iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4622

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: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Housing Prices Prediction using Linear Regression

Mrs. Dhikhi T[1], Saurav Chowdhury[2], Akhil Nair[3], MVS Vashishta[4]
[1]*ASST. Professor,* [2, 3, 4]*Student, SRM Institute of Science and Technology, Chennai*

Abstract: Individuals hoping to purchase home will in general be increasingly traditionalist with their financial plans and market systems. The present structure incorporates figuring of house costs without the fundamental estimate about future market examples and cost increase. It predict the capable house esteeming for land customers concerning their money related plans and needs. By breaking down past market patterns and value ranges, and furthermore up and coming improvements future costs will be anticipated. The entire work will be done using linear regression algorithm of supervised learning. It includes a site which acknowledges client's determinations and after that joins the utilization of various direct relapse calculation of information mining. This application will assist clients with investing in a home without moving toward a specialist. It likewise diminishes the hazard engaged with the exchange.
Keywords: Data mining, Supervised Learning, Prediction.

I. INTRODUCTION

Hotel advertise is of exceptional basic for the economy works out. Hotel improvement and rebuild bolster the economy through a development in the complete utilizations, business and volume of house bargains. They in like manner re establish the enthusiasm for material organizations, for instance, nuclear family durables. The influencing of house costs impacts the estimation of favorable position portfolio for most nuclear families for whom a house is the greatest single asset . A precise figure on the house cost is basic to inevitable property holders, engineers, money related masters, appraisers, charge assessors and other land feature individuals, for instance, contract banks and wellbeing net suppliers . Customary house estimation desire relies upon cost and arrangement esteem examination lacking of a recognized standard and an affirmation strategy. Along these lines, the availability of a house estimation figure show empowers finish off a basic information to gap and improve the viability of the land grandstand . According to the outcomes demonstrate that the houses with more rooms and washrooms are esteemed higher.

A tolerably new house is more expensive than an old house and a house with a yard nursery is assessed higher than one without a greenery walled in area.. Ongoing investigations further legitimize the need of lodging value examination with an end that lodging part assumes a critical job in going about as a main marker of the genuine area of the economy and resources costs help estimate both swelling and yield . Numerous past examinations find exact proof supporting the huge interrelations between house cost and different financial factors, for example, salary, loan costs, development expenses and work showcase factors. Hotel advertise is illiquid and heterogeneous in both physical and topographical perspectives, which makes guaging house estimation a troublesome endeavor. Likewise, the unpretentious joint efforts between house cost and other macroeconomic rudiments make the desire additionally obfuscated.

II. RELATED WORKS

The adjustment in house costs can either mirror a national marvel, for example, the impact of fiscal approach, or be ascribed to nearby factors—conditions that particular to each geographic market [1]. It can either demonstrate the adjustments in the genuine segment factors, for example, work information and creation of products, or be influenced by the exercises in the ostensible part, i.e., money related market advancement [2]. Aside from the above close connect to lodging speculation, house costs have a solid connection with both pay and loan fees – both by means of a standard lodging request work and a lodging supply work. On the interest side, [3] propose a hypothetical model of house value assurance that is driven by changes in pay and loan fees. [4] examined show that in cutting edge economies genuine house costs have changed around an upward pattern at any rate since 1970 at euro advertise, for the most part credited in the writing to rising interest for lodging space – connected to expanding per capita pay just as a developing populace on the interest side. [5] connected fake neural system to assess the present market circumstance amid the world financial emergency in 2008 and anticipated the future execution of property so as to support speculators and other market players in settling on essential choices. Albeit fake neural system has been limitedly utilized for valuation or anticipating property value, ponders did to think about the exactness of direct relapse and fake neural system found that the last has prevalence looked at over the previous. [6] analyzed direct relapse and fake neural system in foreseeing lodging esteem. [7] utilized fake neural system and contrasted its precision and that of straight relapse in anticipating of lodging cost. [8] in their exploration likewise looked at

