predict the house price. All these indicate that house price prediction is an emerging research area of regression which requires the knowledge of machine learning. This has motivated to work in this domain.

There are two major challenges that researchers have to face. The biggest challenge is to identify the optimum number of features that will help to accurately predict the direction of the house prices. Kahn mentions that productivity growth in various residential construction sectors does impact the growth of the housing prices. The model that Kahn worked with shows how housing prices can have an apparently trendy appearance in which housing wealth rises faster than income for an extended period, then collapses and experiences an extended decline. The second major challenge that is faced by the researchers is to find out the machine learning technique that will be the most effective when it comes to accurately predicting the house prices. Ng and Deisenro constructs a cellphone based application using Gaussian processes for regression. Hu et al uses maximum information coefficient (MIC) to build accurate mathematical models for predicting house prices. But now we propose the use of multi variate regression as they are more accurate by highlighting some important attributes.

LITERATURE SURVEY

Property investment as other investment, involves the decision of putting aside for the future benefits. The main distinguishing feature between property investment and other paper investment is it involves huge amount of initial capital which certainly the investor will not take risk by omitting proper decision-making process. The decision to develop or to invest in property, for example lies in the hand of the developers. The developers are responsible to decide what developments will be profitable as well as permissible on that particular price of land. An additional feature of property investment is its market is characterized by low demand and high supply or vice-versa. These are realities in the property business. Improper analysis could lead to overbuilding or glut in the market. The decision to invest or to develop should be accompanied by thorough analysis. The analysis with an aid of contemporary approach should be employed to enhance the reliability of the decision making. The recent mode to property investment is via real estate investment trust. Compared to 'paper investment', direct property investment involves the rights of ownership to a price of land, typically with a building.

Animah, et.al illustrated how Multiple Regression Analysis (MRA) has been used in explaining price variation for selected houses in. Each attribute that theoretically identified as price determinant is priced and the perceived contribution of each is explicitly shown. Their research paper demonstrated how the statistical analysis is capable of analyzing property investment by considering multiple determinants. An old maxim related to the price and rental determination is 'location, location and location'. Nonetheless, time and technology changes have shifted the old price paradigm. As we are aware that all decision makers based their decision on what they believe about all the foregoing element of a decision. The process of decision-making involves the synthesizing stage. Synthesis of candidate, development or investment alternatives has to be undertaken by developers or owners. The set of alternative candidates (a, b, c...) is a fundamental element of a decision situation. he identification and description of the candidate alternatives are designated as synthesis. Next stage involves an analysis of all activities that may be used for the estimation of outcomes.

The most widely used methods to explain and forecast property performance is Multiple Regression Analysis (Chaplin, 1998). The technique is then extended to Hedonic Regression and utilized in Simultaneous Equation System (Thompson and Tsolacos, 2000). Multiple Regression Analysis (MRA) Multiple regression analysis is based on the correlation analysis. Generally correlation and regression can be used to conduct multivariate analysis on fairly small samples. Steven (1992) points out that the strength of the MRA lies primarily in its use as a means of establishing the relative importance of independent variable on the dependent variable (Dependent variable is the phenomenon to be explained and independent variables are factors used to explain dependent variable). The benefits of MRA can only be appreciated with an understanding of basic concepts of MRA. MRA in its classic form is used for prediction. However, in recent practice, MRA is normally used for an explanation of the subject under investigation. Cohen and Cohen (1983) use MRA as a causal model to explain changes in independent variables to explain changes in dependent variables as well as to assess the relative importance of each independent variable. Each regression coefficient estimates the amount of change that occurs in the dependent variable for a unit change in the independent variables in the equation.

PROPOSED SYSTEM

In this proposed system, we focus on predicting the house price using machine learning algorithm like multi variate regression. We proposed the system "House price prediction using Multi variate analysis" we have predicted the house price using the Multiple linear regression. In this proposed system, we were able to train the machine from the various attributes of data points from the past to make a future prediction. We took data from the previous housing prices to train the model. Here, we majorly used two machine-learning libraries to solve the problem. The first one was pandas, which was used to clean and manipulate the data, and getting it into a form ready for analysis. The other was sklearner, which was used for real analysis and prediction and it has containing various inbuilt functions. The data set we used was from the previous years housing and land prices collected from the public database available online, some of the data was used to train the machine and the rest of the data is used to test the data.

