BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 431-438
DOI:10.2478/bjlp-2022-004046

Real Estate Search and Valuation to get best valued sites using Linear Regression Algorithm and Compared with Random Forest Algorithm.

Chekuri.Hemanth
Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode - 602 105.

S.Ashok Kumar
Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode - 602 105.

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: The main aim of this study is to forecast Real Estate Prices using Novel Linear Regression and Random Forest algorithms. Accordingly, A prediction model based on Novel Linear Regression is proposed for prediction of Housing Price as well as in the process of features selection. There are certain factors that influence the price of the houses which includes location, Conditions, Age, square area etc. Compared with other methods, our work can obtain better performance through different experiments. **Materials and Methods:** Linear Regression and Random Forest algorithms are used to predict the Real Estate Prices. Sample size is calculated using G power calculator and found to be 25 per group has been taken and a total of 50 samples are used. Pretest power is 80% and CI of 95% and the significance value is ($p<0.05$). **Results:** Based on the study Linear Regression has significantly more accuracy (85%) compared with Random Forest algorithm (78%) and the significance value $p=0.01$ ($p<0.05$). It shows that there is a significant difference between two groups. **Conclusion:** According to this study Linear Regression has better accuracy than the Random Forest algorithm to predict the Real Estate Prices**.**

Keywords
Real Estate, Price Prediction, Novel Linear Regression, Random Forest, Housing Price, Machine Learning.

INTRODUCTION

Shelter is one of the essential needs of life. So, housing is a prime factor in human resource development of any economy. At one point of life, Everyone has to deal with housing, which is the major investments of life. People have the fortune to buy their dream house. Study of Real Estate Price Prediction is felt critical to help the choices in urban arranging. According to many surveys eight of ten households have their own house. Housing Price Prediction (Tan and Chou 2018). It is because most of the householders are in rural areas, the reason is high Prices of Property in urban areas & the non-deterministic nature of house prices. There are different algorithms that can be utilized for the future esteems. In

this Study, we use the Novel Linear Regression algorithm which predicts the property prices. House price prediction modeling using Machine Learning (Thamarai et al. 2020). Novel Linear Regression predicts the exact numerical target value unlike other models that can only classify the output. Machine Learning models to predict house prices based on home features (Pandiri 2017). Which plays a strong role in predicting the price value of Real Estate property. This is compared with the Random Forest algorithm to improve accuracy. Housing Price prediction using machine learning algorithms: The case of Melbourne, Australia (Phan and The Danh Phan 2018). House price prediction using multiple Linear Regression (Phan and the Danh Phan 2018; Kaushal and Shankar, n.d.).
There are 3000 research articles published based on Real Estate Price prediction based on Linear Regression and K-nearest neighbor algorithm in Science Direct and also 325 research articles in Google Scholar and 22 Research articles were published in IEEE Xplore for house price prediction. The research gap between existing one and for this is four years. The research has been developed using Novel Linear Regression algorithms to improve accuracy. The existing research uses the Random Forest algorithm which has the accuracy value around 85.0. This research is developed with minimum experience gained through learning. The aim of the study is to improve accuracy of the proposed algorithm when compared to the existing algorithm Random Forest, which is the algorithm used for existing research with an average accuracy rate (Zheng 2017). Real Estate Price estimation in french using Machine Learning (Tchuente and Nyawa 2021). Default prediction of Real Estate Properties using Machine Learning (Cowden, Fabozzi, and Nazemi 2019). Housing Prices prediction using Linear Regression (T and Mrs. 2019). Applications of this algorithm are in prediction, Forecasting, Error detection.
Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).This research work has certain client based limitations such as limitations of clients should be user occurrence and less accuracy of client reviews. And may be used through clusters that should be imported through customer allegiance through a website. So that clients may use different alleys for the high quality of nature. The aim of the study is to predict the Real Estate Market that may occur for development of web usage of the clients.

