iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** X **Month of publication:** October 2021

DOI: https://doi.org/10.22214/ijraset.2021.38715

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: 7.429
Volume 9 Issue X Oct 2021- Available at www.ijraset.com

House Price Prediction Using Linear Regression

Samkit Saraf[1], Aryan Verma[2], Shashwat Tare[3], Varun Vasani[4], Dr. Kamal Mehta[5]
[1]*Student,* [2] *Professor, Department of Computer Engineering, NMIMS Deemed to be University, MPSTME, Maharashtra, India*

Index Terms: Regression model, House price prediction, machine learning, housing market, Arima model, artificial neural network, support vector machine, random forest, dataset.

I. INTRODUCTION

The basic need of haven can be satisfied by housing. Housing will likewise be a piece of venture. As human population is increasing day by day so do the need of housing is expanding. There are numerous impacts while choosing the cost of house. For purchasers, venders, and financiers the exact value expectation is consistently voracity. Numerous analysts and researchers had been proposed their work for foreseeing the house cost precisely and as accurate as possible. There are many AI relapse calculations to utilize and numerous algorithms to apply. AI and ML techniques were utilized by a number of analysts for housing value forecast model. Which can be later used as per need to predict housing value to help buyer to make best choice and get the best out of their hard-earned money. For this review paper, we selected various papers on house prediction model and analyzed them. There are many factors considered when predicting the prices of houses like neighboring infrastructure and facilities like park, supermarket, hospital, schools, etc. These are the thing which is most considerable when buying a property. Not many people wants a place in remote area with nothing in surroundings. Builders also consider the factors and make amendments accordingly so as to keep the construction at an optimal level so that budget and quality is maintained.

II. MOTIVATION

Consider moving to a new city and not knowing anything about that area or the city. And the broker you hired is not that selfless, so he may tell you the price that is nowhere close to its actual value. Or even if you are constructing the house, the builder with no proper knowledge may go over your budget. But with proper house price prediction model he can make construction estimation plan with not much difficulty.

III. OBJECTIVE

1) To give correlation of house valuing to clients.
2) To foresee the house value as indicated by the territory.
3) To recommend manufacturer value expectation for the new developments.
4) To ascertain house cost contingent on general climate like railroad station, emergency clinic territory, ATM, school, banks so client can buy level with full offices.

This review paper consists of review of recent research papers for house price prediction, and they are compared on the basis of different criteria like method used, accuracy, efficiency, etc. And summarized to display an overview about the same and help you with direction if you want to implement or research on house price prediction.

IV. LITERATURE SURVEY

In this paper, Machine learning methods are applied to examine verifiable property exchanges in Australia to find valuable models for house purchasers and dealers. There is the high irregularity between house costs in the most costly and most moderate rural areas in the city of Melbourne is appeared in this paper. Additionally, tests show that the blend of Stepwise and Support Vector Machine that depends on mean squared blunder estimation is a serious methodology. This paper looks for helpful models for house value expectation. It likewise gives experiences into the Melbourne Housing Market. Right off the bat, the first information is arranged and changed into a cleaned dataset prepared for examination.

©IJRASET: All Rights are Reserved

Information decrease and change are then applied by utilizing Stepwise and PCA strategies. Various strategies are then actualized and assessed to accomplish an ideal arrangement. The assessment stage demonstrates that the mix of Stepwise and SVM model is a serious methodology. Accordingly, it could be utilized for additional sending [1]. To anticipate Singapore lodging market and to think about the prescient presentation of Artificial neural Network (ANN) model, i.e., the multilayer perceptron, with auto backward incorporated moving normal (ARIMA) model, this paper proposes two calculations. To anticipate the future apartment suite value file (CPI) the more prevalent model is utilized. The lower mean square mistake (MSE) of the ANN models indicated the prevalence of ANN over other prescient apparatuses. The estimates depended on time arrangement information of factors that are accepted to impact the townhouse costs in Singapore. These factors and the CPI were the sources of info and yield to the models, individually [2]. This paper presents how to utilize Linear Regression, Forest relapse, Boosted relapse. The proficiency of the calculation has been additionally expanded with utilization of Neural organizations. The framework will fulfil clients by giving precise yield and forestalling the danger of putting resources into an inappropriate house. Extra highlights for the client's advantage can likewise be added to the framework without upsetting its center usefulness [3].

