iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53273

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

House Price Prediction and Recommendation

Prof. Poonam Patekar[1], Onkar Deshmukh[2], Shantanu Barhanpurkar[3], Aniket Patel[4], Mayur Gawade[5]
[1]*Guide, Department of Computer Engineering, Zeal College of Engineering and Research, Pune*
[2, 3, 4, 5]*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.***

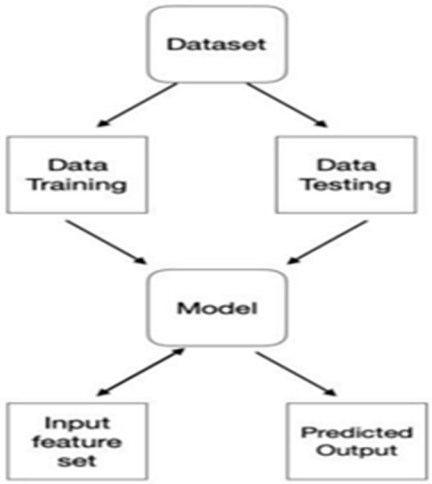
Keywords: House Price Prediction and Recommendation, Machine Learning Model, Light Gradient Boosting Machine, Random Forest Regression.

I. 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.

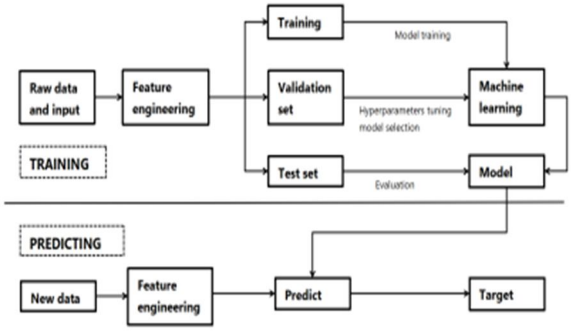


II. LITERATURE SURVEY

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].

III. METHODOLOGY



A. *Cleaning Data*

1) *Data Collection:* Data collection is the methodical process of compiling facts on variables. It supports the pursuit of knowledg, makes excessive hypotheses, and assesses outcomes. Data gathering is done as a means of facilitating social interactions and est imating data on targeted aspects within the pre-existing framework. At this point, related questions can be addressed and the out comes can be assessed.
2) *Data Visualization*: The visual or graphical depiction of data is known as data visualisation. It makes it possible to understand c hallenging ideas or spot novel patterns. This includes developing and researching informational visual representations.
3) *Data pre-processing:* This is how the data is changed before being provided to the algorithm. It is used to transform unclean dat a into a clean data set. Transferring unorganised data into a logical structure is part of this information mining method. Fill up th e blanks with logically organised raw data. The final dataset utilised for preparation and testing is the outcome of data pre-proce ssing.
4) *Data Cleaning:* To increase the value of data, data cleaning is the process of identifying and eliminating inaccuracies. Using dat a processing technologies, data cleaning is accomplished. That is a method for locating and altering records from a record set, ta ble, or database that are inaccurate. It locates the information that is lacking and changes the jumbled information. To make sur e the information is accurate and proper, it is edited.

B. *Regression Model*

1) *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."

2) *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 regularisation, 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, ∑ ( )

3) *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.

4) *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.

C. Evaluation Metrics

The prediction accuracy will be evaluated by measuring the R-Squared (R2), and Root Mean Square Error (RSME) of the model use d in training. R2 will show if the model is over fitted, whereas RSME shows the error percentage between the actual and predicted d ata, 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..

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

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:

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

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.

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

$$\text{(SST),}=1-\frac{\sum(\text{yi}-\hat{}\text{yi})2}{\sum(\text{yi}-\bar{}\text{y})2}$$

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.

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

IV. IMPLEMENTATION

A. Problem Definition

Define the issue, which is forecasting home values and offering advice based on the information.

