Available online at [*http://www.advancedscientificjournal.com*](http://www.advancedscientificjournal.com)
[*http://www.krishmapublication.com*](http://www.krishmapublication.com)
IJMASRI, Vol. 3, issue 1, pp.781- 886, Jan. -2022
https://doi.org/10.53633/ijmasri

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130 **IBI IMPACT FACTOR 1.5** **DOI: 10.53633/IJMASRI**

RESEARCH ARTICLE

HOUSE PRICE PREDICTION

Meenakshi Sharma

B. Tech. Student, Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India.)

Abstract

Machine learning has been growing exponentially in the last decade and has played a significant role in past years in image detection, spam reorganization, standard speech command, product recommendation, and medical diagnosis. Many applications and algorithms evolve in Machine Learning day to day. One such application found in journals is house price prediction. People are careful when buying a new house with their budgets and market strategies. The paper's objective is to forecast coherent house prices for non-householders based on their financial provisions and aspirations. Predicting housing prices with fundamental factors is the main crux of our research project. Here we aim to make our evaluations based on every essential parameter considered while determining the price. The paper involves predictions using different Regression techniques like Multiple linear, Ridge, LASSO, Elastic Net, Gradient boosting, and Ada Boost Regression. House price prediction on a data set has been made by using all the techniques mentioned earlier to find the best among them. The motive of this paper is to help the seller to estimate the selling cost of a house perfectly and to help people to predict the exact time slap to accumulate a house.

Keywords: Machine Learning, House Price, Prediction, LASSO, elastic net, Ridge, Ada Boost, Gradient Boosting

Introduction

Technology is leaping forward daily, and data is becoming the heart of technical innovations, making it possible to achieve desired goals using prediction models. Most people eventually buy/sell a house. This allows us to learn more about the housing market and helps us make more informed decisions. The traditional house price prediction approaches need more capacity for massive data analysis, causing low data utilization.

In this paper, we demonstrate all the possible Regression techniques suitable to our problem. Using several essential attributes related to the property or house and then predicting the most accurate price of that property based on the dataset on which the model is trained.

Evaluating several techniques and models, checking their accuracy and efficiency to find the best fit for the problem statement. Here we aim to make our evaluations based on every essential parameter that is considered while determining the price of the house.

Objective

State the gaps and discrepancies related to our problem statement in the previous literature and research works. To investigate and learn about previously used methods and algorithms to solve such problems and predict house price trends. We evaluate all the algorithms and methods encountered while analyzing the executed solutions on various data sets. The evaluation is based on technologies, complexity, use cases, and accuracy. We are selecting the best combination and algorithm to predict the house price, aiming to increase the accuracy and reduce the complexity and also stating the gaps and discrepancies found in previous related works.

Related work

This section summarises some of the literature's most essential and noteworthy work on house price prediction. A house price prediction system seeks to identify a property's actual value with high accuracy and using an optimized technique. Several concepts and technologies have been utilized to attain specific outcomes. Evaluating them, we learn about the trends in the dataset and optimize our algorithm for maximum accuracy.

We studied several methodologies and concluded that a linear regression model allows us to summarize and study the relationship between two continuous quantitative variables giving an accuracy of 76.15% as shown in Fig.1. Also, using Lasso Regression did not optimize the solution, as it resulted in an accuracy of 76.14% as shown in Fig 2. In comparison, if we opt for Gradient Boosting Regression, a machine that takes in a strategy to relapse. Also, arrangement problems that produce a prediction model in the structure of a group from claiming powerless prediction models Reported an accuracy of 91.27% as shown in Fig 3. A predictive model's exactness might be helped in two ways: Possibly by grasping characteristic building alternately toward applying boosting calculations straight far. There are a significant number of boosting calculations: Gradient Boosting, XGBoost, AdaBoost, Gentle Boost, and several others. Each boosting algorithm needs its underlying math.

Fig. 1. Accuracy achieved by linear regression

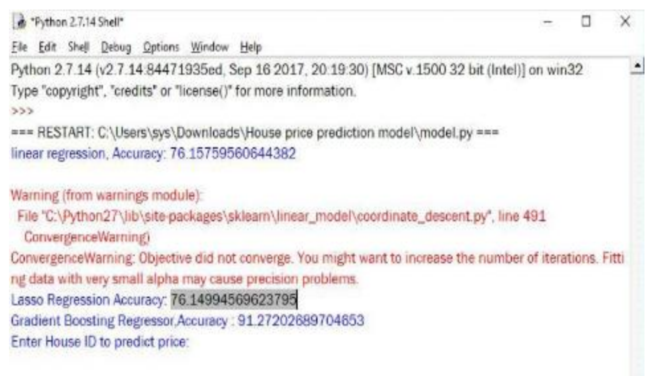

Fig 2. Accuracy achieved by LASSO Regression

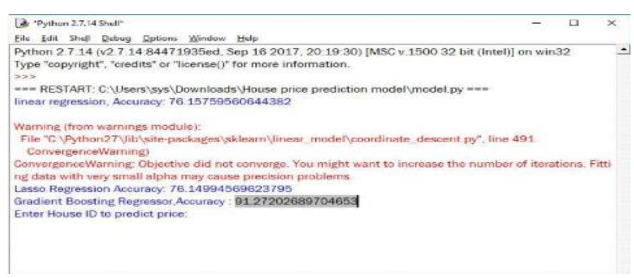

Fig 3. Accuracy achieved by Gradient Boosting Regression

Also, a variety may be watched same time applying them. Boosting calculation will be a standout among those. (Naga Satish et al., 2019).

