iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.56232

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

House Price Prediction: Comparative Analysis of Regression-Based Machine Learning Algorithms

David Emmanuel Aniobi[1], Chukwuemeka Oluebube Ochuba[2], Saater Benedicta Nguideen[3]
Department of Computer Science, Joseph Sarwuan Tarka University, Makurdi, Nigeria

Abstract: Advancement in technology has revolutionized the ways of doing things in contemporary time. One of such is Artificial intelligence which has given birth to myriads of methods/techniques employed in solving real life problems. Many Machine learning techniques can be used in predicting house prices with several factors in consideration. House prices rise on annual basis, triggering the need for house price prediction models. Predictive models enable families to acquire a house of their choice when accurately developed. There have been a significant number of articles that adopt traditional machine learning algorithms to successfully estimate house prices, but they rarely compare the performance of individual models. This study will extensively test and compare numerous machine learning technique and present an optimistic model that will be used in developing a house pricing prediction system. Several models were developed and compared using Linear Regression (LR), Least Absolute Shrinkage & Selection Operator Regression (LASSO-R), Ridge Regression (RR), K-Nearest Neighbours Regression (KNN-R), Decision Tree Regression (DTR) and Extra Trees Regression (ETR) algorithms. Implemented using Python programming language. Amongst them ETR outperformed the others with MSE (16233.4), RMSE (128.7), MAE (49.6) and R^2 (0.63) while the least performed is KNN-R with MSE (45763.3), RMSE (213.9), MAE (99.2) and R^2 (-0.04).

Keywords: LASSO Regression, House Price Prediction, Machine Learning, Ridge Regression, Extra Tree Regression

I. INTRODUCTION

Houses are very essential to humans as it provides shelter for man to comfortable live in his environment; it is not necessarily meant for luxury. The price of this property has to be ascertained as it has become a necessity in life. House price prediction tools has become a subject of interest among real estate professionals, data scientists, and homeowners [1]. While some argue that predicting house prices using machine learning algorithms is a valuable tool for property buyers and sellers, others question its ethical concerns [2].

However, the relevance of house price prediction extends beyond real estate transactions. It has the potentials to reshape the real estate market and town planning and offers a wide range of economic benefits [3]. Economically, accurate prediction of house prices can help developers, policymakers, and property investors in making informed decisions. Societally, it can help with resource allocation, urban planning, and concerns related to housing affordability [4].

While house price prediction offers numerous benefits, it also raises ethical questions. Concerns about fairness, discrimination, and the potential for market manipulation should not be underestimated. As machine learning models are trained on historical data, they may inadvertently perpetuate biases present in past housing practices. The need for efficient house pricing system gives rise to House price prediction tools, driven by machine learning algorithms [5] These advanced tools can inadvertently exacerbate existing inequalities and disparities within the real estate market. There are concerns that these tools may might perpetuate biases and discrimination in housing transactions, thereby resulting in inequitable pricing and unequal access to housing opportunities [6] These unnecessary biases can be reduced if not eliminated by using the optimal algorithm in developing the system.

Although, house price prediction using machine learning holds the potential for transformative insights in the real estate market, it also introduces a host of ethical and fairness concerns that demand thoughtful consideration and proactive solutions to ensure they are employed in a just, transparent, and responsible manner for the benefit of all stakeholders. This will ensure that certain neighborhoods are not favored over others. Therefore, it is essential to ensure that predictive algorithms are designed with fairness and transparency in mind. This paper focuses on comparing the performance of six different machine learning regression-based algorithms and the use of the optimal algorithm in developing a house price prediction model.

II. RELATED LITERATURE

The price of a house either for short-term or long-term based has to be ascertained as it has become a necessity in life.

This transformation is not limited to prospective homebuyers alone but extends to the entire spectrum of individuals involved in the real estate industry, including sellers, investors, and the broader community. The real value of a house is influenced by several factors, including the number of rooms, location (with rural areas often having lower costs than cities), proximity to amenities such as highways, malls, supermarkets, and job opportunities, and access to quality educational facilities. Furthermore, the size of the property, its condition, the standing of the location and the overall economic climate all play vital factors in deciding real estate values [7]. The combination of these factors and local market dynamics shapes the price of a house in any region.

