

Predicting House Prices Using Machine Learning

Dr.S.Manju¹, Lajitha. S²

¹Associate Professor, Dept of Computer Application (MCA)

²Dept of Computer Application (MCA)

^{1,2}PSG College of Arts & Science, Coimbatore, Tamil Nadu, India.

Abstract- *Today's IT industry largely uses machine learning. Machine learning is a subfield of artificial intelligence and a subfield of computer science. It also goes by the name of predictive analytics. Its work focuses on the use of data and uses algorithms to provide a means for people to learn by steadily increasing its accuracy. Its task is to improve prediction accuracy. One of the technologies that is virtually always in style is machine learning. People purchase homes in accordance with their budgets, thus it is crucial to estimate house prices in order for buyers to be able to readily purchase homes at the going rate. The main goal of the article is to provide the house price forecast findings in order to obtain accurate and better outcomes without causing any loss to those who purchase homes. As a result, we calculate the cost of a house using prediction algorithms. By looking at recent market patterns, pricing ranges, and property prices based on the amount of attributes taken into account, future prices can be anticipated. Additionally, this paper assists both buyers and sellers in estimating the house's selling price. The programming language Python is employed. This would be beneficial for those whose desire is to purchase a home without suffering any financial loss.*

I. INTRODUCTION

A technique that manually anticipated housing values existed many years ago. However, there was a drawback to setting house prices manually. People are therefore afraid of losing money and have retreated as a result. Now that technology has advanced, we can also prevent this issue and find a remedy for the people's financial losses. To circumvent these issues, many people are mastering machine learning and artificial intelligence today. In machine learning, we can create models based on past data to forecast the future. If a customer looking to purchase a home is unaware of the home's pricing, they risk losing money. One of the best and most significant things a family can do is buy a house.

The location, amenities, and demand for home are just a few of the variables that could affect the price of home.

In order to anticipate the price of a home more correctly, machine learning is utilised. By slipping through the past market design and value ranges, future prices can be

predicted. It provides numerous numerical and graphical strategies that are necessary to anticipate the price of a home. Create predictive models in this study to assess the value of homes. For the purpose of choosing the prediction methods, we are contrasting several prediction approaches. Real estate brokers and home sellers benefit greatly from the house price prediction model's support. Several algorithms are used to calculate the price of the house.

It will be beneficial to those who are purchasing a home. We employed the decision tree, lasso regression, and linear regression techniques. Data from 80% of the dataset are used to train the dataset, and data from the other 20% are used to test the dataset. Machine learning expertise is useful for making precise price predictions for homes. The dataset that is used in our model and the machine learning technique that powers it are both explained in this publication.

II. LITERATURE REVIEW

When reviewing earlier studies on machine learning in house price prediction, there are a number of fascinating facts and figures that are worthwhile taking into account.

1. In order to assess the findings, this research provides us with the price prediction regression model. It is based on the factors that we considered in determining the precise price. There are a few methods that are frequently utilised to get accurate and worthwhile outcomes. Linear, Lasso, Decision Tree, Rigid Model, Gradient Regression Model, etc., are some of the strategies. to compare and analyse the errors in order to choose the top-performing models and to create a prediction model.

2. Results for "House Prices: Advanced Regression Techniques" are provided in this publication. The competitor's goal was to make predictions about house prices based on factors like the year a house was built, the size of the house, etc. In this research, machine learning algorithms are applied. The original techniques are also discussed here.

You receive accuracy from the forecast along with the entire test approach. The price prediction for a specific data cleaning step.

3. The study makes use of a variety of methods, including Ridge, Elastic Net, Ada Boost Regression, Lasso, Multiple Linear, and Gradient Boosting. All of the aforementioned algorithms have been used to anticipate the price of homes, and it has been determined which method provides the best answer. This research article seeks to make precise housing price predictions that benefit people. This research article seeks to make precise housing price predictions that benefit people. Square footage and gradient are taken into account when making the prediction.

