House Price Prediction Using Machine Learning

Akhilendra Pratap Singh
Computer Science
Galgotias University, Greater
Noida, India
akhisingh2211@gmail.com

Kartikey Rastogi
Computer Science
Galgotias University, Greater
Noida, India
kattu311@gmail.com

Shashank Rajpoot
Computer Science
Galgotias University, Greater
Noida, India
rajpootshashank48@gmail.com

Abstract— Now-a-days everyone wish to live in the large cities but the competition in the market related to all the resources is increasing day by day. A middle-class family can't afford the price of rent, food, water and electricity while surviving his family. The price of the flats in the city is increasing and there is so much of risk to predict the actual price of the house. Our research paper [1] will helps you to predict the price of the house to a good accuracy. The main motive of our research paper is to predict the price [2] of the house by analyzing the customer needs and their financial income. As we see when a client wants to purchase the house in the city he used to see three things within the city location, area and available resources around the society. Our research paper will helps the clients to know the actual price of the house and it will also helps the builders to know about the selling price that will fits the client needs.

Keywords— House Price Prediction, Machine Learning, Decision Tree

I. INTRODUCTION

In Today's era most of the businessmen wants to invest in the company for more and more profit. There are so many sources of investment like one can invest in stock market while other can invest in purchasing land or flat for their family. As we see now most of the people want to make investment in land. They purchase the house in the area where all facilities are available. In India the demand of property has increased this is because we have seen high number of increasing rate in the population [3].

It is very difficult for the engineers to predict and analyze the price of the house. It is possible through machine learning algorithm that will help the user to predict the price of the house using various algorithms and training the dataset on these algorithm [4]. Now-a-days there are so many software and models developed and rolling in the market to predict the price of the house more accurately, with a minor error in predicting that will be neglected. All of the prediction can only be possible by analyzing different part of the house like kitchen, bedroom, balcony area and the land area. Many analyst discover the data and compute the same data of the house by using different techniques to predict the price similar to other that will same as the previous prediction.

Our project has given same data to same weight to analyze to predict the price of the house. This is known as feature features engineering [4]. Most of the organizations have used the real data to predict the price of the house, but sometimes it becomes very difficult while predicting using linear regression, it will help in providing the efficient solution.

We have used another one algorithm that is known as coordinate descent algorithm, that will help in minimizing the time for predicting the data and price of the house. There are so many companies that will be using to predict the price of the house which will be possible through machine learning and artificial intelligence.

Sometimes there are many organization and many circumstances to determine various type of houses price. There are so many models studied previously such as error will be calculated by RMS [6] value calculation. In our model we have also introduced the support vector machine concept to predict the relationship between the cities and locality in the city, it will also predict the density of the area in which the house is to be located. On the other hand there are some other ways to predict the price of the house by considering the dataset taken from different case. Now the prediction of house can be done by implementing regression technique and non-parametric technique models.

II. RELATED WORK

So many factors have been used for the prediction of house price. By developing this type of model it can use so many dataset and features and researcher will use different sources to extract [5] the information about the locality. Feature extraction will be done by analyzing the cities and there resources that will help to predict the price of the house more clearly.

A. Factor that are affecting House prediction

Pricing of the house will be affected by several factors. Main factors affecting the price that belongs to three groups, but mainly includes locality, area and resources. The main conditions for price prediction of house [5] is the size of the area, the number of rooms in the house, available of all resources in the locality. One can also analyze that the environment will be eco-friendly so that there is no risk of pollution and diseases.

B. Hedonic Pricing

It is the best model for the prediction of price of the house. In this we assume that the value associated with the land property is related to the all features in the house. For calculating the price there is one model and equation given below

$$y = a.x_1 + b.x_2 + \ldots + n.x_i \qquad (1)$$

In this equation,

y is known as the value that we have predicted,

x1, x2, and xi will be the symbols or keys used in the house.

b, ... n coffecient for determining the values.

III. DATA SET

Data set contains all the information related to the prediction of the house price. Here we only discuss about the price of the flats in city and the taxes that were imposed on it by building

tax. We have taken the dataset of only some houses to predict the accuracy in the prediction of the house. As we have only very small data of upto 10 houses, our model will gives the best price prediction[7]. We have also assume the data based on empty sets as the price is changing every 3 year of the house, and the price of the material used in the flat or house is also increasing day by day

The data that we have got depends upon various parts and factor to determine the price of the house. It contains the locality name in which the house is available, the date and year in which it was constructed, the price of area and its size is also required for prediction. Our dataset[8] also contains data of the useful resources that are very essential for every person like health, education, water, and electricity. We have taken the data in the area in the range of 1km from the house that no one can suffer with the best facilities within the city.

As we all know that the land is the only part we have to focus while making any building or house. Because if the locality and the area in which we are searching the house is good and all facilities are available within the particular range then it will be easier for every person to survive. It will demands in the increasing price of the house. We are also looking for good restaurant and malls near our house if these facilities are also available the then price of the house will also increases.