B. Data Collection

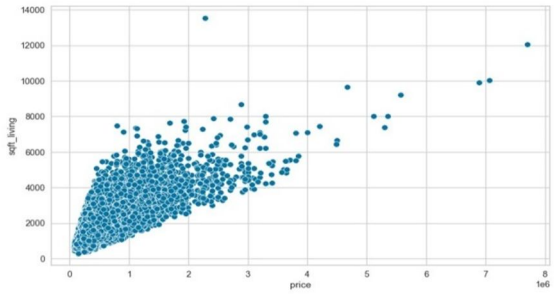
Gather the pertinent information for your prediction model's training. This may include information on housing characteristics and associated costs, such as size, number of rooms, location, etc.

C. Data Preprocessing

To ensure the quality and compatibility of the acquired data with the prediction model, clean and preprocess it. This stage could involve dealing with outliers, missing values, and categorical variable encoding.

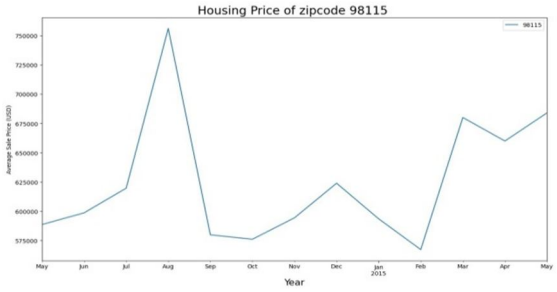
D. Model Development

In this project, we are using a variety of regression models, including Decision Tree Regression, Random Forest Regression, and Light Gradient Boosting Regression. Divide your data into training and testing sets, then use the training data to train the chosen models.



E. Model Evaluation

Employing proper evaluation metrics (such as mean squared error, mean absolute error, etc.) to assess each trained model's performance on the testing set. To determine which model has the highest accuracy, compare their respective accuracies.



F. Model Selection

Selecting the model with the highest accuracy as your final prediction model.

In the following table the highest accuracy in the following regression model is Light Gradient Boosting Regression.

Regression Model	MAE	MSE	RMSE	R2
Light Gradient Boosting Machine	68839.2918	1623.4938	127379.1577	0.8781
Lasso Regression	119369.1562	3686.8000	191887.8877	0.7243
Random Forest Regression	7121.2060	1824.8560	134803.6375	0.8639
Decision Tree Regression	103452.0416	3849.6384	195174.5096	0.7133

G. Model Saving

Save the Light Gradient Boosting Regression (LGB) model to a file (e.g., using the different library) for future use.

H. Flask App Development

Create a Flask application to serve as the backend for your house price prediction and recommendation system. This involves defining routes, handling HTTP requests, and connecting to the prediction model.

I. Deployment Setup

Install and configure XAMPP, which provides a local web server environment for running your Flask application. Set up the necessary configurations, such as defining the server port and ensuring proper file placement.

J. Flask App Deployment

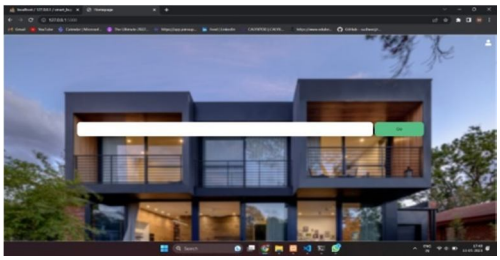
Deploy your Flask application by placing the `main.py` file (containing your Flask app code) into the appropriate directory within XAMPP, such as the `htdocs` folder.

K. Run XAMPP Server

Start the XAMPP server, which will run your Flask application locally. Ensure that the server is running and accessible.

L. Access the Web Application

Open a web browser and navigate to the appropriate URL (e.g., `http://localhost:8080`) to access your deployed house price prediction and recommendation system.