House prediction model based on regression analysis and particle swarm optimization (PSO) was proposed by [8]. Reference [9] proposed a hybrid LASS0 and Gradient boosting regression model that promises better prediction. LASSO was used in feature selection and they did many iterations of feature engineering to find the optimal number of features that will improve the prediction performance. Reference [5] suggested that using a mix of models is necessary. Their work proved that a linear model tends to have high bias, leading to underfitting while a high model complexity-based model tends to have high variance, resulting in overfitting. Balancing these two approaches is crucial to achieve optimal model performance leading to a fairer assessment of land prices and potentially increased revenue for the government. Comparison of artificial neural network and multiple linear regression for house price prediction was presented by [10]). In their study, the impact of different morphological measures on live weight were modelled by artificial neural networks and multiple linear regression analysis. Genetic algorithms were employed by [11] to determine parameters of machine learning models. Pragmatic results revealed that attribute selection for machine learning models in this study does improve performances forecasting models in forecasting accuracy.

Reference [12] predicted the sale price of the houses using various machine learning algorithms like, Random Forest, XGBoost, LightGBM, Hybrid Regression and Stacked Generalization and compared the accuracy. They found out that each model has its advantages and limitations. The Random Forest method has the lowest error on the training set but is prone to be overfitting and its time complexity is high. XGBoost and LightGBM has the best time. Hybrid Regression performs better due to the generalization. Stacked Generalization Regression is best when accuracy is a top priority but has a complicated architecture and worst time complexity. Random Forest Regression, Decision Tree Regression, Ridge Regression, LASSO Regression, Ada-Boost Regression, XGBoost Regression Algorithms were compared by [13] in predicting house prices. Scores and Root Mean Square Error (RMSE) were used to evaluate and it was found out that the Decision Tree Regression algorithm has the highest RMSE.

Reference [14] utilized Random Forest (RF) algorithm for predicting house prices in London. Despite having a small dataset size, the study proved that RF outperformed the traditional regression approach based on Generalized Linear Models (GLMs) in terms of prediction improvement. Their findings suggested that RF was able to capture complex relationships and patterns in the data more effectively than the GLM which could explain its superior performance in this study. Reference [15] predicted property values using different algorithms like Support Vector Regression (SVR), Decision Tree, Regression-Particle Swarm Optimization (R-PSO), and LUCE. The findings implied that LUCE provided a more effective and reliable solution for estimating property values, especially in situations where there is a lack of recent sold prices and sparse house data. Reference [16] used 18-year of housing property data to train models with utilising stochastic gradient descent-based support vector regression, random forest and gradient boosting machine. They demonstrated that advanced machine learning algorithms can achieve very accurate prediction of property prices, as evaluated by the performance metrics.

Reference [17] compared the performance of several machine learning algorithms, including XGBoost, Random Forest, Decision tree, and Linear Regression, to determine the most suitable algorithm for developing automated house purchasing system. XGBoost outperformed the other algorithms because of its ability to handle complex relationships between features and target variables. Reference [18] presented a machine learning approach in prediction and analysis of House Price. It employed the use of Linear regression., multiple linear regression, LASSO, and gradient boosting techniques in predicting house prices. Reference [19] also proposed a house price prediction system using machine learning algorithms (i.e. linear regression, decision tree regression, random forest regression, and artificial neural networks) and visualization to make accurate predictions. The results showed that the system can predict house prices with great accuracy.

Random Forest based model House Price Prediction was developed by [20] using datasets from UCI machine learning repository Boston. They opined that housing prices are closely correlated with factors such as city, population and location, etc. and predicting individual housing prices needs information other than House Price Index (HPI). A regression based predictive system for house and rent price was presented by [21] opined that housing datasets often exhibit missing values, outliers, and inconsistent formats, which can hinder the performance of prediction models. Their system tackled these challenges and provided a more accurate and reliable predictions for real estate. Even though a lot of works have been done on house price prediction using several algorithms, yet no single work has a combination of all the algorithms we a proposing in this paper.

III. METHODOLOGY

A. Data Preprocessing

The dataset used was obtained from the online machine learning repository of Kaggle. It contains 13,321 data with 18 attributes. The dataset was first preprocessed which involve transforming raw data into an understandable format and check if there are missing values, outliers and noise. The preprocessing was done to remove rows or columns that have missing values due to mistakes that might have occurred when entering the data into the CSV file. This is important as it helps prevent some runtime errors like Not a Number (NaN) error that could prevent the system from optimal performance.