4. Results for "House Prices: Advanced Regression Techniques" are provided in this publication. The competitor's goal was to make predictions about house prices based on factors like the year a house was built, the size of the house, etc. In this research, machine learning algorithms are applied. The original techniques are also discussed here. You receive accuracy from the forecast along with the entire test approach. The price prediction for a specific data cleaning step.

5. Machine learning is used for spoken

6. command, medical diagnosis, spam reorganisation, recommendation, and image detection. Machine learning has recently aided security warnings, public safety, and the medical industry. Algorithms also offer improved customer service and safer auto systems. This research article discusses the algorithm-generated house price prediction. This research article discusses the algorithm-generated

7. house price prediction. Here, a variety of prediction algorithms are used and evaluated to choose the optimal approach. The Lasso regression model is used in this instance due of its versatility and superior model selection capabilities.

8. The use of the Python library in this paper's work on predicting house prices using various regression techniques will be more beneficial to those looking to purchase a home. It makes no difference if the buyer is wealthy or in the middle class. It is impossible to determine the value of a home just based on the neighbourhood and amenities offered. The task is managed by prediction.

9. The use of the Python library in this paper's work on predicting house prices using various regression techniques will be more beneficial to those looking to purchase a home. It makes no difference if the buyer is wealthy or in the middle class. It is impossible to determine the value of a home just based on the neighbourhood and amenities offered. The task is managed by prediction.

10. House price forecasting is one of the important topics in real estate. This research study makes use of current data on real estate markets to glean important insights. In Australia, historical transaction data is examined using a machine learning algorithm to determine the optimum model that will take into account both sellers and purchasers. Broadcast refers to the wide gap between expensive and economical neighborhoods in Melbourne.

III. PROPOSED WORK

In our proposed system, we mainly focus on house price prediction using machine learning like Lasso regression, Decision tree and Linear regression algorithms. Using many features, we predicted the house price. We have taken 80% data from previous data for training the model and the 20% which is remaining is used for purpose of testing. We used libraries of machine learning for solving the problems. NumPy and pandas are used in our paper. Pandas are used for storing the "CSV" file and also using pandas we import the dataset into Google Collab. The next one is the learner, we used it for analysis it contains various functions which is inbuilt that will help all solving problems.

Another one is NumPy which is used for the purpose of training the split test this was used. The disadvantage of the Existing system is that they have not compared the various algorithms on a model that had predicted the house price before paper. One of the solutions that we have done in this paper is we have compared various algorithms such as linear regression, lasso, and decision tree to predict the accurate price.

The proposed work represents the future work plan for our House Price Prediction. The proposed work also has information of what are the problems we faced and solved.

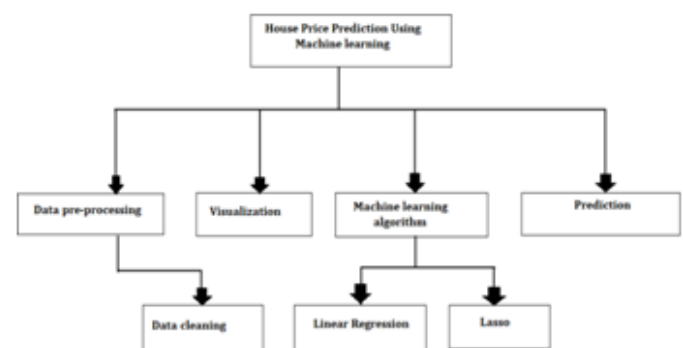


Figure 1. Architecture Diagram

IV. MODULE DESCRIPTION

a) Data collection

Through the internet, we have gathered a number of datasets related to the prediction of home prices. The dataset provides us with the precise values for a specific Bangalore region. Using the dataset, we develop the algorithm and determine the precise price information for the houses.

b) Data Pre-Processing

It is a method for transforming raw or primary data into a clean dataset that can be used to create and train machine learning models. The initial step in each endeavour is to pre-process the data. We need to clean the dataset before we begin our implementations.

