International Journal of Advanced Research and Interdisciplinary Scientific Endeavours, Vol. 1(3), 2024
DOI: 10.61359/11.2206-2414

The Machine Learning Based Regression Models Analysis for House Price Prediction

Ritesh Tandon *

Email Correspondence*: ritesh.tandon@gmail.com

* Independent Researcher, India

Abstract:

House Price Prediction focuses on the development of methods that use ML algorithms to accurately predict house prices. Machine learning has been expanding at a faster rate this decade. There are a lot of new algorithms and uses for Machine Learning every day. Among these applications, one may find journal articles on home price prediction. The need to model home price prediction has arisen due to the fact that property prices are growing annually. Customers may utilise these built models to choose the perfect home for their needs. This study explores the feasibility in predicting house prices concerning the 2019 data obtained from the property listings in Kuala Lumpur by using machine learning algorithms on the gathered data set. Data cleaning, data transformation, and data feature selection were performed sequentially on the dataset to get a clean dataset ready for training the models. These include RF, SVM, LightGBM, and GBR models to assess their prediction model. Hence, the Gradient Boosting Regressor model proved to be the best model since it produced the lowest MSE, RMSE, and the highest R2 score compared to the other models. The experimental results show the GBR model gets a higher 90% accuracy for house price prediction in terms of other models. The findings of this study can be used to inform stakeholders in the real estate, urban planning and investment markets regarding the need for high-quality data combined with suitable algorithms to assist in real estate price prediction.

Keywords: Regression Models, GBR, House Prediction, Correlation Score, Machine Learning.

1. Introduction

The real estate market has grown rapidly in our nation in recent years, and many individuals are putting their money into it because of the potential for large profit gains. Similarly, the condition of real estate development in a certain area may be inferred from investment in real estate. If investment is brisk, for instance, it might indicate imbalanced development and a lack of supply. An indication of a steady real estate market in recent times is a lack of investment activity. Additionally, it reflects a city's development [1][2]. It does not, however, follow that more investment would result in better real estate development or greater benefits for investors. Negative conditions might arise, and some locations have suffered as a result. Furthermore, real estate investment (which has failed miserably in many areas due to a lack of demand) is a waste of money and reckless with regard to people's lives.

Unpaid labour will also sometimes arise, leading to social unrest [3][4]. It is crucial to maintain control over real estate investments in order to avoid these kinds of circumstances. The government should actively develop realistic controls to manage the urbanization process and guarantee that the percentage of

[*] Independent Researcher, India.

Article Timeline: Received: Jul 30, 2024; Revised: Aug 28, 2024; Published: Aug 30, 2024

investments in fixed assets stays at about 25%. This restriction is also not set in stone. Additionally, it is dependent on how the city develops [5][6].

This ratio may be suitably lowered after the first phase of urban growth since there is now no demand for as many homes. Trends in housing prices not only affect buyers and sellers but also provide insight into the state of the economy [7][8][9]. The number of bedrooms and bathrooms are just two of many variables that affect home values. A house's value may be influenced by its location, particularly if it's close to major thoroughfares, schools, retail centres, and places of employment in the area. Many methods have been created for the purpose of predicting home prices since manual house prediction has become challenging. Precise estimation of property values has always captivated purchasers, vendors, and financial institutions alike. Numerous scholars have already attempted to understand the enigmas surrounding the forecasting of housing values [10][11]. Building these models from historical data and using them to forecast new data is the basic process of ML. Our population is growing at an accelerated rate, which is driving up market demand for homes. Due to a shortage of employment, people are moving to rural areas for financial reasons. Consequently, there is a growing need for housing in cities. People incur financial losses because they are unaware of the true cost of that specific home. Several ML methods are used in this research to anticipate the house's price.[12][13][14]. Numerous problems face an organization when it lacks an accurate sales forecast model. Retailers, distributors, and manufacturers may all benefit from sales forecasts. Long-term projections facilitate business expansion, while short-term forecasts help with inventory control and production scheduling.

In industries where goods have a limited shelf life, sales forecasting is essential to reduce revenue loss during periods of excess and scarcity [15][16][17]. The real estate market in Kuala Lumpur, like many metropolitan areas, is dynamic and influenced by a myriad of factors, including economic conditions, urban development, and demographic trends. For a variety of stakeholders, including prospective purchasers, real estate investors, legislators, and urban planners, an accurate assessment of housing costs is essential. Traditional methods of property valuation often fall short due to their inability to handle large datasets and complex relationships among variables. The advent of machine learning (ML) provides a robust alternative by leveraging data-driven approaches to uncover patterns and making precise predictions. The goal of this project is to employ cutting-edge ML approaches to improve the accuracy and dependability of home price forecasts.

A. Contribution of the Study

In the field of predicting housing prices via the use of regression models, this work provides a number of important advances. It is important to note the main contributions, which are as follows:

- The research meticulously gathers and preprocesses data from many sources, including Kaggle and Google Maps, for Kuala Lumpur property listings. By considering amenity distances, which are crucial for home price prediction, this detailed technique improves the dataset's analytical usefulness.
- The study compares the performance of several ML algorithms, including SVM, LightGBM, and GBR, revealing the superiority of GBR in predicting house prices.
- The research uses correlation scores to determine the most important home price forecast elements. This method helps determine how characteristics affect model performance and directs future feature selection.
- Model performance is assessed using RMSE, MAE, and R2. This detailed review compares model accuracy and predictive capability, with GBR being the most successful.