The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data. We used the python panda library for data processing which combined different datasets into a data frame. The raw data makes us to prepare the data for feature identification. The attributes were id, date, price, bedrooms, sqft_living, condition, grade and yr_built. According to the given attributes and data it predicts the price by using the linear regression and multiple linear regression for the house. We also quantified the accuracy by using the predictions for the test set and the actual values. The proposed system gives the Predicted price.

SYSTEM ARCHITECTURE

The first step in prediction of housing price is collection of raw data from the various sources and adding them to the database. The dataset can be of any historical data of the organization. From the raw data we can extract the attributes which are used for the prediction and determine the dependent variables and independent variables. Calculate the intercepts and coefficients using multiple linear regression and it gives the predicted price as output to us.
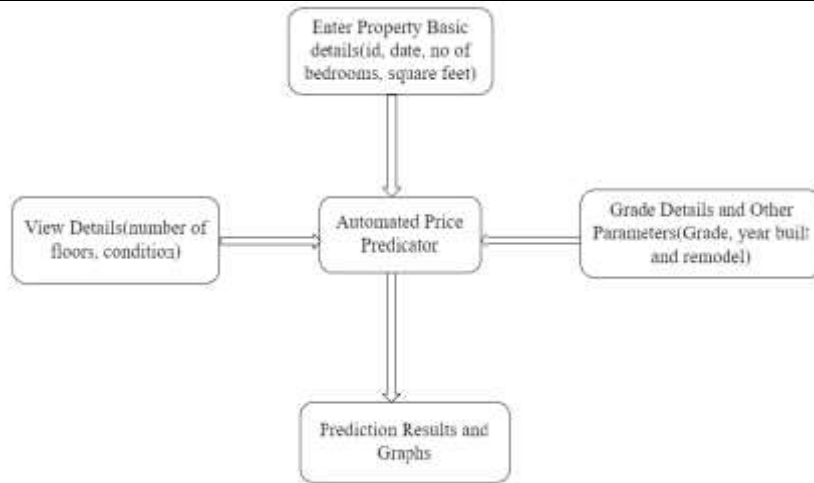
ALGORITHM:

MULTIPLE LINEAR REGRESSION

We used the python panda library for data processing which combined different datasets into a data frame. The raw data makes us to prepare the data for feature identification. The attributes were date, open, low, high and the closing price for a particular day. We used all these features to train the machine on Decision Tree model and predicted the object variable, which is the price for a given day. We also quantified the accuracy by using the predictions for the test set and 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. Calcute total housing price using intercepts and coefficients.
5. Calculate using Multiple regression Formula Y =a + b1x1 + b2x2 + b3x3 +...bnxn
6. dot file generation:
7. from sklearner 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*

Firstly, Dataset can be collected from various sources of any organization. The right dataset helps for the prediction and it can be manipulated as per our requirement. Our data mainly consists of the previous year house prices. The data can be collected from the organization based on the location, area in square feet and number of rooms. By collecting these it makes accurate in prediction.

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*

After the data has been prepared and transformed, the next step was to build the linear model using the multiple linear regression. The regression technique was selected because the construction of graphs using linear regression does not require any domain knowledge, we can done by using the linear_model.Linearregression() which is in the sklearner library. By using the attribute we have considered in the dataset we train the model by using the algorithm. The training sets are used to tune and fit the models.

IV. *Deploying the Model*

The multi variate analysis rules are generated from the multiple linear regresssion algorithm. The trained data can be used for the Testing the data. It helps to give the output or accurate Predicted price of the house using this model.

CONCLUSION

This paper entitled "House price Prediction using multi variate analysis" has presented an approach to predict house prices based on multi variate regression analysis on given data set. The paper is very useful to the new investors of the real estate market, as it gives in time suggestions to them. It is also useful to the house owners, who wants to sell their house, as it predicts the price.This finally leads to the improved quality of transactions in the Real estate market.

FUTURE WORK

This paper is currently working accurately for the dataset corresponding to a US Housing data set. The applicability of this paper to some other sectors, preferably in different countries, is yet to be explored.

REFERENCES

1. Ayush Varma, Abhijit Sarma, 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.

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