

House Price Prediction Using Machine Learning

Shriya, Prabhjot Kaur, Gurnirdesh Singh (Assistant Professor)

Date of Submission: 15-05-2024

Date of Acceptance: 25-05-2024

ABSTRACT: In moment's period the trend of machine literacy is on hike. It is used in each and every field like in medical wisdom, real assiduity, entertainment industry and multitudinous further. Machineknowledgeprovides us variety of libraries to perform numerousfunctions. In this paper we have to bandy about how machine literacy algorithms are used to find the prices of houses and helps the buyers to prognosticate accurate price of the house without involving any real estate agent. This prophecy will be done by usingalgorithms like Linear Retrogression, Decision Tree Regression and Random Forest Regression. The purpose of this model is to enable the client to prognosticate factual value of the house which he/she wants to buy according to his budget and conditions.

I. INTRODUCTION

Like other introductoryconditionslike food, water and numerous other factors house is anintroductory need in a person's life.Buying a home these days is veritably important task for a person because he has to invest his entire plutocraton a house according to his conditions and installationslike the position or society had parking area, cleanliness and many more. Now a days, machine learning a trendy technology.Machine Learning is used in various fields like healthcare, ecommerce or in real estate industry andnumerous further. The main motive of house price prediction model is to create a model which is used to produce houseprices. Data is the core of Machine Literacy. All assiduity sectors are moving towards the booming technology of Machine Learning algorithm. But without data it is not possible to train or assay any model. To do this, we have to use formerdatasetffrom various sites to make a new model for new dataset. The demand of houses is addingextensivelywith the increase in population. Everyone wants to buy a house in good position and society according to his budget and his conditions i.e. price of the house, number of apartments in house, parking installation,medical installations and numerous features like crime rate

and air pollution. But people do not have important knowledge about thefactual price of that houses and suffers from the loss of plutocrat.To handle this problem, a house price vaticination model is designed where former datasets are used to assay the prices of new houses, so that the person who wants to buy a house can use this model to prognosticate the factual and accurate price of house. In this model, we use different Machine Learning algorithms like Linear Retrogression, Decision Tree Regression, K-Means Retrogression and Random Forest Regression. Also to produce house price vaticination model we use copious python libraries like NumPy, Pandas and matplotlib etc. After preparing or cleaning the data we will use 80% data for training the Machine Learning model and 20% data for testing the Machine Learning model. This work appliescopiousways such as features, markers, reduction waysand metamorphosisways similar as trait combinations, set missing attributes as well as looking for new correlations. This all shows that house price vaticination model can be used for exploration in Machine Literacy and give the results accordingly

II. LITERATURE REVIEW

There are number of factors that affects the prices of homes. The prices of houses increase or decreases according to metropolises, countries and position. The priceof houses increases if the house is near the trace, boardwalk, super market, good health installations, good job openings etc.Also, the price of houses will also depend upon the land area of a specific house, number of apartments available in the house, kitchen area, number of bottoms in house, deck area and parking area. The main ideal of this model is to produce a sock price vaticination model that provides us accurate result.

1. Sushant Kulkarni. (2021) Test the dataset using copiousretrogression algorithms- Velicet Lasso Regression.

2. Neelam Shinde and Kiran Gawande suggests Decision Tree, Logistic Regression and Support Vector Regression in their study.
3. For prognosticating the resale price of houses, P. Durganjali suggests bracket algorithm. The selling price of house can be calculated using copious bracket styles like Decision Tree, K-Means, Linear Regression and Random Forest. The prices of houses are varied due to its geographical area, type of profitable conditions of a particular state and position. Then RMSE is use as a performance matrix for different datasets, and make the best model that gives better and accurate result.
4. Sifei Lu, suggests a mongrel regression method for house price vaticination. With limited datasets and data features, creative point engineering system is used. The thing of this model is to prognosticate the price of houses for customers according to their budget and precedences.

III. SYSTEM DESIGN AND ARCHITECTURE

Phase 1: - Collection of Data

Data collection is the most important step of house price vaticination model. In this phase we have to collect data from copious applicable sources and real estate websites. While collecting data, always take care that data is structured, well-organised and distributed. The data may include features like position, size, number of bottoms, number of floors etc. To produce a sock price vaticination model data is must bear for prognosticating accurate value of the new

dataset. Dataset validity is must else there is no point in analysing the data.

