PREDICTION OF HOUSE PRICING USING MACHINE LEARNING WITH PYTHON

R.JAGADEESH[1], VADDI SRIVALLIDEVI[2]

[1]MCA Student, B V Raju College, Kovvada, Andhra Pradesh, India.

[2]Assistant Professor, B V Raju College, Kovvada, Andhra Pradesh, 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.

Keywords: Data set, machine learning, python libraries.

1. INTRODUCTION:

House/Home are a basic necessity for a person and their prices varying 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 of 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 the difficult task to predict the accurate values of house pricing. Our proposed model would make it possible to predict the exact prices of houses. Machine Learning is a field of Artificial Intelligence which enables PC frameworks to learn and improve in execution with the assistance of information. It is used to study the construction of algorithms that make predictions on data. Machine learning is used to perform a lot of computing tasks. It is also used to make predictions with the use of computers. Machine learning is sometimes also used to devise complex models. The principle point of machine learning is to permit the PCs to learn things naturally without the assistance of people.

Machine learning is very useful and is widely used around the whole world. The process of machine learning involves providing data and then training the computers by building machine learning models with the help of various algorithms. Machine learning can be used to make various applications such as face detection application, etc. Machine Learning is a field in software engineering that has changed the way of examining information colossally.

OBJECTIVE:

This project is proposed to predict house prices and to get better and accurate results. The stacking algorithm is applied on various regression algorithms to see which algorithm has the most accurate and precise results. This would be of great help to the people because the house pricing ids 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. To accomplish this task, the python programming language is used. Python is a high level programming language for general purpose programming. It enables clear programming on both small and large scales. It is an easily readable language.

2. LITERATURE SURVEY:

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.

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.

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

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.

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 aerospace. 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.

3. 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

4. IMPLEMENTATION WITH RESULTS EXPLANATION

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.

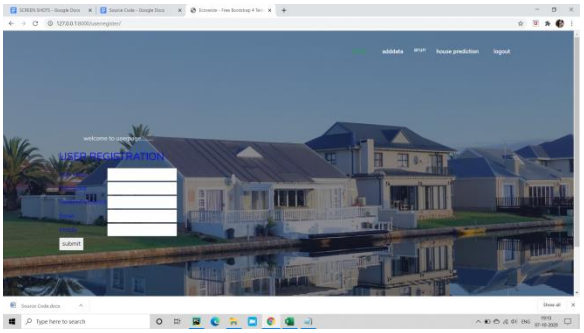

Fig.4.1. User Register.

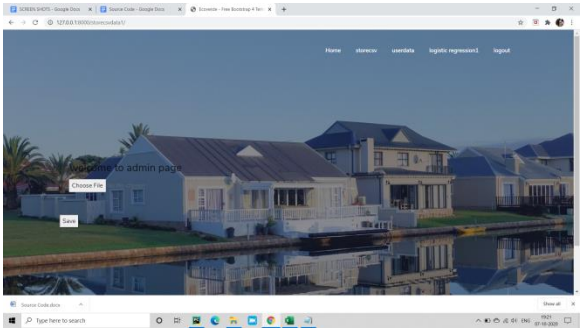

Fig.4.2. Store-Csv data.

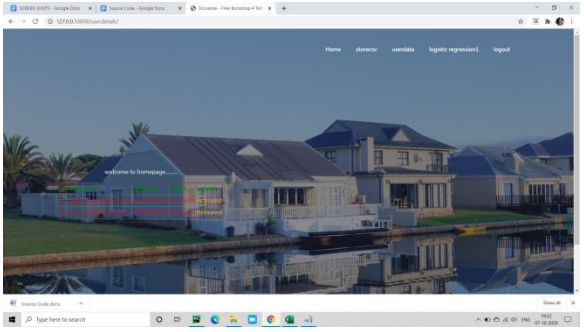

Fig.4.3. User Data.

5. CONCLUSION:

The sales price for the houses are 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 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.

Further Enhancement:

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. 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. 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).

[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.

[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.