IV. PROPOSED SYSTEM

As we see that everyone is shifting from rural area to the urban areas. The main motive of our project is to minimize the time of the individuals that is going to occur through various things. In today's world the user went to the brokers to suggest him a best apartment that fulfill his need and make his investment[9] in that house. But I don't think that this method will work with great accuracy there will be a fraud activity sometime in investing our money in the house. So to overcome and solve this problem we come up with and automated machine learning model that will predict the price of the house for the customers.

The model that we have designed will requires the initial step as the part of data scraping. By using this concept of data scraping user can easily find out or extract the data from different sources and make a dataset of all the useful data in the form of text file. We are going to use the platform known as UIPATH STUDIO to make the flowchart in a well format.

Now we have got the useful data in the form of text file as obtained from the data extraction, we are going to clean the data which will be very useful for the algorithm to use only the exact data that is of use and remove the unwanted data. We have to check each features and attribute of the dataset that will be cleaned after the data extraction. This is an important step of our proposed system.

Now the cleaning part is completed, we will move to the algorithm part where these data will be trained. In machine learning there will be many algorithm that will accurately predict the rate of the house easily without any human efforts. One of them is Light GBM.

Random forest will also used as it is an important part of the decision tree for ensemble of the data. There are so many parts in random forest based on the attribute customer will free to choose the part as per there need. It will be good algorithm

of machine learning in training the data that is small in size and predict the good result with more accuracy.

Fig 1. Block Diagram of Our System

V. RESEARCH METHODOLOGY

A. ALGORITHM THAT WE HAVE STUDIED--

When we are designing this system model, there are so many algorithm that we have studied some of them are

- Decision Tree
- Support Vector Machine
- KNN[11]
- Random Forest
- Linear Regression

The data which we have got is trained with all the algorithm. But one algorithm that is best is known as Decision Tree Regression. This algorithm [10] gives the result with high accuracy and suggests the best price for the house to the customer. It is mostly depend on the data and who we are training the data.

B. DECISION TREE REGRESSOR

It is only the decision tree regressor that will predict the price of the house to the great accuracy. In this algorithm we are going to tain the model in the form of the tree that will helpful for all the prediction comes in the future also. It gets knowledge from analyze the data of the system in the form of max and min.

To deal with very efficiently there is also one way known as grid search CV. It will show the result on max depth and predict the result forming the grid. By using the grid search CV our decision tree is constructed at a maximum depth.

C. FLASK INTEGRATION

Now when we have build our model and got the result, we have to mix or integration with the user interface, we have to use one part known as flask for this purpose. It consists of all tools and libraries that is very essential to build the application or the webpage for our model to predict the price of the house. It is used with the python mostly to build the system.

VI. IMPLEMENTATION

A. DATA PROCESSING

As some value are missed when we are processing then it will be handled with floor parameter. Sometime there will be the drop occurs in the dataset for the value that we are going to targeted. We have used the library known as pandas for our data processing. We have find all the value based on the calculation of the dataset against SD, mean of the value that is targeted and min-max. While data processing we have to divide out dataset into two main parts for prediction one is for the training set [11] that is 70% another one is of 20% which is used for the testing purpose.

B. Max Depth

As we have already studied and discuss about the concept of Grid search CV [12] which is used in construction of the decision tree for max depth value. We have used many concepts to shows this as graphically below are the sample-
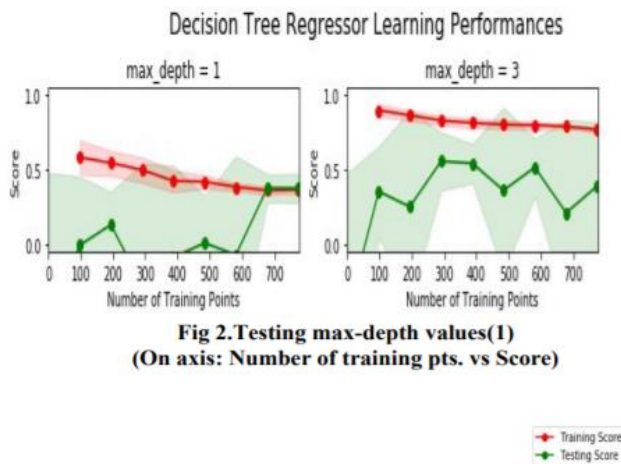

Fig 2.Testing max-depth values(1)
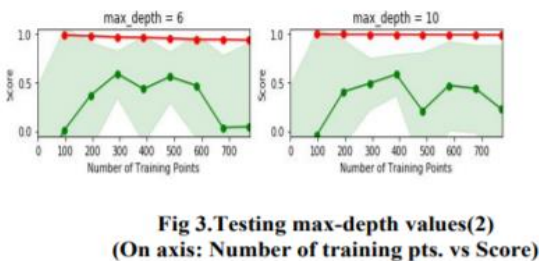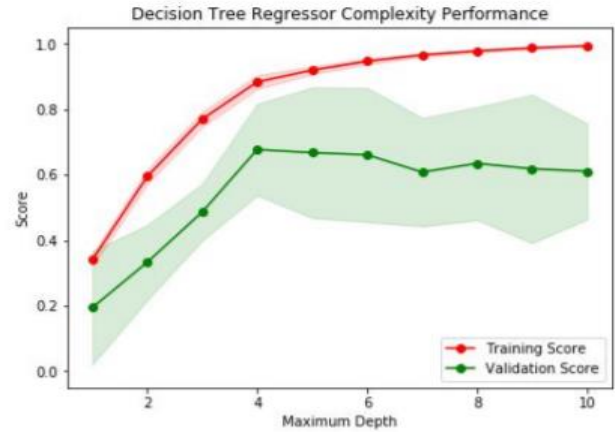**(On axis: Number of training pts. vs Score)**

Fig 3.Testing max-depth values(2)
(On axis: Number of training pts. vs Score)

C. FITTING THE MODEL

Here we see that that decision tree plays an important role in predicting the price of the house and for training the data set also. When we got the dataset of training data after we have to test the result for the model.

