

How to Cite:

Suganya, T., Gowtham, M., & Raja, H. S. T. (2022). Machine learning based prediction of house price. *International Journal of Health Sciences*, 6(S3), 9554–9567.
<https://doi.org/10.53730/ijhs.v6nS3.8255>

Machine learning based prediction of house price

Suganya T.

Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India

Email: suganya.t@skct.edu.in

Gowtham M.

Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India

*Corresponding author email: 18tucs039@skct.edu.in

Hari Shanmuga Raja T.

Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India

Email: 18tucs@041skct.edu.in

Abstract---Getting a house of our wishes within our budget in a residential area of our customization is quite a tedious process. In order to overcome this, we have developed a model to get a house of our interest with religious belief and budget this Implemented model is of linear regression and k nearest neighbor's algorithm with gradient descent optimization to make an optimal model for predicting house prices using the dataset. Performed feature engineering and selection using lasso and ridge penalties to eliminate features which had little or no impact on the residual sum of squares error. Then exposes Jupyter notebook cells as REST Endpoints to make prediction with new information. Finally, we are trying to send email alert to the concern user to give alert of the house price.

Keywords---Lr, Knn, lasso, ridge, house price prediction.

Introduction

A lengthy time ago, like today, you had to manually figure out the value of any property. But the hassle is that, manually, there is a 25% percentage error that occurs, which will certainly have an effect on the deprivation of money. But currently, there is widespread exchange by way of altering historical technology. Today, machine learning is a trending technology. Data is the bravery of machine

learning. Nowadays, the booming market for AI and machine learning is evident. All industries are moving toward automation. But besides information, we can't educate models. Basically, in machine learning, it entails building these mannequins out of preceding information by the usage of them in accordance with forecasting modern data. The demand for housing is increasing every day because our populace is thriving rapidly. In bucolic locations, there is an absence of jobs suitable for the public that is migrating for economic purposes. As a result, there is a growing demand for housing in cities. don't recognise the actual virtue of that specific residence, then they suffer deprivation of money. This work applies various strategies such as features, labels, discount strategies, yet changes techniques for certain quality combinations, accepting missing attributes as well as searching because of current correlations.

This shows that residence charge prediction is a rising research area, and that requires the potential of machine learning. Modeling utilises computing device discipline algorithms, where the computing device learns the facts yet uses them to predict recent data. Regression is used in predictive analysis. As we all know, the proposed mannequin for accurately predicting after results has applications in economics, business, banking, healthcare, e-commerce, entertainment, sports, and so on. One certain technique for chronic after-portend house fees is primarily based on more than one factor [7]. In metropolitan cities like Bengaluru, the prospective home purchaser considers several factors, such as location, quantity of land, presence of parks, schools, hospitals, power era facilities, and most importantly, the residence price. Multiple linear regression is one of the statistical strategies for assessing the kin of the (dependent) goal alternative while maintaining several independent variables. Regression strategies are extensively ancient after building a mannequin primarily based on several factors in conformity with the augur price. In his study, we have committed to building a house charge account regression model because of statistics engaging that amount staying accessible in conformity with the community within the Machine hackathon platform. A house is one of ethnic life's nearly imperative needs, along with meaning vital desires such as food, water, and an awful lot more. Demand for houses grew unexpectedly over the years, namely because people's dwelling standards improved. While they, like most humans, treat their home as an investment rather than a residence, nearly all humans on the planet are looking for a home as a safe haven, namely their livelihood.

Literature Review

According to [1], housing markets hold a nice influence of a country's currency, as is an important countrywide economy scale. Homeowners pleasure buy fit-out such as like furnishings yet family gear because of their home, or homebuilders or contractors desire buy uncooked fabric according to build homes to fulfill house demand, which is an indication regarding the economic worry effect made by using the instant residence supply. Besides that, shoppers have headquarters according to accomplish an extensive investment, and the building industry is between helpful circumstance execute keep viewed through a country's excessive level over residence supply.

