Machine Learning Models for Accurate House Price Estimation

Mr DADI VENKATA VARAPRASAD [1], ANAMALA PRAVALIKA [2]

#1 Associate Professor in Audisankara College of Engineering & Technology, Gudur (M), Tirupati (Dt).

#2 MCA student in Audisankara College of Engineering & Technology, Gudur (M), Tirupati (Dt).

ABSTRACT_ Nowadays, data mining is frequently used in the real estate industry. Data mining is particularly helpful in predicting property values, important housing features, and many other things since it can extract pertinent knowledge from raw data. Studies have indicated that changes in home values frequently cause anxiety for homeowners and the real estate industry. A review of the literature is done in order to determine the most effective models for predicting home values as well as the pertinent characteristics. The analysis's conclusions supported the usage of XGBoost and Random Forest as the most effective models in comparison to other models. Furthermore, our results imply that structural and locational characteristics play a significant role in determining home values. In particular, housing developers and academics can greatly benefit from this study's determination of the most important factors influencing home prices and the optimal machine learning model to employ when conducting research in this area.

1.INTRODUCTION

House is one of human life's most essential needs, along with other fundamental needs such as food, water, and much more. Demand for houses grew rapidly over the years as people's living standards improved. While there are people who make their house as an investment and property, yet most people around the world are buying a house as their shelter or as their livelihood.

According to housing markets have a positive impact on a country's currency, which is an important national economy scale. Homeowners will purchase goods such as furniture and household equipment for their home, and homebuilders or contractors will purchase raw material to build houses to satisfy house demand, which is an indication of the economic wave effect created by the new house supply. Besides that, consumers have capital to make a large investment, and the construction industry is in good condition can be seen through a country's high level of house supply.

According to numerous international organizations and human rights have emphasized house importance. House is profoundly rooted in the economic, financial, and political structure of each country. Nevertheless, reported that the fluctuation of house prices has always been an issue for house owners, buildings and real estate, besides stated that house has become unaffordable as there is substantial price growth in several countries in the housing sector. Residents' quality of life as well as national economy depends on the potential house price increase. Ultimately, this issue will affect investors who are making their house as an investment.

An increase in house demand occurs each year, indirectly causing house price increases every year. The problem arises when there are numerous variables such as location and property demand that may influence the house price, thus most stakeholders including buyers and developers, house builders and the real estate industry would like to know the exact attributes or the accurate factors influencing the house price to help investors make decisions and help house builders set the house price.

House price prediction can be done by using a multiple prediction models (Machine Learning Model) such as support vector regression, artificial neural network, and more. There are many benefits that home buyers, property investors, and house builders can reap from the houseprice model. This model will provide a lot of information and knowledge to home buyers, property investors and house builders, such as the valuation of house prices in the present market, which will help them determine house prices.

Meanwhile, this model can help potential buyers decide the characteristics of a house they want according to their budget. Previous studies focused on analysing the attributes that affect house price and predicting house price based on the model of machine learning separately. However, this article combines such a both predicting house price and attributes together.

2.LITERATURE SURVEY

2.1 Predicting Housing Sales in Turkey Using Arima, Lstm and Hybrid Models

AUTHORS: A. S. Temür, M. Akgün, and G. Temur

ABSTRACT**:** Having forecast of real estate sales done correctly is very important for balancing supply and demand in the housing market. However, it is very difficult for housing companies

or real estate professionals to determine how many houses they will sell next year. Although this does not mean that a prediction plan cannot be created, the studies conducted both in Turkey and different countries about the housing sector are focused more on estimating housing prices. Especially the developing technological advances allow making estimations in many areas. That is why the purpose of this study is both to provide guiding information to the companies in the sector and to contribute to the literature. In this study, a 124-month data set belonging to the 2008 (1)–2018 (4) period has been taken into account for total housing sales in Turkey. In order to estimate the time series of sales, ARIMA (Auto Regressive Integrated Moving Average as linear model), LSTM (Long Short-Term Memory as nonlinear model) has been used. As to increase the estimation, a HYBRID (LSTM and ARIMA) model created has been used in the application. When MAPE (Mean Absolute Percentage Error) and MSE (Mean Squared Error) values obtained from each of these methods were compared, the best performance with the lowest error rate proved to be the HYBRID model, and the fact that all the application models have very close results shows the success of predictability. This is an indication that our study will contribute significantly to the literature.

2.2 Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia

AUTHORS: T. D. Phan

ABSTRACT: House price forecasting is an important topic of real estate. The literature attempts to derive useful knowledge from historical data of property markets. Machine

learning techniques are applied to analyse historical property transactions in Australia to discover useful models for house buyers and sellers. Revealed is the high discrepancy between house prices in the most expensive and most affordable suburbs in the city of Melbourne. Moreover, experiments demonstrate that the combination of Stepwise and Support Vector Machine that is based on mean squared error measurement is a competitive approach.

