

House Price Prediction Using Machine Learning -A Survey

Bharath P, Harshith V, Mohan Kumar G, Prema N S

Department of Information Science and Engineering, Vidyavardhaka College of Engineering Mysuru, India

Abstract - Methods for calculating the sale price of houses in cities remain a difficult and time-consuming task. The purpose of this article is to forecast the coherence of non-house prices. Using Machine Learning, which can intelligently optimize the optimum pipeline fit for a task or dataset, is a key technique to simplify the difficult design. Predicting the resale price of a house on a long-term temporary basis is vital, particularly for those who will be staying for a long time but not permanently. Forecasting house prices is an important aspect of real estate. The literature tries to extract relevant information from historical property market data. The price of real estate causes land price bubbles to expand, causing macroeconomic instability. The reasons that drive up real estate prices are important investigating so that the government may use them as a guide to help stabilize location, and various economic elements influencing at the time are all factors that influence the house selling price.

Key Words: Machine learning, House price, Prediction, regression.

1. INTRODUCTION

The value of a home is well known to be based on a wide range of factors. As a result, predicting the value of a home involves a unique set of issues. Houses are a need for society and rates vary depending on the amenities offered, such as size, area, location, and so on. Predicting the exact values of house pricing is a tricky process. This project is being suggested in order to better estimate property prices and provide more accurate results. This would be extremely beneficial to the people because house pricing is a problem that many individuals, rich and poor, are concerned about because one cannot gauge or predict the price of a property based on the location or amenities provided.

Also, Professional appraisers are commonly used to anticipate house prices in the past. However, due to a huge interest from the people, house broker, buyer, or seller, an appraiser is likely to be biased. So as a result, an automated prediction system can be useful as an objective party source that is less biased. The price of a house is a time series. Various methods for estimating property prices have been offered. A house price prediction model seeks to figure out what elements influence price changes in a certain area. Clearly, the factors that influence housing prices are complicated and intertwined processes that typical statistical methodologies overlook. Despite the fact

that the hedonic price model has gained widespread acceptance in recent years, it has been criticized for model assumptions and estimation, as well as for tackling nonlinear problems, global regression, and local clustering.

To anticipate the variance in house prices, nonlinear machine learning and fuzzy logics were applied. In, a neural network was used to forecast property values. The Support Vector Machine was used with optimization techniques like the Generic Algorithm and Particle Swarm Optimization. Repeated Incremental Pruning to Produce Error Reduction, Nave Bayes, and Ada Boost were among the machine learning techniques studied in. In terms of estimating property price, the RIPPER algorithm surpasses other models, according to the study. Linear regression, decision trees, and nearest neighbor were used to estimate house prices. In addition, the study found that Nave Bayes was the most consistent classifier for unequal frequency distributions.

Multiple linear regressions is a statistical approach for determining the relationship between numerous independent variables and the (dependent) target variable. The use of regression techniques to develop a model based on numerous criteria to forecast price is common. Predicting house prices is a difficult task. On the one hand, the factors that influence housing prices are complicated and vary nonlinearly, resulting in large forecast errors in standard models. On the other hand, the real estate market's daily data is massive and growing at a quick pace. The majority of recent research has focused on dismantling the distraction of house cost prediction. As a result of the analysis work done by various researchers all across the world, several theories have emerged.

2. Literature Survey

Lu *et.al* proposed a hybrid prediction model; the study looked at the impact of land financing and household spending on real estate prices in 33 major Chinese cities. The implementation of Panel data validation of fixed-effects model regression findings our proposition After establishing control of the city's local people, the rate of growth, per capita GDP, and the number of students enrolled in regular classrooms are all things to think about. Institutions of higher education, gender ratio, and consumer pricing Higher education institutions, gender ratios, and consumer pricing urban population density, land finance, and urban development are all indices to look

at. People's consumption levels will have a positive impact on real estate. It can formulate policies for the government, provide constructive opinions when planning to sell land, and prevent the local government from relying excessively on land revenue while attempting to expand by confirming that land transfer has a significant impact on the real estate price and the promotion mode of the factor and the house price. Land finance encourages economic growth, which leads to skyrocketing real estate values. This article indicates that citizens' consumption levels are a significant element influencing real estate price fluctuations, allowing the government to employ various information channels and data to forecast the real estate market's prospects and design suitable policies[1].