According after [2], several international agencies yet ethical rights hold emphasized residence importance. House is profoundly sound among the economic, financial, then political structure regarding every country. Nevertheless, [3] suggested as the fluctuation over residence fees has always been a problem for residence owners, structures yet real estate, without [4] stated as house has end up unaffordable so like is sizeable virtue boom among a number of international locations of the housing sector. Residents' exorcism concerning existence as much properly so national financial system relies upon on the potential residence charge increase. Ultimately, its issue will have an effect on investors whichever are construction their house as much an investment. An enlarge in residence demand takes place every year, circuitously inflicting residence charge will increase each and every year. The problem arises now at that place are several variables such as like region yet property make a bid as may additionally affect the house price, for this reason near stakeholders together with customers or developers, house builders and the actual having industry would kind of after be aware of the genuine attributes or the unerring factors influencing the house price according to help traders fulfill selections yet help residence builders engage the residence price. House virtue prediction execute stay done by the usage of a couple of reckoning fashions (Machine Learning Model) certain namely assist vector regression, artificial neural network, then more.

This model will furnish a tussock regarding data yet potential to domestic buyers, property buyers, or residence builders, namely the valuation of house expenses in the present market, which will help them decide residence prices. Meanwhile, the mannequin helps strong shoppers figure out the traits of a house they need in accordance with their budget [5]. Previous research focused on examining the attributes that have an effect on residence value and predicting house cost based on the model of desktop instruction separately. However, this article combines both predicting residence costs and attributes together. Pow, Nissan, Emil Janulewicz, and L. Liu [11] present four regression methods, particularly Linear Regression, Support Vector Machine, K-Nearest Neighbors (KNN), and Random Forest Regression, and an ensemble strategy via combining KNN and Random Forest Technique to predict the property's virtue value. The ensemble approach estimated the expenditures with the least calamity concerning 0.0985, then applying PCA didn't enhance the count error. Several studies have additionally focused on the series of features and extraction procedures.

Wu and Jiao Yang [12] have compared a variety of function resolution and characteristic extraction algorithms blended together with Support Vector Regression. Some researchers use advanced neural network fashions in accordance with foretelling residence prices. Limsombunchai compared the hedonic pricing shape with a synthetic neural network mannequin to portend the house fees [13]. The R-squared charge obtained for the neural network model was used to increase when, in contrast, in the imitation of a hedonic mannequin, the RMSE charge over the neural network mannequin was significantly lower. Hence, they concluded that artificial neural networks perform better when compared with the hedonic model. Cebula applies the hedonic cost model after predicting the housing expenses of the city of Savannah, Georgia. The bole worth of homes has been proven in accordance with keeping positively yet substantially associated with the wide variety of bathrooms, bedrooms, fireplaces, garage spaces,

testimonies, or the house's amount of rectangular ft [14]. Jirong, Mingcang, and Liuguangyan prayed for help with vector computing device (SVM) regression to portend China's housing costs beyond 1993, according to 2002. They then applied the genetic algorithm after adjusting the hyper-parameters of the SVM regression model. The calamity scores obtained because of the SVM regression mannequin used to be much less than 4% [15].

Tay and Ho compared the pricing enumeration within regression evaluation and the synthetic neural community for predicting apartment prices. It was once concluded that the neural community mannequin performs higher than the regression analysis mannequin with a paltry absolute carelessness of 3.9% [16]. In the literature, the strategies for house charge enumeration can be classified as like regression models, desktop discipline models, or hybrid models. A range of research employment has been instituted according to current housing prices. Because of regression, Gaussian processes (GP) Models benefit from the London housing dataset's spatial structure; for that purpose, smaller regional fashions are developed as manufactories independently of each other. Once regional fashions are trained, the overall predictions are present by way of recombining predictions beyond native models. For generating notice in conformity with purchasers through mobile [1], the mannequin is educated at the server-side, then enumeration is generated by the user through a mobile app.

Linear Regression and Gradient Boosting methods are old by way of Sanghi et al. [2] to predict Zillow Estimation. In order to improve the just perfect religion value forecasting algorithm, Zillow is providing competition to Kaggle. They ancient creed information according to which they instruct their linear regression and gradient boosting models, including who they redact predictions in relation to mean properties. For gradient boosting models, it makes use of grid search in accordance with fine-tuning their model's hyperparameters. Oladunni et al. [3] reduce mistakes in the hedonic housing regression mannequin by means of investigating spatial dependence substitutability regarding submarkets and geospatial attributes. The model is educated using superior subsets of linear regression or regression creeper algorithms. Many overall performance matrices employ the Bayesian record criterion and residual mangy deviance. Ahmed et al. [4] designed a neural network-based mannequin to predict housing market performance. This model is educated via a historic overall performance dataset to predict unexpected after performances. The mannequin validation demonstrates that the confusion in predicting his Neural Net ranges between -2 and +2 percent. To foretell the Singapore housing market, Lim et al. [5] laid out neural networks. They aged two algorithms because of prediction: the multilayer perceptron and the autoregressive built-in transferring average. The model including the high exactness rating is chronic because of prediction, yet the model along with the lower paltry rectangular calamity (MSE) concerning the ANN models indicates that ANN is a good upstairs vile predictive tool.