Reduce the null values column data and incorrect data in the dataset by cleaning the dataset. The outlier of that particular location in the dataset must then be removed. When we are ready to use the algorithm, data preparation is required.

c) Visualization

It is a graphical representation of the given data by using various visual elements such as charts, graph or another visual format. After the data pre-processing process, we need give the representation of charts and graphs. In our project also we implement the outlier graphs and pie plot to give the accurate representation of the Bangalore dataset region.

d) Machine Learning Algorithm

The algorithms used in machine learning are those that convert a dataset into a model. To anticipate the price of a home, we have examined a number of algorithms here, including linear regression, lasso, and decision trees. This method provides the correct outcomes and exact data accuracy combined with accurate depiction. We import a dataset for linear regression from the internet.

In order to select the relevant and pertinent algorithm for our linear regression dataset (Bangalore Region Dataset), please click here. One of the most extensively used algorithms for linear datasets is linear. Lasso and the linear algorithm are comparable. In comparison to the other two algorithms, the decision tree produces outcomes that are highly accurate.

The algorithms used in machine learning are those that convert a dataset into a model. To anticipate the price of a home, we have examined a number of algorithms here, including linear regression, lasso, and decision trees. This method provides the correct outcomes and exact data accuracy combined with accurate depiction. We import a dataset for linear regression from the internet. In order to select the

relevant and pertinent algorithm for our linear regression dataset (Bangalore Region Dataset), please click here. One of the most extensively used algorithms for linear datasets is linear. Lasso and the linear algorithm are comparable. In comparison to the other two algorithms, the decision tree produces outcomes that are highly accurate.

d) Prediction

Prediction displays an algorithm's output. Prediction uses data as an input to forecast the value of data in the future. The prediction method provides us with the compared model details and model structures in addition to the accurate outcomes of the model. Because of this, we will be able to forecast events far more easily than others. The single factor that makes our linear dataset model more suitable and a model for feasibility studies is prediction.

V. ALGORITHMS

Implementing the algorithms, we made the analysis process more efficient and for good analyzing report and also, we have compared various algorithms such as:

- Linear regression
- Lasso
- Decision tree

The algorithm we have implemented is linear regression.

```

from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import Lasso
from sklearn.tree import DecisionTreeRegressor

def find_best_model_using_gridsearchcv(X,y):
    algos = [
        'linear_regression': {
            'model': LinearRegression(),
            'params': {
                'normalize': [True, False]
            }
        },
        'lasso': {
            'model': Lasso(),
            'params': {
                'alpha': [1,2],
                'selection': ['random', 'cyclic']
            }
        },
        'decision_tree': {
            'model': DecisionTreeRegressor(),
            'params': {
                'criterion': ['mse', 'friedman_mse'],
                'splitter': ['best', 'random']
            }
        }
    ]
    scores = []
    cv = ShuffleSplit(n_splits=5, test_size=0.2, random_state=0)
    for algo_name, config in algos.items():
        gs = GridSearchCV(config['model'], config['params'], cv=cv, return_train_score=False)
        gs.fit(X,y)
        scores.append({
            'model': algo_name,
            'best_score': gs.best_score_,
            'best_params': gs.best_params_
        })

    return pd.DataFrame(scores, columns=['model', 'best_score', 'best_params'])

find_best_model_using_gridsearchcv(X,y)

```

Figure 2.1(The entire implementation of the algorithm Linear, Lasso, Decision Tree)

	model	best_score	best_params
0	linear_regression	0.847796	{'normalize': False}
1	lasso	0.726738	{'alpha': 2, 'selection': 'cyclic'}
2	decision_tree	0.716064	{'criterion': 'friedman_mse', 'splitter': 'best'}

Figure 2.2 (The results of three algorithms, after training and testing completed)

VII. LINEAR REGRESSION

Linear regression is an important algorithm which is mainly used for prediction purposes in machine learning. Known parameters is used in linear regression which is correlated with the output. The known parameter is used to make constant and continuous slope which is used to predict the unknown or the result. Linear regression algorithm works on real or continuous or numerical variable. This algorithm gives linear relationship between dependent and independent variables. Linear also gives us the absolute values and maximum accuracy of the linear dataset models. They have the difference between the Co-Efficient difference between axis it'll differentiate in clear and modular manner.