straight relapse and counterfeit neural system in the mass evaluation setting. [9] recreated a speculation in connection to esteeming land an incentive in Madrid. Determining has some level of vulnerability. Be that as it may, a high level of modernity has been grown as of late, with a scope of cutting edge quantitative and subjective methodology utilized by institutional financial specialists in property estimating, including judgemental systems, causal or econometric techniques, and time arrangement and pattern investigation strategies [10]. This examination expected to consider between MLR show and Neural Network model to anticipate the house costs in New York. Helper data from 1047 houses in New York is used in fake neural framework to anticipate the house cost and choose if the desire is extraordinary or not. The discretionary data was assembled in year 2012. The data involve house estimation, living region, number of rooms, number of bathrooms, package size and time of house [11]. The living zone, number of rooms, number of washrooms, parcel size and time of house will be in information layer while house cost will be in yield layer. There is a sum of 1047 information focuses in which 70% was utilized for preparing, 15% for approval and another 15% for testing. Every one of the 1047 test informational collections are separated for preparing, approval and testing. Utilizing Neural Network Toolbox (nntool) in MATLAB, diverse system arrangement with various number of concealed neurons is prepared and their execution is checked. There are 733 informational indexes are utilized for preparing, 157 informational indexes for approval and 157 informational indexes for testing.

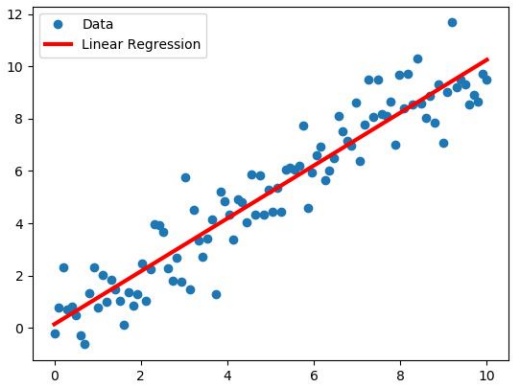
A. Support Vector Machine

Support Vector Machine" (SVM) is a directed AI calculation which can be utilized for both arrangement or relapse difficulties. In any case, it is generally utilized in order issues. In this calculation, we plot every datum thing as a point in n-dimensional space (where n is number of highlights you have) with the estimation of each element being the estimation of a specific facilitate. At that point, we perform characterization by finding the hyper-plane that separate the two classes great

III. PREDICTION

In insights, straight relapse is a direct way to deal with displaying the connection between a scalar reaction (or ward variable) and at least one illustrative factors (or autonomous factors). The instance of one logical variable is called basic straight relapse. For more than one illustrative variable, the procedure is called different straight relapse. This term is unmistakable from multivariate straight relapse, where different associated subordinate factors are anticipated, as opposed to a solitary scalar variable. In straight relapse, the rselationships are demonstrated utilizing direct indicator works whose obscure model parameters are assessed from the information. Such models are called Most usually, the contingent mean of the reaction given the estimations of the informative factors (or indicators) is thought to be an affline capacity of those qualities; less ordinarily, the restrictive middle or some other quantile is utilized. Like all types of relapse investigation, direct relapse centers around the restrictive likelihood didstribution of the reaction given the estimations of the indicators, as opposed to on the joint likelihood circulation of these factors, which is the area of multivariate examination.

In this project, we will apply basic machine learning concepts on data collected for housing prices in the Boston, Massachusetts area to predict the selling price of a new home. We will first explore the data to obtain important features and descriptive statistics about the dataset. Next, we will properly split the data into testing and training subsets, and determine a suitable performance metric for this problem. We will then analyze performance graphs for a learning algorithm with varying parameters and training set sizes. This will enable us to pick the optimal model that best generalizes for unseen data. Finally, we will test this optimal model on a new sample and compare the predicted selling price to our statistics.



