JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

House Price Prediction

[1]Tushar Rao ,[2]Himanshi Sharma ,[3]Kalika Gupta, [4]Ishan Kumar
[1]Student, [2]Student, [3]Student, [4]Assist. Professor
[1]School of Computer Science and Engineering , Lovely Professional University, Phagwara, Punjab, India

***Abstract*:** This research is based on application which helps the buyers and sellers to predict house prices so that they are able to make a right choice based on different factors such as locations, areas, connectivity of the property and their dream of buying a wonderful house in their budget comes true without any difficulty. This research not only focus on location apart from that our project make sure to aware them about each and every details regarding features , roads, nearby facilities and basically it asks the user to provide / add accurate information about number of rooms , kitchen , bathrooms in the desired property and it enables the customers or buyers to choose the size according to their needs and number of rooms according to their family . This research also helps many buyers who are new in the business , we are making sure to provide them the estimated price based on their requirements. So basically to get to know the price of a property /land / house we want that this research helps user or customers never made a wrong decision depends on different prices based on linking of road , availability , health facilities . The main reason of being our application a customer friendly because it provides optimistic result for a property.

***IndexTerms:* Regression , Machine Learning , Random Forest**

1. INTRODUCTION

House is one of human life's most essential need. Demand for houses grew rapidly over the years as people's living standards improving day by day. While there are people who make their house as an investment and property, yet most people around the world are buying a house as their shelter or as their livelihood.

There are many benefits of house price prediction for the buyers, property investors, and house builders can reap from the house-price model. This research will provide a lot of information and knowledge to home buyers, property investors and house builders, such as the predict price of house in the present market, which will help them to determine the factors are available to their nearby area.

Meanwhile, this model can also help potential buyers to decide the characteristics of the house they are looking for. The sales price of a property can be predicted in various ways, but is often based on regression techniques. All regression techniques essentially involve one or more predictor variables as input and a single target variable as output.

selling price of houses based on a number of features such as the area, the number of bed- and bathrooms and the geographical position.

2. LITERATURE SURVEY

In this research paper we have to assay the different Machine Learning algorithms for better training Machine Learning model. Trends in casing cost show the current profitable situation and as well as to directly concern with buyers and merchandisers. Factual cost of house is depending on so numerous factors. They include like no of bedrooms, number of bathrooms, and position as well. In pastoral area cost is low as compare to megacity. The house price grate with like near to trace, boardwalk, job openings, good educational installations etc. Over many times agone, the real estate companies trying to prognosticate price of property by manually. In company there is special operation platoon is present for vaticination of cost of any real estate property. They are decide price manually by analysing former data. But there 25% of error is passed on that prediction.so there is loss of buyers as well as merchandisers. Hence there are numerous systems are developed for house price prediction. The main motive of this research is to make a model which give us 87% approx. prediction based on various features [1],[2].