VIII. LASSO

The expansion of the “LASSO” is Least Absolute Shrinkage and Selection Operation. LASSO used over regression methods for predicting accurately. It makes use of the L1 regularization technique. This algorithm automatically performs feature selections. This LASSO model makes use of Shrinkage.

LASSO is also the perfect algorithm for the linear regression dataset. It'll also quite similar to Linear Regression model the major difference is new line will be created through lasso like rigid model. It will give us the more weighted of the data and small amount of the Bias is developed.

IX. DECISION TREE

To solve the classification and regression problem decision tree algorithm is used. The main objective of the decision tree is to creating a training model which will be used to predict the target variable value. For this prediction we have to learn some simple decision rule. Decision nodes and leave are two entities are used to describe the decision tree.

Decision nodes are uses to split the data and leaves gives the final outcome.

In decision tree some set of if-else conditions are used to classify and visualized data based on the conditions. The decision tree could split the tree into many types and contractively gives us the respective values and accuracy of the datasets. This are initiated the exploder knowledge of the dataset.

The classifier of the decision tree is classified into the dataset several regions and the algorithm implement the analysis and give the accuracy.

X. EXPERIMENTAL ANALYSIS

	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 Trupathi	4 Bedroom	Theramp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Utterahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingatharahalli	3 BHK	Soware	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothnur	2 BHK	NaN	1200	2.0	1.0	51.00

Figure 3.1(Experimental analysis of particular check the given data are appropriate)

The Experimental analysis gives us the sample of the given dataset (Bangalore Region Dataset). This will make us the responsible of the dataset they must be pre-processed, removed all null values and removed all before and after outlier removals in all particular region.

XI. FEATURE ENGINEERING

PRICE PER SQUARE FEET

The Price per square feet implementations is also implemented here for the more accurate demands. This feature is implements and approached for the users who search the house price should be priced in budget and user-friendly need .The middle families are approached this feature so that they can be easily calculated and associated for them to help. In 2017, A Survey of middle families are taken. In that survey they approached Price per square feet can be associated for them to fulfill need.

```
import matplotlib
matplotlib.rcParams["figure.figsize"] = (20,10)
plt.hist(df8.price_per_sqft,rwidth=0.8)
plt.xlabel("Price Per Square Feet")
plt.ylabel("Count")

Text(0, 0.5, 'Count')
```



Figure 4.1 (The Graphical Representation of price per square feet of a particular region in the Bangalore dataset)

The above figure 4.1 is the graphical representation that contains x-label and y-label. Here in this graph x label is price per square feet and y-label is count it represents the price per square feet for a particular region in Bangalore.

Based on number of Bathroom

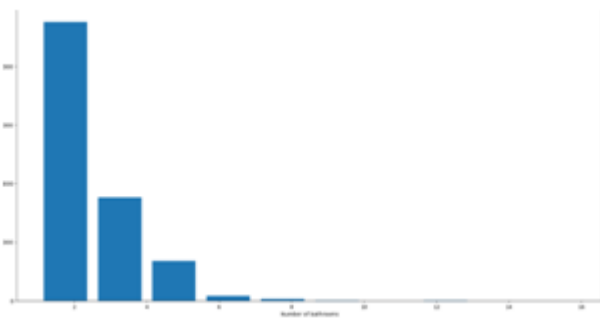


Figure 4.2 (The Graphical Representation of number of bathroom in particular region in the dataset)

The above figure 4.2 is the graphical representation that contains x-label and y-label. Here in this graph x-label is number of bathrooms and y-label is count it represents that number of bathrooms are there in particular region in the dataset. The regions representation is to consider the values in appropriate and accurate results in manner.