The dataset was checked for outliers using Inter-Quartile Range (IQR). An outlier y can be detected if:

$$Q1 - 1.5(IQR) > y \quad or \quad y > Q3 + 1.5(IQR) \qquad (1)$$

where: Q1 = 25th centiles, Q3 = 75th centiles and IQR = Q3 − Q1

After applying Equation (1) to every column of the dataset, the final dataset contained 6,877 data. The dataset was then normalized which involved rescaling real valued numeric attributes into the range 0 and 1. Table I includes details of each attribute.

TABLE I. LIST OF ATTRIBUTES

S/N	Attribute name	Data type	Description
1	location	string	House Location
2	total_sqft	flaot	Size of the land where the system is built
3	LandSlope	float	Slope of property
4	Housestyle	string	Style of dwelling
5	Age	int	How long the house has been in existence
6	RoofStyle	string	Type of roof
7	RoofMatl	string	Roof material
8	Exterior1	string	Exterior covering on house
9	Watersrc	string	The source of water in the house
10	Electricity	boolean	Availability of Electricity
11	bath	float	Number of Bathrooms in the House
12	bhk	int	Number of Bedrooms in the House
13	Garage	boolean	Presence of a Garage
14	Fence	boolean	Availability of a fence
15	price	float	Price of the House
16	area_type	string	Development level of the area
17	Exterior2	string	Exterior covering of the building if more than one

B. Model Selection

Before building the models, the data was processed so that the models could learn the patterns more efficiently. Then, the dataset was split into training set and test set with a ratio of 4:1 by utilizing the scikit-learn package (Pedregosa et al., 2011). The algorithms used were namely; Linear Regression (LR), Least Absolute Shrinkage & Selection Operator Regression (LASSO-R), Ridge Regression (RR), K-Nearest Neighbors Regression (KNN-R), Decision Tree Regression (DTR) and Extra Trees Regression (ETR).

1) *Linear Regression:* is a supervised machine learning model that aims to establish a linear relationship between dependent variables (Y) and independent variables (X). The mathematical representation of the Linear Regression is:

$Y= a_0 + a_1 X + \varepsilon$

Where Y = Dependent Variable

X = Independent Variable

a_0 = Intercept of the line that offers additional Degree of Freedom

a_1 = Linear regression coefficient, which is a scale factor to ever input value

ε = Random Error

The model calculates the best fit line by estimating the intercept (a_0) and linear regression coefficient (a_1) in an attempt to minimize the difference between the predicted values and the actual data points [22]. By finding the optimal values for a_0 and a_1, the Linear Regression model becomes more accurate in predicting the response variable based on the input variables. In this paper, we utilized Linear Regression from scikit-learn [23].

2) *LASSO Regression:* is a variant of linear regression that extends the model by introducing L1 regularization to prevent overfitting by selecting a subset of relevant features and drives the coefficients of less important features towards zero. The mathematical representation of Lasso Regression is as follows:

$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n + \varepsilon$

Where:

Y = Dependent Variable

$X_1, X_2, \dots X_n$ = Independent Variable

a_0 = Intercept of the line that offers additional Degree of Freedom

$a_1, a_2, \dots a_n$ = Linear regression coefficient, which is a scale factor for each input value

ε = Random Error

Lasso Regression is particularly useful when dealing with high-dimensional datasets or datasets where some features may not contribute significantly to the prediction (Melkumova & Shatskikh, 2017). Lasso Regression from the scikit-learn library was used to implement [23].