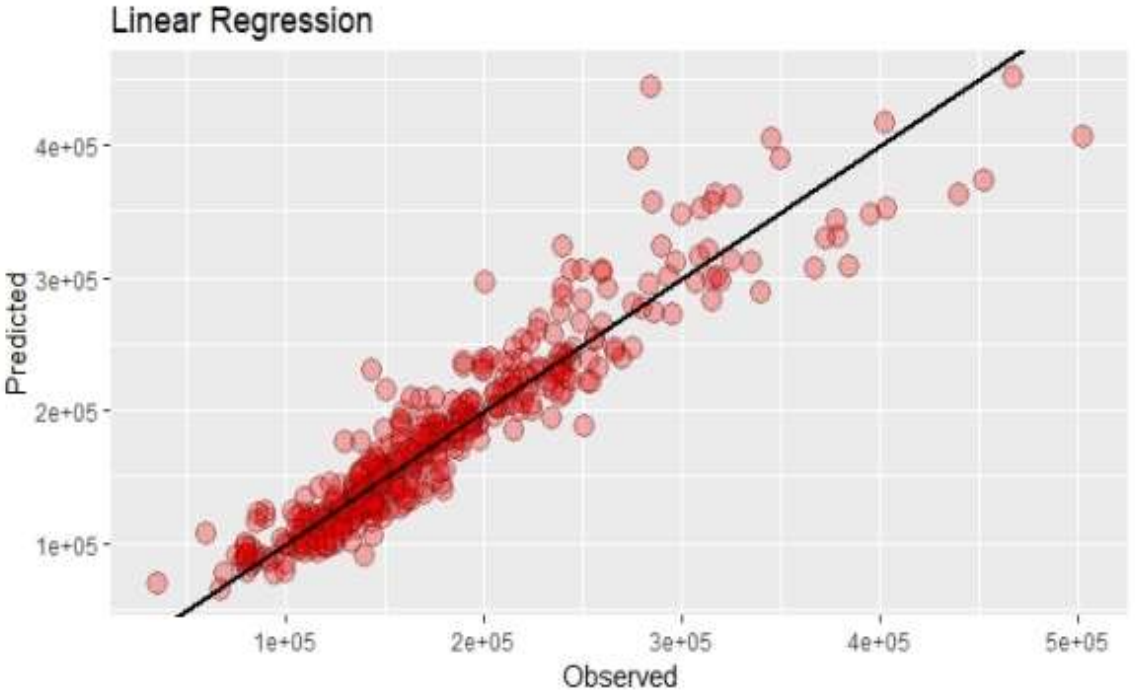

figure:1 house price prediction using linear regression.

3. ATTRIBUTE

Attributes are the type of data that are used in machine learning. Attributes are also known as variables, fields, and predictor. House price prediction can be divided into two categories first by focusing on house characteristics and second by focusing on the model used in house price prediction [3]. According to [4] the component affecting house price can be described into three categories Location, structure, and neighbourhood. [5]

Location : Location is considered to be the most important feature of house price resolution. The location of the property was classified in a fixed locational attribute. All of these studies point to the close association between locational attributes such as distance from the closest shopping centre, or position offering views of hills or shore, and house price variations [3].

Structural : Structural is a feature that people may identify, whether number of bedrooms and bathrooms, or floor space, or garage and patio. These structural attributes, often offered by house builders or developers to attract potential buyers, therefore meet the potential buyers' wishes [3].

Neighbourhood : Neighbourhood feature can be included in deciding house price. According to effectiveness of school, industrial and hospital etc generally improve the worth of a property [3].

4. MACHINE LEARNING MODEL

A machine learning model is an expression of an algorithm that combs through mountains of data to find patterns or make predictions. Fuelled by data, machine learning models are the mathematical engines of artificial intelligence. For example, an ML model for computer vision might be able to identify cars and pedestrians in a real-time video. One for natural language processing might translate words and sentences. The model selection to be used to predict house price is quite critical as varieties of models are available. Most utilized models in this research field is Regression Analysis.

5. REGRESSION ANALYSIS

5.1. DECISION TREE

Decision tree make regression or classification models in the formation of a tree structure. It helps to divide a complex dataset into tiny and compact subsets while at that same time a linked decision tree incrementally developed. The end outcome is going to be a tree with decision nodes and leaf nodes. Decision Tree is most frequently used algorithm used for superintend studying. It can be used to answer the couple of things like Regression as well as Classification [6].

It is basically a tree like structure classifier having three kind of nodes. The *root node* is the opening node which simply constitute entire sample and gets divide further into different nodes. The *interior node* constitute characteristic of a data set as well as the branches constitute the decision rules. Eventually, the *leaf node* constitute the end result. This algorithm is helpful in solving decision-related problems. There are various types of Decision Tree Terminologies present in this algorithm like root node , leaf node , splitting which means splitting of the root node into furthermore nodes as the condition applied . now comes pruning which is simply means removal of the branches that are not required by the tree or may be not useful for the tree. Lastly Parent and child node the base nodes are known as parent node and the nodes that are further derived from the base node are known as child nodes. For predicting dataset ,working of the algorithm begins from root node . First of all it will compare the value of root attribute from real dataset and then comes up to next node. Then it repeatedly compares the value of root with other nodes till it comes to leaf node [7].