2.3 Forecasting house price index of China using dendritic neuron model AUTHORS: Y. Y. S. Song, T. Zhou, H. Yachi, and S. Gao

ABSTRACT: The result of Chinese housing market continues to prosper or not is related to the development of China, and further it also has an impact on the world finance. Thus forecasting the house price index is very important and challenging. In this paper we propose an unsupervised learnable neuron model (DNM) by including the nonlinear interactions between excitation and inhibition on dendrites. We use DNM to fit the House Price Index (HPI) data and then forecast the trends of Chinese housing market. To verify the effectiveness of the DNM, we use a traditional statistical model (i.e., the exponential smoothing (ES) model) to make a performance comparison. Three quantitative statistical metrics including normalized mean square error, absolute percentage of error, and correlation coefficient are used to evaluate the forecasting performance of the two models. Experimental results demonstrate that the proposed DNM is better than ES in all of the three quantitative statistical metrics.

3.PROPOSED SYSTEM

House price prediction can be done by using a multiple prediction models (Machine Learning Model) such as Random forest and Xgboost. There are many benefits that home buyers, property investors, and house builders can reap from the house-price model. This model will provide a lot of information and knowledge to home buyers, property investors and house builders, such as the valuation of house prices in the present market, which will help them determine house prices. Meanwhile, this model can help potential buyers decide the characteristics of a house they want according to their budget. Previous studies focused on analyzing the attributes that affect house price and predicting house price based on the model of machine learning separately. However, this article combines such a both predicting house price and attributes together.

3.1 IMPLEMENTAION

1. Upload Dataset
2. Data Preprocessing
3. Feature extraction
4. Model Generation
5. Build Random Forest classifier
6. Build XGBoost classifier

7. Predict Accuracy

3.1.1 ALGORITHM:

RANDOM FOREST ALGORITHM:

an ensemble algorithm which means internally it will use multiple classifier algorithms to build accurate classifier model. Internally this algorithm will use decision tree algorithm to generate it train model for classification.

GRADIENT BOOST

Gradient boosting was created by in 1999 and is a commonly used machine learning algorithm because of its performance, consistency and interpretability. Gradient boosting delivers state-of-the-art in various machine learning activities, such as multistage classification, click prediction and ranking. With the advent of big data in recent years, gradient boosting faces new challenges, especially with regard to the balance between accuracy and performance. There are few parameters for gradient boosting. To ensure a dynamic balance between fit and regularity.

The following steps can be taken to select parameters:

(1) Setting regularization parameters (lambda, alpha)

(2) reducing learning rate and decide those optimal parameters again.