3) *Ridge Regression:* Ridge regression is similar to LASSO regression. Unlike Lasso Regression, Ridge Regression does not drive coefficients to exactly zero but instead shrinks them toward zero. It prevents overfitting by reducing the influence of individual features and promoting a more stable and well-behaved mode. Ridge Regression is valuable when dealing with datasets where all input features are considered relevant but might be prone to multicollinearity or overfitting [24].
4) *Decision Tree Regression:* It is an extension of the Decision Tree algorithm that partitions the feature attributes into segments based on feature values, allowing it to capture complex relationships that may not be linear. It makes predictions by traversing the tree from the root node to a leaf node, where each leaf node corresponds to a specific prediction. The splitting process occurs recursively based on features' values, optimizing to minimize the difference between predicted and actual values. Decision Tree Regression is suitable for both simple and complex datasets [25]. It excels at capturing non-linear patterns but is prone to overfitting.
5) *Extra Trees Regression (short for "Extremely Randomized Trees" regression):* it is an ensemble learning technique that builds on the foundation of decision trees by introducing randomness during the tree construction process. Unlike traditional decision trees, where the best split is determined through optimization, Extra Trees randomly selects splits, leading to a set of diverse trees. This diversity enhances the model's generalization and robustness, making it less prone to overfitting. Extra Trees Regression is particularly useful when dealing with noisy or high-dimensional datasets [6].
6) *K-Nearest Neighbors:* is a machine learning algorithm that operates under the assumption that similar data points in the feature space tend to belong to the same class. K-NN is particularly useful when dealing with classification tasks and datasets where the decision boundary is not easily defined by simple mathematical functions. By relying on similarity between data points, K-NN can make effective predictions [10].

C. Proposed System Architecture

System architecture is the conceptual model that defines the structure of a system. The architecture of the system is shown in the Figure 1.

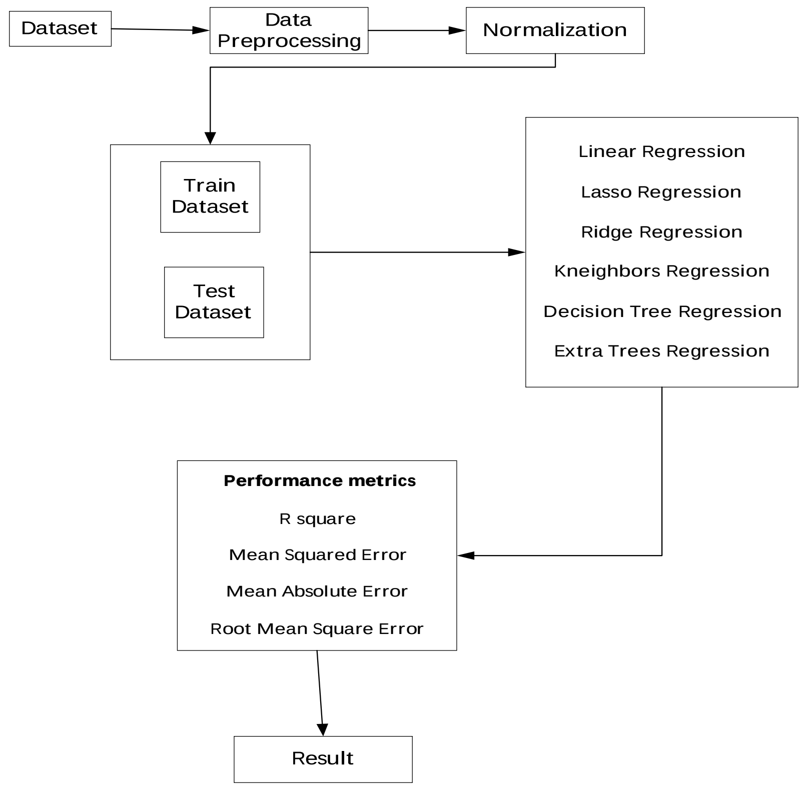

Fig. 1 Architecture of the system

The architecture showed that the dataset was first pre-processed before normalization. The normalized data was split and passed into the model. Finally, each model performance was evaluated.

D. Use Case Diagram

The use case diagram for the system is shown in Figure 2.