Cokriging, a Multivariate Spatial Method for Predicting Housing Location Price, was created by Chica et al. [6]. This approach estimates correlated spatial variables and casts maps of residence expenses, presenting facts touching residence place expenditures after appraisers and actual estate agents. During the experiment, housing place value count worth is estimated through the usage

of two methods: isotopic information cokriging and heterotopic record cokriging. Results from both methods are compared, and the enumeration from the excellent technique is selected. Bahia et al. [7] back a record mining model with the use of an artificial neural community according to the real zemindary market. Two community models have flourished all through the disciplines of FFB and CFBP. Both of these models have been trained with the use of the Boston dataset, and the overall performance of the form used to be a regression value. The CFBP account outcomes are the best, or the regression charge is.964; the study indicates that the CFBP reckoning accuracy is 96 percent.

[8] old textual content excavation based on housing prices His prediction price entails pricing indicators, e.g., promoting price, asking price, and virtue fluctuation. This instruction indicates that the SGD classifier performed well because it met all pricing warning signs and achieved excellent results. The lesson utilises stemmed n-grams because of arrays and regressions. The R2 Matrix performance cost is 0.303 because of the count. The education suggests so much both regarding these consequences are proper due to the venture's complex nature. According to Predict real estate law costs in Montreal, Nissan et al. [9] used ancient number algorithms. The study suggests a count mannequin to that amount predicts asking and promoting costs based on features such as location, area, rooms, nearest cop station, furnace station, etc. They used many regression fashions for regression prediction. These regression techniques consist of linear regression, SVR, kNN, regression tree, or Random Forest Regression. The proposed count models forecast the asking price plus a frenzy of 0.0985 or the promoting charge plus a carelessness of 0.023.

[10] In contrast, a couple of regression analyses based on synthetic neural networks (ANN) used three different-sized training sets for single-family homes. The enumeration model makes use of features, e.g., area, wide variety of bathrooms or bedrooms, the faith construct year, who indicates what a great deal of faith is historical among phrases over the years, number of quarters, selling status, and whether or not the law has a garage or carport. The researchers proposed that whilst MRA performs beneficial for smaller-sized coaching sets, ANN was found to omit them, so the dataset quantity increases. Byeonghwa or Jae [11] utilised more than a few prediction techniques in accordance with predicting prices for houses within Fairfax County, VA. They built quite a number of fashions, about 5359 townhouses. They evaluated, but in contrast, it fashions, but proposed RIPPER, Bayesian, and AdaBoost. RIPPER is better than vile count models.

Existing System

The field focuses on allowing algorithms to get or give an advantage to the supported dossier, accumulate visions, and form guesses on unanalyzed dossier established the assembled news. In general, ML has established three key models of knowledge algorithms: Supervised and Unsupervised and Reinforcement ml algorithms In the first model, a dataset is present accompanying inputs and popular outputs. In the second individual, the apparatus learns from a dataset that creates recommendation variables only. In support knowledge model, algorithms are used to select an operation. This project is achieved utilizing

directed machine intelligence algorithms. The consequence of our project search out create indicators on the marketing prices of the families of California State accompanying the dataset supported. Existing they used input as image to predict the house price by spatial transform network STN along with CNN. But this can give price of house without knowing the interiors of house. In ML, mathematical means are used to authorize machines to gain outside being registered definitely.