Parameter 'max_depth' is 4 for the optimal model.

Fig 4. Max-depth value for optimal model

VII. RESULT

The following diagram shows the actual price vs predict price of the house using our training dataset [13], which will gives the accuracy of 90% .
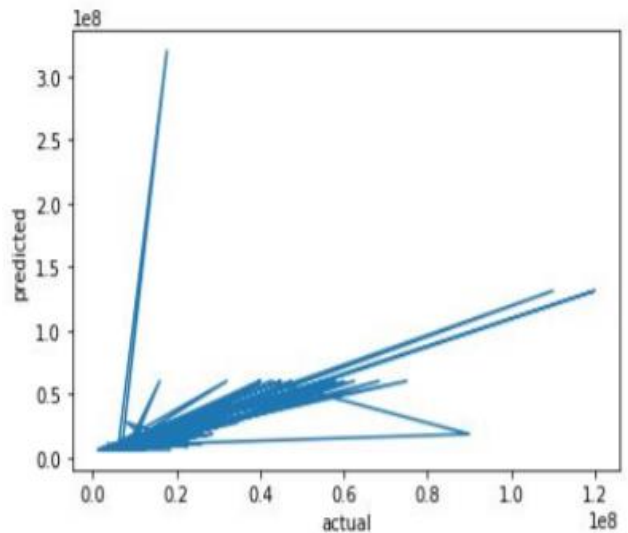

Fig 5. Actual vs predicted price graph based on the dataset

VIII. CONCLUSION

The main motive of our research paper is to predict the price of the house by analyzing the customer needs and their financial income. In our research paper, the machine learning algorithms are used to predict the price of the house, but we have used only one algorithm known as Decision Tree [4]. Our system will help the customers to make a right choice based on their financial income [5]. The prices will vary with the resources and facility around the flat in the society. Our system will give you the best accuracy result with 90% of the accuracy in the test data. In this research paper prediction of house is done by decision tree regressor [12]. All of the prediction can only be possible by analyzing different part of the house like kitchen, bedroom, balcony area and the land area. Many analysts discover the data and compute the same

data of the house by using different techniques to predict the price similar to other that will same as the previous prediction.

REFERENCES

[1] Lakshmi, B. N., and G. H. Raghunandhan. "A conceptual overview of data mining." 2011 National Conference on Innovations in Emerging Technology. IEEE, 2011.

[2] Manjula, R., et al. "Real estate value prediction using multivariate regression models." Materials Science and Engineering Conference Series. Vol. 263. No. 4. 2017..

[3] Li, Li, and Kai-Hsuan Chu. "Prediction of real estate price variation based on economic parameters." 2017 International Conference on Applied System Innovation (ICASI). IEEE, 2017.

[4] Alexander N. Bogin , William M. Doerner, "Property Renovations and Their Impact on House Price Index Construction",

[5] N. N. Ghosalkar and S. N. Dhage, "Real Estate Value Prediction Using Linear Regression," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5.

[6] Pow, Nissan, Emil Janulewicz, and Liu Dave Liu. "Applied Machine Learning Project 4 Prediction of real estate property prices in Montréal." Course project, COMP-598, Fall/2014, McGill University (2014).

[7] Banerjee, Debanjan, and Suchibrota Dutta. "Predicting the housing price direction using machine learning techniques." 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). IEEE, 2017.

[8] Manjula, R., et al. "Real estate value prediction using multivariate regression models." Materials Science and Engineering Conference Series. Vol. 263. No. 4. 2017..

[9] Li, Li, and Kai-Hsuan Chu. "Prediction of real estate price variation based on economic parameters." 2017 International Conference on Applied System Innovation (ICASI). IEEE, 2017.

[10] Nihar Bhagat, Ankit Mohokar, Shreyash Mane "House Price Forecasting using Data Mining" International Journal of Computer Applications,2016.

[11] David C. Ling, Joseph T.L. Ooi and Thao T.T. Le, "Explaining house price dynamics: Isolating the role of non-fundamentals", Journal of Money, Credit and Banking, vol. 47, Issue S1, pp. 87-125, March/April 2015

[12] Marco Helbich, Wolfgang Brunauer, Eric Vaz, Peter Nijkamp, "Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria", Urban Studies, vol. 51, Issue 2, Feb. 2014

[13] Stephen Law, "Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan London", Cities, vol. 60, Part A, pp. 166–179, Feb. 2017