Prediction of House Pricing Using Machine Learning with Python

Satya Mohan Chowdary G[1], A. Jaswanthi[2], P. Jaya Sri[3], S. Rajendranadh[4], N. Snehith[5]
[1,2,3,4,5]*Pragati Engineering College*

Abstract **- House prices are increasing every year which has necessitated the modeling of house price prediction. 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 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.**

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. 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 have 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.

MOTIVATION

As customer have come across some problems of the existing system which motivated us to develop this application. House Price Prediction model helps buyer to choose the house based on their financial status and also saves the time of searching the house. This helps the seller to know better value of their property and avoids the brokers.

II. PROBLEM DEFINITION

The project we had built is to solve the house price prediction problem, based on certain features of the house, such as the longitude, latitude, number of rooms, number of bedrooms, ocean proximity. We have to predict the estimated price of the house.

III. LITERATURE STUDY

1. Analysis of Factors Affecting Infant Mortality Rate Using Decision Tree in R Language.

AUTHORS: Parul Kalra, Deepti Mehrotra

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. using the decision

tree technique in R tool which illustrated trees of two different decades separately and analyzed the factors affecting the mortality rate with their contribution in driving its rate, and also the summary of decision tree will indicate its accuracy and kappa factor to judge the authenticity of the factors chosen. This will be useful to the governing bodies to get to know about the factors and work upon them for the decrease of the infant mortality rate.

2. International Journal of Housing Markets and Analysis (Int J Hous Market Anal) AUTHORS: RJ Bolton, DJ Hand.

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.

3. House Price Prediction: Hedonic Price Model vs Artificial Neural Network. AUTHORS: Zhang, Xinwei; abHan, Yaocia, WeiXu, WangQilia

The objective of this study is to empirically compare the predictive power of the hedonic 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

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

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. We discuss some of the challenges in the development and validation of such techniques and outline the path to clinical adoption.

5. A New Tool for CME Arrival Time Prediction Using Machine Learning Algorithms: CAT-PUMA

AUTHORS: Jiajia Liu, Yudong Ye, Chenlong Shen, Yuming Wang, Robert Erdélyi

Coronal Mass Ejections (CMEs) are arguably the most violent eruptions in the Solar System. CMEs can cause severe disturbances in the interplanetary space and even affect human activities in many respects, causing damages to infrastructure and losses of revenue. Fast and accurate prediction of CME arrival time is then vital to minimize the disruption CMEs may cause when interacting with geospace. In this paper, we propose a new approach for partial-/full-halo CME Arrival Time Prediction Using Machine learning Algorithms (CAT-PUMA). Via detailed analysis of the CME features and solar wind parameters, we build a prediction engine taking advantage of 182 previously observed geo-effective partial-/full-halo CMEs and using algorithms of the Support Vector Machine (SVM). We demonstrate that CAT-PUMA is accurate and fast. In particular, predictions after applying CAT-PUMA to a test set, that is unknown to the engine, show a mean absolute prediction error.

5.9 hours of the CME arrival time, with 54% of the predictions having absolute errors less than 5.9 hours. Comparison with other models reveals that CAT-PUMA has a more accurate prediction for 77% of the events investigated; and can be carried out very fast, i.e. within minutes after providing the necessary input parameters of a CME. A practical guide containing the CAT-PUMA engine and the source code of two examples are available in the Appendix, allowing the community to perform their own applications for prediction using CAT-PUMA.

III. EXISTING SYSTEM

In the Existing system xgboost is used 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. Cat boost/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.

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.

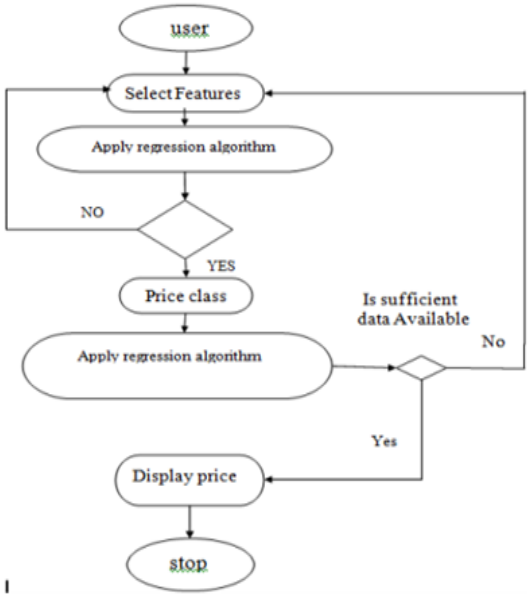
This technique splits the dataset into a number of subsets. At that point, it has been attempted for preparing on the entirety of the subsets; however, leave one (k-1) subset for the assessment of the prepared model. This strategy emphasizes k times with an alternate subset turned around for the preparation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

V. DATA FLOW DIAGRAM

- Food, clothing, and shelter are the essential needs of life. Availability of these needs increases the physical effectiveness and productivity of the people. So housing is a factor of prime importance in human resource development of any economy. At one point in life, everybody has to deal with the housing dilemma. For many people housing is one of the major investments of their life, people pay a fortune to buy their Dream House.
- Data is at the heart of technical innovations, achieving any result is now possible using predictive models. Machine learning is extensively used in this approach. Machine learning means providing valid dataset and further on predictions are based on that, the machine itself learns how much importance a particular event may have on the entire system based on its preloaded data and accordingly predicts the result.
- Various modern applications of this technique include predicting stock prices, predicting the possibility of an earthquake, predicting company sales and the list has endless possibilities.



VI. MODULES

There are two modules:

- User
- Admin

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.

VII. CONCLUSION

The sales price for the houses is calculated using different algorithms. The sales prices have been calculated with better accuracy and precision. To achieve these results, various data mining techniques are utilized in python language. The various factors which affect the house 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. To improve the accuracy of our regression algorithms, a separate stacking algorithm is proposed. House Price Prediction model helps buyer to choose the house based on their financial status and also saves the time of searching the house. This helps the seller to know better value of their property and avoids the brokers.

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,‖ 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., Uhmn, 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