

House Price Prediction Using Machine Learning

Mrs.Niteesha Sharma -Assistant Professor.

Varshika Anantha, Arshitha Polina, Hrithik Manchi.

Department of Information Technology, Anurag University, Hyderabad, India.

ABSTRACT:

This paper 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. House Price Index (HPI) is commonly used to estimate the changes in housing price. Since housing price is strongly correlated to other factors such as location, area, population, it requires other information apart from HPI to predict individual housing price. There has been a considerably large number of papers adopting traditional machine learning approaches to predict housing prices accurately, but they rarely concern about the performance of individual models and neglect the less popular yet complex models.

KEY WORDS: CSV, CSS, GUI, ML

Date of Submission: 11-03-2023

Date of acceptance: 25-03-2023

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 expands 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. 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. Many prediction models (Machine Learning Models), including support vector regression, artificial neural networks, and others, can be used to predict house prices.

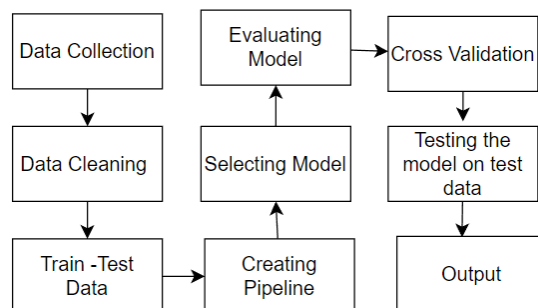


Fig1: Research Data flow Diagram

A- Motivation:

Having lived in India for so many years if there is one thing that I had been taking for granted, it's that housing and rental prices continue to rise. Since the housing crisis of 2008.

However, in the 4th quarter of 2016, I was surprised to read that Bombay housing prices had fallen the most in the last 4 years. In fact, median resale prices for condos and coops fell 6.3%, marking the first time there was a decline since Q1 2017. The decline has been partly attributed to political uncertainty domestically and abroad and the 2014 election.

B. Research Goal

Our project includes estimating the price houses without any expectations of market prices and cost increment. The objective of the project is prediction of residential prices for the customers considering their financial plans and needs. This project means to predict house prices in Pune city with various regression techniques. It will help clients to put resources into a web-based application without moving toward a broker. It also provides a brief about various graphical and numerical techniques which will be required to predict the price of a house. Our project contains what and how the house pricing model works with the help of machine learning and which dataset is used in our proposed model.

II. LITERATURE SURVEY:

Analysis of Factors Affecting Infant Mortality Rate Using Decision Tree[1]:

This is a study done for the social cause that was increasing at an alarming rate and was creating a situation of panic among the people of the world, Mortality Rate. This situation was analyzed by analyzing various factors such as birth rate, literacy rate, number of health centers, etc.

International Journal of Housing Markets and Analysis (Int J Hous Market Anal) [2]:

From both a local and international perspective, housing remains the most common form of land use. In recent times housing in many countries has rapidly evolved as an investment medium for private and institutional investors, which in turn is

now demanding a higher level of research. The International Journal of Housing Markets and Analysis aims to provide an international forum for the interchange of information and ideas relating to housing, housing markets and the interaction thereof.

House Price Prediction: Hedonic Price Model vs Artificial Neural Network. AUTHORS: Zhang, Xinwei; Han, Yaia, Wei Xu, Wang Qilia[3]:

The objective of this study is to empirically compare the predictive power model with an artificial neural network model on house price prediction. A sample of 200 houses in Christchurch, New Zealand is randomly selected from the Harcourt website.

Factors including house size, house age, house type, number of bedrooms, number of bathrooms, number of garages, amenities around the house and geographical location are considered. Empirical results support the potential of artificial neural network on house price prediction, although previous studies have commented on its black box nature and achieved different conclusions.

Lung cancer prediction using machine learning and advanced imaging techniques AUTHORS: Kadir, Fergus Gleeson[4]:

Machine learning based lung cancer prediction models have been proposed to assist clinicians in managing incidental or screen detected indeterminate pulmonary nodules. Such systems may be able to reduce variability in nodule classification, improve decision

making and ultimately reduce the number of benign nodules that are needlessly followed

or worked-up. In this article, we provide an overview of the main lung cancer prediction approaches proposed to date and highlight some of their relative strengths and weaknesses.

III. EXISTING SYSTEMS:

In The Existing system used xgboost for house price prediction. This study aims to explore the important explanatory features and determine an accurate mechanism implement spatial prediction of housing prices in Beijing..., based on the housing price and features data in Beijing, China. Our result shows that compared traditional hedonic methods, machine learning methods demonstrates significant improvements 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.

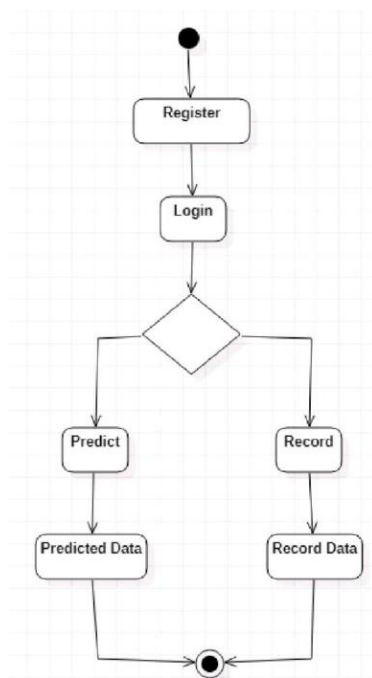


Fig 2: Representation of workflow

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.

IV. 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. The error free prediction provides better planning in the prediction of house price and other industries. 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. With starting a house price prediction business, there is truly never a dull moment. Your job offers a lot of variety and allows you to meet interesting people from all walks of life. Starting A House Price Prediction Business allows you to control every aspect of your life and make your own dreams come true every day. One reason for the strong rise in house prices in recent years have been cheap money — lower mortgage interest rates has made it easier to buy homes and has pushed up prices all around the country.

IMPLEMENTATION:

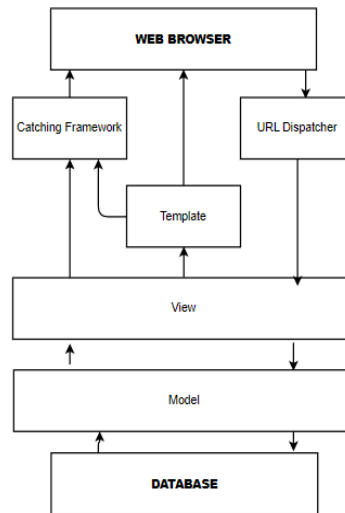


Fig 3: Django Architecture

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

Django's primary goal is to ease the creation of complex, database-driven websites. Django emphasizes reusability and "pluggability" of components, rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models. The dropout feature is the target. The procedure is carried out using the Pandas library. Use Matplotlib in Python to visualise a dataset's graph. function. After that, try to identify some attribute combinations and fill in any values that are lacking. We divided the data in half, using 80% for training and the remaining 20% for testing.

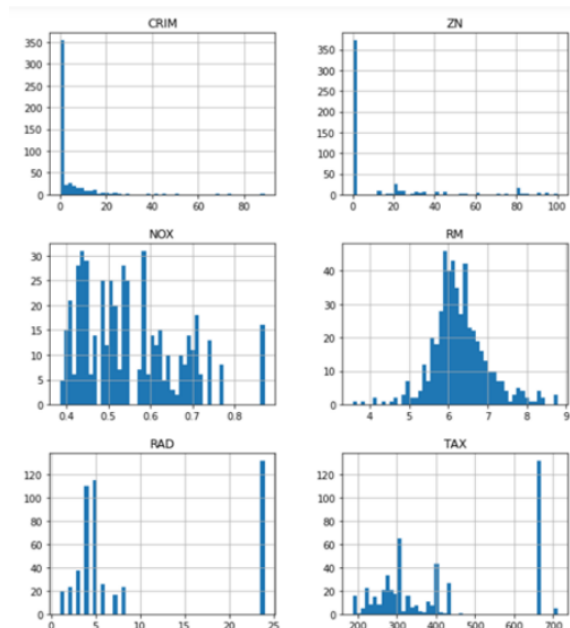


Fig 4: Visualization of dataset

V. RESULTS:

In this paper, we used regression model to make a prediction. In this model we used linear regression. There is multiple linear regression such as simple regression and multivariable regression. Primarily we can use simple regression to predict the price of the house easily. However, you can use multivariable regression to predict

the results more accurately using different variables. We used complete dataset that has accurate information regarding the houses.

This linear regression model works very well. The accuracy we attained by using this model is 87.8%. It has the highest accuracy. We are able to predict the accurate prices of the houses using this model.

VI. CONCLUSIONS:

We created this project for mute and deaf people, sometimes known as physically disabled people, by merging Python with Open CV and other technologies. They will be able to converse without difficulty with this program. Hand gesture recognition will have a very broad reach in the future as the number of users grows every day. In order to find an alternative to using a person to transmit their sentiments, they are willing to utilize software like this. As a result, this initiative will be of more assistance to them. Most of these physically impaired people are dependent on sign language translators to express their thoughts to rest of the world which may make them feel uncomfortable. Machine learning has assisted to complete our 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 house 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.

S.No	Machine Learning Algorithms	Accuracy Percentage
1.	Linear Regression	87.8

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 houses.

FUTURE ENHANCEMENTS:

Future work on this study could be divided into seven main areas to improve the result even further. Which can be done by: - The used pre-processing methods do help in the prediction accuracy. However, experimenting with different combinations of pre-processing methods

to achieve better prediction accuracy. - Make use of the available features and if they could be combined as binning features has shown that the data got improved. - Training the datasets with different regression methods such as Elastic net regression that combines both L1 and L2 norms. In order to expand the comparison and check the performance. - The correlation has shown the association in the local data. Thus, attempting to enhance the local data is required to make it rich with features that vary and can provide a strong correlation relationship.

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]. R. A. Rahadi, S. K. Wiryono, D. P. Koesrindartotoor, and I.B. Syamwil, Factors influencing the price of housing in Indonesia, *J. Hous. Mark. Anal.*, v1no. 8. 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.