An ARIMA Model and Deep Learning Approach for House Price Prediction House prices and influential factors have a complex and nonlinear relationship. And one of the most common house price forecasting approaches is the absence of capacity for huge projects. Data examination A housing price index was established to address these issues. ARIMA is a deep learning prediction approach based on ARIMA. A model is proposed in this study. There are numerous elements that influence the price of a home. Some explanatory elements were chosen to be the important factors of house price in order to objectively depict the changing rules of house price. The raw housing data is obtained initially from the internet. The raw data is then transformed into outputs that may be easily used as inputs in data modelling via a data preparation procedure. The experimental findings suggest that the proposed strategy predicts individual property prices better than the SVR method. When making short-run predictions, the expected house price trend is essentially consistent with the real data[2].

A 6-layer BP neural network based on the Keras deep learning framework employing 12 macro parameters that have a substantial impact on property prices in Shanghai. When the Keras deep learning framework's elu and liner activation functions are paired with the RMSprop optimization method, the BP network performs better. By comparing the error between the test set's actual output and the expected output, the model's validity is confirmed[3].

The BP network performs better when the elu and liner activation functions of the Keras Keras deep learning framework are combined with The RMSprop optimization method. The model's Validity is confirmed by comparing the error Between the test set's actual output and the Expected output.

An empirical experiment that used an actual data-set of houses in Petaling, Jaya, Selangor, Malaysia, to demonstrate different approaches to hyperparameter tweaking with Python modules like Scikit-Learn and TPOT. The Python codes for utilizing conventional machine

learning with manual configurations of the five selected algorithms are longer than those for using the AML TPOT, but the best result of the AML forecast did not decrease at all in the supplied data-set, and actually increased slightly. This discovery offers up new avenues for AML research in the future, such as examining a larger data set and other GP parameterize settings[4].

Lim *et al* purposed useful models for predicting property prices. It also provides details on the Melbourne housing market. To begin, the raw data is cleaned and transformed into a readable data-set. The data is then reduced and transformed using Stepwise and PCA techniques. Following that, a variety of tactics are implemented and evaluated in order to arrive at the optimal solution. According to the evaluation phase, combining Step-wise and SVM models is a competitive strategy. As a result, future deployments may include it. This research can also be extended to transitional datasets from other sections of the Australian property market. The studies were run on a Windows system using the R programming language. Both the train and assessment datasets Mean Squared Error (MSE) are shown. The baseline for model comparison will be linear regression, as discussed previously. Each model's evaluation ratio is equal to its evaluation MSE divided by Linear regression's evaluation MSE. The higher the accuracy of the model's forecast, the lower the evaluation ratio[5].

Patel and Upadhyay[6] studied different pruning techniques and their characteristics, and thus pruning efficiency is assessed. They also assessed the accuracy of the glass and diabetes datasets using the WEKA tool and various pruning factors. The ID3 algorithm divides attributes according to their entropy. The TDIDT algorithm creates a set of classification rules using an interactive model of a decision tree[7, 8]Formalized paraphrase. Fan et al [9]used a decision tree approach to determine house resale prices based on significant characteristics. In this paper, a hedonic-based regression method is used to identify the relationship between house prices and significant characteristics. Ong et al. [10], as well as Berry et al. [11], used hedonic-based regression for house price forecasting based on significant attributes.