The Predicted house prices using the decision tree regression classifier for the test data are given in Table 1. From the Table, it is clear that five test data record prices are predicted with lesser deviations. For example, record number 25 is predicted accurately, and record numbers 10 and 34 are predicted with less deviation.

Record No.	Actual Price (Lakhs)	Predicted Price (Lakhs)
4	22	24
28	17	20
29	19	14
33	27	24
34	24	23
25	11	11
10	18	17
22	12	10

Table 1. Predicted price values

Comparatively, the performance of multiple linear regression is better than the decision tree regression in predicting house prices. (Thamarai, and Malarvizhi 2020).

In order to find out the efficient regression technique for prediction, we require specific parameters to perform a comparison among the techniques. The parameters chosen for the comparison are Scores of the algorithm, [MSE] Mean Square Error, and [RMSE] Root Mean Square Error. Table 2 represents the resultant summary of the parameters when the above techniques are implemented practically.

Algorithm	Score	MSE	RMSE
Multiple Linear Regression	0.732072	391875744 48.88446	197958 51699
Ridge Regression	0.732164	391740496 29.73141	197924 35330
LASSO Regression	0.732072	391875537 34.32263	197958 46466
Elastic Net Regression	0.665228	489642930 85.00798	221278 76781
Ada Boost ing Regr ession	0.7801099	32161481 079.94242	179336 22355
Gradient Boosting Regression	0.9177022	12037006 088.27804	109713 90390

Table 2. Comparison of algorithms

From the above Table, we can efficiently perform comparisons of different algorithms clearly to find the best among them. Figure 2 below is used to visualize the performance of various techniques in a graphical format based on their scores. In Fig 4, the x-axis represents the various regression techniques considered for the study, and the y-axis represents the score values observed. (Raga Madhuri *et al.*, 2019)

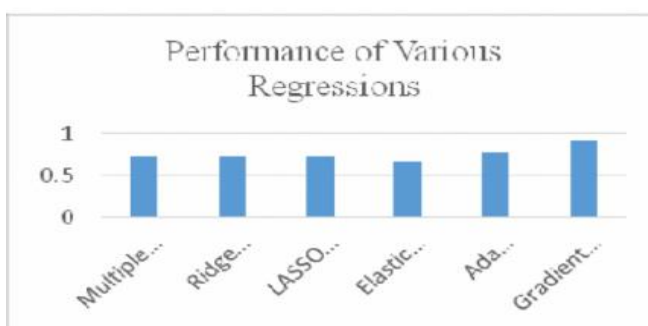

Fig 4. Performance of various regression models

Furthermore, the results are fed as input to the neural network. We use a neural network applied with boosted regression to increase the accuracy of the result. A neural network compares all the predictions and computes them to display the most accurate result. The system makes optimal use of Linear Regression, Forest regression, and Boosted regression. The algorithm's efficiency has been further increased with the use of Neural networks. The system will satisfy customers by providing accurate output and preventing the risk of investing in the wrong house. (Varma et al., 2018).

After further study, we came across the ARIMA (Autoregressive Integrated Moving Average) model, a well-known time series forecasting method. The main advantages of the ARIMA model include that it requires data on the time series in question only and it has good short-run forecasting ability. Therefore, the ARIMA model is employed in this paper to predict the house price trend. ARIMA model is the combination of the AR (autoregressive model), MA (moving average model), and ARMA models, where AR uses previous data to predict future change, MA tries to reduce the forecasting errors, AR and MA are mixed to be the ARMA model. P order AR and Q order MAcan be denoted in Fig 5.

$$u_t = c + \sum_{i=1}^{p} \phi_i u_{t-i} + \varepsilon_t ,$$

$$u_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t ,$$

Fig. 5. ARIMA model's Formulae

Where c and u are two constants, p is the order of the AR model, φ_i is the AR coefficient, ε_i is the white noise series with a mean value of zero and a variance of δ^2, q is the order of the MA model, θ_i is MA coefficient.

The ARIMA and SVR experimental results are shown in Figs. 6 - 7, where the solid and dashed curves denote the ground truth and predict values, respectively.