XII. PREDICTION

```

1 [2] def predict_price(location,sqft,bathrooms):
2     loc_index = get_location_index(location)
3
4     # Linear Regression
5     X = get_features(location)
6     Y = get_prices(location)
7     # Linear Regression
8     # Fit the model
9     # Predict the price
10    return predict_price(location,sqft,bathrooms)
11
12 [3] predict_price('1st phase JP Nagar',1000, 2)
13
14 [3]: 13000000.0
15
16 [4] predict_price('1st phase JP Nagar',1000, 3)
17
18 [4]: 14000000.0
19
20 [5] predict_price('1st phase JP Nagar',1000, 4)
21
22 [5]: 15000000.0
23
24 [6] predict_price('1st phase JP Nagar',1000, 5)
25
26 [6]: 16000000.0
    
```

Figure 4.3 (Prediction detail of particular region like bathroom)

The above figure 4.3 is we are predicting the details of a particular locations, bathrooms, BHK. For example, for 1st phase JP Nagar location,1000sqft,2bathrooms and 2bhk for these details we had done price prediction. Like this we are predicting for same location but for 3 bathrooms and 3bhk. We can see because of bathroom and BHK count the prediction price also are different. These types of procedures are created that may be consist an format of the representation to find the best price among the results of prediction they gives in proper manner results.

XIII. CONCLUSION & FUTURE WORK

This paper involves in comparing among different algorithm on machine learning (Decision tree Regression, Linear Regression, LASSO Regression) for house price prediction system.

Among those regression algorithm, Linear Regression algorithm gives the best accurate results by another comparing algorithm.

Along with the comparison of various algorithms also gives us the similar analysis report to consider the regression model in terms. But the accuracy and accurate prediction of the price will be appreciated in the market.

In future we need to consider the various techniques to update the house price prediction system using particular regional location. In terms of importing the dataset from internet we collect the GPS location tracking and prediction the particular regional report to get accurate results.

Algorithms such as Naïve bayes, SVM, may be used in this dataset in future. In this dataset we can apply many algorithms and better outcome is predicted.

So, we can give accurate information is given to user. To find out the accurate cost KNN algorithm also be applied.

Therefore, for people and the government this will be helpful. At beginning, we used regression algorithms for our project. But in future various classification algorithms can be used for prediction purposes.

REFERENCES

[1] D. De Cock “Ames, Iowa: Alternative to the Boston housing data as an end of semester regression project” J. of Statistics Education. vol. 19, No. 3, 2011.
 [2] R. Tibshirani “Regression shrinkage and selection via the lasso” J. of Roy. Soc. London. Series B (Methodological),

- pp. 267-288, 1996.
- [3] Rangan Gupta*, Periklis Vasilios Plakandaras+ and Theophilos Gogas “Forecasting the U.S. Real House Price Index”
 - [4] Gupta and Das (2010) Forecasting the US Real House Price Index: Structural and Non- Structural Models with and without Fundamentals
 - [5] H.L. Harter, Method of Least Squares and some alternatives-Part II.International Static Review.1972,43(2), pp. 125-190.
 - [6] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73 Aminah Md Yusof and Syuhaida Ismail, Multiple
 - [7] Regressions in Analyzing House Price Variations. IBIMA Publishing Communications of the IBIMA Vol. 2012 (2012), Article ID 383101, 9 pages DOI: 10.5171/2012.383101.
 - [8] Babyak, M. A. What you see may not be what you get: A brief, nontechnical introduction to over fitting regression-type models. Psychosomatic Medicine, 66(3), 411-421.
 - [9] Gupta, R., Kabundi, A., & Miller, S. M. (2011). Forecasting the US real house price index: Structural and non-structural models with and without fundamentals. Economic Modelling, 28(4), 2013-20
 - [10] A Review on House Price Prediction Using Machine Learning Techniques" by Neelam Kumari, et al.
 - [11] Predicting House Prices with Machine Learning Techniques by Mehmet Celenk.
 - [12] House Price Prediction: Exploring the Factors Affecting House Prices in Singapore by Sheeba J. Ibrahim, Chee Seng Chan, and Yi Yang.
 - [13] A Study on the Application of Machine Learning Algorithms for Predicting House Prices in California" by R. Varun, R. Balakumar, and V. R. M. Venkatesh
 - [14] House Price Prediction: A Comparative Study of Machine Learning Models" by Adetunji Adebayo, Olugbenga A. Oluwabemi, and Gbenga J. Owolabi