With restricted dataset and information includes, a functional and composite information pre-processing, imaginative element designing strategy is inspected in this paper. To foresee singular house value half and half Lasso and Gradient boosting relapse model were utilized in this paper. Relapse calculations with boundary in excess of 10000 cycles are applied by thinking about the homoscedasticity confirmation. In any case, the outcome is controlled by the homoscedasticity between preparing information and test information. Linearity of each component is the measurement key of relapse calculation, consequently, numerous changes are applied to upgrade the linearity of info highlights [4].

This paper considers the issue of changing house cost as a characterization issue and applies AI procedures to foresee whether house costs will rise or fall. This work applies different element choice strategies, for example, fluctuation impact factor, Information esteem, guideline part examination and information change procedures, for example, anomaly and missing worth treatment just as box-cox change methods. Random Forest gives more exactness anyway simultaneously this specific sort of classifier additionally inclined to over fitting along these lines the presentation of Support Vector Machine classifier can have supposed to be solid and steady over the remainder of the two classifiers [5].

In this paper, the investigation of land esteem is felt basic to help the decisions in metropolitan masterminding. The land system is a shaky stochastic cycle. Monetary authority's decisions rely upon accessible examples to secure most outrageous returns. Fashioners are interested in knowing the future examples for their fundamental initiative. To absolutely check land costs what's more, future examples, immense proportion of data that effects show up cost is required for assessment, exhibiting, and deciding [6].

In this paper, there is conversation on which highlight extraction method is best by thinking about different kinds of highlights and highlight extraction strategies. In the event of character acknowledgment application, we will allude highlights and highlight extraction techniques in this paper. Utilizing this paper, a fast thought of highlight extraction procedures might be got, and it very well may be concluded what include extraction strategy will be better for the work to be done dependent on kind of picture [7].

This paper suggests that the housing market is not like any other market for goods or services. The housing market has the following characteristics A house fills in as a product and a speculation.

1) A house typically makes up the most noteworthy segment of a person's total assets.
2) A house has a significant expense of flexibly (Building and assessment costs).
3) A house is exceptionally solid and can keep going for a very long while.
4) Houses are heterogeneous.
5) Houses are enduring.
6) Houses can be utilized as insurance against credits.
7) A house can be sold for more cash second hand. This makes speculation openings.
8) Houses can produce rental salary.

This suggests the general housing market is an assortment of a few associated submarkets [8].

In many examinations, value displaying is utilized to clarify the variation in cost. The attribute of a property is utilized as explanatory factors. The principal hypothesis in this field is hedonic value hypothesis. This hypothesis can be followed back to the work complete by [9] in 1966 and [10] in 1974. It is the main hypothesis of reference to clarify cost in the housing market. In this hypothesis the cost of a decent, in this case a house, relies upon its attributes. Since this theory uses the fundamental attributes of a house, for example, number of room and size, various exploration has been distributed on its impediments.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue X Oct 2021- Available at www.ijraset.com