5.2. RANDOM FOREST

Random Forest is a ensemble technique used for representing regression as well as classification tasks having multiple decision trees and a technique known as Bootstrap and Aggregation, usually called as bagging. The simple strategy used in this is to combining many decision trees so in the end it will help in determining the outcome rather than depending on particular trees. It has many decision trees as there base models. We randomly carry out row sampling as well as feature sampling. This is known as Bootstrap. We just need to use the Random Forest regression like any other machine learning technique. The Decision Tree is simple and basic interpreted algorithm that is used for single tree purposes but we can't use it for models that include models that's why we required something different to overcome this Random forest regression is used. Random Forest is known as Tree-like algorithm that simply uses the features of multiple Decision Trees that helps in making decisions. Therefore, it is known as a 'Forest' of trees and since from then it is named as "Random Forest". The word 'Random' simply denotes the fact that algorithm is a forest of 'Randomly created Decision Trees' [8] [9].

5.3. GRADIENT BOOST

Gradient boosting was created in 1999 and is a commonly used machine learning algorithm because of its performance, consistency and interpretability [10]. Gradient boosting delivers state-of-the-art in various machine learning activities, such as multistage classification, click prediction and ranking. With the advent of big data in recent years, gradient boosting faces new challenges, especially with regard to the balance between accuracy and performance. There are few parameters for gradient boosting. To ensure a dynamic balance between fit and regularity, the following steps can be taken to select parameters: Setting regularization parameters (lambda, alpha), reducing learning rate and decide those optimal parameters again [3]

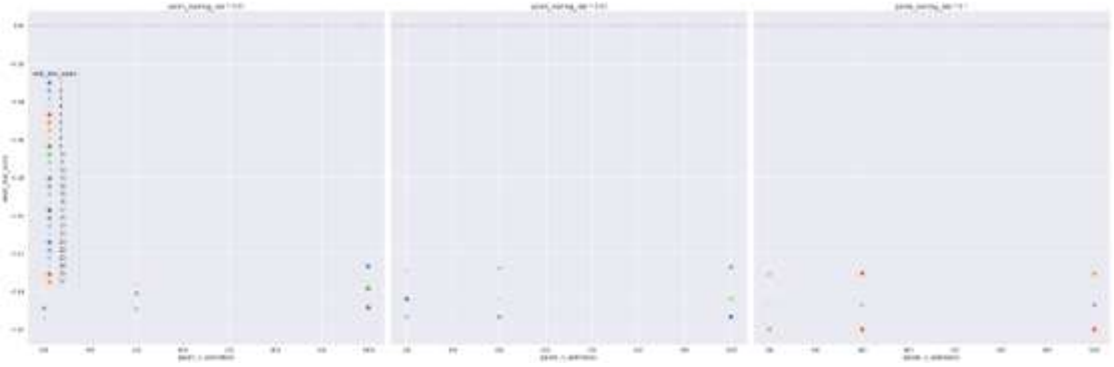

figure 2: gradient boosting

6. METHODOLOGY

6.1. DATA PRE-PROCESSING

A dataset containing more than 20,000 data with different attributes representing expected details in today's era. These variables which served as different features of the dataset, which we used to predict the estimated price of the house based on the details of property / land / house.

The next step was to investigate and clean the data. There are majorly variables which are missing from the dataset. Any observation which consists of some missing or incomplete variables were removed and thus the data the dataset and to clean the data we took these engineering processes steps: -

Removing the attributes, the number of lawns and different features because of incomplete data in the dataset.

Set the number of House Area from 1BHK to 4BHK.

Split the area into different features based on the area covered.

The equation for the multiple regression analysis is: -

$$Y^\wedge = a + b1x1 + b2x2 + b3x3 + \ldots . + bkxk.$$

where $x1 \ldots . xk$ are independent variable.

