Regression Modelling For Real Estate Pricing

*

1st Sayama Kazi

Department of Computer Engineering

D. Y. Patil College of Engineering

Pune, India

sayamakazi786@gmail.com

2nd Snehal Sarangi

Department of Computer Engineering

D. Y. Patil College of Engineering

Pune, India

ssarangi@dypcoeakurdi.ac.in

Abstract—**The real estate market is a standout the most attentive about pricing and keeps changing. The housing market is a champion among the most engaged with respect to valuing and continues to change. It is one of the superb fields to apply the thoughts of AI on instructions to improve and anticipate the expenses with high precision. There are three factors that impact the cost of a house which incorporates states of being, ideas and area. The current structure incorporates assessing the cost of houses with no assumptions for market costs and cost increase. The goal of the paper is to construct a predictive model for evaluating the cost based on the aspects that affect the price of houses.**

Keywords—**Linear regression, Random Forest Regressor, XGBoost Regressor, SVM Regressor.**

I INTRODUCTION

This paper refer together for predicting the house prices with different advanced regression algorithm for more accuracy. The main motivation of the project is to predict real estate pricing by using appropriate algorithm and finding which is more suitable for predicting the price with low error rate. Food, attire, and safe house are the fundamental requirements of life and its necessary. Thus housing is a main importance in human resource development of any economy. So in this article we try to demonstrate some regression techniques to predict real estate price. In this system uses some regression techniques like linear regression, support vector Machine regression, and boosting algorithms like Extreme Gradient Boost Regression (XGBoost). These models are used to build a predictive model, and for best performing model with more accuracy Here, the attempt is to develop a predictive model for

evaluating the price of real estate based on different factors that affects the price.

II Related Work

Our main aim is to develop a model on predicting the house pricing with more accuracy as per the users requirement or their interests. Our model analyse different parameters according to their interest of customer and predict the ideal price. It uses Machine Learning technique known as Regression and their different algorithms.[6]

Yu, Jiafu Wu. has predicted real estate prices using various regression techniques including Ridge, Random Forest regression and SVM regression, Naive Bayes, SVM classification, logistic regression and Random Forest classification.

Nissan Pow, Emil Janulewicz and Liu "Applied Machine Learning Project for Prediction of real estate property prices in Montreal" They predicted selling prices as well as buying prices of real estate based on different features such as geographical location, number of rooms, living area, etc. It was able to predict geographical features like nearest police station ,fire station. They used Random Forest and method of KNN.

Jiao Yang Wu "Housing Price prediction Using Support Vector Regression" used SVM regression to predict the housing prices with more accuracy. The prediction used various features like longitude, latitude, separate square feet area of each rooms. Ridge and Lasso were used for Feature

to select the most important features. Different models of SVR with Linear, Poly, RBF were created.[5]

III PROPOSED SYSTEM

Methods clarify all the basic steps taken into consideration to get the desired output. We took the Kaggle dataset of Boston then pre-processed the dataset to remove useless features and data. Analyzing the data before applying to algorithm. And used different regression algorithm like (XG Boost, SVM, Random forest) for the appraisal. We use different Machine Learning Algorithms and tested our system using different regression techniques. We carried out that with manual examples and this gave us great results in terms of accuracy while prediction.[6]

A. Data Analysis

Before giving the dataset or data to any model we have to assure that all given data is correct or accurate and ready to use. We analysed our dataset based on characteristics, features and their relation among the all related features. From the examination we found

Fig1.Visualizing the differences between actual prices and predicted values.
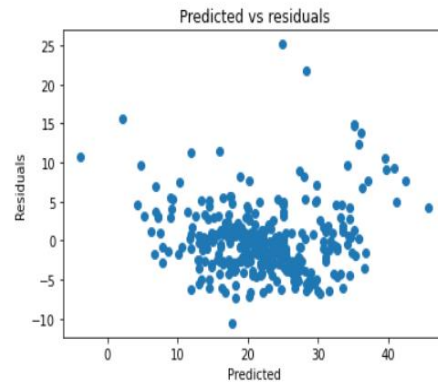

