a
Journal of Engineering Sciences

Vol 13, Issue 06, June/2022
ISSN NO:0377-9254

Real Estate Price Prediction Using Machine Learning Algorithm

Mr. K. Chandra Sekhar Reddy,[1] Uppugunduru Lokesh [2]

[1]Asst. Professor, Department of Computer Science and engineering

[2]Student, Department of Computer Science and engineering

[1,2]Vikas College Of Engineering & Technology

Abstract- Linear Regression (LR) and other Machine Learning algorithms are used to forecast the price of real estate. Therefore, it is possible to utilise the expected price to assist property/house sellers set their selling prices. The concept that housing prices are influenced by characteristics such as location, distance, and region is known as price prediction. Because of this, the prices they expect to pay are vastly different from what they really do. A regression model was constructed to forecast the price of residential houses. Entering the data into the website will get the desired outcome. Then, the input data will be subjected to the regression procedure. The user's input will be fed into the model, and the estimated sale price of the property will be shown on the website in a matter of seconds.

Keywords— Houses, Price, prediction, Location, Real Estate, Machine Learning.

I. INTRODUCTION

There are several basic requirements in existence, such as a place to live and food and drink, that cannot be ignored. In recent years, people's living standards have continued to rise, resulting in an increased need for housing. The vast majority of individuals in the globe purchase a home to live in or to utilise as a source of income or as a place to save their money. Every year, there is a rise in demand for housing, which results in an increase in home prices. Because of the many variables that can affect a home's price, such as its location and the demand for specific types of property, it can be difficult to accurately determine the exact attributes or factors that influence a home's final selling price [1]. This makes it difficult for investors to make informed decisions and for home builders to accurately set a home's final selling price. Predicting fluctuations in the price of a home is a frequent practise in the real estate industry. Because of the high correlation between home prices and other variables like location, area, and population. Traditional machine learning algorithms have been used in a significant number of studies to forecast housing values [2]. But We opted to apply the Regression method in our project. This kind of supervised learning is called linear regression, and it is used to do regression. It is mostly used for the purpose of establishing a connection between two otherwise unrelated variables. Users may input square feet, bedrooms, location, and the anticipated price on the website to make it more user-friendly.

Using a predictive model approach, this study proposes a method for predicting the price. The regression approach is chosen based on the attenuation caused by the expected price. We used the Regression method in the course of working on this project. One or more independent variables are used in regression analysis to represent the connection between a dependent (target) and an independent (predictor) variable [3]. Regression analysis, in particular, enables us to understand how the value of a dependent variable changes as a function of an independent variable while other independent variables remain constant. Forecasts real-world data like as temperature (temperature), age, income, and price. Linear regression is one of the most often used Machine Learning methods. It's a statistical technique for making predictions. Linear regression predicts continuous/real or quantitative variables such as sales, salary, age, product price, and so on.. When a dependent (y) variable and one or more independent variables are

shown to have a linear relationship, it is referred to as linear regression. It determines how the dependent variable's value changes as a function of the independent variable's value since linear regression indicates a linear connection [4]. It is used for regression testing. Based on a set of independent factors, regression calculates a predicted value for a target variable. It is mostly used for determining the correlation between variables and predicting. The kind of link between dependent and independent variables that each regression model considers, as well as the quantity of independent variables that it employs, make each unique [5].

II. RELATEDWORKS

The most important phase in the software development process is a literature review. Before creating the tool, it is essential to determine the time, cost, and strength of the firm. After this is done, the following 10 stages are to choose which software and language will be used to construct the tool. Once the programmers begin constructing the tool, they will need a great lot of assistance from other sources [6]. Senior programmers, a book, or websites provide this assistance. The above-mentioned considerations are taken into account before the suggested system is built. There aren't any comprehensive frameworks out there that take into account all the many ways that ideas might operate together. Many studies on real estate price prediction using regression analysis have relied heavily on generic characteristics. Several scientists collaborated on this endeavour.