MATERIALS AND METHODS

This research work is done in the Department of Computer Science and Engineering, Saveetha School of Engineering (SSE), Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai. In this study two sample groups were taken. Group 1 was Novel Linear Regression algorithm and group 2 was the Random Forest algorithm. Sample size is calculated using Gpower , consider the pre-test power to be 80%. This is mainly dependent on two algorithms, which have the sample sizes of Linear Regression (258) and Random Forest (258)which is a total of 516. The work has been carried out with 3000 records which is taken from the kaggle dataset.The accuracy is predicted using two different groups. Here the data is from the kaggle website House Price Prediction (ammar 2019).
The model is tested on the setup with hardware requirements i7 processor, 16GB RAM and 256 SSD by using Hp laptop. The software configuration is windows 10.The tool which is used to execute the process is jupyter notebook version 6. Algorithm is implemented using the python3 code and accuracy of both groups is determined based on the dataset.

Linear Regression

Linear Regression is used to predict the value of a variable based on another variable. The variable which we want to predict is the value id dependent variable. The variable used to predict other variables' value is the independent variable. Housing Prices prediction using Linear Regression (T and Mrs. 2019). Linear Regression has an equation $Y=a+bX$. Algorithm used in this study was Novel Linear Regression which plays a major role in predicting Real Estate Prices. The algorithm is calculated based on certain parameters such

as MSE (Mean squared error) $MSE=1/n\sum^{n}_{t=1}e^{2}_{t}$, RMSE (Root mean squared error) and MAE (Mean absolute error) $MSE=1/n\sum^{n}_{t=1|et|}$.

Algorithm Steps
Step 1: First initialize the parameters.
Step 2: Predict the value of a dependent variable by giving an independent variable.
Step 3: Calculate the error in prediction for all data points.
Step 4: Calculate the partial derivative.
Step 5: Calculate the sum for each number.

Random Forest Algorithm
The existing algorithm compared with the Random Forest algorithm. Random Forest comes under the type of supervised Machine Learning algorithm. It is easy, and flexible to use, which produces greater and better accuracy. (Adetunji et al. 2022). Random Forest has its diversity to use both in classification and regression problems. Model trees and decision trees are subsets of Random Forest which is formed by growing trees determined on a random vector which has symbol (theta), The training set is independently drawn from distribution of random vector X and Y.The MSE is $EX,Y(Y-h(X))^2$.

Algorithm Steps
Step 1:Random Forest has N number of random records that are taken from the dataset having k no.of records.
Step 2: Individually decision trees are constructed for each sample.
Step 3: Each decision tree will generate an output.
Step 4: Final output is considered based on the averaging for classification.
The tool used to execute the process in jupyter notebook version 6. Algorithm is implemented using python code and accuracy of both groups is determined based on the dataset.

Statistical Analysis
The analysis is done using IBM SPSS software. Independent sample t test is carried out for the analysis. Independent variables are the dataset and dependent variables are accuracy. The independent sample t-test analysis is carried out in this research work. Independent variables are PID, House style, Street, Lot area, Number of bedrooms, Garage, Year Built, Sale price. House price prediction using Random Forest Machine Learning Technique (Adetunji et al. 2022). The Statistical analysis of Experimental Data (Mandel 2012).

RESULTS

In this study, Machine Learning algorithms are used for prediction of Housing Price, as everything related to the Real Estatemarket. We test for the performance of these algorithms how precisely a technique can predict the prices. Two algorithms are selected and tested for which algorithm produces the highest rate of accuracy. Table 1 explains the pseudocode of the Novel Linear Regression algorithm. Table 2 explains the pseudocode of the Random Forest algorithm. Table 3 represents the dataset with attributes.
Table 4 explains the group statistics of the algorithm by comparing the algorithm and accuracy using sample values of =50 for Linear Regression and also 50 sample values for Random Forest, Mean=85.00 for Linear Regression and Mean=78.00 for Random Forest. std.Deviation=7.360 for both models.
Table 5 explains about the independent variables, which defines the equal variances assumed and equality of means with significance value p=0.01 ($p<0.05$) for both assumed and non assumed variances and mean differences=7.000 for both assumed and non assumed variances and 95% of confidential value respectively.
Figure 1 represents the comparison of mean accuracy between proposed and the existing algorithm. The accuracy of the proposed algorithm is found to be 85% and the proposed

algorithm gives better results compared to the existing algorithm which has accuracy of 78%.