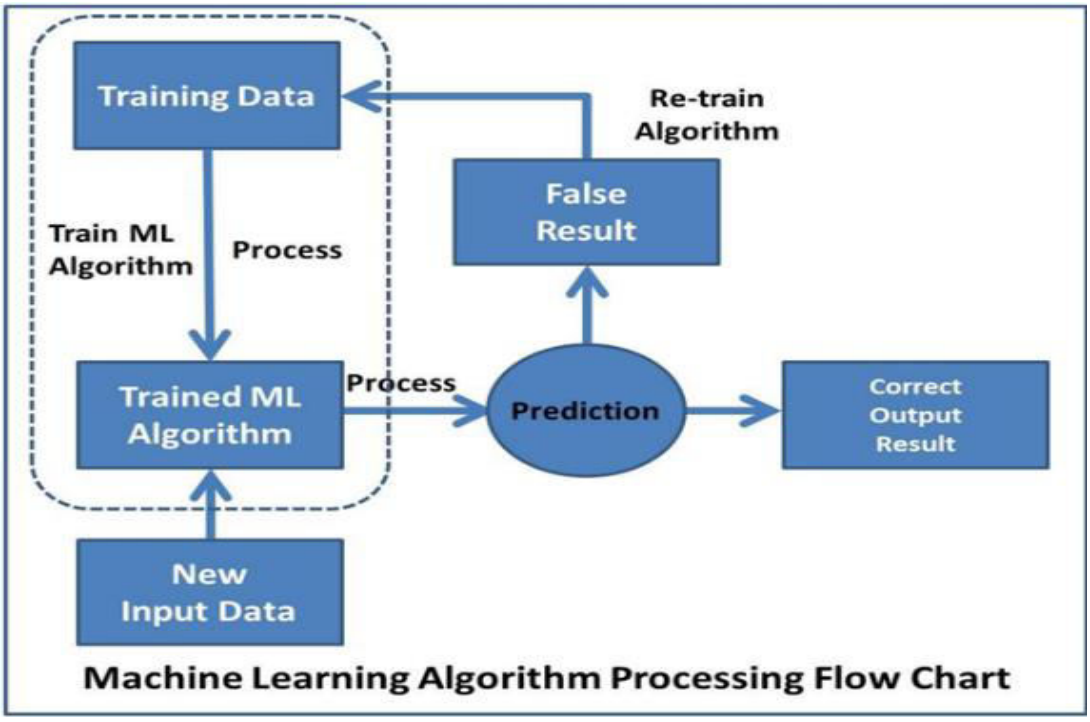
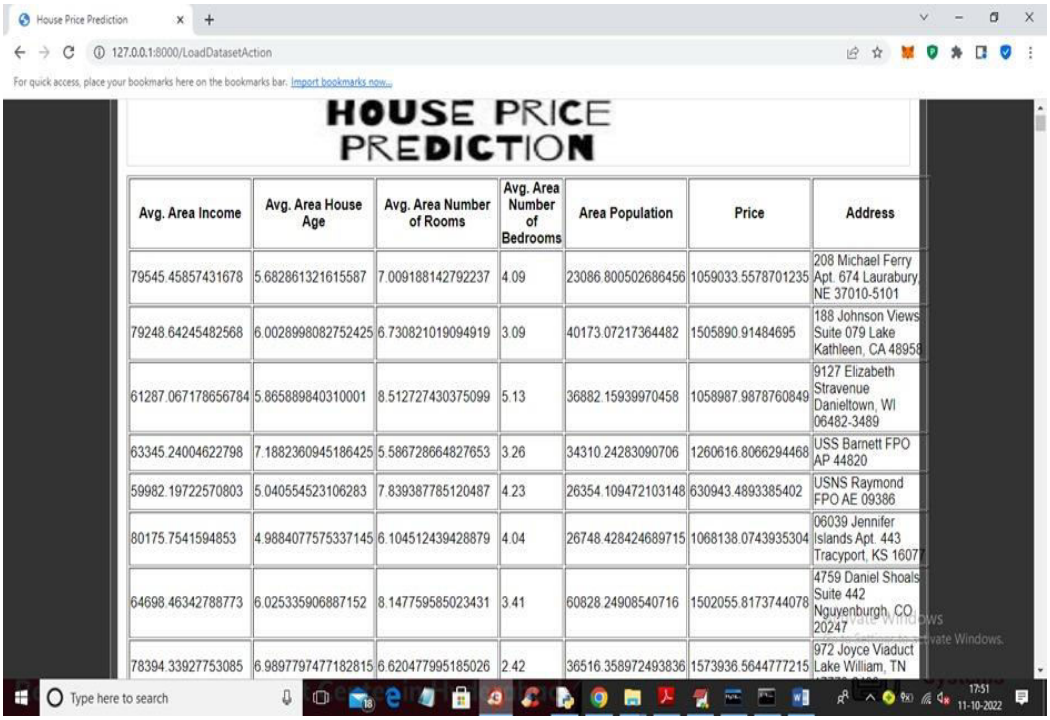
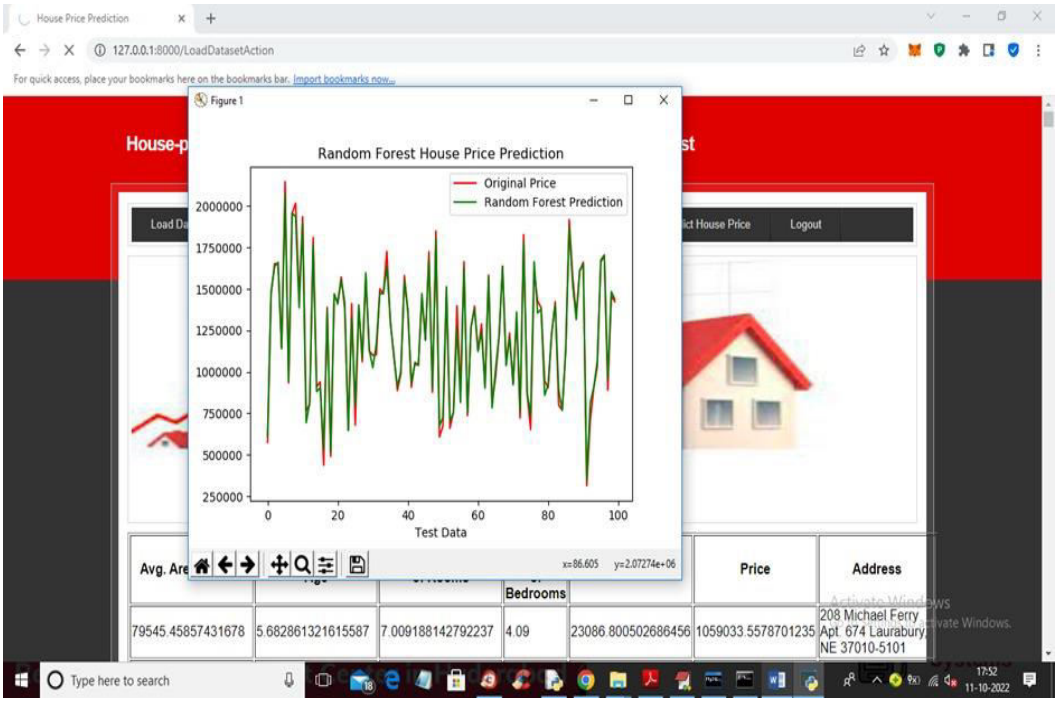


