

PREDICTION OF HOUSE PRICING USING MACHINE LEARNING WITH PYTHON

¹PANTHAM SARIKA, ²S.VIJAY KUMAR

¹MCA Student, ²Assistant Professor

DEPARTMENT OF MCA

SREE CHAITANYA COLLEGE OF ENGINEERING, KARIMNAGAR

ABSTRACT

This project provides an overview about how to predict house costs utilizing different regression methods with the assistance of python libraries. The proposed technique considered the more refined aspects used for the calculation of house price and provided the more accurate prediction. It also provides a brief about various graphical and numerical techniques which will be required to predict the price of a house. This paper contains what and how the house pricing model works with the help of machine learning and which dataset is used in our proposed model.

I. INTRODUCTION

House/Home are a basic necessity for a person and their prices vary from location to location based on the facilities available like parking space, locality, etc. The house pricing is a point that worries a ton of residents whether rich or white collar class as one can never judge or gauge the valuing of a house based on area or offices accessible. Buying a house is one of the greatest and significant choices of a family as it expends the entirety of their investment funds and now and again covers them under loans. It is a difficult task to predict the accurate values of house pricing. Our proposed model would make it possible to predict the exact prices of houses.

In today's society, medical care problems have become a hot topic, and problems such as the unbalance and insufficient allocation of medical resources has become increasingly apparent. In this situation, the application of ML has become the unavoidable trend in the current

development of medical care. As early as 1972, the scientists in the University of Leeds in the UK had been trying to use artificial intelligence (ANN) algorithms to judge abdominal pain. Now, more and more researchers are committed to the combination of ML and medical care. The methods of pathological diagnosis of tumors, lung cancer, etc. by ML has gradually entered the field of vision. Some companies, such as Alibaba, Amazon, and Baidu have established their own research team working for it. This introduction of ML in medical care has greatly saved medical resources and provided a new way for citizens to see a doctor and facilitate people's lives. At the same time, the demand of people also provides a new impetus for the research and development of ML, with promoting its continuous improvement. B.

II. SYSTEM ANALYSIS

EXISTING SYSTEM:

In The Existing system used xgboost for house price prediction. This study aims to explore the important explanatory features and determine an accurate mechanism to implement spatial prediction of housing prices in Beijing..., based on the housing price and features data in Beijing, China. Our result shows that compared to traditional hedonic methods, machine learning methods demonstrate significant improvements on the accuracy of estimation despite that they are more time-costly. Moreover, it is found that XGBoost is the Less accurate model in explaining and predicting the spatial dynamics of housing prices in Beijing.

DISADVANTAGES OF EXISTING SYSTEM:

- IN Xgboost, you have to manually create dummy variable/ label encoding for categorical features before feeding them into the models. Catboost/Lightgbm can do it on their own, you just need to define categorical features names or indexes.
- Training time is pretty high for larger datasets.
- Moreover, it is found that XGBoost is the Less accurate model in explaining and predicting the spatial dynamics of housing prices in Beijing.

Algorithm: XGBOOST.

PROPOSED SYSTEM:

The proposed method is based on the linear regression. This project is proposed to predict house prices and to get better and accurate results. The data for the house prediction is collected from the publicly available sources. In validation, training is performed on 50% of the dataset and the rest 50% is used for testing purposes.

This technique splits the dataset into a number of subsets. At that point, it has been attempted for preparing on the entirety of the subsets; however, leave one (k-1) subset for the assessment of the prepared model. This strategy emphasizes k times with an alternate subset turned around for the preparation reason each time.

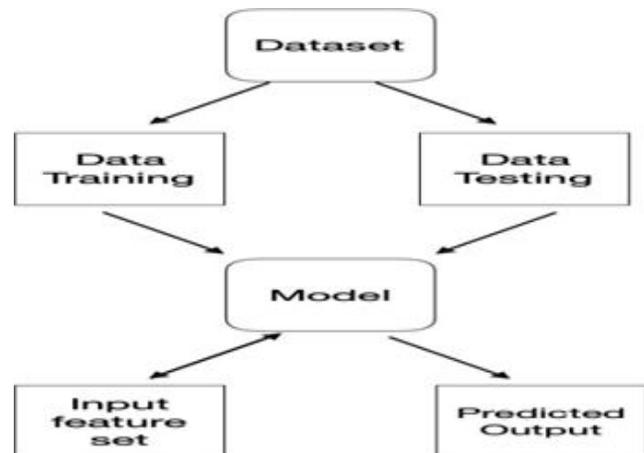
ADVANTAGES OF PROPOSED SYSTEM:

- The error free prediction provides better planning in the prediction of house price and other industries.
- This would be of great help for the people.

- This would be of great help to the people because the house pricing is a topic that concerns a lot of citizens whether rich or middle class as one can never judge or estimate the pricing of a house on the basis of locality or facilities available.
- Linear Regression is simple to implement and easier to interpret the output coefficients
- The ability to determine the relative influence of one or more predictor variables to the criterion value

Algorithm: Linear Regression (LR)

SYSTEM DESIGN



III. IMPLEMENTATION

MODULES DESCRIPTION:

User:

The User can register the first. While registering he required a valid User email and mobile for further communications. Once the User registers, then the admin can activate the User. Once the admin activates the User then the User can login into our system. After login User will add the data to predict house values.

Admin:

Admin can login with his credentials. Once he logs in he can activate the users. The activated users only login in our applications. The admin will store csv data into our database. we can implement logistic algorithm to predict house and also we can perform cross validation.

Machine learning:

Machine learning refers to the computer's acquisition of a kind of ability to make predictive judgments and make the best decisions by analyzing and learning a large number of existing data. The representation algorithms include deep learning, artificial neural networks, decision trees, enhancement algorithms and so on. The key way for computers to acquire artificial intelligence is machine learning. Nowadays, machine learning plays an important role in various fields of artificial intelligence. Whether in aspects of internet search, biometric identification, auto driving, Mars robot, or in American presidential election, military decision assistants and so on, basically, as long as there is a need for data analysis, machine learning can be used to play a role.

IV. CONCLUSION

The sales price for the houses is calculated using different algorithms. The sales prices have been calculated with better accuracy and precision. This would be of great help for the people. To achieve these results, various data mining techniques are utilized in python language. The various factors which affect the Real Estate pricing should be considered and work upon them. Machine learning has assisted to complete out task. Firstly, the data collection is performed. Then data cleaning is carried out to remove all the errors from the data and make it clean. Then the data preprocessing is done. Then with help of data visualization, different plots are created. This has depicted the distribution of data in different forms. Further, the preparation

and testing of the model are performed. It has been found that some of the classification algorithms were applied on our dataset while some were not. So, those algorithms which were not being applied on our Real Estate pricing dataset are dropped and tried to improve the accuracy and precision of those algorithms which were being applied on our house pricing dataset. To improve the accuracy of our classification algorithms, a separate stacking algorithm is proposed. It is extremely important to improve the accuracy and precision of the algorithms in order to achieve better results. If the results are not accurate then they would be of no help to the people in predicting the sales prices of lands. It also made use of data visualization to achieve better accuracy and results. The sales price is calculated for the houses using different algorithms. The sales prices have been calculated with better accuracy and precision. This would be of great help for the people.

REFERENCES

- [1] Jain, N., Kalra, P., & Mehrotra, D. (2019). Analysis of Factors Affecting Infant Mortality Rate Using Decision Tree in R Language. In *Soft Computing: Theories and Applications* (pp. 639-646). Springer, Singapore.
- [2] [2] R. A. Rahadi, S. K. Wiryono, D. P. Koesrindartotoor, and I.B. Syamwil, —Factors influencing the price of housing in Indonesia,|| *Int. J. Hous. Mark. Anal.*, vol. 8, no. 2, pp. 169–188, 2015
- [3] V. Limsombunchai, —House price prediction: Hedonic price model vs. artificial neural network,|| *Am. J. ...*, 2004
- [4] Kadir, T., & Gleeson, F. (2018). Lung cancer prediction using machine learning and advanced imaging techniques. *Translational Lung Cancer Research*, 7(3), 304-312.

- [5] Liu, J., Ye, Y., Shen, C., Wang, Y., & Erdélyi, R. (2018). A New Tool for CME Arrival Time Prediction using Machine Learning Algorithms: CATPUMA. *The Astrophysical Journal*, 855(2), 109.
- [6] Velankar, S., Valecha, S., & Maji, S. (2018, February). Bitcoin price prediction using machine learning. In *Advanced Communication Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE Xplore Part Number: CFP20V66-ART; ISBN: 978-1-7281-4108-4 978-1-7281-4108-4/20/\$31.00 ©2020*
- [7] Malhotra, R., & Sharma, A. (2018). Analyzing Machine Learning Techniques for Fault Prediction Using Web Applications. *Journal of Information Processing Systems*, 14(3).
- [8] Choo, M. S., Uhm, S., Kim, J. K., Han, J. H., Kim, D. H., Kim, J., & Lee, S. H. (2018). A Prediction Model Using Machine Learning Algorithm for Assessing Stone-Free Status after Single Session Shock Wave Lithotripsy to Treat Ureteral Stones. *The Journal of urology*.
- [9] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L., & Farahmand, M. (2018). A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. *Biocybernetics and Biomedical Engineering*, 38(1), 1-15.
- [10] Fan, C., Cui, Z., & Zhong, X. (2018, February). House Prices Prediction with Machine Learning Algorithms. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing* (pp. 6-10). ACM.
- [11] Zhou, J., Zhang, H., Gu, Y., & Pantelous, A. A. (2018). Affordable levels of house prices using fuzzy linear regression analysis: the case of Shanghai. *Soft Computing*, 1-12.
- [12] Jang, H., Ahn, K., Kim, D., & Song, Y. (2018, June). Detection and Prediction of House Price Bubbles: Evidence from a New City. In *International Conference on Computational Science* (pp. 782-795). Springer, Cham.
- [13] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7), 1145-1159.
- [14] Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, 42(6), 2928-2934. [15] Harrison, D., and D. L. Rubinfeld. 1978. "Hedonic Housing Prices and the Demand for Clean Air." *J. Environ. Econ. Manag.* 5 (1): 81–102.