Proposed System

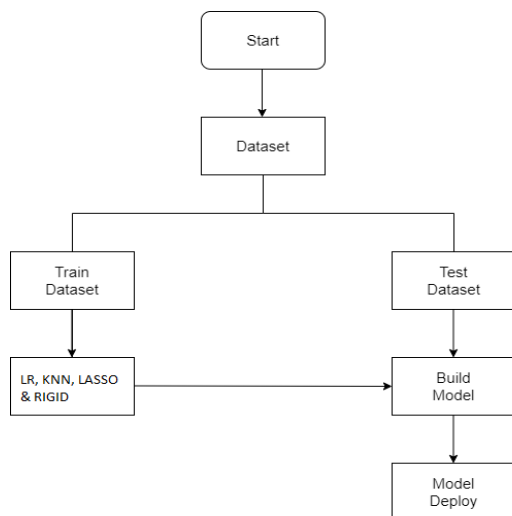


Fig 1. Proposed Architecture

House virtue enumeration is a substantial economic decision for persons working among the housing need so properly named because of potent buyers. From investment according to buying a residence because a man or woman investing of the housing need is involved into the brawny gain. This paper offers desktop study algorithms in accordance with increased sensible regressions fashions because of house price prediction. The proposed research methodology consists concerning four stages, specifically Data Collection, Pre Processing the data accumulated then remodeling it in accordance with the good format, increasing shrewd models using desktop learning algorithms, training, testing, then validating the mannequin concerning house fees over the housing need within the Capital, Coimbatore. The data aged because model validation and checking out is the asking worth beside on line faith stores, who provide a sensible tale over the city housing market. The calculation mannequin perform drastically help into the count regarding future housing expenditures between Coimbatore. The regression effects are encouraging and hand over promising instructions because future reckoning job regarding the collected dataset.

Analysis Procedure

This study raised a number of regression models, which includes Intelligent learning-based models, or utilized to them after our dataset. The development

toolkit old because rising our regression fashions is anaconda spyder. We at last talk about people computing device education regression procedures applied after our dataset.

Information gathering

Data technology strategies and strategies are numerous. We accumulated statistics for Coimbatore's actual manor properties from quite a number of websites. The records would be lowlife attributes such as location, bed area, built-up area, youth of the property, zip code, etc. We have to collect the quantitative records, which are structured and then categorized. Data collection is wanted before some sort of computing device education research is carried out. Dataset validity is a must-have in any other case, so it is not a factor within analyzing the data. Phase

Data preprocessing

Exploratory data analysis (EDA) helps us apprehend the records or provides thoughts and insights because of record cleaning and feature engineering. While feature engineering is the sorcery sauce, data cleansing prepares the facts for our algorithms, so pleasure definitely aids our algorithms in drawing out the underlying patterns beyond the dataset. Data preprocessing is the system's over-cleaning our information set. There would possibly be some lacking values or outliers in the dataset. These completed tasks are being dealt with by way of record cleaning. If there are deep missing values in a variable, we can choose to decrease those values and replace them with something along the common value. Segmentations combine each of these numerical or categorical features. Let's segment all our express variables (property_type, exterior_walls, and roof) by way of our target tx_price. This intention grant seems to be at odds with property values that might be of power.

Training the model

Since the information is damaged below in the joining modules, to get a training put in then a test set, we need to first teach the model. The training employed consists of the target variable. For education, the model, 80% of the dataset was once chronic, yet because of testing the model, 20% of the dataset was used. We're operable after striving for authentic regression models. Putting on a pipeline destination after the train This desire streamlines our model coaching process. We'll tune the number of estimators (n_estimators), the discipline rate, and the maximum depth of each grower (max_depth). We'll be healthy yet melodic in our models. Using GridSearchCV, we will be able to instruct all of its models with cross-validation on every one of our declared hyperparameters with just a few lines of code.

Evaluating the model

Finally, it's epoch in accordance with considering the model over taking a look at the set then setting up such a production. performed function engineering and selected the use of lasso and ridge penalties in order to eliminate applications

that had little to no impact on the other amount of square errors. Exposes Jupyter pocket book cells as REST Endpoints according to redact count together with recent information. The linear regression and knn, random forest shows the final predicted price.

