



# Implementation of Linear Regression Algorithm In Machine Learning For Real Estate Price Prediction

Mohit S Amode<sup>1</sup>, S. R. Jadhav<sup>2</sup>, S. P. Dolli<sup>3</sup>, Pragati S Patil<sup>4</sup>, Mrunmai S Hiwarale<sup>5</sup>

Department of E&TC, Smt. Kashibai Navale College of Engineering, Pune, SPPU

<sup>1</sup>mohitamode@gmail.com

<sup>2</sup>sonali.jadhav\_skncoe@sinhgad.edu

<sup>3</sup>sudershandolli@gmail.com

<sup>4</sup>pragatipa45@gmail.com

<sup>5</sup>mrunmaihiwarale2000@gmail.com

**Abstract**—The Real estate market is one of the prime field for investment and with continuous development prices keep on fluctuating. In this paper, machine learning is used to accurately predict the value of the property based on different factors such as location, total area, number of bedrooms and number of bathrooms. Here, we have compared accuracy of three machine learning algorithm such as linear regression, lasso regression and ridge regression using the R-squared value. Based on the result, we used linear regression in our project to predict the value of the property in Pune City. The user will need to enter their requirements like location, total area, number of bedrooms and number of bathrooms in the application to get an idea of the price. This paper gives an idea on best suitable model for price prediction.

**Keywords**—Machine Learning, Price Prediction, Linear Regression, Lasso Regression, Ridge Regression.

## I. INTRODUCTION

Machine learning is very important aspect of present day business and research. People and real estate agencies buy or sell houses. Over valuation or under valuation has always been an issue and there is lack of proper decision taken. House price prediction can be done using various predicting models. There are many benefits that home buyer, property investors can get from house price model. This model will provide a lot of information and lot of knowledge to home buyers and property investors which will help them to determine house prices. Also this model helps potential buyers to decide characteristics of house they want according to their budget. The objective of this model is prediction of market value of real estate property. This helps find a starting price for a property based on geographical variables. By breaking down past market patterns and value ranges and coming advancement price will be predicted. This examination means to predict house prices in Pune city with linear regression. It will help clients to add or give choices for their house and predict it's price. This project makes it easier as we don't need to take help from broker.

## II. BACKGROUND



FIG. 1 BLOCK DIAGRAM

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. We have proposed an advanced house prediction system using linear regression. This system aim is to make a model which can give

us a good house pricing prediction based on other variables. We are going to use Linear Regression for this dataset and hence it gives a good accuracy.

This house price prediction project has two modules namely, Admin and User. Admin can add location and view the location. Admin has authority to add density on the basis of per unit area. User can view the location and see the predicted housing price for the particular location.

### III. NEED AND MOTIVATION

Having lived in India for so many years if there is one thing that I had been taking for granted, it's that housing and rental prices continue to rise. Since the housing crisis of 2008, housing prices have recovered remarkably well, especially in major housing markets. However, in the 4th quarter of 2016, I was surprised to read that Bombay housing prices had fallen the most in the last 4 years. In fact, median resale prices for condos and coops fell 6.3%, marking the first time there was a decline since Q1 of 2017.

The decline has been partly attributed to political uncertainty domestically and abroad and the 2014 election. So, to maintain the transparency among customers and also the comparison can be made easy through this model. If customer find the price of house at some given website higher than the price predicted by the model, so he can reject that house.

### IV. BLOCK DIAGRAM

The dataset is passed in preprocessing state where unwanted data or null values are removed. Later on in next step the features are been extracted using linear regression method and the model is trained which is used to compare the features from input data. Depending on basics of feature the output price is been predicted.

The working of our project is mainly categorized into five blocks that are data collection, pre-processing, data analysis, application of algorithms and evaluating the models.

- Data Collection: Here we collect data from the user about what kind of property user is looking for. Example - locality, residential, bungalow/flat, no. of bedrooms, no. of baths, etc.
- Pre-Processing : Data pre-processing is the first and very important step in ML, here the data is cleaned and stored in a specific format for the ML algorithms to use that data.
- Data Analysis: The data is analyzed to identify patterns in our case to analysis the prices in the area the user has entered to predict the price.
- Application of Algorithms: Here the linear regression algorithm is applied to the data to predict the price of the property. Then further the result are evaluated and final result is displayed.

### V. LINEAR REGRESSION

<sup>[7]</sup>Linear Regression is a machine learning algorithm based on supervised learning. Linear Regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). It is a statistical method that is used for predictive analysis. (Ex : Price, Age, Sales, Salary) A linear regression line has an equation of the form 'Y = a + bX'.

Advantage: A linear model can include more than one predictor as long as the predictors are additive. the best fit line is the line with minimum error from all the points, it has high efficiency but sometimes this high efficiency created.

Disadvantage: Linear Regression Is Limited to Linear Relationships. Linear Regression Only Looks at the Mean of the Dependent Variable. Linear Regression Is Sensitive to Outliers. Data Must Be Independent

### VI. LASSO REGRESSION

<sup>[7]</sup>Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

We define the Lasso Regression mathematically as:

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Lasso Regression uses L1 regularization technique. It is used when we have more number of features because it automatically performs feature selection.