DISCUSSION

The data evolution was performed using IBM SPSS software version 21. To analyze data for performing independent sample t-test and group statistics be carried out. Which represents the comparison of two algorithms with their accuracy percentages of 85% for Novel Linear Regression and 78% for Random Forest.
There are many studies similar to this study of proposed research where the findings are, House price prediction using Machine Learning algorithms (Zhou 2020). Predicting log error for House price using Machine Learning (Kanani 2019). Spatial analysis with applications on Real Estate market price prediction (Zheng 2017). Machine Learning models to predict house prices based on home features (Pandiri 2017). House price prediction using Linear Regression (T and Mrs. 2019). Understanding the Real Estate Price prediction using Machine Learning(Tripathi 2021). (Singeisen 2019). (Kolbe et al. 2021), (Cesaroni et al. 2020), (Wang and Wolverton 2002; Mooya 2016). From the above results it is concluded that Linear Regression has more accuracy. So the clients may search for the high quality in finding the websites.
Main limitation is the assumption of linearity between the dependent and independent variables. Assumes that there is a straight line relationship between dependent and independent variables which is incorrect many times. Non linearity of prediction relationships. The future scope of this study explains how it is useful for the clients with improved Accuracy. Feature selection techniques are used in this algorithm. To simplify the model. To get the best priced houses. The feature selection algorithm can be used to reduce the computation time and improve the classification accuracy of classifiers.

CONCLUSION

In the proposed work, accuracy percentage of the Novel Linear Regression algorithm is improved to 85% compared to Random Forest having 78%. It proves that Novel Linear Regression is an efficient algorithm compared to Random Forest. Independent sample T-test result is done with confidence interval as 95% and significance level as 0.01 (Linear Regression appears to perform significantly better than Random Forest with the value of $p<0.05$).

DECLARATION

Conflicts of Interest

No conflicts of interest in this manuscript.

Authors Contribution

Author CH was involved in data collection and analysis. Author SAK was involved in the action process, data verification and validation process.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS) for providing necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support to complete this study.

1. Oracle Tech Solutions Pvt.Ltd.
2. Saveetha Institute of Medical and Technical Sciences (SIMATS).
3. Saveetha University.
4. Saveetha School of Engineering.

REFERENCES

Adetunji, Abigail Bola, Oluwatobi Noah Akande, Funmilola Alaba Ajala, Ololade Oyewo, Yetunde Faith Akande, and Gbenle Oluwadara. 2022. "House Price Prediction Using

Random Forest Machine Learning Technique." *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2022.01.100.

ammar. 2019. "House Price Prediction." Kaggle. March 10, 2019. https://kaggle.com/ammar111/house-price-prediction-an-end-to-end-ml-project.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

Cesaroni, Giulia, Giorgia Venturini, Lorenzo Paglione, Laura Angelici, Chiara Sorge, Claudia Marino, Marina Davoli, and Nerina Agabiti. 2020. "[Mortality inequalities in Rome: the role of individual education and neighbourhood real estate market]." *Epidemiologia e prevenzione* 44 (5-6 Suppl 1): 31–37.

Cowden, Chad, Frank J. Fabozzi, and Abdolreza Nazemi. 2019. "Default Prediction of Commercial Real Estate Properties Using Machine Learning Techniques." *The Journal of Portfolio Management*. https://doi.org/10.3905/jpm.2019.1.104.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

Kanani, Swapnilkumar. 2019. *Predicting Log Error for House Price Using Machine Learning*.

