House Price Prediction Using Different Machine Learning Techniques

Frency Chauhan, Nirva Sangani, Yagnik Hingrajiya, Vijay Rana
Ahmedabad University, frency.c, nirva.s, yagnik.h, vijay.r@ahduni.edu.in

***Abstract* - The decision making for investment in real estate can be a task with lots of thinking. With the frequent fluctuating prices and fulfilling the expectation of the investor, there arises the need for a mathematical model that helps the user with their decision making process. In this work, the house price prediction is done by using different machine learning techniques like linear regression, gradient boosting, polynomial regression and random forest regression algorithm.**

Key Words - House Price, Prediction, Linear Regression, Maximum Likelihood Estimation, Gradient Boosting, Polynomial Regression, Random Forest Algorithm

I. INTRODUCTION

With the increase in population the demand for buying houses is increasing as well. The price prediction for a house depends on various parameters like the location of the house, area, number of bedrooms etc. As price varies with these parameters, a challenge to make a model that accurately predicts the price emerges. This paper explores techniques like linear regression with maximum likelihood estimation and gradient boosting polynomial regression and random forest regression algorithm to predict the price of house in Mumbai city.

II. LITERATURE SURVEY

Manasa, Gupta and Narahari [1] have used house price data of Bengaluru city and predicted price by using various regression techniques such as Lasso regression, Ridge regression, Support vector regression (SVR) and Extreme gradient boosting (XGBoost) regression and have used the evaluation metrics as root mean square error (RMSE), R-square and adjusted R-square and root mean squared logarithmic error (RMSLE).
Sawant, Jangid, Tiwari, Jain and Gupta [2] have used house price data of Pune city and predicted price by using decision tree and random forest algorithms.
Durganjali and Pujitha [3] have used various classification algorithms like Logistic regression, Decision tree, Naive Bayes and Random forest.
Varma, Sarma, Doshi and Nair [4] have used Linear regression, random forest regression, boosting regression algorithm and neural networks approach to predict the house price.

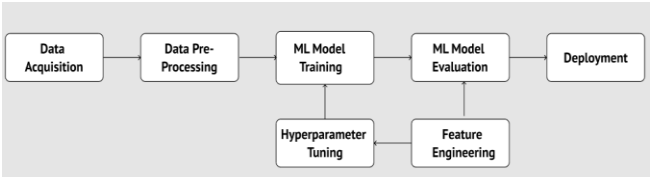
III. IMPLEMENTATION

1. DATA SET: We have acquired house price data of Mumbai city from Kaggle website. The data set consists of different parameters like price of the house, area in square feet, locality, number of bedrooms and bathrooms etc.
For all the models used, the evaluation metric for train and test data is accuracy in percent.

TABLE 1: Specification of data set

Number of columns	19
Number of rows	6348
Type of problem	Regression
Target variable	Price
Missing of value	NIL
Choice of evaluation metrics	Accuracy in percent

2. FLOW OF IMPLEMENTATION:



3. MODELING: The different machine learning based models used to predict the price of the house are:

- Simple linear regression with gradient boosting
- Linear regression with maximum likelihood estimation
- Polynomial regression (degree = 2)
- Random forest regression

IV. RESULTS

TABLE 2: Accuracy Results of train and test data set for different techniques used

Technique	Accuracy	
	Train Data	Test Data
Simple Linear Regression	55%	47%
Simple Linear Regression with Gradient Boosting	93%	49%
Linear Regression with Maximum Likelihood Estimation	52%	47%
Polynomial Regression	64%	63%
Random Forest Regression	92%	52%

V. CONCLUSION

Observing the table 2, it is evident that gradient boosting technique for linear regression gives higher accuracy than simple linear regression and maximum likelihood estimation for linear regression. The random forest regression gives almost same accuracy as gradient boosting for train data but both the models give comparatively lower accuracy for test data. This is happening due to overfitting problem. Polynomial regression gives higher accuracy than simple linear regression and linear regression with maximum likelihood estimation. In future work, we will address the problem of overfitting and tune hyperparameters more effectively.

VI. REFERENCES

[1] 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), Bangalore, India, 2020, pp. 624-630, doi: 10.1109/ICIMIA48430.2020.9074952.

[2] R. Sawant, Y. Jangid, T. Tiwari, S. Jain and A. Gupta, "Comprehensive Analysis of Housing Price Prediction in Pune Using Multi-Featured Random Forest Approach," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697402.

[3] P. Durganjali and M. V. Pujitha, "House Resale Price Prediction Using Classification Algorithms," 2019 International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2019, pp. 1-4, doi: 10.1109/ICSSS.2019.8882842.

[4] 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), Coimbatore, India, 2018, pp. 1936-1939, doi: 10.1109/ICICCT.2018.8473231.