Linear Regression

IV. ALGORITHM

The algorithm works as follows:

1) Read n //total number of points
2) Read x, y //x and y co-ordinates of points
3) Initialize diffx[n], diffy[n]
4) Initialize diffxy, diffx2 to 0
5) for i = 1 to n do calculate the mean of x : xm mean of y : ym diffx[i] = x[i] – xm //find the difference values between each x and mean of x diffy[i] = y[i] – ym //find the difference values between each y and mean of y diffx2 = Σ(diffx[i])2 //calculate the summation of all the difference values of x diffxy = Σ((diffx[i]) * (diffy[i])) //compute the product Of diff values of x and y end for 6. m = diffxy / diffx2 //the slope value is obtained by this Formula
6) c = ym – (m * xm) //the intercept value is obtained with this Formula 8. Equation complete: y = (m * x) + c
7) Stop. By substituting the value of x in the obtained equation the respective y value can be found.

A. Input

The modified Boston housing dataset consists of 489 data points, with each datapoint having 3 features. This dataset is an advanced view of the Boston Housing dataset found in UCI MACHINE LEARNING RESPRITORY.

B. Features

1) *RM:* normal number of rooms per abiding
2) *LSTAT:* level of populace considered lower status
3) *PTRATIO:* pupil-teacher ratio by town
4) *MEDV:* value of homes

The sample data can be stated as

RM	LSTAT	PTRATIO	MEDV
6.575	4.98	15.3	504000
6.421	9.14	17.8	453600
7.185	4.03	17.8	728700
6.998	2.94	18.7	701400
7.147	5.33	18.7	760200
6.43	5.21	18.7	602700
6.012	12.43	15.2	480900
6.172	19.15	15.2	569100
5.631	29.93	15.2	346500
6.004	17.1	15.2	396900
6.377	20.45	15.2	315000
6.009	13.27	15.2	396900
5.889	15.71	15.2	455700
5.949	8.26	21	428400
6.096	10.26	21	382200
5.834	8.47	21	417900
5.935	6.58	21	485100

We will apply multiple linear regression to predict the required values using the above mentioned dataset. We will use gradient descent algorithm to reduce the cost function.

The cost function equation is expressed as C(x)= FC + V(x), where C equals total production cost, FC is total fixed costs, V is variable cost and x is the number of units.

Gradient Descent Algorithm is an advancement calculation used to limit some capacity by iteratively moving toward steepest drop as characterized by the negative of the inclination. In AI, we use angle plunge to refresh the parameters of our model.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 \quad [1.0]$$

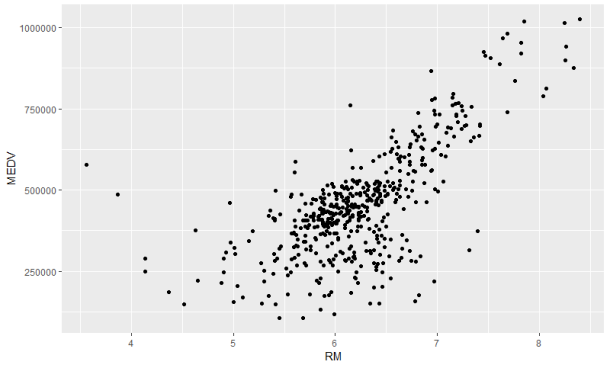
$$\frac{d}{d\theta_0}J(\theta_0,\theta_1) = \frac{d}{d\theta_0}\left(\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2\right) \quad [1.1]$$

$$= \frac{1}{m}\sum_{i=1}^{m}\frac{d}{d\theta_0}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 \quad [1.2]$$