Kaushal, Anirudh, and Achyut Shankar. n.d. "House Price Prediction Using Multiple Linear Regression." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3833734.

Kolbe, Jens, Rainer Schulz, Martin Wersing, and Axel Werwatz. 2021. "Real Estate Listings and Their Usefulness for Hedonic Regressions." *Empirical Economics*, January, 1–31.

Mandel, John. 2012. *The Statistical Analysis of Experimental Data*. Courier Corporation.

Mooya, Manya M. 2016. *Real Estate Valuation Theory: A Critical Appraisal*. Springer.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Pandiri, Venkat Shiva. 2017. *Machine Learning Models to Predict House Prices Based on Home Features*.

Phan, The Danh, and The Danh Phan. 2018. "Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia." *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*. https://doi.org/10.1109/icmlde.2018.00017.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.

Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.

Singeisen, Julien. 2019. *Real Estate Valuation with Hedonic Regression Models*.

Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.

Tan, Wei Peng, and Tsung-Nan Chou. 2018. *Housing Price Prediction*.

Tchuente, Dieudonné, and Serge Nyawa. 2021. "Real Estate Price Estimation in French Cities Using Geocoding and Machine Learning." *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-03932-5.

Thamarai, M., Sri Vasavi Engineering College, Andhra Pradesh, India, and S. P. Malarvizhi. 2020. "House Price Prediction Modeling Using Machine Learning." *International Journal of Information Engineering and Electronic Business*.

https://doi.org/10.5815/ijieeb.2020.02.03.

T, Mrs Dhikhi, and Dhikhi T. Mrs. 2019. "Housing Prices Prediction Using Linear Regression." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2019.4622.

Tripathi, Elika. 2021. "Understanding Real Estate Price Prediction Using Machine Learning." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2021.33720.

Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.

Wang, Ko, and Marvin L. Wolverton. 2002. *Real Estate Valuation Theory*. Springer Science & Business Media.

Zheng, Yujing. 2017. *Spatial Analysis with Applications on Real Estate Market Price Prediction*.

Zhou, Yichen. 2020. *Housing Sale Price Prediction Using Machine Learning Algorithms*.

Tables and Figures

Table 1. Pseudocode for Linear Regression Algorithm.

Input:
Start
Read Number of Data (n)

For i=1 to n:
  Read Xi and Yi
  Next i

Initialize:
  sumX = 0
  sumX2 = 0
  sumY = 0
  sumXY = 0

Calculate Required Sum
  For i=1 to n:
  sumX = sumX + Xi
  sumX2 = sumX2 + Xi * Xi
  sumY = sumY + Yi
  sumXY = sumXY + Xi * Yi
  Next i

Calculate Required Constant a and b of y = a + bx:
  b = (n * sumXY - sumX * sumY)/(n*sumX2 - sumX * sumX)
  a = (sumY - b*sumX)/n

Display value of a and b **Stop**

Table 2. Pseudocode for Random Forest Algorithm.

Input: Test data.
Predict and store the outcome of each randomly created decision tree on the given test data**.** Compute the total votes for the individual class.

Declare the majority class as the final outcome class.
Output: Final predict class.

Table 3. Data description used for comparison of the existing and the proposed algorithms. Represents the sample dataset calculating the proposed algorithm to get the accuracy value.

S.No	Attribute	Value	Description
1.	No. of observation	Integer	The number of data used in the system.
2.	Co-ordinates	Integer	The x and y axis coordinates of the eye.

Table 4. Statistical analysis of mean, standard deviation and standard error of Accuracy for Novel Linear Regression and Random Forest algorithms. There is a statistically significant difference between the groups. Novel Linear Regression has a higher mean (85) than the Random Forest (78). The group statistics of the algorithm by comparing the algorithm and accuracy using sample values of 50 for Linear Regression and also 50 sample values for randomforest, Mean=85.00 for Linear Regression and Mean=78.00 for Random Forest, Standard.Deviation=7.360 for both models.