### VII. RIDGE REGRESSION

<sup>[7]</sup>Ridge regression is a model tuning method that is used to analyze any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.

We define the Ridge Regression mathematically as:

$$Y = XB + e$$

Where Y is the dependent variable, X represents the independent variables, B is the regression coefficients to be estimated, and e represents the errors are residuals.

Advantages :

- It protects the model from over fitting.
- It does not need unbiased estimators.
- There is only enough bias to make the estimates reasonably reliable approximations to the true population values.
- It performs well when there is a large multivariate data with the number of predictors (p) larger than the number of observations (n).
- The ridge estimator is very effective when it comes to improving the least-squares estimate in situations where there is multicollinearity.
- Model complexity is reduced.

Disadvantages :

- It includes all the predictors in the final model.
- It is not capable of performing feature selection.
- It shrinks coefficients towards zero.
- It trades variance for bias.

## VIII. IMPLEMENTATION AND RESULTS

### Data Loading and Cleaning

Data cleansing or data cleaning is the process of detecting and correcting corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

Data cleansing may be performed interactively with data wrangling tools, or as batch processing through scripting.



Fig 2: Cleaned data

Fig 3: Results

### Comparison of Different Algorithms

After cleaning the data we applied three algorithms : Ridge, Lasso and Linear Regression on the cleaned data. We compared the accuracy of all the three algorithms, they all have similar accuracy hence we decided to use Linear Regression Algorithm.

### Result

Here the user needs to select the location/area, enter the required number of bedrooms (BHK), enter the required number of bathrooms and the size of the house in square feet. Then click on Predict Price button to know the price of the house according to inputs provided by the user.

Fig 4: Output

## IX. CONCLUSION

In this project, we have compared the R-squared value of three machine learning algorithms - Linear Regression, Lasso Regression and Ridge Regression. We noticed that all three values were almost same for our data set hence we used linear regression for prediction. We have developed real estate price predictor for Pune city based on different factors like location, number of bedrooms, number of bathrooms and total area of the property. So that the user can know the prices of the property in any area around Pune city and select the location which suites their budget.

## X. REFERENCES

- [1] House Price Prediction Via Improved Machine Learning Techniques (2019 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2019))
- [2] House Price Prediction Using Machine Learning (International Journal of Innovative Technology and Exploring Engineering (IJITEE), July 2019)
- [3] Literature Review on Real Estate Value Prediction Using Machine Learning by Akshay Babu, Dr. Anjana S Chandran, International Journal of Computer Science and Mobile Applications, Vol.7 Issue. 3, March- 2019
- [4] HOUSE PRICE FORECASTING USING MACHINE LEARNING BY ALISHA KUVALEKAR, SHIVANI MANCHEWAR and SHIVANI MANCHEWAR, DEPT. OF INFORMATION TECHNOLOGY DATTA MEGHE COLLEGE OF ENGINEERING AIROLI, INDIA
- [5] Employing Machine Learning for House Price Prediction by Nikita Malik, Vidhu Gaba and Priyansh at Dept. of Computer Applications, Maharaja Surajmal Institute, New Delhi, India
- [6] VALUATION OF HOUSE PRICES USING PREDICTIVE TECHNIQUES by NEELAM SHINDE, KIRAN GAWANDE, International Journal of Advances in Electronics and Computer Science, Volume-5, Issue-6, Jun.-2018
- [7] Review of Regression Analysis Models by Aviral Gupta, Akshay Sharma and Dr. Amita Goel, Associate Professor Department of Information Technology, Maharaja Agrasen Institute of Technology New Delhi. Published by : International Journal of Engineering Research & Technology (IJERT), August - 2017
- [8] Multiple Regressions in Analysing House Price Variations by Aminah Md Yusof and Syuhaida Ismail, University Teknologi Malaysia, Johor, Malaysia.
- [9] Predicting Public Housing Prices Using Delayed Neural Networks by Lipo Wang, Fung Foong Chan, Yaoli Wang, and Qing Chang. School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
- [10] A Hybrid Regression Technique for House Prices Prediction by Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick Siow Mong Goh. Institute of High Performance Computing (IHPC) Agency for Science Technology and Research (A\*STAR), Singapore
- [11] House Price Prediction using a Machine Learning Model: A Survey of Literature by Nor Hamizah Zulkifley, Shuzlina Abdul Rahman. I.J. Modern Education and Computer Science, 2020
- [12] Real estate value prediction using multivariate regression models by R Manjula, Shubham Jain, Sharad Srivastava and Pranav Rajiv Kher, School of Computing Science and Engineering, VIT University, Vellore.
- [13] Real Estate Price Prediction by Smith Dabreo, Shaleel Rodrigues, Valiant Rodrigues, Parshvi Shah, International Journal of Engineering Research & Technology, April-2021.