

HOUSE PRICE PREDICTION AND RECOMMENDATION

1. Mayur Gawade 2. Aniket Patel , 3. Shantanu Barhanpurkar, 4. Onkar Deshmukh

Department of Computer Engineering, Zeal College of Engineering and Research, Pune

Guide name: Prof. Poonam Patekare , Department of Computer Engineering, Zeal College of Engineering and Research, Pune

Abstract:- Determining how much a house will sell for in a city is still a challenging and time-consuming task. This article's goal is to make predictions about the coherence of non-housing prices. A crucial method to ease the challenging design is to use machine learning, which can intelligently optimize the best pipeline fit for a task or dataset. For individuals who will be residing in a home for an extended period of time but not permanently, it is essential to predict the selling price. Real estate forecasting is a crucial part of the industry. From historical real estate market data, the literature seeks to extract pertinent information. Land price bubbles grow as a result of real estate prices, which leads to macroeconomic instability. The government should look into the variables that drive up real estate prices so that it can use them as a guide to assist stabilize the area. There are many economic circumstances that are in play at the time also have an impact on the selling price of a home.

KEYWORDS- House Price Prediction and Recommendation, Machine Learning Model, Light Gradient Boosting Machine, Random Forest Regression.

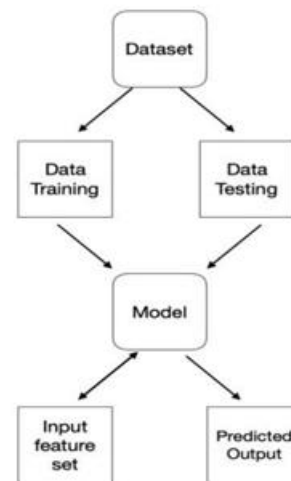
Introduction

A place to call home is among a person's most basic needs, along with other items like food, water, and many other things. As people's living standards climbed over time, so did the need for housing. The majority of people purchase homes for occupancy or as a source of support, however some people buy homes as investments or as real estate.

It's commonly recognized that a number of different elements affect how much a home is worth. As a result, estimating a home's worth includes a special set of problems. House prices vary depending on the

facilities they offer, such as size, area, location, and other factors. It might be difficult to forecast the precise prices of houses. To more accurately estimate property prices and deliver outcomes, this initiative is being proposed. Because one cannot measure or forecast the price of a property based on the location or amenities given, house pricing is a subject that many people, rich and poor, are concerned about. For the people, this would be really beneficial.

The literature review in this article focuses on predicting house prices using a machine learning model and analyzing attributes that were predominantly employed in prior studies that affect house prices. The structure of this essay is as follows: the first section provides a summary of the entire study. The second portion discussed the universal characteristics that are utilized to forecast housing prices everywhere. A brief overview of the machine learning model employed in an earlier study to predict home prices came next. The entire impacts of the present house price prediction model are discussed in the next section. The description and conclusion of this thorough literature analysis are presented in sections 5 and 6, respectively.



Literature Survey

A Deep Learning and ARIMA Model for Predicting House Prices
The relationship between housing prices and determining factors is complicated and nonlinear. The absence of capacity for large developments is another of the most popular methods for predicting home prices. data analysis To deal with these problems, a house price index was created. A deep learning prediction method based on ARIMA is called ARIMA. In this research, a model is suggested. The cost of a home is influenced by a variety of factors. In order to accurately show the shifting rules of housing price, some explanatory components were picked as the significant

determinants. The initial source of the raw housing data is the internet. A data preparation technique is then utilized to change the raw data into outputs that can be quickly used as inputs in data modelling. According to the experimental results, the proposed strategy outperforms the SVR method in predicting the price of a single property. When making short-run predictions, the expected house price trend is essentially consistent with the real data[1].

Shinde and Gawande compared the efficacy of different machine learning algorithms for forecasting the sale price of homes, including lasso, SVR, logistic regression, and decision trees. A technique for forecasting home prices combining regression and particle swarm optimization (PSO) was created by Alfiyatin et al. [2].