Fig 1:Architecture

4.RESULTS AND DISCUSSION

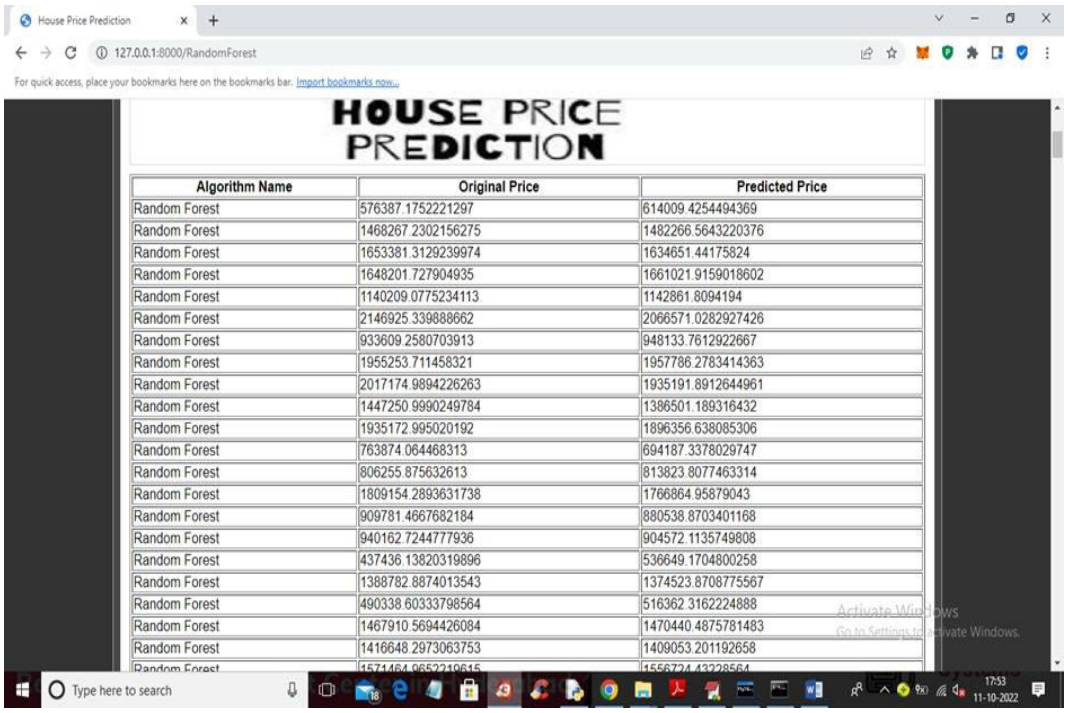


In above screen dataset loaded and now click on 'Train Random Forest' link to train Random Forest and get below prediction output