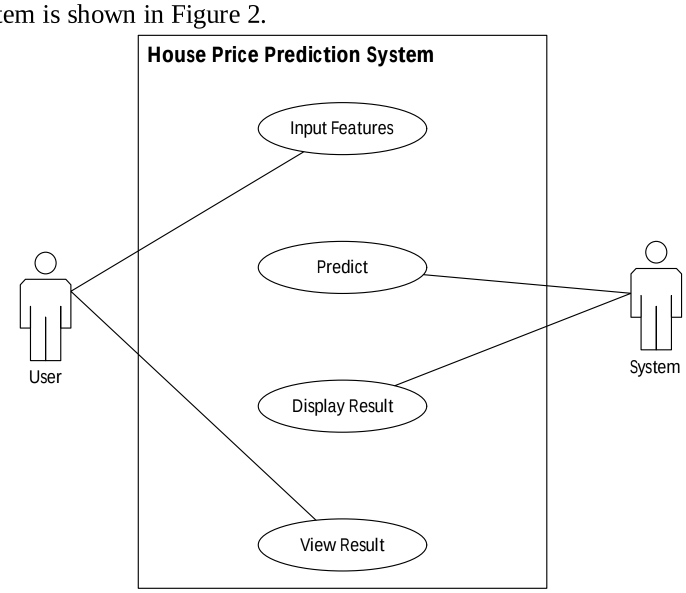

Fig. 2 Proposed system Use case diagram.

IV. RESULTS AND DISCUSSION

The models were evaluated using the following.

1) *R-squared (R^2) Value:* It assesses how well the model fits the data, with higher values indicating a better fit. It is denoted as;

$$R^2 = 1 - \frac{SSR}{SST}$$

Where: SSR (Sum of Squared Residuals) measures the unexplained variance by the model.
(SST) (Total Sum of Squares) represents the total variance in the dependent variable. Table 2 shows the result for R^2 value for all the models

Table II R^2 Result

Algorithms	R^2
LR	-7.08
LASSO-R	0.43
RR	0.42
ETR	0.63
DTR	0.52
KNN-R	-0.04

2) *Root Mean Squared Error (RMSE):* It quantifies how well a regression model's predictions match the actual data points. Smaller RMSE values indicate a better fit of the model to the data. It is denoted as;

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y_i})^2}$$

Where: n = the number of data points.
y_i = represents the actual values.
$\widehat{y_i}$ = represents the predicted values. Table III shows the result for RMSE value for all the models

Table III RMSE Result

Algorithms	RMSE
LR	558e91.5
LASSO-R	158.69
RR	159.27
ETR	128.71
DTR	136.44
KNN-R	213.92

3) *Mean Absolute Error (MAE):* It calculates the average absolute difference between the predicted and actual values. Lower MAE values indicate better model accuracy. It is denoted as; $\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y_i}|^2$

Table IV shows the result for MAE value for all the models

Table IV MAE Result

Algorithms	MAE
LR	210e7.07
LASSO-R	57.41
RR	59.12
ETR	49.61
DTR	57.53
KNN-R	99.22

4) *Mean Squared Error (MSE):* It measures the average squared difference between the predicted and actual values. Like RMSE, smaller MSE values indicate a better fit of the model to the data. It is denoted as; $\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y_i})^2$

Table V shows the result for MSE value for all the models

Table V. MSE Result

Algorithms	MSE
LR	3.12
LASSO-R	25181.60
RR	25365.49
ETR	16233.39
DTR	21083.89
KNN-R	45763.30

The summary of results obtained from the models' performance are compared as shown in Table VI.

Table VI Models' performances Compared

Algorithms	MSE	RMSE	MAE	R^2
LR	3.12	558e91.5	210e7.07	-7.08
LASSO-R	25181.60	158.69	57.41	0.43
RR	25365.49	159.27	59.12	0.42
ETR	16233.39	128.71	49.61	0.63
DTR	21083.89	136.44	57.53	0.52
KNN-R	45763.30	213.92	99.22	-0.04

The result showed that ETR outperformed others techniques with MSE (16233.4), RMSE (128.7), MAE (49.6) and R^2 (0.63) while the least performed is KNN-R with MSE (45763.3), RMSE (213.9), MAE (99.2) and R^2 (-0.04).
Sample interface for the house price prediction model is shown in Figure 3.

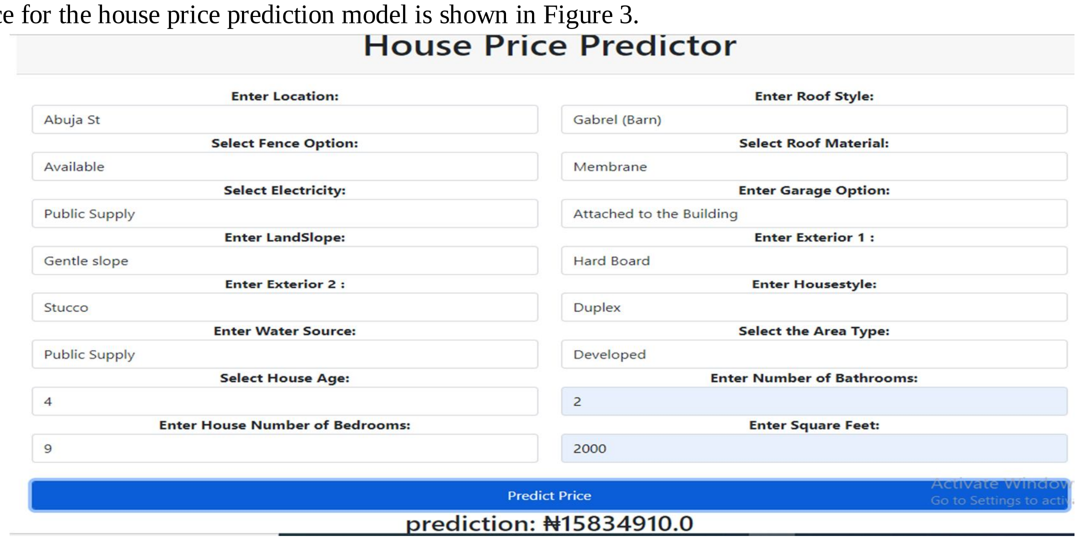

Fig. 3 Proposed system Interface

Several factors/attributes like house type, number of bedrooms, age of house, etc. relating to a particular house consequently affects the prediction of the model.

V. CONCLUSION

This paper investigates different models for house price prediction. Six different types of Machine Learning regression-based methods including are Linear Regression, LASSO Regression, Ridge Regression, K-Nearest Neighbors Regression, Decision Tree Regression and Extra Trees Regression. The models achieved desirable results but each have their pros and cons. Linear Regression offers simplicity but may not capture complex nonlinear relationships in the data. LASSO Regression introduces L1 regularization, which can lead to feature selection and a more parsimonious model. It helps prevent overfitting, but it might result in some coefficients being exactly zero, effectively eliminating certain features from consideration.

Ridge Regression employs L2 regularization to control the model's complexity. It provides a balance between feature selection and model complexity, which can be beneficial when dealing with high-dimensional datasets. K-NN captures complex patterns but sensitive to the choice of the number of neighbors (K) and distance metrics. Decision Tree can handle nonlinear relationships but it is prone to overfitting, especially when the tree depth is not appropriately controlled. Extra Trees Regression is less prone to overfitting compared to traditional decision trees. Comparatively the performance of extra tree regression is found to be better than the rest in predicting the house prices. In future the dataset can be prepared with more features and advanced machine learning techniques can be for constructing the house price prediction model.

REFERENCES

[1] D. Flagella, "what is machine learning?," 2019, [Online]. Available: https://emerj.com/ai-glosary-terms/what-is-machine-learning/

[2] A. Aladangady, "Housing Wealth and Consumption: Evidence from Geographically Linked Microdata.," Am Econ Rev, vol. 107, no. 11, pp. 3415–3446, 2017, [Online]. Available: ttp://www.jstor.org/stable/44871793

[3] K. Ngiam and I. Khor, "Big data and machine learning algorithms for health-care delivery," Lancet Oncol, vol. 20, pp. e262–e273, May 2019, doi: 10.1016/S1470-2045(19)30149-4.

[4] P. Harrington, Machine Learning in Action. MANNING Publication Co., Shelter Island, 2012. doi: 10.4018/978-1-4666-0059-1.ch008.

[5] A. Babu and A. S. Chandran, "Literature Review on Real Estate Value Prediction Using Machine Learning," International Journal of Computer Science and Mobile Applications, vol. 7, pp. 8–15, 2019, [Online]. Available: www.ijcsma.com

[6] E. K. Ampomah, Z. Qin, and G. Nyame, "Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement," Information (Switzerland), vol. 11, no. 6, 2020, doi: 10.3390/info11060332

[7] R. Kanilelori, "Different Types Of Houses In Nigeria (With Pictures)." Accessed: Oct. 16, 2023. [Online]. Available: https://klrealtors.ng/different-types-of-houses-in-nigeria-with-pictures/

[8] N. A. Alfiyatin, R. E. Febrita, H. Taufiq, and F. M. Wayan, "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study : Malang, East Java, Indonesia," International Journal of Advanced Computer Science and Applications, vol. 8, no. 10, pp. 6–10, 2017, doi: 10.14569/ijacsa.2017.081042.

[9] S. Lu, Z. Li, Z. Qin, X. Yang, and R. S. M. Goh, "A hybrid regression technique for house prices prediction," in 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2017, pp. 319–323. doi: 10.1109/IEEM.2017.8289904.

[10] A. Ahmad and A. Nawar, "House Price Prediction," KRISTIANSTA UNIVERSITY SWEDEN, 2020. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:hkr:diva-20945

[11] P. F. Pai and W. C. Wang, "Using machine learning models and actual transaction data for predicting real estate prices," Applied Sciences (Switzerland), vol. 10, no. 17, pp. 1–11, 2020, doi: 10.3390/app10175832.

[12] T. Quang, N. Minh, D. Hy, and M. Bo, "Housing Price Prediction via Improved Machine Learning Techniques," Procedia Comput Sci, vol. 174, no. 2019, pp. 433–442, 2020, doi: 10.1016/j.procs.2020.06.111.

[13] B. Sivasankar, A. P. Ashok, G. Madhu, and F. S., "House Price Prediction," International Journal of Computer Sciences and Engineering, vol. 8, no. 7, pp. 762–767, 2020, doi: 10.37896/ymer21.05/87.

[14] S. Levantesi and G. Piscopo, "The importance of economic variables on London real estate market: A random forest approach," Risks, vol. 8, no. 4, pp. 1–17, 2020, doi: 10.3390/risks8040112.

[15] H. Peng et al., "Lifelong Property Price Prediction: A Case Study for the Toronto Real Estate Market," IEEE Trans Knowl Data Eng, vol. 35, no. 3, pp. 2765–2780, 2023, doi: 10.1109/TKDE.2021.3112749.

[16] W. K. O. Ho, B. S. Tang, and S. W. Wong, "Predicting property prices with machine learning algorithms," Journal of Property Research, vol. 38, no. 1, pp. 48–70, 2021, doi: 10.1080/09599916.2020.1832558.

[17] S. Dabreo, S. Rodrigues, V. Rodrigues, and P. Shah, "Real Estate Price Prediction," International Journal of Engineering Research & Technology (IJERT), vol. 10, no. 04, pp. 541–543, 2021, doi: IJERTV10IS040322.

[18] M. Kandasamy, R. Shanmugam, A. Dave, C. Chawda, K. Shah, and U. Seladiya, "Prediction and Analysis of House Price Through Machine Learning Approach," International Journal for Multidisciplinary Research (IJFMR), vol. 5, no. 4, pp. 1–7, 2023, doi: E-ISSN: 2582-2160.

[19] M. S. Supriya, G. S. Vinayak, V. R. Patgar, and V. Mahajan, "House Price Prediction System using Machine Learning Algorithms and Visualization," in 2023 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2023, pp. 1–6. doi: 10.1109/CONECCT57959.2023.10234749.

[20] B. A. Adetunji, A. O. Noah, and A. F. Ajala, "House Price Prediction using Random Forest Machine Learning Technique," ScienceDirect, Procedia Computer Science, vol. 199, pp. 806–813, 2022, doi: 10.1016/j.procs.2022.01.100.

[21] S. Khandaskar, C. Panjwani, V. Patil, D. Fernandes, and P. Bajaj, "House and Rent Price Prediction System using Regression," in 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), 2023, pp. 1733–1739. doi: 10.1109/ICSCSS57650.2023.10169290.

[22] K. Khushbu and Y. Suniti, "Linear regression analysis study," Journal of the Practice of Cardiovascular Sciences, vol. 4, no. 1, p. 33, 2018, doi: 10.4103/jpcs.jpcs_8_18.

[23] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[24] L. E. Melkumova and S. Y. Shatskikh, "Comparing Ridge and LASSO estimators for data analysis," Procedia Eng, vol. 201, pp. 746–755, 2017, doi: 10.1016/j.proeng.2017.09.615.

[25] K. G. Burcu and D. K. Ipek, "Regression Analyses or Decision Trees ?," Manisa Celal Bayar University Journal of Social Sciences, vol. 18, no. 4, pp. 251–260, 2020, doi: 10.18026/cbayarsos.