To do so, one must be able to estimate the cost of a home. We can accurately anticipate the value of a home by using the right method. In this work, the feature engineering approaches that were used have been described in the literature study. In addition, assessment measures are used to assess the algorithms' performance [7]. Factors utilised in the local dataset are also included. There are several basic requirements in existence, such as a place to live and food and drink, that cannot be ignored. In recent years, people's living standards have continued to rise, resulting in an increased need for housing. The vast majority of individuals in the globe purchase a home to live in or to utilise as a source of income or as a place to save their money. The purpose of statistical analysis is to shed light on the interrelationships between various aspects of a home's design and its eventual sale price. Predicting fluctuations in the price of a home is a frequent practise in the real estate industry [8]. Because of the high correlation between home prices and other variables like location, area, and population. Traditional machine learning algorithms have been used in a significant number of studies to forecast housing values. Regression, on the other hand, was used in our project. This kind of supervised learning is called linear regression, and it is used to do regression. It is mostly used for the purpose of establishing a connection between two otherwise unrelated variables. Users may input their preferences, such as the number of bedrooms and square footage, on a website that provides an estimate of the final cost. Product demand, qualities, and current real estate trends are all taken into account by an algorithm to anticipate a product's price. Next, the algorithms that use machine learning determine a pricing that is both attractive to buyers and likely to result in a high level of sales [9].

The purpose of this statistical study is to shed light on the link between house attributes and the factors used to estimate the price of a home. In this case, we used the most recent technologies, such as Regression. As a starting point, we use the "Real estate price forecast using machine learning" website to train our model [10].

There is a great deal of effort being put into developing models that can discover patterns in large datasets and extrapolate those trends into future output. Researchers have used a variety of machine learning algorithms and pre-processing data approaches in their studies [11].

The current approach does not take into account future market trends and price increases when calculating home values. Using artificial neural networks vs. multiple linear regression for forecasting, Suna Akkol, Ash Akilli, and Ibrahim Cemal conducted a research in 2017. Their research used artificial neural networks and numerous linear regression analyses in order to examine how changes in morphological measurements affect live weight. They employed Levenberg-Marquardt, Bayesian regularization, and Scaled conjugate back-propagation methods for ANN. In their experiment, they found that ANN outperformed multiple linear regression in terms of accuracy. They found that the Regression method didn't work out well for them [12].

III. PROPOSED SYSTEM ARCHITECTURE

There are several basic requirements in existence, such as a place to live and food and drink, that cannot be ignored. In recent years, people's living standards have continued to rise, resulting in an increased need for housing. The vast majority of individuals in the globe purchase a home to live in or to utilise as a source of income or as a place to save their money. The purpose of this statistical study is to shed light on the link between house attributes and the factors used to estimate the price of a home. Predicting future property prices may be done in a plethora of methods. However, one method we're considering is creating a model based on data from the Bangalore area to estimate property prices. In this case, we used the most recent technologies, such as Regression. As a result, we create a model for real estate price prediction using machine learning that is fed a dataset and has access to current input information for the city in question. Minimize the discrepancy between the projected and actual rating when estimating the home price using two alternative models which is shown in Fig.1. This system's goal is to calculate a home's value based on the many characteristics that a user provides as input. These attributes are fed into the ML model, which then makes a prediction based on how they impact the label. The first step is to choose a dataset that meets both the developer's and the user's requirements. Data cleaning is performed after the dataset has been finalised and the raw data has been converted to a.csv file, in order to remove any data that is no longer required. For further pre-processing, missing data will be handled and if necessary, labels will be encoded in the data. Additionally, it will be transformed into a NumPy array before being transmitted to the model's training environment. Various machine learning techniques will be employed to train the model, and their error rates will be retrieved, resulting in a final algorithm and model that can make correct predictions. You may log in and fill out a form about the many aspects of your property that you'd want to know how much it will cost. Additionally, the form will be sent after a comprehensive selection of qualities. The user's input will be sent to the model, and the estimated price of the property will be shown to them in a matter of seconds.