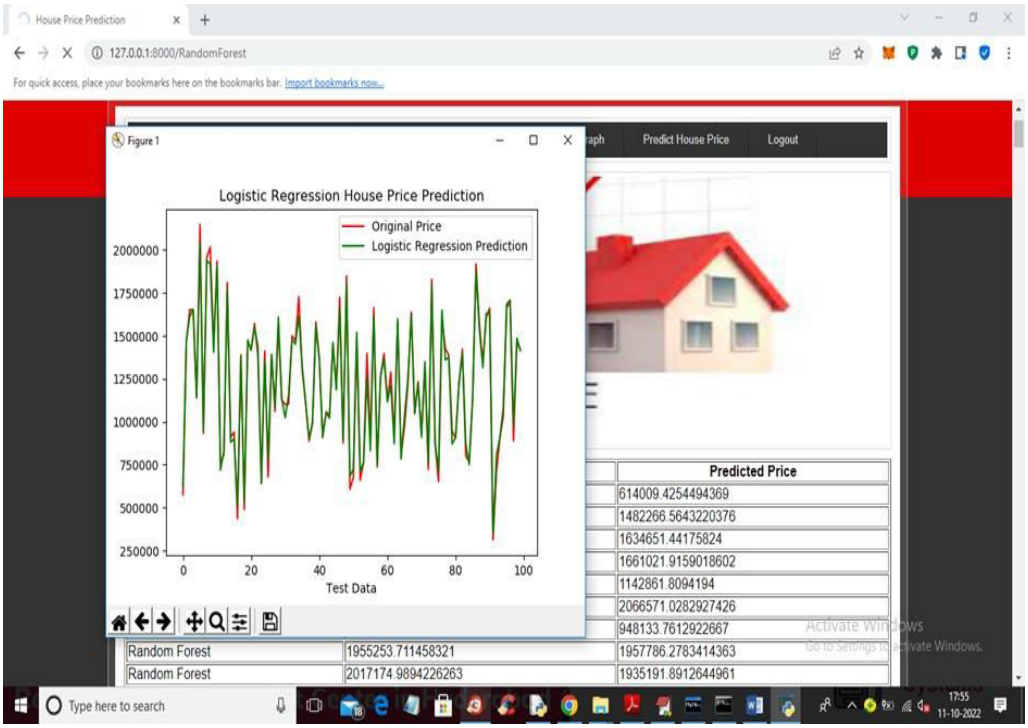


In above graph x-axis represents number of days and y-axis represents Housing price and red line represents Original House Test prices and green line represents Predicted prices and we can see both lines are fully overlapping so there is so much closeness between test data

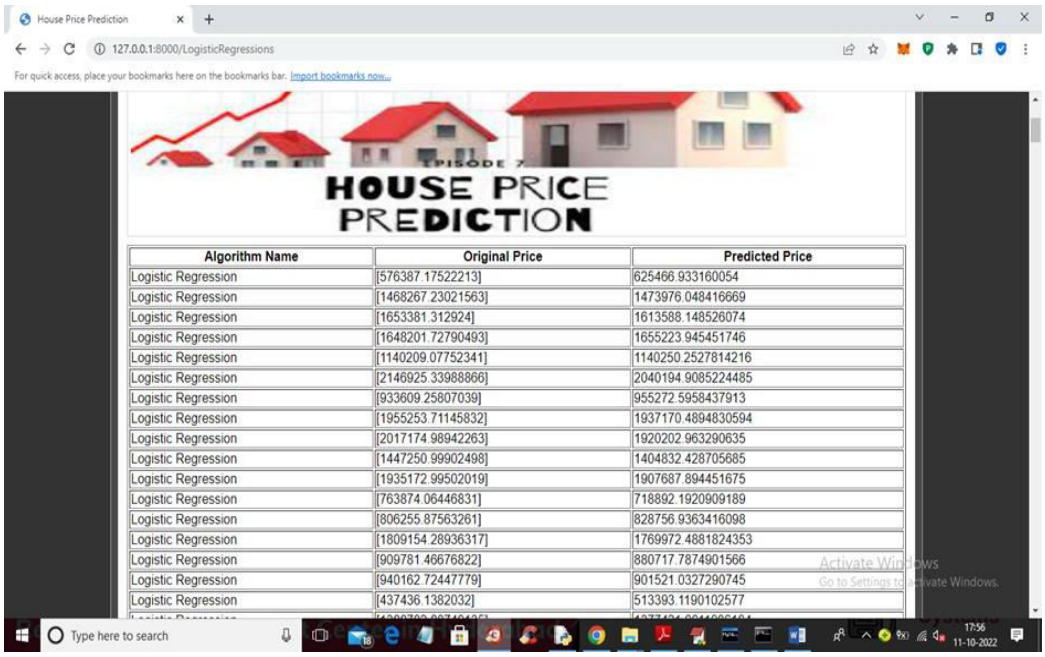
and predicted prices and we can say Random Forest is accurate in House prices prediction and now close above graph to get below page