Shinde and Gawande [11], compared the accuracy of various machine learning algorithms such as lasso, SVR, Logistic regression, and decision tree in predicting the sale price of houses. Alfiyatin et al.[12] developed a system for predicting house prices using Regression and Particle Swarm Optimization (PSO).

II. Proposed Model

Proposed system is an real time application used in real time. It is a browser based application which is meant for real estate business. Proposed system is a generic application which can be accessed from different locations.

System over comes all different drawbacks that we have in the existing system and come up with the solution. System major objective is to predict the price of house. Proposed system uses machine learning technique to predict price. We use supervised learning technique for prediction. From online training data-sets downloaded, data collected from the sources such as www.kaagle.com, www.data.world.com, and www.data.gov.in. we use many parameters for house price prediction.

III. CONCLUSION

In this paper an optimal model isn't always the same as a robust model. A model that uses a learning strategy that isn't appropriate for the data structure at hand on a frequent basis. Although the data may be too noisy or contain insufficient samples to allow a model to adequately reflect the target variable, the model is nonetheless fit. We can see that the evaluation metrics for advanced regression models behave similarly when we look at them. When we look at the evaluation metrics for advanced regression models, we can see that they behave similarly. In comparison to the basic model, we can choose any one for house price forecast. We can look for outliers with the use of box plots. If outliers are present, we can eliminate them and evaluate the model's performance to see if it can be improved.

3. CONCLUSIONS

Several approaches have been proposed related to this issue in many papers which we have mentioned above. We have discussed above how various algorithms and methods are used to predict the disease. But in none of the papers they have done a real time application. We have proposed a model, which will be a real time application to predict the prices.

REFERENCES

- [1] S. Lu, Z. Li, Z. Qin, X. Yang, and R. S. M. Goh, "A hybrid regression technique for house prices prediction," in 2017 IEEE international conference on industrial engineering and engineering management (IEEM), 2017, pp. 319-323.
- [2] M. F. Mukhlshin, R. Saputra, and A. Wibowo, "Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor," in 2017 1st International Conference on Informatics and Computational Sciences (ICICoS), 2017, pp. 171-176.
- [3] P. Durganjali and M. V. Pujitha, "House resale price prediction using classification algorithms," in 2019 International Conference on Smart Structures and Systems (ICSSS), 2019, pp. 1-4.
- [4] R. E. Febrita, A. N. Alfiyatin, H. Taufiq, and W. F. Mahmudy, "Data-driven fuzzy rule extraction for housing price prediction in Malang, East Java," in 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2017, pp. 351-358.
- [5] W. T. Lim, L. Wang, Y. Wang, and Q. Chang, "Housing price prediction using neural networks," in 2016 12th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), 2016, pp. 518-522.
- [6] N. Patel and S. Upadhyay, "Study of various decision tree pruning methods with their empirical comparison in WEKA," International journal of computer applications, vol. 60, 2012.
- [7] R. Quinlan, "4.5: Programs for machine learning morgan kaufmann publishers inc," San Francisco, USA, 1993.
- [8] J. R. Quinlan, "Induction of decision trees," Machine learning, vol. 1, pp. 81-106, 1986.
- [9] G.-Z. Fan, S. E. Ong, and H. C. Koh, "Determinants of house price: A decision tree approach," Urban Studies, vol. 43, pp. 2301-2315, 2006.
- [10] S. E. Ong, K. H. D. Ho, and C. H. Lim, "A constant-quality price index for resale public housing flats in Singapore," Urban Studies, vol. 40, pp. 2705-2729, 2003.
- [11] N. Shinde and K. Gawande, "Valuation of house prices using predictive techniques," Journal of Advances in Electronics Computer Science, vol. 5, pp. 34-40, 2018.
- [12] A. N. Alfiyatin, R. E. Febrita, H. Taufiq, and W. F. Mahmudy, "Modeling house price prediction using regression analysis and particle swarm optimization," International Journal of Advanced Computer Science and Applications, vol. 8, pp. 323-326, 2017.