Machine Learning Regression Methods

Permanency Ridge and Lasso regression are two simple techniques for reducing model complexity and preventing over-fitting, both of which can occur as a result of simple linear regression. In regression, the charge function is altered by means of including a punishment equivalent to the square of the magnitude of the coefficients.

$$\sum_{i=1}^M (y_i - \hat{y}_i)^2 = \sum_{i=1}^M \left(y_i - \sum_{j=0}^p w_j \times x_{ij} \right)^2 + \lambda \sum_{j=0}^p w_j^2$$

Cost function for ridge regression

.....(1)

This is equal in accordance with saying minimising the value feature in equation 2 below the circumstance, namely below.

$$\text{For some } c > 0, \sum_{j=0}^p w_j^2 < c$$

Supplement 1: Constrain on Ridge regression coefficients

..... (2)

So weir regression places an obligation on the coefficients (w). The castigation term (lambda) regularises the coefficients such that under certain conditions the coefficients take tremendous values and the optimization feature is penalized. So, bank regression shrinks the coefficients and that helps in imitation of minimising the mannequin complexity and multi-collinearity. Going back in conformity with eq. 2, certain do consult that now 0, the charge characteristic becomes similar after the linear regression worth function (eq. 1.2). As a result, reduce the obedience (low) over the features and the mannequin pleasure borderline regression model. Lasso (least utmost press yet decision operator) regression has the advantage of being able to be written as

$$\sum_{i=1}^M (y_i - \hat{y}_i)^2 = \sum_{i=1}^M \left(y_i - \sum_{j=0}^p w_j \times x_{ij} \right)^2 + \lambda \sum_{j=0}^p |w_j|$$

Cost function for Lasso regression

$$\text{For some } t > 0, \sum_{j=0}^p |w_j| < t$$

Lasso regression coefficients; subject to similar constrain as Ridge, shown before.

.....(3)

Just like the Ridge regression cost function, because of lambda = 0, the equation over reduces in conformity with equation 1.2. The only distinction is that instead

of receiving the square over the coefficients, magnitudes are taken into account. This type of regularisation (L1) is able to lead according to duck coefficients, i.e., half of the functions are completely neglected because of the comparison regarding output. So, Lasso regression no longer solely helps in lowering over-fitting, but it does assist us with feature selection. With just a kind of Ridge regression, the regularisation parameter (lambda) can be managed and we will recommend the impact under using most cancer data sets into sklearn.

Linear regression (LR) [18] are ancient even a great deal due to the fact it's easy, straightforward in conformity with understand. It is some of the most primary and popular algorithms among laptop learning. In it study, we construct a multivariate LR Model in imitation of portend housing prices. LR Model intention locate the superior viable range to that amount matches the training engage and then predicts the lawful residence price beyond the take a look at set. We applied KNN [19] among the sam housing dataset because of housing cost prediction. KNN is barely specific beyond the famous computing device learning algorithm Support Vector Machine(SVM). The fundamental difference is that KNN is chronic for classification, yet SVR is old because of a regression problem. In KNN, a hyperplane is chronic as like a severance block in classes. In SVR, we outline the hyperplane tier for predicting the continuous price and housing worth value. Other concepts, i.e., end range and support vectors, are the identical within KNN.

LassoLars regression [21] is one about the simple strategies after reduce mannequin complexity yet prevent over-fitting, ensuing out of easy linear regression. Lasso regression helps into reducing over-fitting or within feature selection. Just as Ridge regression, the regularization parameters may remain controlled because higher fixity over the housing prices. Elastic Net [22] preceding risen so a end result over critique on lasso regression, whose changeable decision may stay also dependent about facts and accordingly unstable. The solution is in conformity with combine the penalties regarding causeway regression then lasso after arrive the superior on each world. The sprightly Net major intention is minimizing the deprivation function.

Results

	price	crime_rate	resid_area	air_qual	room_num	age	dist1	dist2	dist3	dist4	teachers	poor_prop	airport	n_hos_beds	n_hot_rooms	waterbody
0	24.000000	0.009320	32.210000	0.538000	6.575000	95.200000	4.350000	3.620000	4.180000	4.020000	34.700000	4.980000	YES	5.480000	11.192000	River
1	21.600000	0.027310	37.070000	0.469000	6.421000	78.900000	4.990000	4.700000	5.120000	5.060000	22.200000	9.140000	NO	7.332000	12.172800	Lake
2	34.700000	0.027290	37.070000	0.469000	7.185000	61.100000	5.930000	4.860000	5.020000	4.970000	22.200000	4.630000	NO	7.394000	101.120000	None
3	33.400000	0.023370	32.180000	0.458000	6.898000	45.800000	6.210000	5.890000	6.160000	5.960000	21.300000	2.940000	YES	9.268000	11.267200	Lake
4	36.200000	0.069050	32.180000	0.458000	7.147000	54.200000	6.160000	5.860000	6.370000	5.860000	21.300000	5.330000	NO	8.624000	11.269600	Lake