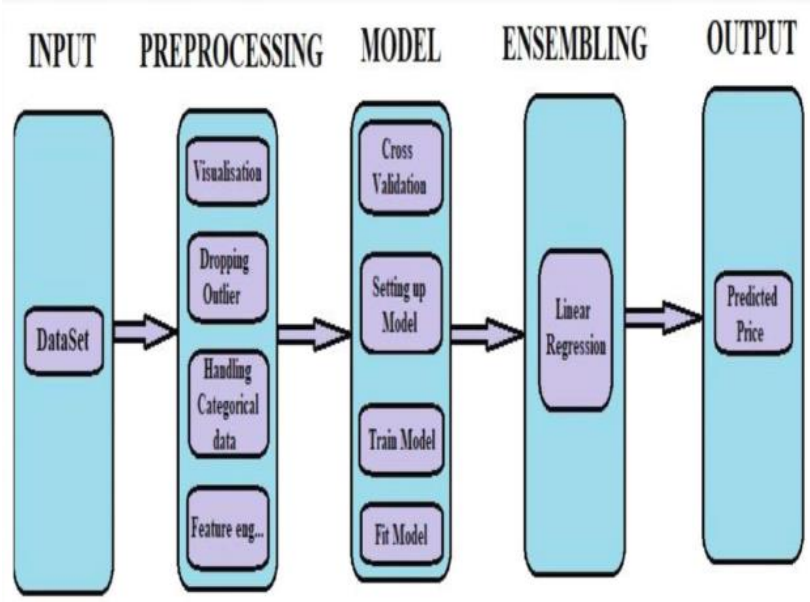

Fig.1 Proposed Methodology

The accompanying flowchart demonstrates in detail how our project receives input, processes it, and then outputs the results. Here, we go over it step-by-step. Various datasets must be gathered to provide input to the system or project. As a more developed city with greater resources, "Banglore dataset" was one of the datasets we used. Additionally, we gather a dataset of 13320 entries with 9 columns. Data is preprocessed once it is received as input. Understanding historical data and applying its results to fresh information is the primary function of machine learning algorithms. It verifies the accuracy of the dataset as a whole. Dropping outliers is the process of removing outliers from the data set. It gets rid of everything that's gone astray. Numeric data is required for all input and output variables of machine learning models. In order to fit and assess a model, you must convert your categorical data to numbers first. Extracting features from raw data using domain expertise is known as feature engineering. The goal is to enhance the quality of machine learning outputs by including these additional characteristics. Afterwards, the model undergoes a series of phases. When evaluating a trained model against a testing dataset, model validation is known as cross validation. When we talk about "setup," we're talking about creating a template for teaching and testing. To train the model, utilise just around 80% of your dataset, and only about 20% of your dataset to test it.

After training and testing, the model goes through regression approach and compares it with decision trees and lasso regression to ensure that the individual parameters of our machine learning model are best suited to address our unique real world business challenges with a high degree of accuracy. The model will then forecast the price depending on the input variables.

The data obtained over time helped us train our regression approach model. The data was broken down into many categories, such as the number of bedrooms, the square footage, the location, and the number of bathrooms. There were a total of 13320 training examples utilised. Our project's end objective is to anticipate a price that can be used to estimate a price with ease, which is an important next step. The Banglore dataset is a part of our project. Because Banglore is a developing metropolis, this data set was taken into consideration. All of the unnecessary columns and rows will be eliminated. Using a website, the criteria may be entered and the website will display the outcome. An efficient model for predicting a price is built using the data that was learned in the first module. For the sake of both testing and forecasting, the data is split

in half. Algorithm has to be trained so that it can anticipate home prices based on current information, and this is done by first providing the housing factors as input and testing the data. We can quickly import the necessary packages for the graphical representation of the data from the matplotlib library, which is a library for Python. For the sake of clarity, we've included some graphs below. We'll use this graph to show the elimination of outliers from the data. There is a graph between the price per square foot and the graph after deleting the unnecessary data, which is shown. The dataset used in our work is shown in Fig.2

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	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

Fig.2 Dataset used in our research

IV. RESULTS AND DISCUSSION

The output and graphs obtained after executing the implementation code is shown from Fig. 3 to Fig.7.