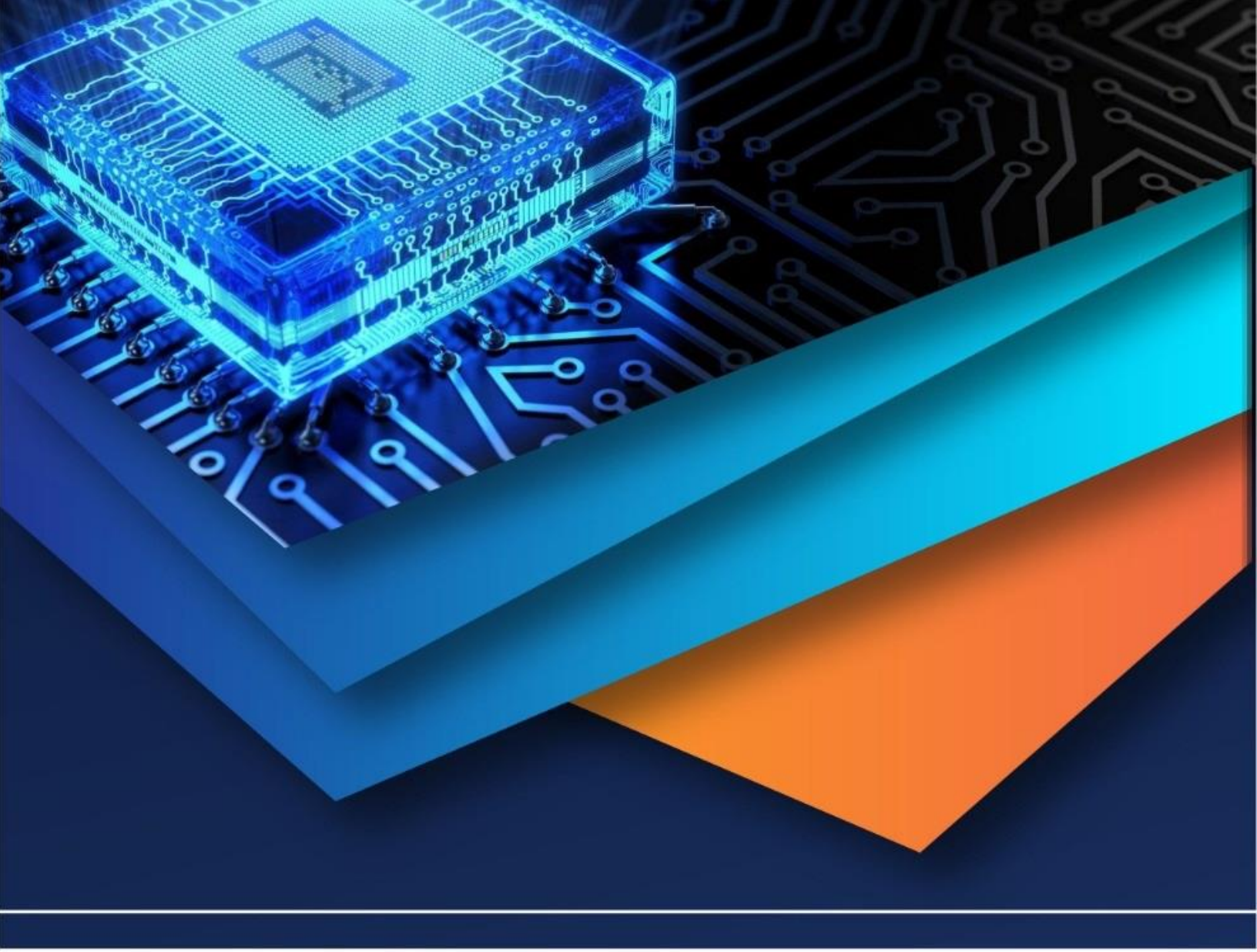
These steps is helping in developing and deploying your project. We are making sure to test the application thoroughly to ensure it works as expected.

V. 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. Mukhlishin, 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.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY