

Prediction and Analysis of House Price Through Machine Learning Approach

**Manivel Kandasamy¹, Raju Shanmugam², Ayush Dave³,
Charvin Chawda⁴, Krish Shah⁵, Umang Seladiya⁶**

^{1,2}Unitedworld School of Computational Intelligence, Karnavati University, Gandhinagar, Gujarat, India
^{3,4,5,6}B. Tech- CSE, Unitedworld School of Computational Intelligence, Karnavati University,
Gandhinagar, Gujarat, India

Abstract:

In recent years, picture detection, spam restructuring, natural speech command, product recommendation, and medical diagnosis have all benefited greatly from machine learning. The current machine learning algorithm aids in improving medical advancements, public safety, and security alerts. In the current work, we talk about the machine learning algorithm-generated forecast of future home prices. We compare and investigate numerous prediction methods for the selection of prediction methods. We employ lasso regression as our model because to its flexible and probabilistic model selection process We construct a housing cost prediction model in the considering such machine learning algorithm models as XG Boost, lasso regression, and neural systems on look at their order precision execution. At that point, we urge a real estate agent or home seller to employ a housing cost estimates model to assist them with better information based on property valuation. In regards to precision those tests show that the lasso regression algorithm constantly beats alternative models when utilized to predict costs associated with housing.

Keywords: Lasso Regression Approach and Neural Network-Based Machine Learning Algorithm, Hosing Cost Prediction

1. Introduction

Pricing is an important trouble in the real estate industry as it is so ridiculously broad and competitive. Real estate agents, buyers, and sellers all rely upon accurate property price predictions. It requires assessing a sizable quantity of data from several sources, making house price prediction an intimidating commitment. However, human error results in very high predicting inaccuracies when done manually. Since a home is a fundamental human need, many individuals who wish to acquire one find it difficult to determine where to do so. They visit any agents as a result, which costs them money. Therefore, there has to be a suitable technique for determining future pricing. A number of research have been carried out to create precise prediction models using both artificial intelligence and machine learning, which has emerged as a potent tool for forecasting home values. This study also tries to identify the best machine learning algorithm to predict home price or plot price using home price index (HPI) location, number of rooms, and other factors. the previous years. In recent years, a variety of ML techniques, including Support Vector Machines, Random Forest, and Linear Regression, have been employed to estimate home values. These algorithms have been used in a number of academic papers and real-world applications, showing

promise. However, the truth is that most users are reluctant to do so for security reasons, therefore this software is more reliable and safer. The results of this study will offer insightful information about the usage of ML algorithms for housing price prediction, and they can help those working in the real estate sector make decisions. ML benefits real estate, therefore we're going to employ it.

2. Objective

Clients are having a busy period buying and selling property. Dealing with it in the conventional manner, where the client is required to pay the real estate agent a commission. In addition, the buyer is unsure of the future profitability of the property. In answer to such issues, we created a website where clients can visit and check out properties without paying a fee. using different regression techniques, data analytics, data mining, and machine learning. The entire procedure gets carried out using the Python programming language for machine learning. The stacking algorithm has been applied to many regression algorithms to determine which one delivers the most exact and accurate results.

3. Literature Review

Author Name	Methodology Proposed	Research Outcome
Anand G. Rawool ; Dattatray V. Rogye ; Sainath G. Rane ; Dr. Vinayak A. Bharadi	Random Forest Regression, Machine Learning	Predict the accurate prices of house without any loss, considering factors like budget and priorities. The model is used to help the allocate resources without any relying on brokers with maximum accuracy.
G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu	Machine learning algorithm, lasso regression process and neural system, hosing cost prediction	This particular research algorithms to develop the housing price prediction models using XGBoost model, neural system techniques and lasso regression and also recommended for different house venders and selling mediators for satisfactory information based on property estimate with better accuracy.
Ayush Varma, Abhijit Sarma , Sagar Doshi and Rohini Nair	Linear, forest and boosted regression Model	several regression techniques, using WM results for honest results. The research approach produces maximum accuracy and minimum error related to every previous individual algorithm. Also, they propose using real-time neighborhood details using Google maps for precise calculations.
Rapach Furthermore strauss (2007)	auto regressive dispersed slack (ARDL)	They discover that ARDL models tend should beat a benchmark AR model.
Yiyang Luo.et.al	Random Forest and support vector machine	R-squared for both methods turned out above 0.9

4. Methodology

4.1. Linear Regression

A technique for predicting the value of one element based on another is simple linear regression. The independent variable is referred to as the one you are forecasting, and the dependent variable is the one you are predicting.

$$Y=A(x) + C \tag{1}$$

4.2. Multiple Regression

A type of statistical technique called multiple linear regression (MLR) involves use of a number of explanatory factors to predict the outcome of the response variable.