- The study finds research gaps including the need for more advanced algorithms and localized datasets by comparing models and noting their shortcomings. It stresses the need to use varied data sets and powerful machine learning to enhance home price projections.

The structure of the paper is as follows. Background research on house prediction utilising the house price prediction dataset is presented in Section 2, Materials and methods are presented in Section 3, and suggestions for the comparison findings are covered in Section 4. Section 5 contains the study's conclusion and future directions.

2. Literature Review

This section offers a summary of the literature on the research done on regression analysis based on DL and ML in order to forecast home prices. In this study, Sagala and Cendriawan, (2022) offer a model that might be used to forecast home prices for businesses and help customers make business decisions. The process consisted of seven main steps: modelling, assessment, data standardisation, data cleansing, business understanding, and data understanding.

This research uses data from the Maribel Ajar firm to create and evaluate a linear regression model for predicting housing prices. To achieve this, two pipelines are established on the Azure Platform. Two pipelines: One is for testing, and the other one is for training. That is followed by Power company Intelligence, which helps to visualise the data in the right manner for analysis of company performance. According to the results of the experiment, it got RMSE = 0. 0334 and R coefficient = 0. 7 in the model. There is perhaps a possibility of applying a more refined method of ML to improve the multiple linear regression method employed in this work for estimating and evaluating house prices based on data analysis and testing conducted for this work [18].

This research investigates, Budiman et al., (2023) the problem of attempting to predict property prices based on the mentioned six classification algorithms, namely KNN, RF, GBM, XGBoost, and LR. Consequently, the ROC assessment showed that RF has the best result of 85. 32% for the R2 value and 0.93 for AUC in estimating property prices. This Research shows that RF is comparable to other methods such as GBM, XGBoost, LR, and KNN and that the algorithm is useful in dealing with difficult and curved data sets [19]. In this paper, Ahtesham, Bawany and Fatima, (2020) Housing prices are predicted using the Gradient Boosting Model XGBoost.

A publicly accessible dataset of 38,961 information about Karachi city was obtained from Pakistan's Open Real Estate Portal. Numerous studies have been conducted to forecast home prices in other nations; however, very little research has been done to forecast home prices in Pakistan. With 98% accuracy, our suggested model can forecast home prices [20].

This paper, Thakur et al., (2022) aims to overcome the difficulties by thoroughly comprehending the elements that influence the nightly rates and offering a three-layered DNN solution with two outputs that has a 74.43% accuracy rate for pricing properties [21]. Based on the details the renter supplied about his home, these outputs show the lowest and highest sum he may charge [22].

This paper, Lin et al., (2020) explains how AItalk supports cyber IoT devices, including scikit-learn and Tensor Layer, and expands AItalk for use in non- IoT applications. They illustrate how AItalk may flexibly include the aspects that significantly affect home prices using an example of property appraisal. They demonstrate that an additional 38% may be added to the forecast (value) accuracy by using variables other than house profile characteristics. Additionally, they looked at the distributed AItalk structure's 3.7% communication overhead for valuing a single property [23].

Table-1 Presents comparative research on the prediction of house prices using a regression model and machine learning approaches

Refere nces	Methodology	Dataset	Performance	Limitations and Future Work
[4]	Linear Regression	Maribel Ajar Company data	RMSE = 0.0334, R coefficient = 0.7	Model could be improved with more advanced machine learning Approaches.
[5]	Various ML methods (RF, GBM, XGBoost, LR, KNN)	House price data	Random Forest: R^2 = 85.32%, AUC = 0.93	Other methods also show competitive performance, suggesting potential for further exploration of alternative algorithms.
[6]	XGBoost	Karachi city dataset	Accuracy = 98%	There is a need for further validation and regional comparison notwithstanding the paucity of work done in Pakistan to forecast housing values.
[7]	Three-layered deep neural network	Property rental data	Accuracy = 74.43%	Models can dynamically adjust prices based on various factors. further improvement is needed for higher accuracy.
[8]	AI tools (scikit-learn, Tensor Layer) in AItalk framework	House valuation data	Improved accuracy by 38%	Future work includes investigating communication overhead and extending the model with more features for better performance.

A. Research Gaps

Numerous research gaps are exposed by comparing several housing price forecast models. While more sophisticated methods like Random Forest and XGBoost have demonstrated better performance in managing non-linear and complicated datasets, classic linear regression models only offer mediocre prediction accuracy. Still lacking, though, are thorough studies that use deep learning techniques, which could be able to capture more complex market dynamics. Prediction accuracy has also been demonstrated to be greatly increased by utilising non-traditional characteristics, such as those included by AI technologies in Internet of Things applications. However, there hasn't been much study done on this strategy in the mainstream for predicting home prices. More localised datasets and models are required, as evidenced by the few research projects that concentrate on certain regional markets. To close these gaps and raise the precision and dependability of home price forecasts, enhanced machine learning techniques, the incorporation of varied feature sets, and localised market research are all urgently needed.