Fig 2. Checking Residuals
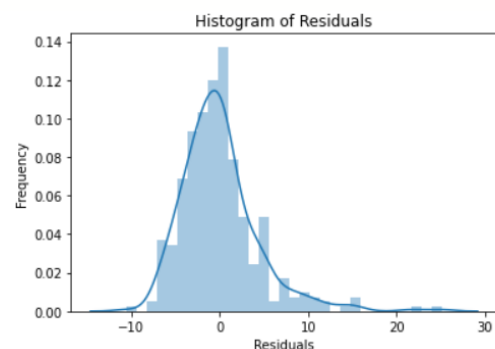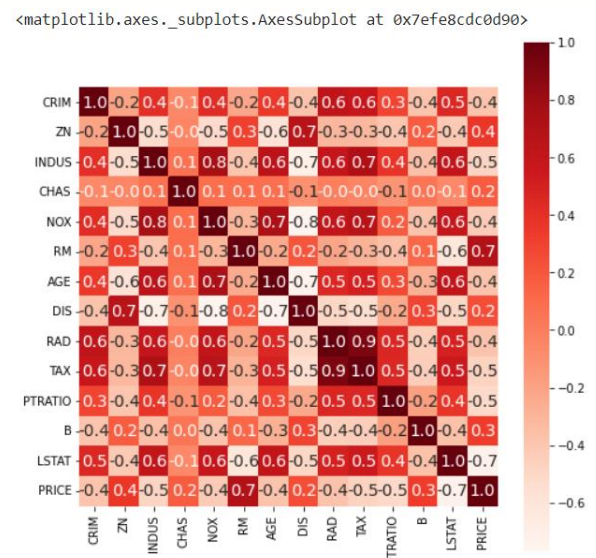

Fig 3. Checking Normality of Errors

<matplotlib.axes._subplots.AxesSubplot at 0x7efe8cdc0d90>

Fig 4. Correlation Matrix

Comparison of different regression model on our system to get more accurate result.

	Model	R-squared Score	
2	XGBoost	84.948947	
1	Random Forest	82.446049	
0	Linear Regression	71.218184	
3	Support Vector Machines	59.001585	

Fig 5. Evaluation of all models.

B.   Algorithm used

i)   Linear Regression

Linear regression is the simple algorithm for prediction. It uses two things as variables which are the predictor variable and the variable which is the most crucial one first whether the predictor variable.

These regression algorithm are used to explain the relationship between one dependent variable and one or more independent variables of model.

ii)   Random Forest Regression

In Random forest regression multitude decision trees at training time and outputting the class that is the classification or regression of the individual trees.[6] The random forest regression comes under the type of ensemble learning. It works as a begging method in which we add several base learners to make a strong model.[3]

iii)   Support Vector Machine

Support vector machines (SVM) are used for classification and regression analysis. It is best suited for regression as well as classification. SVM constructs hyperplanes or set of hyperplanes for performing classification and regression analysis.

iv)   XG Boost Regressor

XG Boost is comes under type of ensemble learning which needs to combine weak learners to from strong learners to increase accuracy of model.

XG Boost is an advanced version of gradient boosting method and it uses to design large and complicated datasets.[3]

IV FUTURE WORK

This system can be helpful to many people once it deployed but this system uses limited amount of data and in future The dataset used is bound to get outdated after sometime due to several reasons such as changes in government rules, geographical changes in that area and therefore needs to be constantly updated. After sometime the given dataset will be out dated so its necessary to use updated data and many more advanced techniques for more result accuracy with low error rate.

V CONCLUSION

This paper predicts real estate properties prices using different machine learning methods, namely, support vector machine regression, random forest regression, XG Boost regressor and linear regression for more accurate prediction. In this system we used Kaggle dataset of Boston. This system will satisfy customers by providing accurate output and preventing the risk of investing in the wrong house and to make right decision.

REFERENCES

[1]. 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), 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697639.

[2]. A. Varma, A. Sarma, S. Doshi and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018, pp. 1936-1939, doi: 10.1109/ICICCT.2018.8473231.

[3]. V. S. Rana, J. Mondal, A. Sharma and I. Kashyap, "House Price Prediction Using Optimal Regression Techniques," 2020 2nd International Conference on Advances in Computing,

Communication Control and Networking (ICACCCN), 2020, pp. 203-208, doi: 10.1109/ICACCCN51052.2020.9362864.

[4]. C. R. Madhuri, G. Anuradha and M. V. Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study," 2019 International Conference on Smart Structures and Systems (ICSSS), 2019, pp. 1-5, doi: 10.1109/ICSSS.2019.8882834.

[5]. C. R. Madhuri, G. Anuradha and M. V. Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study," 2019 International Conference on Smart Structures and Systems (ICSSS), 2019, pp. 1-5, doi: 10.1109/ICSSS.2019.8882834.

[6]. A. Varma, A. Sarma, S. Doshi and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018, pp. 1936-1939, doi: 10.1109/ICICCT.2018.8473231.

[7]. J. Manasa, R. Gupta and N. S. Narahari, "Machine Learning based Predicting House Prices using Regression Techniques," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020, pp. 624-630, doi: 10.1109/ICIMIA48430.2020.9074952.