Fig.3 Input/Output

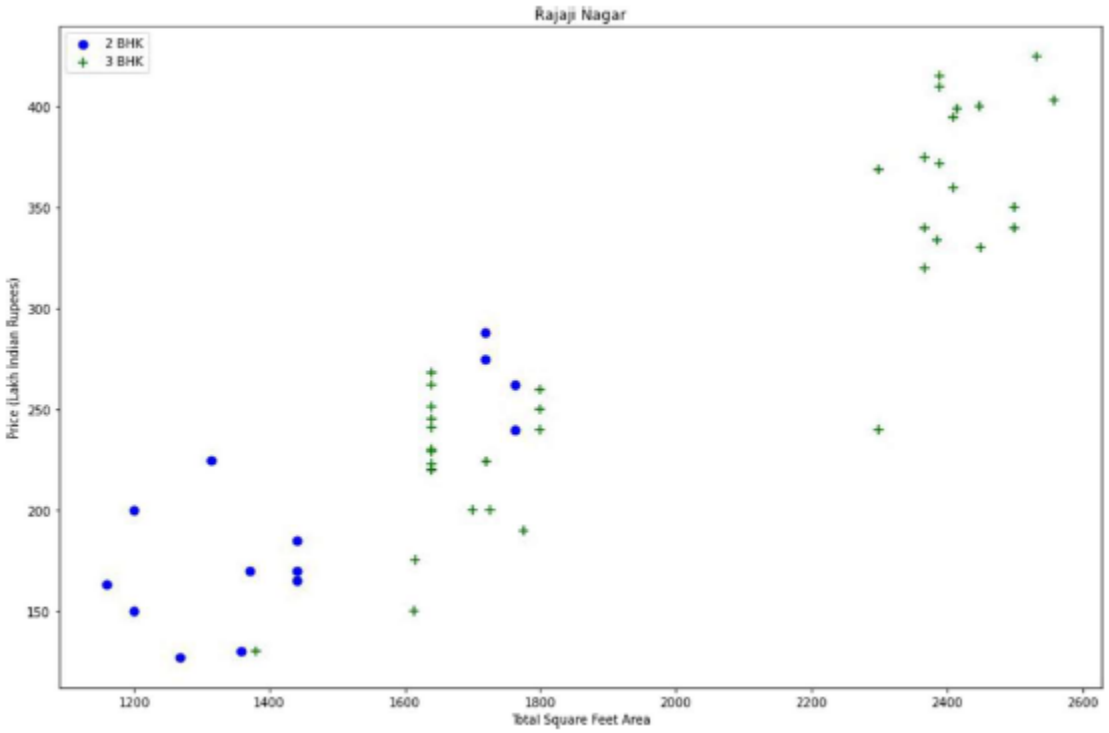

Fig.4 Price graph for Rajaji Nagar

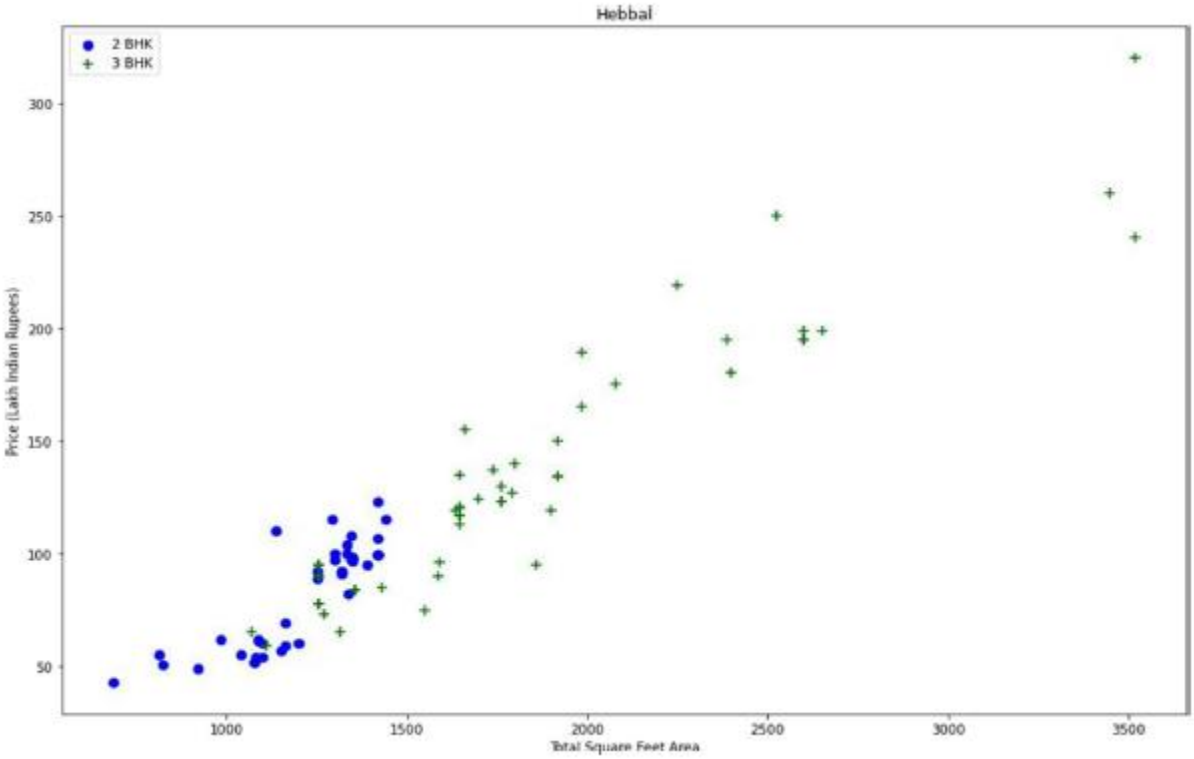

Fig.5 Price graph for Hebbal

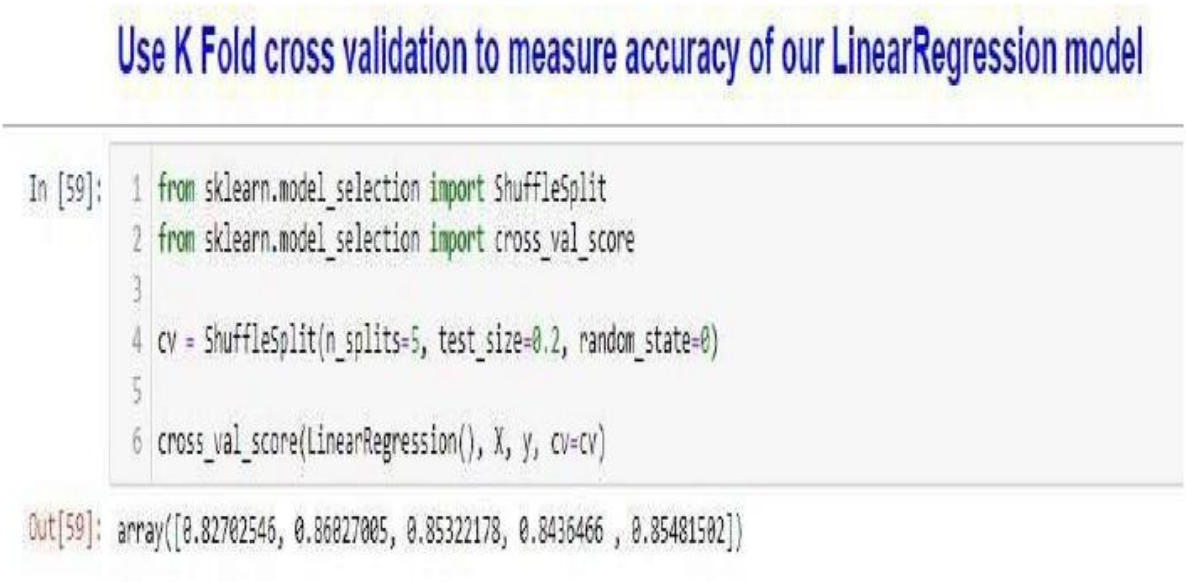

Fig.6 Accuracy Using Linear Regression Model

Fig.7 Price Prediction for Indira Nagar and JP Nager

V. FUTURE SCOPE AND CONCLUSION

Predicting fluctuations in the price of a home is a frequent practise in the real estate industry. Investors or homebuyers might use this reliable prediction model to figure out what a house should cost, and house developers could use it to figure out what a house should cost. Prospective purchasers will benefit from this data since they will have a better understanding of the property. They will then be able to get better bargains from real estate agents as a result. Regression is the approach of choice since it is straightforward. A swimming pool, parking, and other amenities may be added to our project to make it more complete. These projects let prospective buyers get a sense of the pricing. Modifying and adding new features to this