3. Methodology

In this section, the methodology that has been employed for the study is described. It is important to note that this section has many subsections. Additionally, it provides a concise summary of the dataset as well as the gathering methods.

A. Research Design

When aiming at creating a strong foundation for your house price prediction model, you started with data collection from Kaggle and Google Maps, with data specifically referring to property listings of Kuala Lumpur from the year 2019. In the preprocessing phase, it was necessary to do table, record, and attribute selection, data cleaning and attribute generation, wherein all these tasks were done in an iterative manner. This process caused the deletion of 21,995 rows with missing values in them and the remaining 31,899 rows were fit for the next analysis. Some of the data transformation processes I performed include normalisation, aggregation and one hot encoding on the dataset. Feature importance was assessed using correlation scores, ranking features based on their absolute correlation with the target variable. The dataset was then split into training and testing subsets to ensure the model's ability to generalise. Finally, multiple machine learning algorithms—such as SVM, LightGBM, and GBR—were employed to develop and compare the predictive models. This whole process is shown in Figure 1.

The following is a detailed description of the stages that make up the data flow chart. Figure 1:

1) Data Collection

The systematic process of collecting information or measurements from a variety of sources in order to build a dataset for analysis is referred to as data collection. The data collection involved using a dataset house price prediction from Kaggle and Google Maps, initially scraped from Kuala Lumpur property listings in 2019.

2) Data Preprocessing

An essential component of text mining methods and applications is the preprocessing approach. The text mining process starts with this phase. The final dataset was created during the data preparation stage by performing a number of operations, often repeatedly and in a random sequence. Among the duties were selecting tables, records, and attributes; cleaning and creating attributes; and transforming data for modelling tools. In this study, the initial dataset of real estate listings underwent cleansing and transformation. It was essential to discover and fix data cleaning problems, such as eliminating missing information. The dataset was enhanced in value for the research by data transformation. With Python, a total of 21,995 rows with missing data were found and eliminated, mostly in more than two columns, leaving 31,899 rows available for examination.

Data Transformation: A database's tables and columns may be arranged with the least amount of redundancy by using data normalisation. Processing time and complexity are reduced as a result.

One Hot Encoding: One Hot Encoding is the process of converting numerical features from categorical variables so that they may be fed into algorithms for Deep Learning and Machine Learning. A binary vector is used to represent a categorical variable, and the length of the vector corresponds to the number of unique categories in the variable.

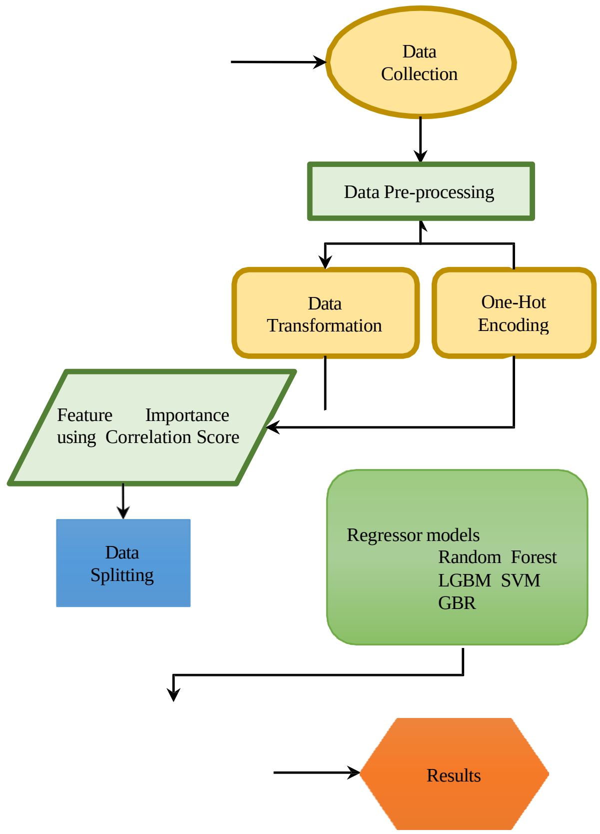

Figure-1 Flowchart of house price prediction

3) Feature Importance using Correlation Score

The process of using correlation scores to order features according to their absolute values involves first calculating a correlation coefficient among every feature and a target variable. Relationship strength is indicated by higher absolute correlation values. Correlations for the target variable may be extracted, sorted, and the correlation matrix can be computed in Python. This is because the absolute value of correlation gives the measure of relevance and the closer it is to unity, the more significant the characteristic is.

4) Data Splitting

Data splitting is the process in which the dataset is divided into different parts for testing and training of the ML model. Often the data set is split into two parts; test data set, which accounts for about 20 percent to test the model and the training data set which accounts for about 80 percent to train the model. This increases the chances of the model's ability to generalise to new unseen data it has not seen before.