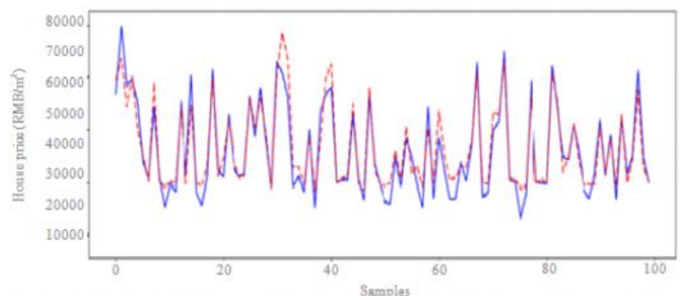

Fig 6. The ground truth and predicted value of the ARIMAmodel

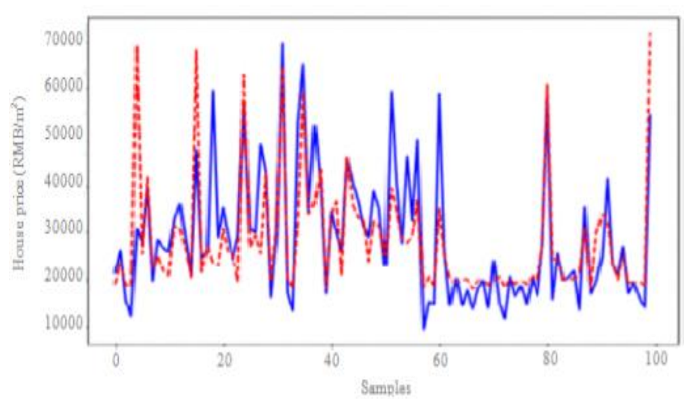

Fig 7. The ground truth and predicted value of the SVR model

The root mean square error (RMSE) and mean relative error (MRE) are used to evaluate the prediction performance of the two models. Table 3 shows the performance comparison of the two models on the training and testdatasets.

Model	RMSE /Training	RMSE /Test	MRE /Training	MRE /Test
Proposed	5393	7671	0.19	0.22
SVR	9024	10216	0.20	0.28

Table 3. Comparison of algorithms

From Figs. 6 - 7, it can be found that both the proposed model and the SVR model fit the training data well. Moreover, the fitting effects of the two models on the test dataset are inferior to the training data, which are consistent with the change curves of training loss and test loss. The predicted value of the ARIMA model is consistent with the actual value. At the same time, there is a significant deviation between the predicted value and the actual value of some data for the SVR model. As shown in Table I, both the RMSE and MRE indicators of the ARIMA model are smaller than those of the SVR model on the training dataset and test dataset, especially the RMSE indicator. Therefore, the ARIMA model achieves better prediction performance than the SVR model. (Wang *et al.*, 2019).

Challenge

Techniques absent in prior implementations include optimal way of stoping the model training at the right instant, a Keras callback that stops training when a monitored measure no longer improves.

Others have not employed stacked regression techniques. Base models and meta models on top of those base models could increase the efficiency and accuracy much higher.

Previous techniques did not incorporate the layers to improve model accuracy and optimise CNN logic effectively.

Conclusion

Several gaps were identified during the analysis of previously deployed procedures and approaches. Techniques such as linear regression, Multi regression, LASSO regression, Gradient boosting regression, SVR, ARIMA model paid off nicely in terms of improving model implementation and accuracy. However, ARIMA model proved to be the best out of them. Several holes were discovered throughout this evaluation. There was no way to terminate training the model if it attained a given outcome or reached a plateau. Implementing the Keras layers API would also aid in improving accuracy. Using stacked regression models in conjunction with top performing models from our study will assistin achieving better outcomes.

Reference

1. Raga Madhuri, C. H., Anuratha, G and Vani Pujitha. M. (2019). House Price Prediction Using Regression Techniques: A Comparative Study". IEEE 6th International Conference on smart structure and systems. ICSSS 2019, DOI: 10.1109/icsss.2019.8882834.
2. Thamarai M and Malarvizhi, S.P. (2020).“ Houose Price Prediction Modeling Using Machine Learning”, II J. Information Engineering and Electronic Business, 2020, 2, 15-20, DOI: 10.5815/ijieeb.2020.02.03.
3. Naga Satish, G., Ranghavendran, V. Sugnana Rao, M.D. Ch. Srinivasaulu (2019). House Price prediction Using Machine Learning” International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN : 2278-3075, Volume-8.Issue-9. July 2019, DOI: 10.35940/ijitee. I7849.078919.
4. Varma, A., Sarma, A. Doshi, S and Nair, R. (2018). “House Price Prediction Using Machine Learning and Neural Networks”

2018. Second international Conference on inventive Communication and Computational Technologies (ICICCT), DOI: 10.1109/icicct.2018.8473231

5. Wang,F.,Zou, Y. Zhang, H and Shi, H. (2019). House Price Prediction Approach based on Deep Learning and ARIMA Model". 2019. IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT). DOI: 10.1109/iccsnt47585.2019.8962443.

*****