project will make it better. A swimming pool, parking, and guest rooms may all be added to the property. The user must be able to log in to access the information. Create this project so that you may experiment with a variety of algorithms simultaneously. Finally, allowing users to use the app through their mobile devices. Finally, we want to make our system more efficient so that it may be used on Android smartphones.

REFERENCES

1. B. Trawinski, Z. Telec, J. Krasnoborski et al., "Comparison of expert algorithms with machine learning models for real estate appraisal," in Proceedings of the 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, Poland, July 2017.
2. V. Kontrimas and A. Verikas, "The mass appraisal of the real estate by computational intelligence," Applied Soft Computing, vol. 11, no. 1, pp. 443–448, 2011.
3. M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," Information Fusion, vol. 16, pp. 3–17, 2014.
4. M S. Borde, A. Rane, G. Shende, and S. Shetty, "Real estate investment advising using machine learning," International Research Journal of Engineering and Technology (IRJET), vol. 4, no. 3, p. 1821, 2017.
5. J. R. Barr, E. A. Ellis, A. Kassab, C. L. Redfearn, N. N. Srinivasan, and K. B. Voris, "Home price index: a machine learning methodology," International Journal of Semantic Computing, vol. 11, no. 1, pp. 111–133, 2017.
6. W. J. McCluskey, M. McCord, P. T. Davis, M. Haran, and D. McIlhatton, "Prediction accuracy in mass appraisal: a comparison of modern approaches," Journal of Property Research, vol. 30, no. 4, pp. 239–265, 2013.
7. S. Rosen, "Hedonic prices and implicit markets: product differentiation in pure competition," Journal of Political Economy, vol. 82, no. 1, pp. 34–55, 1974.
8. E. Lughofer, B. Trawiński, K. Trawiński, O. Kempa, and T. Lasota, "On employing fuzzy modeling algorithms for the valuation of residential premises," Information Sciences, vol. 181, no. 23, pp. 5123–5142, 2011.
9. H. Kusan, O. Aytekin, and I. Özdemir, "The use of fuzzy logic in predicting house selling price," Expert Systems with Applications, vol. 37, no. 3, pp. 1808–1813, 2010.
10. O. Bin, "A prediction comparison of housing sales prices by parametric versus semiparametric regressions," Journal of Housing Economics, vol. 13, no. 1, pp. 68–84, 2004.
11. Y. Kang, F. Zhang, W. Peng et al., "Understanding house price appreciation using multi-source big geo-data and machine learning," Land Use Policy, vol. 2020, Article ID 104919, 2020.
12. A. Din, M. Hoesli, and A. Bender, "Environmental variables and real estate prices," Urban Studies, vol. 38, no. 11, pp. 1989–2000, 2001.