V. FLOWCHART

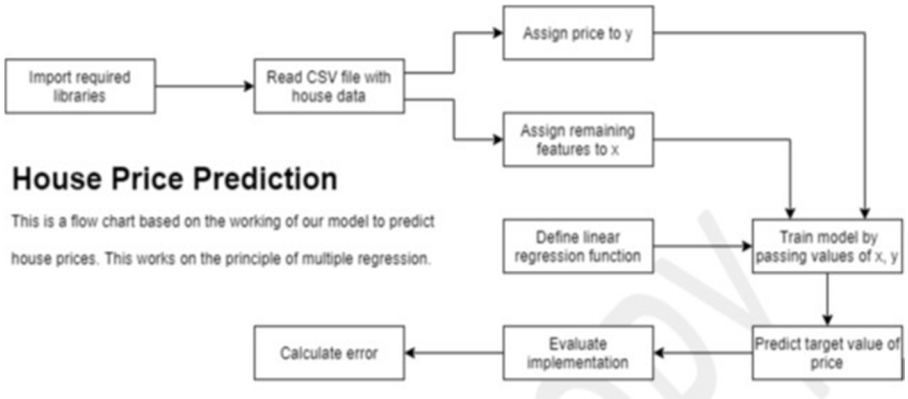


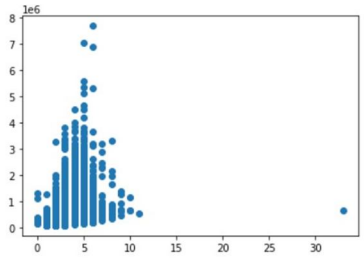
Figure 1. Flowchart of process of house price prediction using multiple linear regress

VI. DATASET

gra de	sqf t_a bov e	pri ce	be dr oo ms	bat hro om s	sqf t_li vin g	sqf t_l ot	flo ors	sqft_ base ment	yr_ bui lt	zip co de
7	118 0	22 19 00	3	1	11 80	56 50	1	0	19 55	98 17 8
7	217 0	53 80 00	3	2.2 5	25 70	72 42	2	400	19 51	98 12 5
6	770	18 00 00	2	1	77 0	10 00 0	1	0	19 33	98 02 8
7	105 0	60 40 00	4	3	19 60	50 00	1	910	19 65	98 13 6
8	168 0	51 00 00	3	2	16 80	80 80	1	0	19 87	98 07 4
11	389 0	1.2 3E +0 6	4	4.5	54 20	10 19 30	1	1530	20 01	98 05 3
7	171 5	25 75 00	3	2.2 5	17 15	68 19	2	0	19 95	98 00 3
7	106 0	29 18 50	3	1.5	10 60	97 11	1	0	19 63	98 19 8
7	105 0	22 95 00	3	1	17 80	74 70	1	730	19 60	98 14 6

©IJRASET: All Rights are Reserved

VII. IMPLEMENTATION




```
In [7]: #fitting model
        linear_regression.fit(x, y)

Out[7]: LinearRegression()

In [8]: #predicting the value
        y_predict = linear_regression.predict(x)
        y_predict

Out[8]: array([300696.03726121, 667168.11036987, 230032.70518315, ...,
               113462.21517047, 409422.44139761, 117604.65272381])
```

```
In [2]: #importing libraries
        import pandas as pd
        import numpy as np
        from sklearn.linear_model import LinearRegression
        import matplotlib.pyplot as plt

In [3]: #reading csv file
        df = pd.read_csv('C:/Users/arvee/Downloads/washington_state_housing_data.csv')
        df.head()

Out[3]:
```

	grade	sqft_above	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	sqft_basement	yr_built	zipcode
0	7	1180	221900.0	3	1.00	1180	5650	1.0	0	1955	98178
1	7	2170	538000.0	3	2.25	2570	7242	2.0	400	1951	98125
2	6	770	180000.0	2	1.00	770	10000	1.0	0	1933	98028
3	7	1050	604000.0	4	3.00	1960	5000	1.0	910	1965	98136
4	8	1680	510000.0	3	2.00	1680	8080	1.0	0	1987	98074

```
In [4]: y = df['price']
        x = df[['bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors', 'grade', 'sqft_above', 'sqft_basement', 'yr_built', 'zipco

In [5]: #defining linear regression model
        linear_regression = LinearRegression()

In [6]: #example for visualization of data
        plt.scatter(x['bedrooms'], y)

Out[6]: <matplotlib.collections.PathCollection at 0x1679f0a9dc0>
```

VIII. CONCLUSION

For simplicity, many people solve problem with linear regression with one feature and one target. But in reality, most problems solved with help of linear regression model require more than one feature. To get the best result there are many parameters to take into consideration.