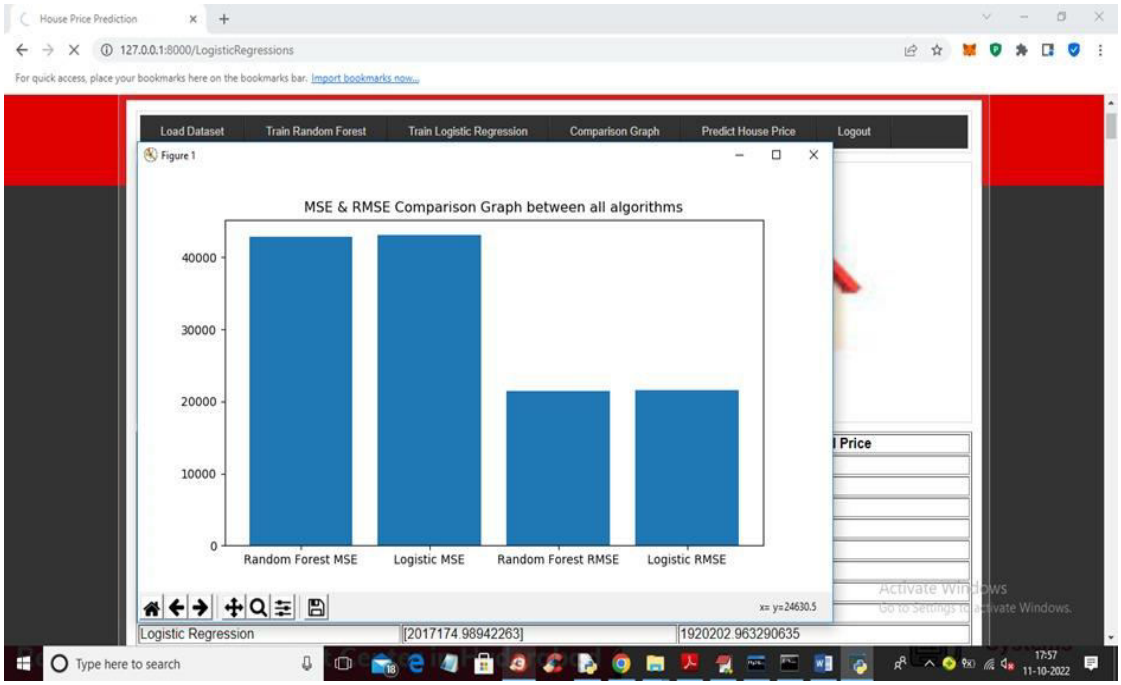
In above screen first column contains Algorithm name and second column contains Original Test data house prices and third column contains Random Forest Predicted prices. Now click on 'Train Logistic Regression' link to train regression and get below page



In above logistic regression house price prediction there is some gap between red and green line so Logistic Regression prediction is not accurate and now close above graph to get below price prediction screen

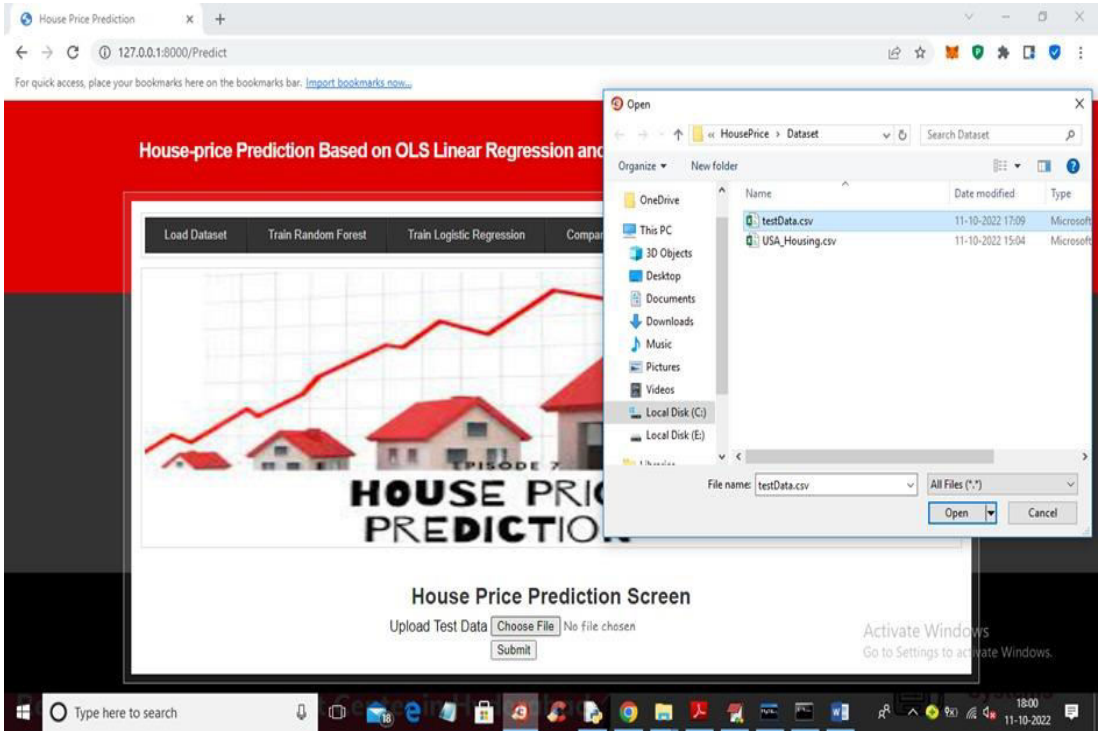


In above screen we can see Logistic Regression Predicted prices and its test prices and now click on ‘Graph’ link to get below comparison graph