5) Classification Models

This section presents different ML algorithms that have been discussed to build house price prediction models. These models are made using the regressor models which include RF, LGBM, SVM, and GBR, and can be used to compare the results on the dataset used to predict the price of a house.

a) Random Forest (RF) Regressor Model

It is a collection of individual distinct decision trees that are combined to act as one model called the random forest. Each individual tree in a random forest returns to the predicted class, and the model selects the most popular one. This means that, where the combined models' elements are disconnected, the ensemble makes more accurate predictions than each coupled model's prediction. Cross-validation produces results with higher accuracy. A RFC can handle missing values while retaining the accuracy of a significant percentage of the data [13][24]. It is capable of managing a sizable data set with a higher degree of dimensionality.

b) Light Gradient Boosting Machine (LightGBM) Regressor Model

Classification and regression are two applications of gradient boosting, a type of ML algorithm. Combining models from various algorithms generates a novel iterative algorithm. Gradient boosting is a common ML algorithm that is widely used for its efficiency and accuracy. AdaBoost, or adaptive boosting, was the first algorithm and method to be created. From then, several algorithms and approaches were developed, including GBM and Model-Based Boosting (MBoost), followed by Cat Boost, XGBoost, and LightGBM [25]. IoT networks comprise of diverse devices with constrained resources. For this reason, any high-performance security solution for IoT networks should use the lightweight regressor method. Algorithm for ML LightGBM [26].

c) Support Vector Machine (SVM) Regressor Model

The primary use of SVM is as a classifier for identifying hyperplanes in an n-dimensional space (the number of features) that may effectively divide objects into distinct categories by optimising the distance among each category's data points and dividing hyperplane [27][28]. The features may be projected onto a higher dimensional plane, or the categorisation can be done inside the same feature space. Implies that because there is a kernel function that allows every data set to be linearly divided, data are easier to separate in higher-dimensional spaces. However, finding the kernel function is usually a matter of trial and error.

d) Gradient Boosting Regressor (GBR) Model

A regression technique similar to boosting is called gradient boosting [29][30]. Finding an approximation for the function $F*(x)$, which minimises the expected value of a specific loss function and thus connects instances x to their matching output values $y, L\,(y, F(x))$, is the aim of boosting gradient on a given training dataset mathematical equation. According to equation (1), GB creates a weighted sum of functions as an additive estimate of $F*(x)$:

$$Fm(X) = Fm - 1(X) + \rho mhm(X) \tag{1}$$

Where p_m is the weight of the mth function, $h_m(X)$. The ensemble uses these functions as its models. Iterated construction of the estimate is used. To begin with, equation (2) provides a constant approximation of $F*(x)$:

$$F_0(x) = argmin \sum_{i=1}^{N} L(y_i, \alpha) \quad (2)$$

The observed value is denoted by yi while the model that results from merging the weak learners is represented by F^(xi). The dependent variable's model predictions are measured using a loss function. Therefore, minimising a specified error metric between anticipated and observed values is the goal of the regressor algorithm [31].

4. Results and Discussion

In this section, description of a dataset, performance metrics, and classifier statistics are included, which offers the outcomes that were acquired via the evaluation of the dataset that was utilized in this research.

A. Dataset Description

A House price prediction dataset from Kaggle and Google Maps, first collected from real estate listings in Kuala Lumpur, Malaysia in 2019, is used in this investigation. Eight variables made up the original dataset: price, location, size, kind of house, rooms, bathrooms, and furnishings. Using the Geocoder and Google Places Python scripts, four more variables were added: the distances to the closest hospital, school, retail Centre, and public transportation. Price in Ringgit Malaysia (MYR) is the goal variable in the final dataset, which has 12 variables and 53,883 observations. After the data was cleaned and transformed for data preparation, 31,899 rows with missing values were removed, allowing the remaining 21,995 rows to be used for analysis:

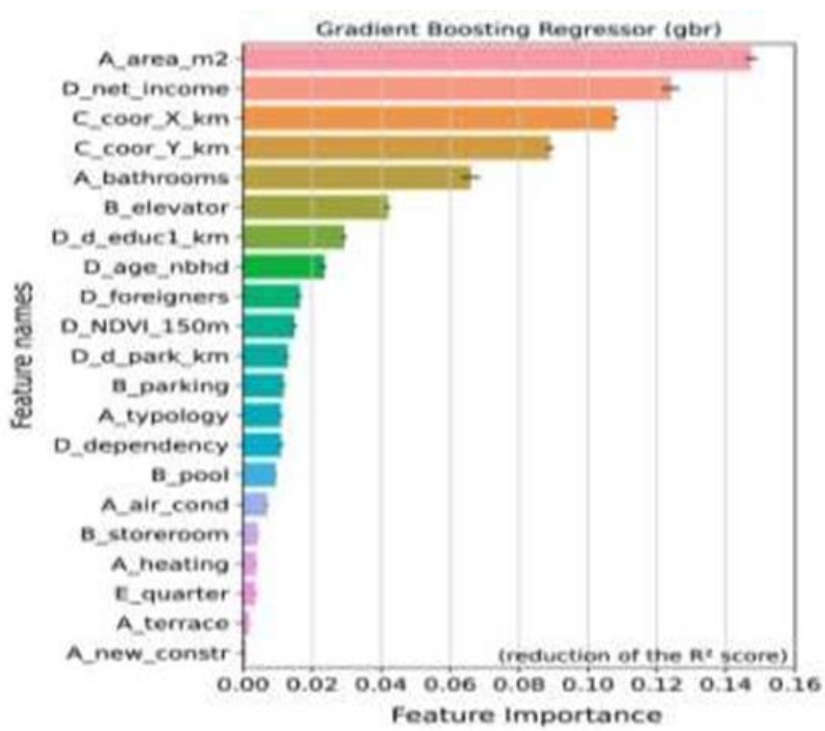

Figure-2 Feature Importance for Property Price Prediction Using Gradient Boosting Regressor

In Figure 2, the bar graph represents a Gradient Boosting Regressor (GBR) to estimate property values, highlighting the significance of each attribute. A_area_m2, D_net_income, and C_coor_X_km are the characteristics that have the most influence, whereas A_new_constr has the least. The x-axis illustrates how each attribute affects the model's performance by displaying the R2 score's decline.