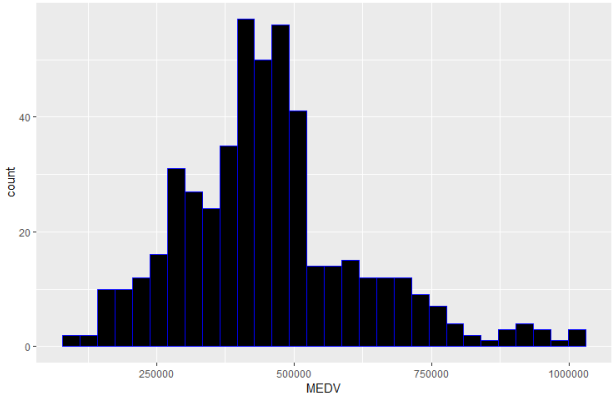
$$= \frac{1}{m}\sum_{i=1}^{m}2\left(h_\theta(x^{(i)}) - y^{(i)}\right)\frac{d}{d\theta_0}\left(h_\theta(x^{(i)}) - y^{(i)}\right) \quad [1.3]$$

$$= \frac{2}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right) \quad [1.4]$$

where h = hypothesies
J()= Cost Function
x= input variables
y= output variables
m= no. of samples



Attribute RM vs MEDV



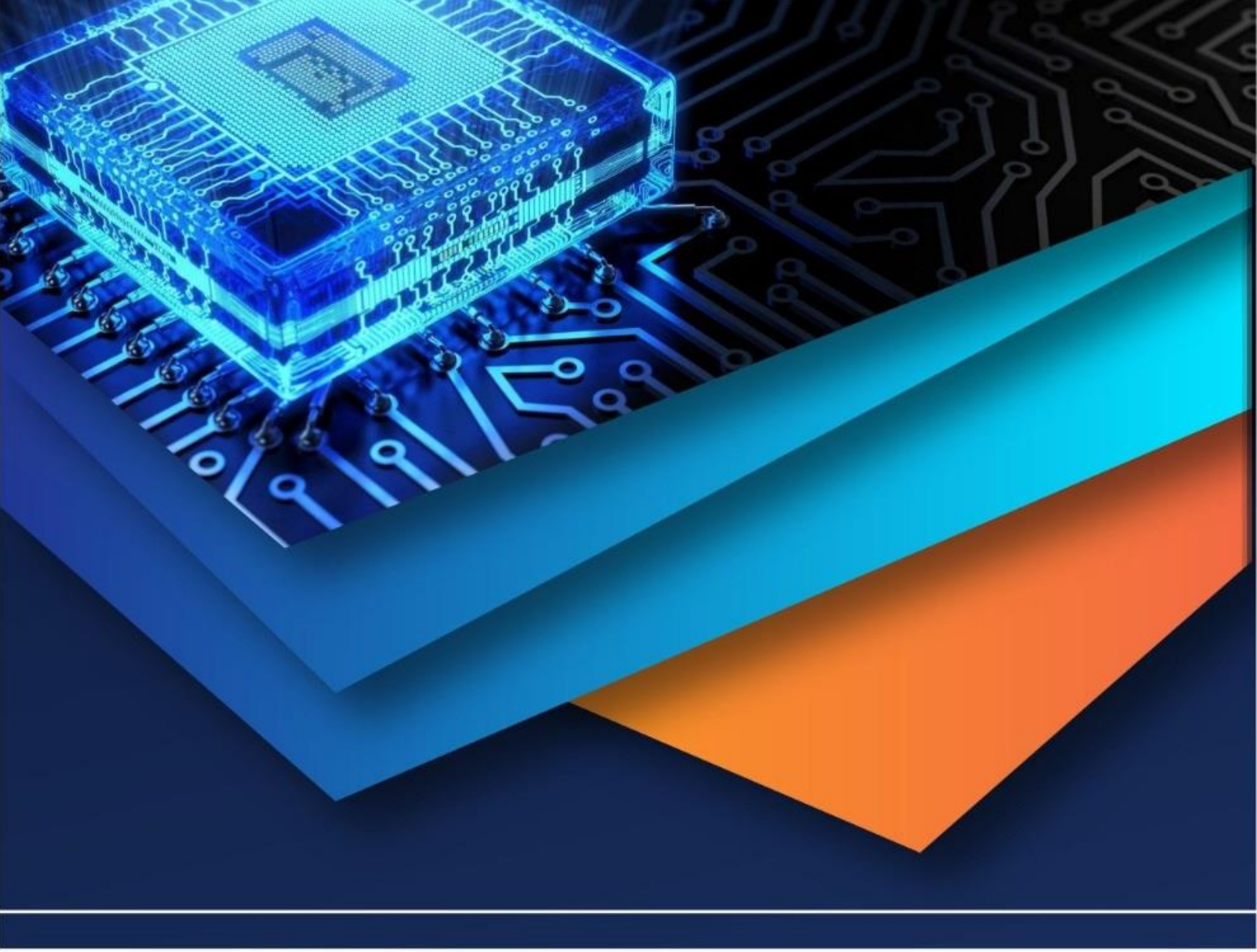
Count of MEDIAN VALUES of Houses.