Methodology

3.1 Regression Model:-

A. Light Gradient Boosting Machine-

Based on the decision tree method, LGBM is a quick, distributed, high-performance gradient boosting framework that may be used for many different machine learning applications, including classification and ranking. It divides the tree leaf-wise with the best fit since it is based on decision tree algorithms, as opposed to other boosting algorithms that divide the tree depth- or level-wise. As a result, in Light GBM, when growing on the same leaf, the leaf-wise method can reduce more loss than the level-wise strategy, which leads to significantly superior accuracy that can only be sometimes attained by any of the existing boosting algorithms. Additionally, it moves astonishingly quickly, hence the word "light."

B. Lasso Regression-

Least Absolute Shrinkage and Selection Operator is referred to as LASSO. One form of linear regression that makes advantage of shrinking is the lasso regression. It is a regression analysis technique that includes both variable selection and regularization, as the name would imply. Only a portion of the available covariates are chosen for use in the final model through lasso regression. The formula for Lasso regression is $\sum ()$

C. Random Forest Regression-

With the aid of several decision trees using a method known as Bootstrap Aggregation, also referred to as bagging, it is an ensemble strategy capable of carrying out both regression and classification tasks. As it is an ensemble technique, the basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

D. Decision Tree Regression-

This regression trains a model in the structure of a tree by observing features of an object to predict data in the future to produce meaningful continuous output. Continuous output denotes the absence of discrete output, i.e., output that is not

only represented by a discrete, well-known set of numbers or values.

3.2 Evaluation Metrics:-

The prediction accuracy will be evaluated by measuring the R-Squared (R2), and Root Mean Square Error (RSME) of the model used in training. R2 will show if the model is overfitted, whereas RSME shows the error percentage between the actual and predicted data, which in this case, the house prices.

1. Mean Absolute Error:-

The average variation between the dataset's significant values is referred to as the MAE..

$$MAE = \frac{\sum_{i=0}^n y_i - x_i}{n}$$

Where n is the number of samples, y are the target values and y are the predicted values. A MAE closer to 0 means that the model predicts with lower error and the predicted is better closer to the MAE is to 0.

2. Mean Squared Error:-

MSE and MAE are comparable, but a term's impact is inversely proportionate to its size. The mean of all squared absolute values of all mistakes is used to get the prediction error, which is:

$$MSE = \frac{\sum_{i=0}^n (y_i - \hat{y}_i)^2}{n}$$

3. Coefficient of Determination(R-squared):-

R-squared is a statistical indicator of how well a regression model fits the data. R-square values range from 0 to 1. R-square is equal to 1 when the model precisely fits the data and the anticipated value and actual value are identical.

$$R2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}}$$

$$(SST), = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

4. Root Mean Squared Error:-

The model developed in this research will be tested using Root Mean Square Error (RMSE). RMSE is used to calculate predicted performance by considering the prediction error of each data.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=0}^n (d_i - p_i)^2}$$

Conclusion

A resilient model isn't always the same as an optimal model in this research. a model that frequently employs a learning approach that is inappropriate for the current data format. The model is fit even when the data may be excessively noisy or have insufficient samples to allow a model to accurately reflect the target variable. When we look at the evaluation metrics for advanced regression models, we can see that they behave similarly. We can see that advanced regression models behave similarly when we examine their assessment metrics. We can pick any one to forecast house prices compared to the fundamental model. Box plots can be used to search for outliers.

If outliers are present, we can eliminate them and evaluate the model's performance to see if it can be improved.

References

- [1] S. Lu, Z. Li, Z. Qin, X. Yang, and R. S. M. Goh, "A hybrid regression technique for house prices prediction," in 2017 IEEE international conference on industrial engineering and engineering management (IEEM), 2017, pp. 319- 323.
- [2] M. F. Mukhlisin, R. Saputra, and A. Wibowo, "Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor," in 2017 1st International Conference on Informatics and Computational Sciences (ICICoS), 2017, pp. 171-176.
- [3] P. Durganjali and M. V. Pujitha, "House resale price prediction using classification algorithms," in 2019 International Conference on Smart Structures and Systems (ICSSS), 2019, pp. 1-4
- [4] R. E. Febrita, A. N. Alfiyatin, H. Taufiq, and W. F. Mahmudy, "Data-driven fuzzy rule extraction for housing price prediction in Malang, East Java," in 2017

International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2017, pp. 351-358.

[5] W. T. Lim, L. Wang, Y. Wang, and Q. Chang, "Housing price prediction using neural networks," in 2016 12th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), 2016, pp. 518-522.