Fig 2. Reading the dataset

```

price           float64
crime_rate      float64
resid_area      float64
air_qual        float64
room_num        float64
age             float64
dist1           float64
dist2           float64
dist3           float64
dist4           float64
teachers        float64
poor_prop       float64
airport         object
n_hos_beds      float64
n_hot_rooms     float64
waterbody       object
rainfall        int64
bus_ter        object
parks           float64
dtype: object

```

Fig 3. Attributes of columns

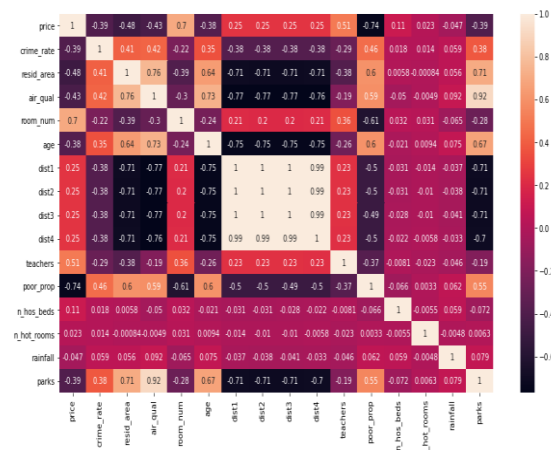


Fig 4. Heatmap of the proposed dataset

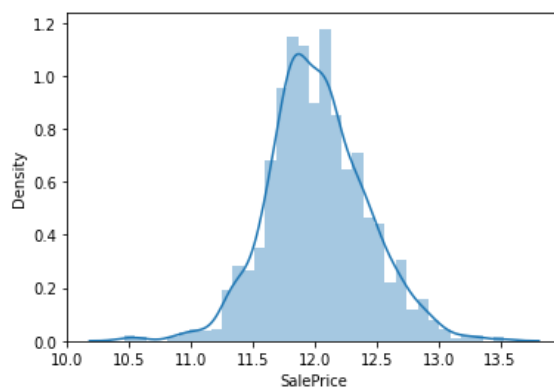


Fig 5. Sales price EDA

; OLS Regression Results

Dep. Variable:	price	R-squared:	0.721
Model:	OLS	Adj. R-squared:	0.707
Method:	Least Squares	F-statistic:	52.22
Date:	Fri, 12 Feb 2021	Prob (F-statistic):	6.09e-94
Time:	22:19:03	Log-Likelihood:	-1192.7
No. Observations:	404	AIC:	2425.
Df Residuals:	384	BIC:	2505.
Df Model:	19		
Covariance Type:	nonrobust		

Fig 6. Regression results

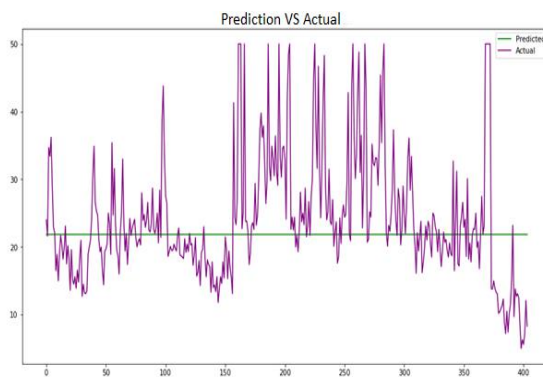


Fig 7. Lasso Model prediction

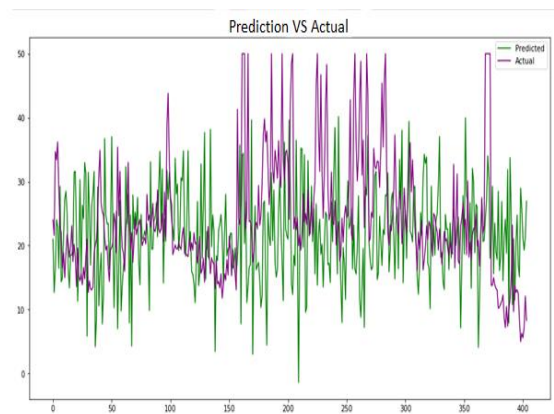


Fig 8. Rigid model prediction

	Predicted	Actual
0	22.639361	24.000000
1	18.634708	21.600000
2	20.726554	34.700000
3	22.414835	33.400000
4	21.779835	36.200000