For example, if we want to calculate eligibility of a person to apply for loan, apart from his income we also need to consider his credit score that is his CIBIL score as well.

So, we have implemented linear regression with multiple features for house price prediction that helps conservative people with their budgets and business strategies, who are looking to purchase a new home. The current method includes the estimation of house prices without predicting future market conditions and price changes as appropriate. The purpose of the implementation is to forecast real estate customers' efficient house pricing regarding their budgets and priorities. Future prices can be estimated by evaluating past industry dynamics and price ranges, as well as future innovations.

This activity includes a simple python program that recognizes the requirements of the customer and then incorporates the implementation of several data mining linear regression algorithms. Without approaching an attorney, this application can assist clients to invest in an estate. It also reduces the possibility of bad investment choices and transaction entails.

©IJRASET: All Rights are Reserved

We used various libraries to ease the work and get better understanding of what has been done. Some libraries used are –

1) *Pandas:* To read and manipulate data as per requirement
2) *Numpy:* To vectorize our data and prepare for use in efficient calculations
3) *Sklearn:* It provides many built-in functions that are required for linear regression. It simplifies code and reduces extra work.
4) *Matplotlib:* To visualize data

First the data is loaded into data frame using pandas library and separated the features and target. Then regression model is defined, and object created for the same. After that, we fit the model with help of sklearn library and train the model with the training set that was split from the main dataset. After constructing the model, we are ready to use that model to predict house prices.

It has become tough to store such items in today's real estate world. Giant data and remove it for one's own demand. Often, the data collected should be useful.