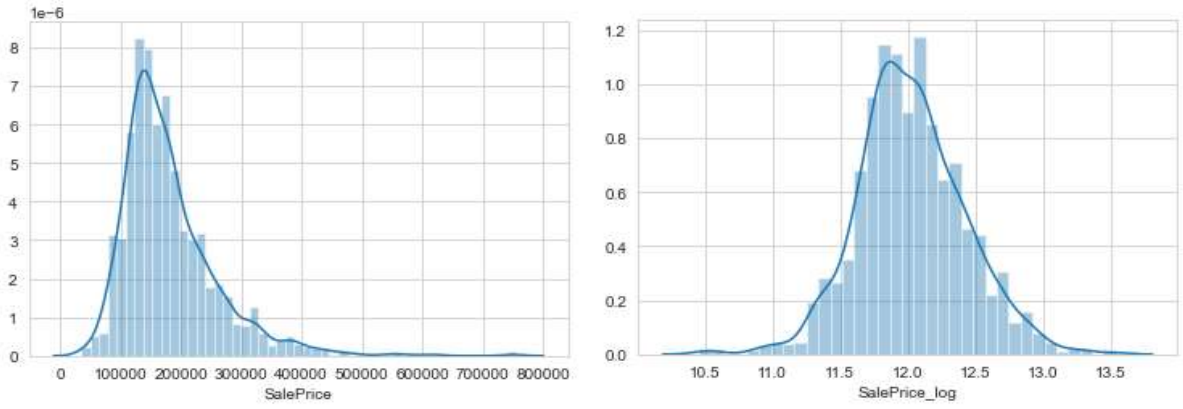

figure 3: sale price histogram before and after log transformation.

3.2. DATA ANALYSIS

So basically, data analysis is an important step before building a regressing model or to work on a dataset. By this way, researchers or users discovers the implicit information of the data, which turns out to be helpful to approach machine learning models.

3.3. TRAINING THE DATA

Training of the data consists of 1,460 examples of houses with different features describing every aspect of the house. We are given sale prices for each house. The training data is what we will use to teach our models/ dataset.

3.4. TESTING

The test data set consists of 1,459 examples with the same number of features as the training data. Our test data set excludes the sale price because this is what we are trying to predict. Once our models have been built we will run the best one the test data and submit it to the Kaggle leader board.