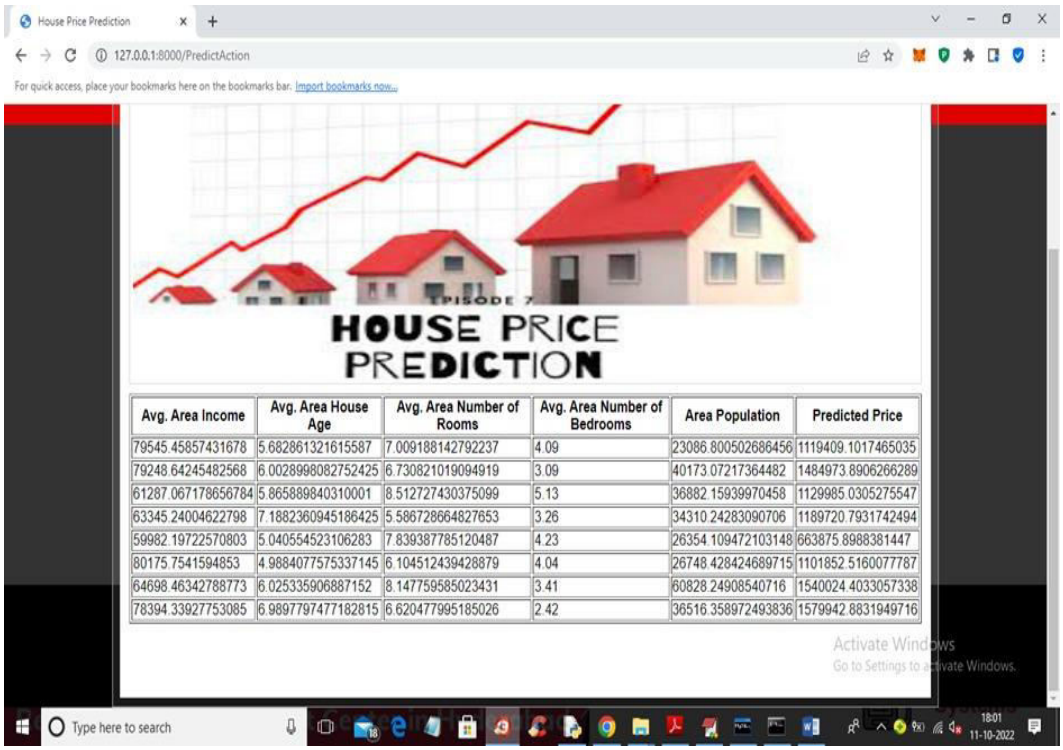


In above screen x-axis represents algorithms names and y-axis represents MSE and RMSE and in both algorithms Random forest got little less MSE and RMSE compare to Logistic

Regression and now close above graph and then click on ‘Predict House Prices’ link to get below page



In above screen I am selecting and uploading ‘testData.csv’ file and then click on ‘Open’ button and then click on ‘Submit’ button to get below prediction output



Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Predicted Price
79545.45857431678	5.682861321615587	7.009188142792237	4.09	23086.800502686456	1119409.1017465035
79248.64245482568	6.0028998082752425	6.730821019094919	3.09	40173.07217364482	1484973.8906266289
61287.067178656784	5.865889840310001	8.512727430375099	5.13	36882.15939970458	1129985.0305275547
63345.24004622798	7.1882360945186425	5.586728664827653	3.26	34310.2428309070 6	1189720.7931742494
59982.19722570803	5.040554523106283	7.839387785120487	4.23	26354.109472103148	663875.8988381447
80175.7541594853	4.9884077575337145	6.104512439428879	4.04	26748.428424689715	1101852.5160077787
64698.46342788773	6.025335906887152	8.147759585023431	3.41	60828.24908540716	1540024.4033057338
78394.33927753085	6.9897797477182815	6.620477995185026	2.42	36516.358972493836	1579942.8831949716

In above screen we can see test data in all columns and last column contains predicted prices.

Similarly you can add new test data in 'testData.csv' file and then upload and predict prices.