B. Model Evaluation

The performance of each model has been evaluated utilising a total of four distinct performance measures, which are as follows: RMSE, MAE, and R2 (R-Square). These various parameters are extensively used in the research studies:

1) RMSE

The RMSE calculates an average size of prediction mistakes without considering their direction. The corresponding equation is shown below in equation (3):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)2} \tag{3}$$

2) MAE

Without considering their direction, the MAE calculates the average size of the mistakes in a series of projections. Accuracy for continuous variables is measured. Equation (4) below displays the MAE metric's notation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{4}$$

3) R2

The R2 coefficient, calculated as the variance ratio in the dependent variable that can be predicted from the independent variable or variables, is expressed as follows in equation (5):

$$R2 = 1 - \frac{(\sum_{i=1}^{n} (y_i - \hat{y_l})2)/n}{(\sum_{i=1}^{n} (\underline{y_i} - \hat{y_l})2)/n} \tag{5}$$

In the following section, an outcome of experiments conducted on the ML regressor models for the house price prediction dataset is presented. Figures, graphs, and tables are the formats in which the results are presented.

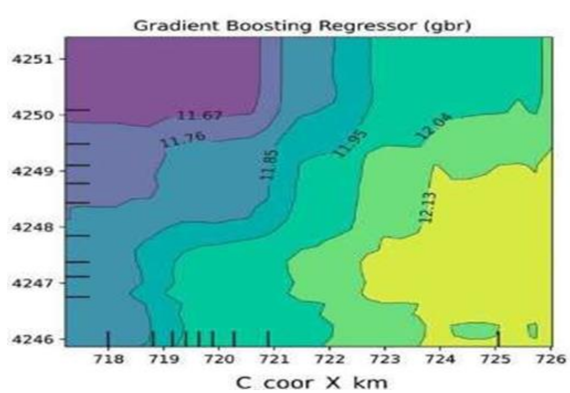

Figure-3 Two-way directional partial dependence plots for gradient boosting regressor model

Figure 3 represents a partial dependence plot in two directions for a Gradient Boosting Regressor (GBR) model. It illustrates the relationship between the target variable and two predictors: 'C coor X km' and an unidentified 'Y' coordinate. The contour lines show places where the model's predictions deviate from actual values. Understanding the interactions between the predictors and how they affect the model's output is made easier with the support of this visualization.

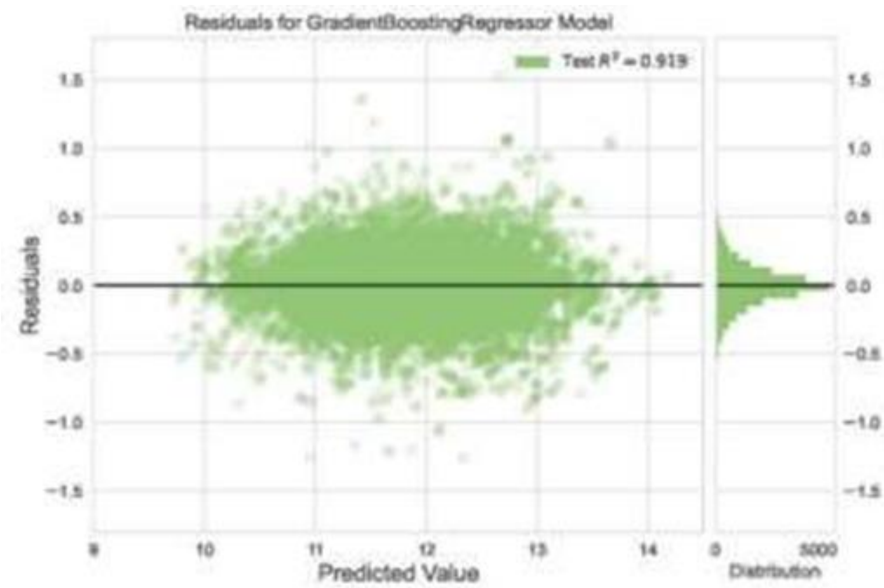

Figure-4 Residual plots of the trained GBR

Figure 4 displays a Residual plot of the trained GBR for house price prediction with 91% of R2: An abscissa axis with predicted prices on errors representing ordinate axis of a natural logarithm. GBR IS the best and fastest tor small dataset prediction.