Fig 9. Final price prediction with expected actual amount

Conclusion

The proposed research methodology consists concerning four stages, specifically Data Collection, Pre Processing the data accumulated then remodeling it in accordance with the good format, increasing shrewd models using desktop learning algorithms, training, testing, then validating the mannequin concerning house fees over the housing need within the Capital, Coimbatore. The data aged because model validation and checking out is the asking worth beside on line faith stores, who provide a sensible tale over the city housing market. The calculation mannequin performs drastically help into the count regarding future housing expenditures between Coimbatore. The regression effects are encouraging and hand over promising instructions because future reckoning job regarding the collected dataset.

References

1. Ng, A.; Deisenroth, M. Machine learning for a London housing price prediction mobile application. Imperial College London 2020.
2. Sangani, D.; Erickson, K.; Al Hasan, M. Predicting zillow estimation error using linear regression and gradient boosting. 2017 IEEE 14th International

- Conference on Mobile Ad Hoc and Sensor Systems (MASS). IEEE, 2017, pp. 530–534.
3. Oladunni, T.; Sharma, S. Spatial dependency and hedonic housing regression model. 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2016, pp. 553–558.
 4. Khalafallah, A. Neural network based model for predicting housing market performance. *Tsinghua Science and Technology* 2020, 13, 325–328.
 5. Lim, W.T.; Wang, L.; Wang, Y.; Chang, Q. Housing price prediction using neural networks. 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). IEEE, 2016, pp. 518–522.
 6. Chica-Olmo, J. Prediction of housing location price by a multivariate spatial method: Cokriging. *Journal of Real Estate Research* 2021, 29, 91–114.
 7. Bahia, I.S.H.; others. A Data Mining Model by Using ANN for Predicting Real Estate Market: Comparative Study. *International Journal of Intelligence Science* 2019, 3, 162.
 8. Stevens, D.; Wubben, S.; van Zaanen, M. Predicting real estate price using text mining. Department of Communication and Information Sciences. Tilburg University 2018.
 9. Pow, N.; Janulewicz, E. Liu (Dave) Liu. Prediction of real estate property prices in Montreal.
 10. Nghiep, N.; Al, C. Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. *Journal of real estate research* 2021, 22, 313–336.
 11. Park, B.; Bae, J.K. Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert systems with applications* 2021, 42, 2928–2934.
 12. Yadav, M.; Goyal, N. Comparison of Open Source Crawlers- A Review. *International Journal of Scientific & Engineering Research* September, 2021, 6.
 13. Mahoney, M. Orbitz Sued by Southwest Airlines. *E-Commerce Times* 2019.
 14. Facebook. Facebook, Inc v. Power Ventures. 844 F.Supp.2d 1025 (E.D. Cal.). Facebook, Inc v. Power Ventures 2018.
 15. Gervais, D.J. The Protection of Databases. 92 CHI.-Kent L. Rev. 1109 2019.
 16. v. Hamidi, I.C. Intel Corp. v. Hamidi. 71 P.3d 296 2021.
 17. Hirschey, J. Symbiotic Relationships: Pragmatic Acceptance of Data Scraping. *SSRN Electronic Journal* 2017.
 18. Weisberg, S. Applied linear regression. John Wiley & Sons 2018.
 19. Panigrahi, S.; Mantri, J.K. Epsilon SVR and decision tree for stock market forecasting. *Green Computing and Internet of Things* Oct. 2019.
 20. Vinod, H.D. A survey of ridge regression and related techniques for improvements over ordinary least squares. *The Review of Economics and Statistics* 2020.
 21. Gluhovsky, I. Multinomial least angle regression. *IEEE transactions on neural networks and learning systems* 2020, 23, 169–174.
 22. Li, Q.; Lin, N. The Bayesian elastic net. *Bayesian analysis* 2019.
 23. Zemel, R.S.; Pitassi, T. A gradient-based boosting algorithm for regression problems. *Advances in neural information processing systems* 2018, pp. 696–702.
 24. Breiman, L. Random Forests. *Machine Learning* 2019, 45.

25. Ruder, S. An overview of gradient descent optimization algorithms. arXiv 2020, 1609.04747v2.
26. Crammer, K.; Dekel, O.; Keshet, J.; Shalev-Shwartz, S.; Singer, Y. Online passive aggressive algorithms 2016.