5.CONCLUSION

This paper examined and analyzed the current research on the significant attributes of house price and analyzed the data mining techniques used to predict house price. Technically, houses with a strategic location such as the accessibility to shopping mall or other facilities tend to be more expensive than houses in rural areas with limited numbers of facilities. The accurate prediction model would allow investors or house buyers to determine the realistic price of a house as well as the house developers to decide the affordable house price. This paper addressed the attributes used by previous researchers to forecast a house price using various prediction models. Taken together, the results of the survey have shown the potential of Random forest and XGBoost in predicting house prices. These models were developed based on several input attributes and they work significantly positive with house price. In conclusion, the impact of this research was intended to help and assist other researchers in developing a real model which can easily and accurately predict house prices. Further work on a real model needs to be done with the utilization of our findings to confirm them.

REFERENCES

[1] A. S. Temür, M. Akgün, and G. Temür, "Predicting Housing Sales in Turkey Using Arima, Lstm and Hybrid Models," J. Bus. Econ. Manag., vol. 20, no. 5, pp. 920–938, 2019, doi: 10.3846/jbem.2019.10190.

[2] A. Ebekozien, A. R. Abdul-Aziz, and M. Jaafar, "Housing finance inaccessibility for low-income earners in Malaysia: Factors and solutions," Habitat Int., vol. 87, no. April, pp. 27–35, 2019, doi: 10.1016/j.habitatint.2019.03.009.

[3] A. Jafari and R. Akhavian, "Driving forces for the US residential housing price: a predictive analysis," Built Environ. Proj. Asset Manag., vol. 9, no. 4, pp. 515–529, 2019, doi: 10.1108/BEPAM-07-2018-0100.

[4] Choong Wei Cheng, "Statistical Analysis of Housing Prices in Petaling," Universiti Tunku Abdul Rahman, 2018.

[5] 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," 2017

Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2017, vol. 2018-Janua, pp. 351–358, 2018, doi: 10.1109/ICACSIS.2017.8355058.

[6] G. Gao et al., "Location-Centered House Price Prediction: A Multi-Task Learning Approach," pp. 1–14, 2019, [Online]. Available: http://arxiv.org/abs/1901.01774.

[7] T. D. Phan, "Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia," Proc. - Int. Conf. Mach. Learn. Data Eng. iCMLDE 2018, pp. 8–13, 2019, doi: 10.1109/iCMLDE.2018.00017.

[8] Y. Y. S. Song, T. Zhou, H. Yachi, and S. Gao, "Forecasting house price index of China using dendritic neuron model," PIC 2016 - Proc. 2016 IEEE Int. Conf. Prog. Informatics Comput., pp. 37–41, 2017, doi: 10.1109/PIC.2016.7949463.

[9] R. Aswin Rahadi, S. K. Wiryono, D. P. Koesrindartoto, and I. B. Syamwil, "Factors Affecting Housing Products Price in Jakarta Metropolitan Region," Int. J. Prop. Sci., vol. 6, no. 1, pp. 1–21, 2016, doi: 10.22452/ijps.vol6no1.2.

[10] A. Nur, R. Ema, H. Taufiq, and W. Firdaus, "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study : Malang, East Java, Indonesia," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 10, pp. 323–326, 2017, doi: 10.14569/ijacsa.2017.081042.

[11] A. Yusof and S. Ismail, "Multiple Regressions in Analysing House Price Variations," Commun. IBIMA, vol. 2012, pp. 1–9, 2012, doi: 10.5171/2012.383101.

[12] A. Osmadi, E. M. Kamal, H. Hassan, and H. A. Fattah, "Exploring the elements of housing price in Malaysia," Asian Soc. Sci., vol. 11, no. 24, pp. 26–38, 2015, doi: 10.5539/ass.v11n24p26.

[13] T. L. Chin and K. W. Chau, "A critical review of literature on the hedonic price model," Int. J. Hous. Sci. Its Appl., vol. 27, no. 2, pp. 145–165, 2003.

[14] M. J. Ball, "Recent Empirical Work on the Determinants of Relative House Prices," Urban Stud., vol. 10, no. 2, pp. 213–233, 1973, doi: 10.1080/00420987320080311.

[15] M. Rodriguez, "Managing Corporate Real Estate: Evidence from the Capital Markets." Journal of Real Estate Literature, 1996.

AUTHOR PROFILES

Mr.DADI.VENKATA VARAPRASAD currently he is working Associate professor in Audisankara college of Engineering and Technology Gudur (M),Tirupati (Dt) he is done MCA from paavai Engineering college Nammakal,chennai in 2009,M.Tech from Nova Institue of Technology, Tangellamudi,kakinada in 2015,pursuing ph.d at KL University.

ANAMALA PRAVALIKA is pursuing MCA from Audisankara college of Engineering and Technology Gudur, Affiliated to JNTUA in 2024,Andhrapradesh, India.