Table-2: GBR model performance for house price prediction

Model	MAE	MSE	RMSE	R2
GBR	0.126	0.033	0.18	0.91

Figure-5 Bar graph of GBR performance of house price prediction

E-ISSN: 3048-7021 (Online)

The performance of the GBR model for house price prediction is summarised in Table 2 and depicted in Figure 5. The model's MAE, which measures how accurate the forecast is, is 0.126. This means that there is not much of a difference on average between the anticipated and actual values of homes. The MSE stands at 0. 033 and is an average squared variance between actual and predicted values, which further illustrates the accuracy of the constructed model. RMSE of 0. 18 reveals the model's ability to accurately predict home values with minimal mistakes and provides the standard deviation of the prediction errors. The relatively high accuracy of the model is, therefore, demonstrated by the R-Square (R2) of 0. 91, which means that it explains 91 percent of the changes to the target variable. These performance measures, taken together, show how accurate and dependable the GBR model is in predicting home prices.

C. Comparative analysis

The comparison to the forecasts of home prices based on machine learning has been made. In terms of performance measures, the following table presents a comparison of the comparative analysis and forecast home price prediction characteristics of a number of various ML regressor models. Table 3 shows the comparison of house price prediction with ML regressor models.

Table-3 Comparing several models to forecast house price prediction

Model	MSE	RMSE	R2
RF[32]	18.31	13.54	0.87
LGBM[33]	0.4	0.21	0.9
SVM[34]	0.1	1.1	0.82
GBR	0.03	0.1	0.9

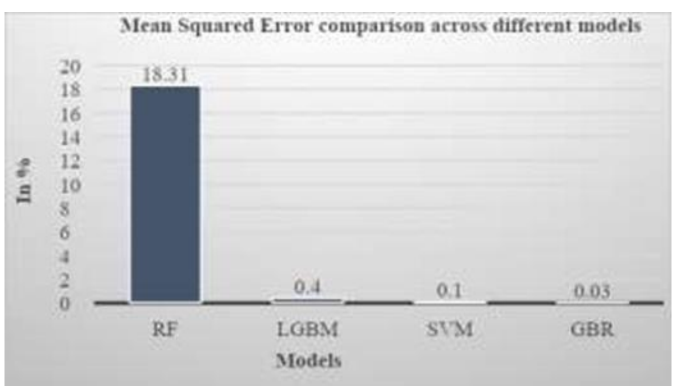

Figure-5 Mean squared error comparison across different models

Figure 5 represents the comparison among MSE models. GBR performs the best with an MSE of 0.03, indicating high predicted accuracy. RF performs the least well with the highest MSE of 18.31, showing the highest prediction errors, which shows the worst results in all models.

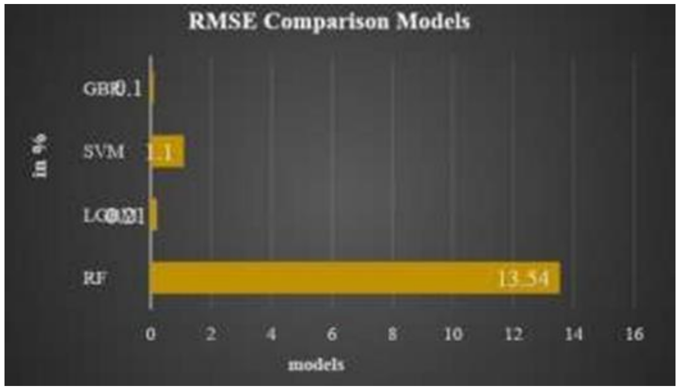

Figure-6 RMSE comparison models

Figure 6 represents the comparison of different models for RMSE. In this, GBR performs the best with an RMSE of 0.1, indicating a smallest RMSE and thus higher predictive accuracy. SVM has an RMSE of 1.1, which suggests less accurate predictions compared to the other models. RF performs the least well with an RMSE of 13.54, indicating the highest prediction errors.