The relationship between two or more independent variables and one dependent variable is computed using MLR.

$$y_i = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \dots + \beta_px_{ip} + \epsilon \tag{2}$$

Where:

- y_i is the predicted or dependent variable
- β_0 is the y-intercept function. That is the value of y when both x_1 and $x_2=0$
- β_1 and β_2 are the coefficients of regression model that represent the change in y relative to a single-unit modification in x_{i1} and x_{i2} , respectively.
- β_p is the slope coefficient for each independent variable

4.3. Lasso Regression

It is a technique for making precise predictions. shrinking is used in this model. Where data values are reduced at the mean's middle. Least Absolute Shrinkage and Selection Operator is also referred to as LASSO.

$$\begin{aligned} \hat{\beta}^{\text{lasso}} &= \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|y - X\beta\|_2^2 + \lambda \sum_{j=1}^p |\beta_j| \\ &= \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \underbrace{\|y - X\beta\|_2^2}_{\text{Loss}} + \lambda \underbrace{\|\beta\|_1}_{\text{Penalty}} \end{aligned} \tag{3}$$

For the above-mentioned equation epitomized least difference between lasso regression and ridge regression model with the penalty term

Further, $\lambda \geq 0$ is a modification of the parameter for maintaining the strength of the penalty

4.4. Estimation of Cost Function

Let's say you are increasing a store's size because you anticipated making more money there. However, despite the hose size expanding, those discounts at that shop didn't really expand that much. Regarding the costs associated with growing the store that offered you experience a bad result. We must therefore reduce that expense.

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \tag{4}$$

5. Execution of the Work

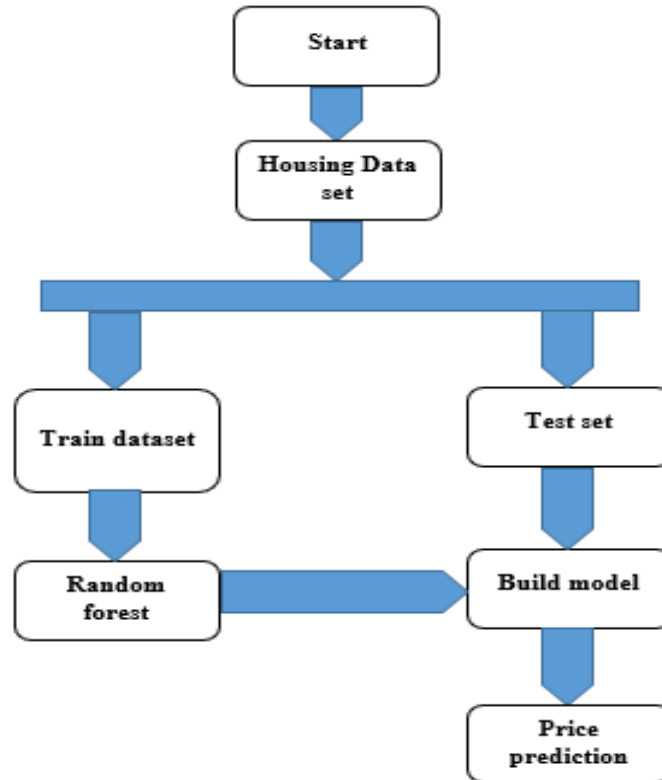


Figure .1. Flow Diagram of the working Model

The proposed diagram to short the house price prediction in a certain region and easy to refer the real estate selling market. After the different functionalities the price predication may summarized effectively and efficiently.

5.1. Data Collection and Cleaning

Data processing employs a variety of techniques and procedures. Information has been taken from tier 1 city real estate websites. Built-up area, location, carpet size, property age, zip code, and other information are included in the data. The first thing we should do is collect structured and organized data. Data collection must come first before a machine learning study is conducted. There is no dataset without no sense of data evaluation.

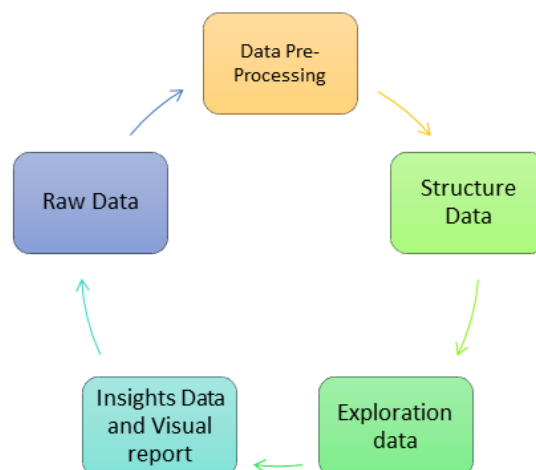


Figure.2. Frame work of House price prediction model

5.2. Data Preprocessing and Analysis

Before using models to forecast house values, the output needs to be pre-processed. A missing data investigation is conducted first. Since they are essential in finding the appropriate methods to handle missing data, a variety of missing patterns are rigorously evaluated. Since it is challenging to impute these missing values with an acceptable degree of accuracy, columns with more than 55% of their values missing are eliminated from the original dataset. Additionally, a significant portion of rows (price) lack values for the result variable. Since imputation of these values can increase bias in the input data, the observation with the missing values in the price column is removed.

6. Data Visualization

Data visualization is employed in many high-quality visual representations due to its ability to successfully study difficult ideas and identify unique patterns.

6.1. Cross Validation and Training Model

Cross validation is used to train a machine learning model using a part of the dataset. When separating data sets into "N" sets for evaluations of the built-in model, training is crucial to obtaining accuracy. Due to the fact that the data is split into two modules—a test set and training set—we must first train the framework. A component of the training set is the target variable. Applying the decision tree regressor method, the training data set is analysed. The decision tree develops a regression model through the use of a tree-like structure.

6.2. Results and Discussion

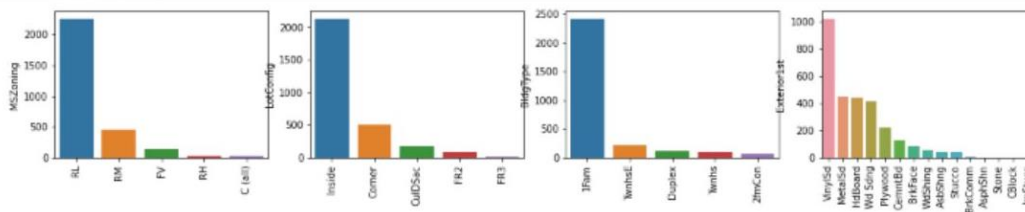


Figure.3. Visual representation of various models

	MSSubClass	MSZoning	LotArea	LotConfig	BldgType	OverallCond	YearBuilt
0		60	RL	8450	Inside	1Fam	5
1		20	RL	9600	FR2	1Fam	8
2		60	RL	11250	Inside	1Fam	5
3		70	RL	9550	Corner	1Fam	5
4		60	RL	14260	FR2	1Fam	5

	YearRemodAdd	Exterior1st	BsmtFinSF2	TotalBsmtSF	SalePrice
0	2003	VinylSd	0.0	856.0	208500.0
1	1976	MetaSd	0.0	1262.0	181500.0
2	2002	VinylSd	0.0	920.0	223500.0
3	1970	Wd Sng	0.0	756.0	140000.0
4	2000	VinylSd	0.0	1145.0	250000.0

Figure.4. Results for House price prediction and analysis

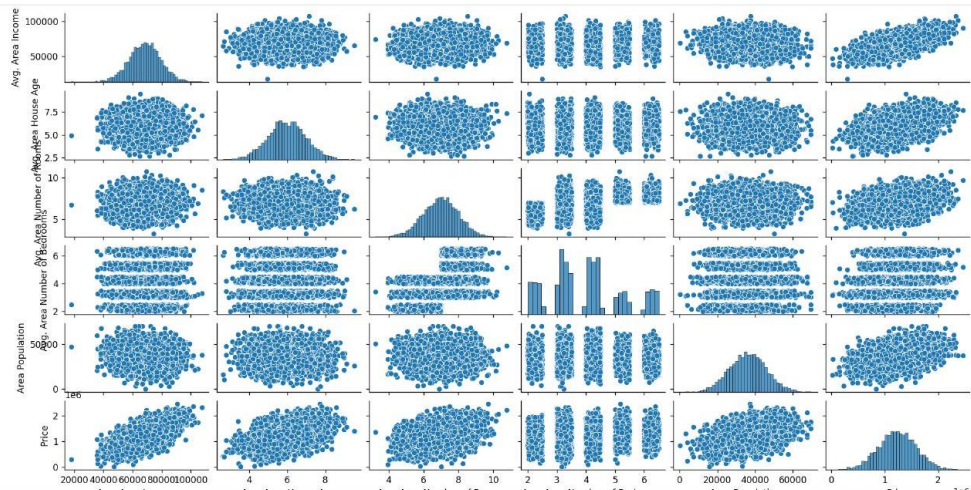


Figure.5. Result of various representation for Relation of price, area, house age

The sales prices have been calculated with more accuracy and precision. The general people would gain a lot from this. Python is used to generate these results using a variety of data mining techniques. The myriad factors that affect property pricing must be considered and taken care of. Machine learning assisted us in completing our job. Data gathering is done first. The data is then cleaned to make it accurate and to eliminate all inaccuracies. After that, the data pre-processing is finished. Then, utilizing data visualization, a number of graphs are created.

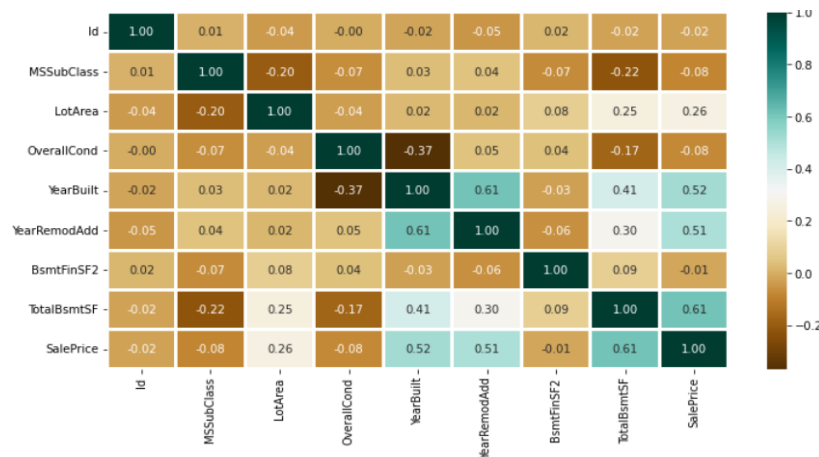


Figure.6. Result of summarised house price value with different region

This has demonstrated the many ways that data is disseminated. The model has also been tested and made ready for usage. It was found that certain classification methods had been applied to our dataset while others had not. In order to improve the accuracy and precision of the algorithms that were already being utilized, the algorithms that were not being employed on our house price dataset were eliminated. We propose a novel stacking strategy to improve the accuracy of our categorization schemes. It is essential to improve the accuracy and precision of the algorithms in order to obtain better results. If the results were incorrect, the people would not be able to predict the sales prices of homes. Additionally, data visualization was employed to enhance accuracy and results. The sales prices of the properties are calculated using a variety of methods. The sales prices have been more carefully and precisely computed. The general people would gain a lot from this.

7. Conclusion

The conclusion of this paper presents a useful technique for determining property values is house price prediction using machine learning. Many techniques have been used to provide precise predictions with comparatively low error rates, including multiple linear regression, LASSO, gradient boosting, and linear regression. These models have proven to be capable of handling and analysing enormous datasets, offering insightful information to real estate investors and purchasers.

Future Scope

Future goals include expanding the dataset beyond Tier-1 cities to include additional Indian states and cities. To make the system even more educational and user-friendly, we'll incorporate a map into it. Property values are influenced by a variety of significant variables. If statistics on things like income, salary, population, local amenities, cost of living, annual property tax, school, crime, and marketing data are accessible, it is a good idea to include more items. Soon, we'll present a contrast between the price predicted by the algorithm and the price from real estate websites like Housing.com, magicbricks.com, and others, for the same user input. In an effort to make things simpler for the consumer, we will also recommend real estate properties based on the projected cost.

References

1. Vachana J Rai, Sharath H M, Sankalp M R, Sanjana S, Bhavya Balakrishnan, "House Price Prediction Using Machine Learning Via Data Analysis", International Research Journal of Engineering and Technology (IRJET), Vol.No.9, Issue.11, 2022
2. Anand G. Rawool , Dattatray V. Rogye , Sainath G. Rane , Dr. Vinayak A. Bharadi "House Price Prediction Using Machine Learning" Iconic Research And Engineering Journals Volume 4 Issue 11 2021 Page 29-33
3. Ayush Varma,et.al, "House Price Prediction Using Machine Learning and Neural Networks", 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 10.1109/ICICCT.2018.8473231, 2018.
4. G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu," House Price Prediction Using Machine Learning", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019.
5. Sifei Lu; Zengxiang Li; Zheng Qin; Xulei Yang; Rick Siow Mong Goh, " A hybrid regression technique for house prices prediction", 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 10.1109/IEEM.2017.8289904
6. Sushant Kulkarni,et.al, "House Price Prediction Using Ensemble Learning", International Journal of Creative Research Thoughts (IJCRT), Volume 9, Issue 5 May 2021 | ISSN: 2320-2882.
7. CH. Raga Madhuri,et.al, "House Price Prediction Using Regression Techniques: A Comparative Study", 2019 International Conference on Smart Structures and Systems (ICSSS), 10.1109/ICSSS.2019.8882834, ISBN:978-1-7281-0027-2