V. CONCLUSION

In the present land world, it has turned out to be hard to store suchimmense information and concentrate them for one's very own necessity. Too, the removed information ought to be helpful. The framework makes ideal utilization of the Linear Regression Algorithm. The framework makes utilization of such information in the most effective way. The direct relapse calculation satisfies clients by expanding the precision of bequest decision and lessening the danger of putting resources into a bequest. A great deal's of highlights that could be included to make the framework all the more broadly adequate. One of the major future degrees is including domain database of more urban areas which will give the client to investigate more bequests and come to an exact choice. More factors like subsidence that influence the house costs will be included. Inside and out subtleties of each property will be added to give sufficient subtleties of an ideal domain. This will assist the framework with running on a bigger dimension.

REFERENCES

[1] L. Gattini, & P. Hiebert, ―Forecasting and assessing euro area house prices through the lens of key fundamentals‖, Working Paper Series, No.124/ October, 2010, 2010.
[2] R. Gupta, S.M. Miller, & D.V. Wyk, ―Financial market liberalization, monetary policy, and housing price dynamics‖. Working paper No. 201009, Dept. of Econ., University of Pretonia. 2010.
[3] K. McQuinn, & G. OReilly, ―A model of cross-country house prices, Central Bank and Financial Services Authority of Ireland‖, Research Technical Paper 5 (July), 2007.
[4] N. Girouard, , M. Kennedy, P. van den Noord, & C. Andr´e, ―Recent house price developments: The role of fundamentals‖, OECD Economics Department Working Paper No. 475, 2006.
[5] A. Khalafallah, ―Neural network based model for predicting housing market performance‖. Tsinghua Science and Technology, 13(1): 325-328, 2008.
[6] N. Nguyen & Al Cripps, ―Predicting housing value: a comparison of multiple regression analysis and artificial neural networks‖, Journal of Real Estate Research, Vol . 22, No.3. 2001.
[7] S. T. Ng, , & M. Skitmoreb, ―Using genetic algorithms and linear regression analysis for private housing demand forecast‖. Building and Environment, 43: 1171-1184, 2008.
[8] J. Gallego, & Mora-Esperanza (2004). ―Artificial intelligence applied to real estate valuation: An example for the appraisal of Madrid‖. Catastro: 255-265.
[9] M. S. Mohd Radzi, , C. Muthuveerappan, N. Kamarudin, & I. S. Mohammad, ―Forecasting house price index using artificial neural network‖, International Journal of Real Estate Studies, Volume 7, Number 1, 2012.
[10] Creative Research Systems. (n.d.). Retrieved December 22, 2013, from Survey System: http://www.surveysystem.com/correlation.htm., 2013.
[11] Mymat. ―Probabilities and Statistics : Guidance to solve House Prices Data Set Statistical Analysis‖. Retrieved May 28, 2014, from Math10.com., 2013.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY