

HOUSE COST PREDICTION USING DATA SCIENCE AND MACHINE LEARNING

Anuj V. Kumar¹, Anshuman Kumar², Satish S. Tiwari³, Sneha G. Gobade⁴, Prof. Amita Suke⁵

^{1,2,3,4,5}CSE, Wainganga College of Engineering and Management, Nagpur, India.

Abstract- This project is based on data science and machine learning to predict the cost of house by collected data for housing price to analyze data and predict the cost of new house. Predicting housing cost based on three important factor which include physical condition, Concept for building and location. First we explore the data for analyzing by plotting different graph and next, split the data into training set and testing set and using different regression method to predict the price. There are various method to such this and each method has pros and cons. I think regression is one of the most important method because it gives all more insight about the data by using different regression technique to improve the performance of the model.

Keywords: data science, machine learning, linear regression, multiple linear regression, MSE, RMSE, polynomial regression, K-NN regression, R² square, adjusted R² square.

I. INTRODUCTION

In day to day life population is increasing, everyone wants house for living. So, this is very difficult to predict the cost of house, so there is a need of the system to predict the cost of price in future. The system help the developer to the determine the selling price of house and help customer to arrange the right time to purchase a house. Regression is best method of prediction to predict the cost of house. To predict the cost of house price person usually tries to locate similar properties his or her locate similar properties his or her neighborhood and based on collected data that predict the cost of house. All these analysis of data collection emerging the area of prediction, which is required for machine learning for create the model of who predicts the cost of house by linear regression, also apply multiple regression which is more accurate predict the cost of house. Also available various regression technique to predict the good price of the data like polynomial regression, KN-regression, importing modules, reading the dataset and defining an evaluation table. This data frame includes root mean squared error (RMSE), R-squared, adjusted R-squared and mean of the R-square value obtained by the k-fold cross validation.

II. RELATED WORK

The biggest challenges faced by researchers is to optimum number of feature that will help to predict the accurate direction of house prices. It can mention that productivity growth in various residential construction sectors does impact the growth of the housing prices. [2]

In datasets various attributes utilizes such as ID, date, price, bedroom, bathroom, sqft-living, sqft-lot, floors, water tank, view, yr-renovated, Zip code, flat, long.

Another biggest challenge that is faced by the researchers is to find out various machine learning technique that will be more effective for accurate predicting the cost of house. The current work is unique compared to other problem from regression perspective. That tries to predict the cost of house but in regression less feature take for finding more accurate result apply different regression technique.

2.1 Install

This project requires anaconda python, because below libraries already available.

- Numpy
- Matplotlib
- Seaborn
- Skit-learn
- Pandas

Also need to have software. Install, run and execute a Jupyter notebook.

Anaconda python provide various features that will really helpful for us such as 'help' one of the keyword that is used to take library help and also learn sum extra features.

2.2 Data

The housing data consist of 21614, Dataset, and 21 feature(attributes). This dataset is a modified version of the housing dataset found on kaggle.

2.3 Features

Id, Date, price, bedroom, bathroom, sqft-living, sqft-lot, floors, water tank, view, grade, sqft-above, sqft-base, Yr-built.

There are number of features available some of important features given below

- Price: Price of each room
- Bedroom: number of bedroom in each house
- Bathroom: Number of bathroom in each house
- Sqft-living: Square fit area for living
- Yr-built: Year of built
- Yr-renovated: Year of renovated

2.4 Data exploration

In this first section of this project you will make a cursory investigation about the housing data and provide your observations.

Familiarizing oneself with the data through discovery. Process is a fundamental practice to help you for better understanding and justify with results. [2]

The main aim of this project is a construct a regression model which is predicting the cost of house.

As different research are done by author using various machine learning algorithm, it is seen then predicting real estate cost is a complex study.[1]

We see various papers and conclude that most of them do not predict the future cost of house. Location is an important factor for predicting the price of a house. This is because the location determine the prevailing land price of the required house .[4]

III. METHODOLOGY

The methodology section of a research gives two main answer to the question how was the data collected and how was it analyzed.

We have undertaken various Machine learning concepts

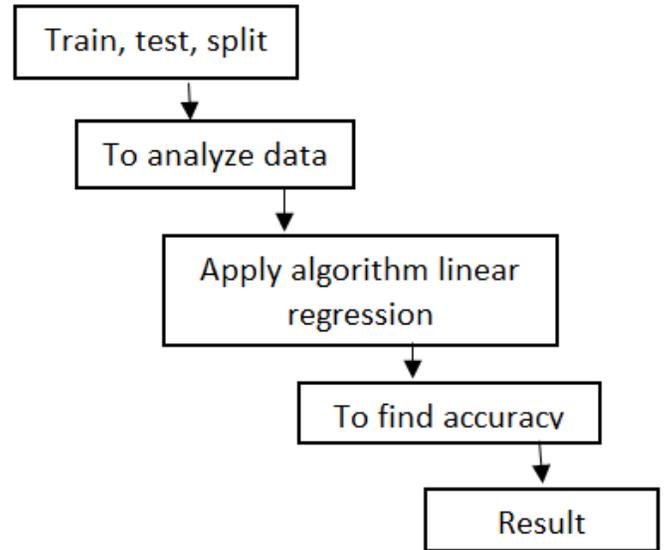


Fig.3.1. Steps of methodology

We are taking row data, preprocessing given dataset after that we analyze the data by plotting different types of graph. Then apply different regression for finding good accuracy.

IV. CONCLUSION

In the present real estate world population is continuously increasing. People are not able predict accurate price of house. In today real world, it has become different to staring. House amount of data. The making of different model applied optimal use of the regression algorithm. The model use such data in more efficient way, the regression algorithm fulfill the customer requirement and reducing the risk of the investment in an estate. A lot of features added make model more widely acceptable. after applying different regression technique we will see which technique can well classifier fit to regression problem.

V. FUTURE WORK

We suggest that working on large data sets, better result can be found. We have taken only regression algorithm for predicting the cost of house applying different regression algorithm to improve the performance of model. This model is useful for development of application for various cities.

VI. REFERENCES

[1].Nehal N Ghosalkar “Real Estate Value Prediction”
nehal.ghosalkar@spit.ac.in.

[2].B. Balakumar, P. Raviraj, S. EssakKiammal “Predicting Housing prices using Machine Learning Techniques”.

[3].An Nguyen “Housing price Prediction” March2018.

[4].Adyan Alfiyatin, Hilman Taufiq, Ruth Febrita, Wayan Mahmudy “Modeling House Price Prediction Using Regression Analysis and Particle Swarm Optimization” International Journal of Advanced Science and Application 2017.

[5]Shantanu Chakraborty, “Application of Incentive Based Scoring Rule Deciding Pricing for Smart Houses”shantanu.chakraborty@nitech.ac.jp.

[6]Danping Ren, Hui Li, Yuefeng Ji, “House energy management system for the residential load control based on the price prediction” 2011.

[7]Rushab Sawant, Saurabh Jain, Tushar Tiwari, Yashwant Jangid, Anita Gupta, “A multi feature based Housing Price Prediction For Indian Market Using Machine Learning” International journal of computer & Mathematical Science December 2017.

[8]Azme Bin Khamis, Nur Khalidah Khalilah Binti Kamarudin, “Comparative Study On Estimate House Price Using Statistical And Neural Network model” International Journal of Science & Technology Research Volume 3.December 2014.