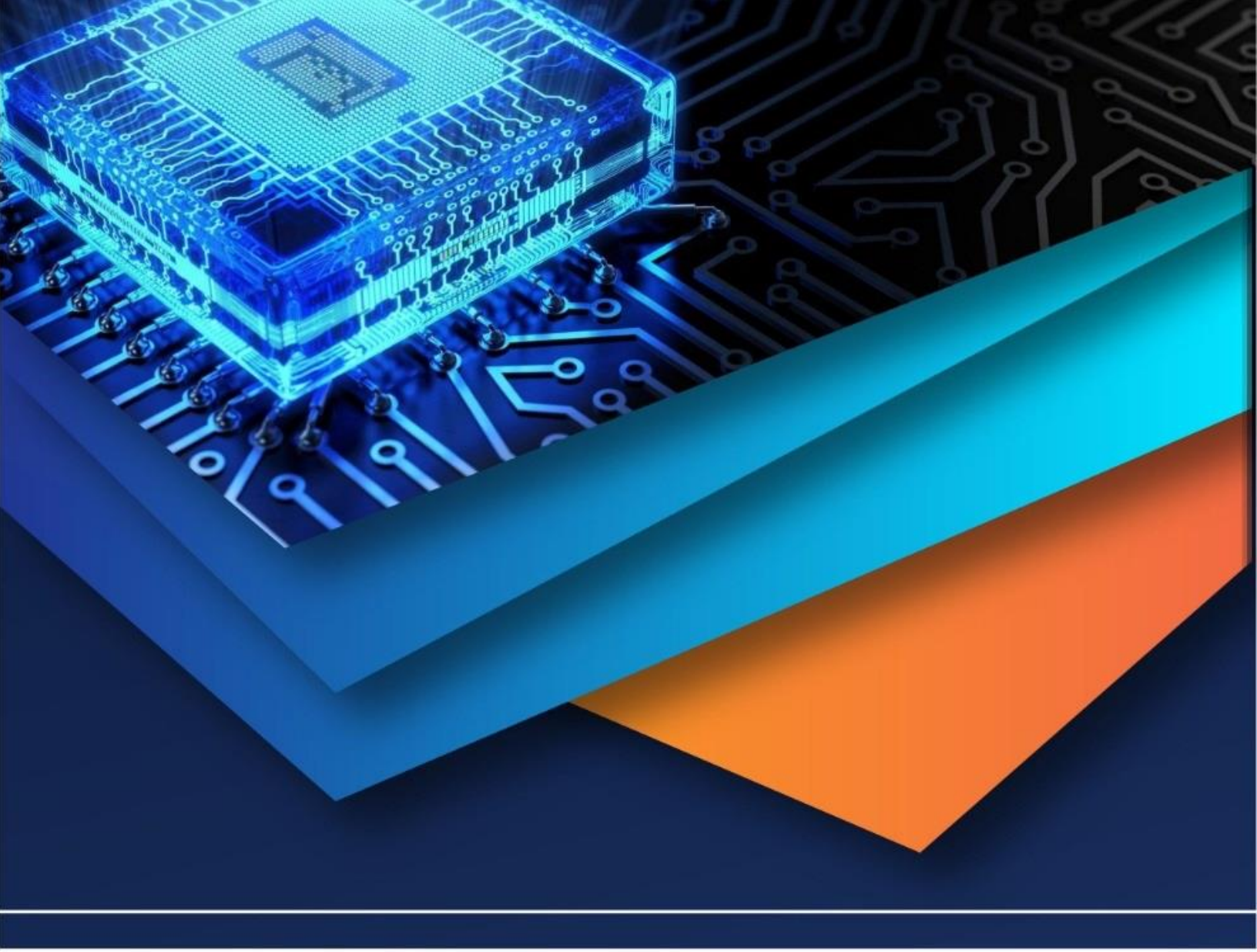
The process renders optimal application of the Algorithm for Linear Regression. Makes the device, in the most efficient way, use of such data. The Linear, the regression algorithm helps to please clients by boosting the reliability of the choice of property and reducing the probability of investing in a farm.

More variables, such as the recession, influence the in-house rates are applied. In-depth knowledge for every property to provide enough descriptions of a desired estate, they will be added. This will allow the system to operate on a larger scale.

REFERENCES AND FOOTNOTES

[1] The Danh Phan, “Housing Price Prediction using Machine Learning Algorithms: The Case of Melbourne City, Australia”, 2018.
[2] Wan Teng Lim, Lipo Wang, Yaoli Wang, and Qing Chang, “Housing Price Prediction Using Neural Networks”, 2016.
[3] Ayush Varma, Abhijit Sharma, Sagar Doshi, Rohini Niar, “House Price Prediction Using Machine Learning and Neural Networks”, 2018.
[4] Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick Siow Mong Goh Institute of High-Performance Computing (IHPC), Agency for Science Technology and Research, Singapore, “A Hybrid Regression Technique for House Prices Prediction”, 2018.
[5] Debanjan Banerjee, Suchibrota Dutta, “Predicting the Housing Price Direction using Machine Learning Techniques” 2018.
[6] Nehal N Ghosalkar, “Real Estate Value Prediction Using Linear Regression”, 2018.
[7] Gaurav Kumar, Pradip Kumar Bhatiya, Fourth International Conference on Advanced Computing & Communication Technologies, “A Detailed Review of Feature Extraction in Image Processing Systems”, 2014.
[8] M, Lacoviello, “House prices and Macroeconomy in Europe. Results from a structural VAR analysis”. Working Paper. No.018. European Central Bank, 2000.
[9] K. Lancaster, "A New Approach to Consumer Theory", Journal of Political Economy, vol. 74, no.2, pp. 132-157, 1966.
[10] S. Rosen, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", Journal of Political Economy, vol. 82, no. 1, pp. 34-55, 1974.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)