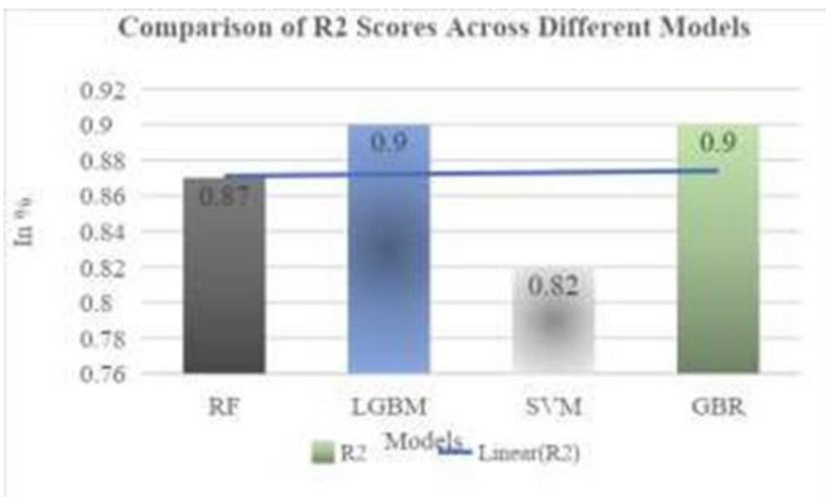

Figure-7 Comparison of R2 Scores Across Different Models

Figure 7 represents the comparison of R2 of various models. In this, LGBM and GBR perform well across all metrics with an R2 of 0.9, indicating strong predictive performance. RF follows closely with an R2 of 0.87. SVM performs relatively less well an R2 of 0.82. Overall, Gradient Boosting Regressor (GBR) is the model that performs the best overall when evaluated across MSE, RMSE, and R2 metrics. With its lowest MSE of 0.03 it shows excellent predictive accuracy. However, GBR has the lowest RMSE of 0. 1, showing a reliable prediction with minimal bias. In addition, GBR has reasonable accuracy in many evaluation metrics and has the highest R2 = 0. 9 along with LGBM, meaning that it captures a large portion of the variability in the dependent variable. Due to higher prediction errors in comparison with the other models, Random forest (RF) yields the highest MSE = 18. 31 and RMSE = 13. 54. Although tested low on MSE and RMSE figures, SVM has an R2 of 0. 82, implying it captures less variance in the data than GBR, LGBM. Consequently, out of all the evaluated metrics in this comparison, GBR proves to be the most accurate and efficient model.

5. Conclusion and Future Scope

House price prediction using RF and GB regressor and more, employed computational methods for the examination of the factors relating to the house price. Therefore, this study finds that machine learning algorithms can predict house prices from Kuala Lumpur real estate listings. The study finally comes to the conclusion that GBR is the most accurate and efficient in house price prediction for the Klang Valley, more precisely Kuala Lumpur Real Estate Market. As far as the evaluation metrics are concerned, the results indicate that the GBR model surpasses other ML models such as RF, SVM, and LGBM. In particular, it is noticed that the GBR gives the least MAE of 0. 126, MSE of 0. 033, and RMSE of 0. 18 with the highest value of R-Square (R2) = 0.91. This denoted that GBRs had a better accuracy predictor and stabilisation compared with others. Although, RF incurred the highest prediction errors followed by LGRE, SVM was less capable of explaining variations in the house prices having low R2 value, whereas, GBR outperformed all models and thereby, could be used reliably to forecast the house prices. The result of this research is that the quality of the data used and the selection of the algorithm are the factors that determine the success of real estate forecasting with machine learning. Decision makers in the industry of real estate, urban planning and investors in particular use such information to get reliable tools to work on changing market conditions. As for future work, socio-economic factors, environmental features, and time dynamics can be incorporated into the framework. Ensemble methods or deep learning models might improve forecast accuracy and resilience, especially for non-linear correlations and complicated housing market dynamics. Expanding the geographical area outside Kuala Lumpur or diversifying property kinds might increase model applicability. In volatile markets, models must be validated and refined against real- time data updates to remain relevant and reliable. Future research may advance real estate predictive analytics and improve global decision-making by enhancing methodology and combining varied datasets.

7. References

[1] B. Yang and B. Cao, "Research on Ensemble Learning-based Housing Price Prediction Model," Big Geospatial Data Data Sci., 2018, doi: 10.23977/bgdds.2018.11001.

[2] M. Thamarai and S. P. Malarvizhi, "House Price Prediction Modeling Using Machine Learning," Int. J. Inf. Eng. Electron. Bus., 2020, doi: 10.5815/ijieeb.2020.02.03.

[3] J. Q. Guo, S. H. Chiang, M. Liu, C. C. Yang, and K. Y. Gou, "Can machine learning algorithms associated with text mining from internet data improve housing price prediction performance?," Int. J. Strateg. Prop. Manag., 2020, doi: 10.3846/ijspm.2020.12742.

[4] A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, and G. Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique," in Procedia Computer Science, 2021. doi: 10.1016/j.procs.2022.01.100.

[5] H. Joshi and S. Swarndeep, "A Comparative Study on House Price Prediction using Machine Learning," Int. Res. J. Eng. Technol., 2022.

[6] S. Mathur., "Supervised Machine Learning- Based Classification and Prediction of Breast Cancer," Int. J. Intell. Syst. Appl. Eng., vol. 12(3), 2024.

[7] F. M. Basysyar and G. Dwilestari, "House Price Prediction Using Exploratory Data Analysis and Machine Learning with Feature Selection," Acadlore Trans. AI Mach. Learn., 2022, doi: 10.56578/ataiml010103.

[8] J. Thomas, "Optimizing Bio-energy Supply Chain to Achieve Alternative Energy Targets," pp. 2260–2273, 2024.

[9] Y. Wu, "Analysis of Various Models for House Price Prediction based on Machine Learning," Highlights Sci. Eng. Technol., 2023, doi: 10.54097/hset.v39i.6680.