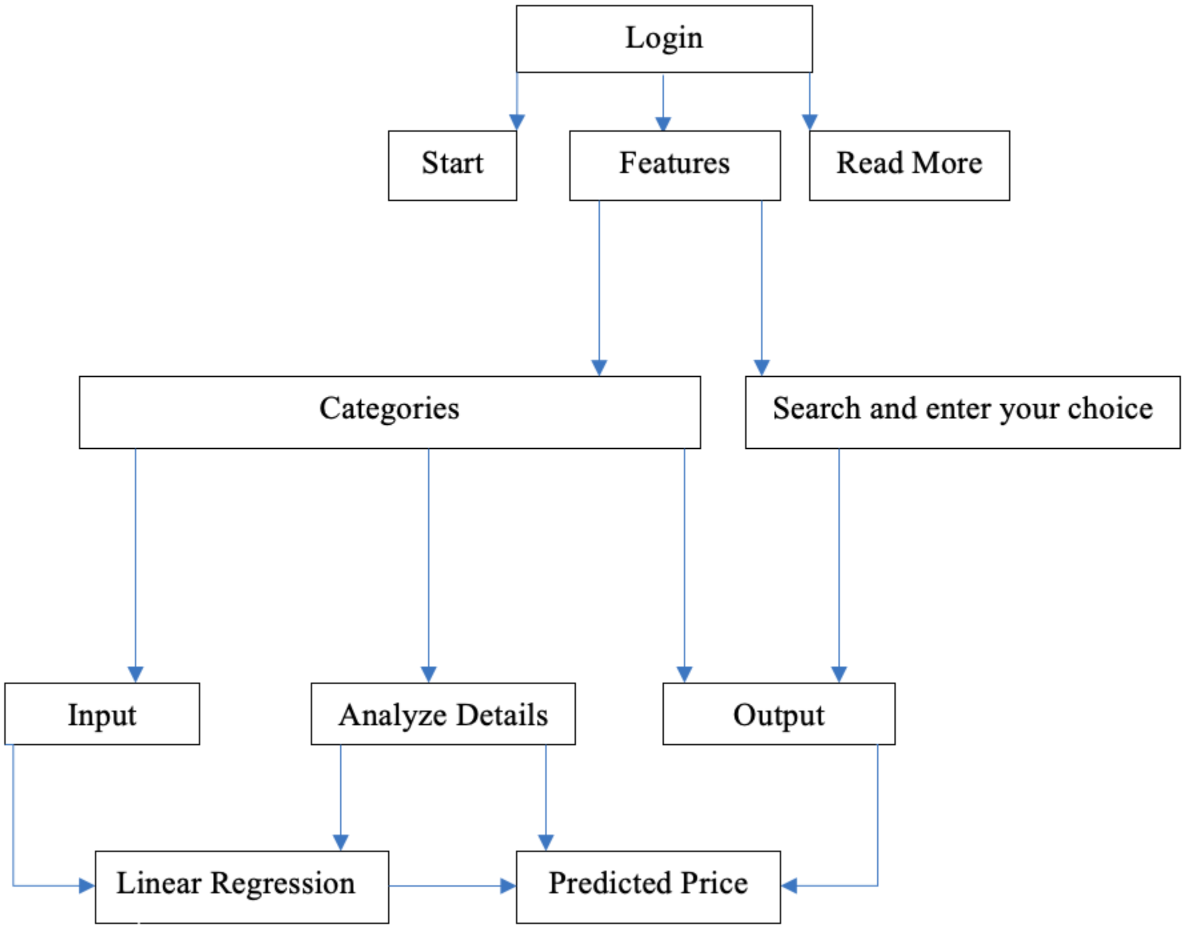

figure 4: architecture of web application

7. CONCLUSION

This paper examined and analysed the current research on the significant attributes of house price and analysed the data mining techniques used to predict house price. Technically, houses with a strategic location such as the accessibility to shopping mall or other facilities tend to be more expensive than houses in rural areas with limited numbers of facilities. The accurate prediction model would allow investors or house buyers to work out the realistic price of a house also because the house developers to make a decision for the affordable house price. This research addressed the attributes used by previous researchers to forecast a house price using various prediction models. These models were developed based on several input attributes and they work significantly positive with house price. In conclusion, the impact of this research was intended to help and assist other researchers in developing a real model which can easily and accurately predict house prices. Further work on a true model must be through with the use of our findings.

(1) Coupling effect of multiple regression.
(2) Re-learn from the past input.
(3) Combination of machine learning and artificial intelligence

8. REFERENCES

[1] A. G. RAWOOL, D. V. ROGYE, S. G. RANE and D. V. A. BHARADI, "House Price Prediction Using Machine Learning," May 2021.

[2] A. Babu and D. A. S. Chandran, "Literature Review on Real Estate Value," Akshay Babu et al, International Journal of Computer Science and Mobile Applications, 2019 March.

[3] N. H. Zulkifley, S. . A. Rahman and U. . N. Hasbiah, "House Price Prediction using a Machine Learning Model: A Survey of Literature," International Journal of Modern Education and Computer Science, 2020.

[4] N. A. , E. R., T. H. and F. W., "House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study : Malang, East Java, Indonesia," 2017.

[5] "Oracle," [Online]. Available: https://docs.oracle.com/en/database/oracle/machine-learning/oml4sql/21/dmprg/about-attributes.html#GUID-7AAB55D5-6711-4BE5-A0CE-B2A6B68ED689.

[6] G. M K, "Machine Learning Basics: Decision Tree Regression," july 2020.

[7] "JavaPoint," [Online]. Available: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm.

[8] "Geeksforgeeks," [Online]. Available: https://www.geeksforgeeks.org/random-forest-regression-in-python/.

[9] G. M K, "Machine Learning Basics: Random Forest Regression," 2020 july.

[10] J. H. Friedman, "Scochastic gradient boosting," 1999.

[11] H. Patel and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING, 2018.

[12] J. Ali, R. Khan and N. Ahmad, "Random Forests and Decision Trees," 2012.