Phase 2: - Data Pre-processing

In data pre-processing phase, the first task is to clean the raw data. It means that it involves transferring the raw data into meaningful and logical information. Occasionally missing values also exist in the dataset. There are three ways to fill our missing values:

- Clear out missing data points from the dataset.
- Clear out whole trait.
- Set the value to some value (0, mean or standard).

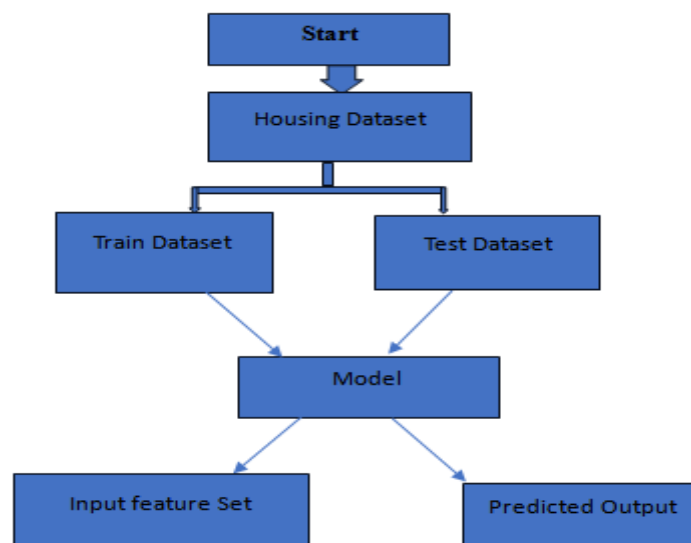
Phase 3: -Data Cleaning

In this phase of data cleaning, errors and missing values are removed from the dataset to increase the performance and delicacy of data. There are three ways to set missing values: -

- Clear out missing data point.
- Clear out whole trait.
- Set the value to some value (0, mean or median)

Phase4: - Training and Testing the Model

In this phase of training and testing, the data we want to assay will be trained first. In this phase data is divided into two corridors i.e.training and testing. In training phase data is trained by using colorful machine learning algorithms to get the better and efficient result. After training of the dataset, the trained model is ultimately applied to test dataset and house price predicted. The trained model is saved by using ‘.joblib’.



IV. MODELS USED

Linear Retrogression: -

- It is a supervised machine knowledge algorithm.
- It's used for prognosticating the values of variables grounded on independent variables.
- It's generally used for prognosticating out connections between variables and soothsaying.
- To use this, we first make the model by using direct retrogression and find the delicacy score nearly about 80% to 81%.

Lasso Retrogression: -

- Lasso Regression is a regularization fashion. This model is used over retrogression styles for farther delicacy.
- It stands for Least Absolute Shrinkage and Selection Operator.
- This model uses loss.
- The lariat model encourages simple, meager models.
- The delicacy score of Lasso

- Regression is approximate about 68.74%.

Ridge Retrogression: -

- Ridge retrogression is one of the types of direct retrogression in which small quantum of bias is introduced so that we can get better long-term prognostications.
- It is a regularization fashion, which is used to reduce the complexity of the model. It is also known as L2 regularization.
- It helps to break the problems if we have further parameters than samples.
- The delicacy score of Ridge regression algorithm is about 80%.

V. RESULTS AND CONCLUSION

House Prices of colorful areas and metropolises are reckoned by using colorful machine knowledge algorithm. To achieve the delicacy of house price vaticination model originally, we have to clean data means that we have to remove indistinguishable data, fill or remove null spaces and outliers.

Dataset of House Price Prediction Model

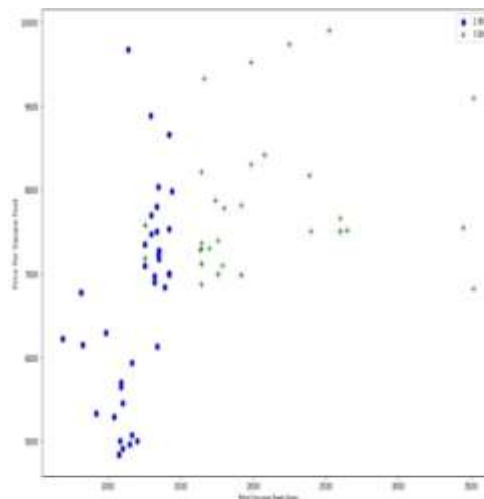
Before Setting Missing Attributes. This type of dataset contains null values, duplicates and outliers

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Comee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

After setting missing attributes the from the dataset. In these all-missing values, duplicates and outlier are removed by removing null spaces, outliers and duplicates from the dataset.

	total_sqft	bath	price	bhk	price_per_sqft
count	10153.000000	10153.000000	10153.000000	10153.000000	10153.000000
mean	1502.725758	2.471880	91.104765	2.573525	5669.621675
std	873.997461	0.971938	86.241338	0.893533	2274.739750
min	300.000000	1.000000	10.000000	1.000000	1250.000000
25%	1110.000000	2.000000	49.000000	2.000000	4242.424242
50%	1283.000000	2.000000	67.000000	2.000000	5183.823529
75%	1650.000000	3.000000	100.000000	3.000000	6451.612903
max	30400.000000	16.000000	2200.000000	16.000000	24509.803922

	bath	balcony	price
count	13247.000000	12711.000000	13320.000000
mean	2.692610	1.584376	112.565627
std	1.341458	0.817263	148.971674
min	1.000000	0.000000	8.000000
25%	2.000000	1.000000	50.000000
50%	2.000000	2.000000	72.000000
75%	3.000000	2.000000	120.000000
max	40.000000	3.000000	3600.000000



This graph shows Scatter Plot which represents the relationship between Total Forecourt bases Area and Price Per Forecourt Feet of a particular area after removing all the outliers from the dataset.

	Model	Best Score
1	Linear Regression	0.805701
2	Lasso	0.7971081
3	Ridge	0.805701

This table shows the delicacy score of different algorithms.

A comparison of algorithms was done and the algorithm that have loftiest delicacy is stationed for erecting the prophecy model.

The above table will show you the comparison between different algorithms.

	Algorithm	Accuracy	Advantages
1	Linear Regression	80%	1. Simple Preparation. 2. Performs best on Linear Data.
2	Lasso Regression	79%	1. Capability to perform point selection. 2. It is suitable for high dimensional datasets.
3	Ridge Regression	80%	1. It protects the model from overfitting. 2. It does not need unprejudiced estimators.

VI. FUTURE SCOPE:

-The compass of house price vaticination model in future is veritably vast. Due to the advancement of machine literacy in future more data is available and this will affect in further delicacy invaticination of these kinds of models. Also, integrating real-time data sources, similar as market trends and profitable pointers, can further enhance their prophetic power. Likewise, there'seventuality for incorporating more advanced ways like deep literacy and ensemble styles to ameliorate model performance. As the house market evolves, so will the styles used to prognosticate prices, offering openings for nonstop improvement and invention.

REFERENCES: -

- [1]. Anand G.Rawool, Dattatray V.Rogye, Sainath G.Rane, DR.Vinayk A. Bharadi - "House Price Prediction Using Machine Learning 2021, IRE.
- [2]. Ayush Varma, Abhijit Sarma, Sagar Doshi, Rohini Nair- "House Price Prediction Using Machine Learning and Neural Networks 2018", IEEE.
- [3]. M. Jagan Chowhaan, D.Nitish, G. Akash, Nelli Sreevidya, Subhani Shaik- "Machine Learning Approach for House Price Prediction 2023", Asian Journal of Research in Computer Science.
- [4]. Kumari Sandhya, Sarwar Siddiqui- "House Price Prediction Using Machine Learning 2022", iJRASET.
- [5]. Puja A. Chavan, Prem Shejole, Rituraj Sharma, Shafaque Shaikh, Dheeraj Shelke, Pratham Sharma- "House Price Prediction Using ML 2022" YMER.
- [6]. Bindu Sivasankar, Arun P. Ashok, Gouri Madhu, Fousiya S. - "House Price Prediction 2020"
- [7]. Vaishnavi Aghav, Dhanashree Anhad, Supriya Nanaware, Rakesh Gudekar, Prof. Mahendra Pawar' "House Price Prediction Using Machine Learning 2023" iRJMETS.