[10] H. Li, "House price prediction based on machine learning," Appl. Comput. Eng., 2023, doi: 10.54254/2755-2721/4/2023362.
[11] J. Thomas, "Optimizing Nurse Scheduling: A Supply Chain Approach for Healthcare Institutions," pp. 2251–2259, 2024.
[12] R. T. Mora-Garcia, M. F. Cespedes-Lopez, and V. R. Perez-Sanchez, "Housing Price Prediction Using Machine Learning Algorithms in COVID-19 Times," Land, 2022, doi: 10.3390/land11112100.
[13] A. S. Hassan Younas, Zohaib Ur Rehman Afridi, Kaleem Ullah, Sahib Gul Afridi, "Development of Photocatalytic Ultrafiltration Membranes Technology for Enhanced Removal of Carbamazepine: Optimization, Mechanistic Insights, and Engineering Applications."
[14] A. C. Kin, Z. A. B. A. Salam, and K. B. Nowshath, "Machine Learning based House Price Prediction Model," in International Conference on Edge Computing and Applications, ICECAA 2022 - Proceedings, 2022. doi: 10.1109/ICECAA55415.2022.9936336.
[15] R. Odegua, "Applied Machine Learning for Supermarket Sales Prediction," no. January, 2022.
[16] S. Mann, A. Balyan, V. Rohilla, D. Gupta, Z. Gupta, and A. W. Rahmani, "Artificial Intelligence-based Blockchain Technology for Skin Cancer Investigation Complemented with Dietary Assessment and Recommendation using Correlation Analysis in Elder Individuals," Journal of Food Quality, 2022. doi: 10.1155/2022/3958596.
[17] V. Rohilla, S. Chakraborty, and R. Kumar, "Car Automation Simulator Using Machine Learning," SSRN Electron. J., 2020, doi: 10.2139/ssrn.3566915.
[18] N. T. M. Sagala and L. H. Cendriawan, "House Price Prediction Using Linear Regression," in 2022 IEEE 8th International Conference on Computing, Engineering and Design, ICCED 2022, 2022. doi: 10.1109/ICCED56140.2022.10010684.
[19] B. Budiman, N. Alamsyah, T. P. Yoga, R. Y. R. Alamsyah, and C. Habibi, "Machine Learning-Based Classification of House Prices: A Comparative Study," in 2023 5th International Conference on Cybernetics and Intelligent Systems, ICORIS 2023, 2023. doi: 10.1109/ICORIS60118.2023.10352211.
[20] M. Ahtesham, N. Z. Bawany, and K. Fatima, "House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan," in Proceedings - 2020 21st International Arab Conference on Information Technology, ACIT 2020, 2020. doi: 10.1109/ACIT50332.2020.9300074.
[21] V. Rohilla, S. Chakraborty, and M. Kaur, "An Empirical Framework for Recommendation-based Location Services Using Deep Learning," Eng. Technol. Appl. Sci. Res., 2022, doi: 10.48084/etasr.5126.
[22] N. Thakur, R. Jain, A. Mahajan, and S. M. N. Islam, "Deep Neural Network based Data Analysis and Price Prediction framework for Rio de Janeiro Airbnb," in 2022 IEEE 7th International conference for Convergence in Technology, I2CT 2022, 2022. doi: 10.1109/I2CT54291.2022.9824383.
[23] Y. W. Lin, Y. B. Lin, C. Y. Liu, J. Y. Lin, and Y. L. Shih, "Implementing AI as Cyber IoT Devices: The House Valuation Example," IEEE Trans. Ind. Informatics, 2020, doi: 10.1109/TII.2019.2951847.
[24] V. Rohilla, D. S. Chakraborty, and D. R. Kumar, "Random Forest with Harmony Search Optimization for Location Based Advertising," Int. J. Innov. Technol. Explor. Eng., 2019, doi: 10.35940/ijitee.i7761.078919.
[25] A. Sayed, "Face mask detection model based on deep CNN technique using AWS," Int. J. Eng. Res. Appl., vol. 13, no. 5, pp. 12–19, 2023.

[26] J. Su, V. Danilo Vasconcellos, S. Prasad, S. Daniele, Y. Feng, and K. Sakurai, "Lightweight Classification of IoT Malware Based on Image Recognition," in Proceedings - International Computer Software and Applications Conference, 2018. doi: 10.1109/COMPSAC.2018.10315.
[27] W. S. Noble, "What is a support vector machine?," Nature Biotechnology, 2006. doi: 10.1038/nbt1206-1565.
[28] P. Khare, "Enhancing Security with Voice: A Comprehensive Review of AI-Based Biometric Authentication Systems," vol. 10, no. 2, pp. 398–403, 2023.
[29] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Stat., 2001, doi: 10.1214/aos/1013203451.
[30] S. Arora and S. R. Thota, "Ethical Considerations and Privacy in AI-Driven Big Data Analytics," no. May, 2024.
[31] P. Khare and Sahil Arora, "Predicting Customer Churn in SaaS Products using Machine Learning," Heal. Inf. Sci. Syst., 2020, doi: 10.1007/s13755-019-0095-z.
[32] A. S. Prof. J. Kalidass, T. Dharshalini, R. Nivetha, "House Price Prediction Using Machine Learning," Proc. - 2021 3rd Int. Conf. Adv. Comput. Commun. Control Networking, ICAC3N 2021, pp. 203–206, 2021, doi: 10.1109/ICAC3N53548.2021.9725552.
[33] S. Abdul-Rahman, S. Mutalib, N. H. Zulkifley, and I. Ibrahim, "Advanced Machine Learning Algorithms for House Price Prediction: Case Study in Kuala Lumpur," Int. J. Adv. Comput. Sci. Appl., 2021, doi: 10.14569/IJACSA.2021.0121291.
[34] W. K. O. Ho, B. S. Tang, and S. W. Wong, "Predicting property prices with machine learning algorithms," J. Prop. Res., 2021, doi: 10.1080/09599916.2020.1832558.

8.Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

9.Funding

No external funding was received to support or conduct this study.