Algorithm		N	Mean	Std. deviation	Std. Error Mean
Accuracy	**LR**	25	85.00	7.360	1.472
	RF	25	78.00	7.360	1.472

Table 5. The significance value p=0.01 (p<0.05) shows that two groups are statistically significant..

	Levene's test for equality of variables		T-test for Equality of Mean						
	F	Sig	t	df	Sig(2-tailed)	Mean difference	Std.Error difference	95% confidence interval of the difference	
								Lower	Upper
Accuracy									

Equal variances assumed	.000	0.01	3.363	48	.002	7.000	2.082	2.815	11.185
Equal variances not assumed			3.363	48.000	.002	7.000	2.082	2.815	11.185

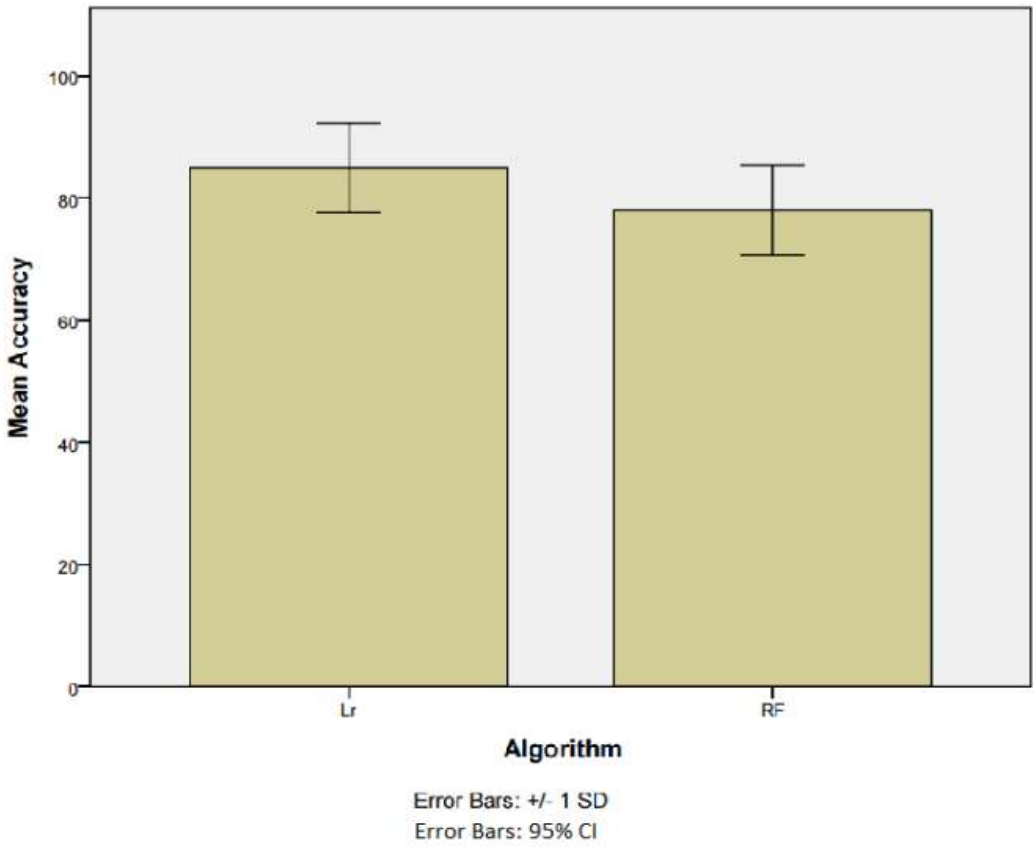

Fig. 1. Bar chart representation of the comparison of mean accuracy of the proposed and the existing algorithm. In this fig Y-Axis represents Mean Accuracy and X-Axis represents the Algorithms. The accuracy of the prediction of the proposed algorithm is found to be 85% and the proposed algorithm gives better results compared to the existing algorithm that has accuracy of 78% the mean accuracy detection is ±1SD.