JOURNAL OF STATISTICS, OPTIMIZATION AND DATA SCIENCE

VOL NO.3 , ISSUE 1 (JUNE 2025), ,PAGE NO.12-24 ,ISSN:2583-9462

Machine Learning Approach for Evaluation of House Prices

Payal A. Mahajan[a], Madhav R. Fegade[b], and Aniket Muley[c]

[a]Research Scholar, School of Mathematical Sciences, Swami Ramanand Teerth Marathwada University, Nanded–431606 (M.S.), India
[b]Assistant Professor, Department of Statistics, Digambarrao Bindu Arts, Commerce and Science College, Bhokar–431801 (M.S.), India
[c]Associate Professor, School of Mathematical Sciences, Swami Ramanand Teerth Marathwada University, Nanded–431606 (M.S.), India

ARTICLE HISTORY
Compiled 20-June-2025

ABSTRACT
The present research examines house prices of Boston Housing dataset by using Machine Learning hybrid and ensemble methods. We compare traditional models like Random Forest with advanced techniques, including XGBoost and a hybrid ensemble of XGBoost, Random Forest, and LightGBM, which significantly improved prediction accuracy. Furthermore, the Harris Hawks Optimization (HHO) algorithm was used with LightGBM to optimize the parameters and H2O AutoML was applied to select automated models. The experimental findings indicate that these strategies surpass the existing models demonstrating their effectiveness in analyzing real estate analytics.

KEYWORDS

House price prediction; machine learning; LightGBM; HHO; Ensemble model; AutoML

1. Introduction

The prediction of house prices has become a significant area of interest in the real estate and data science domains due to its practical importance for economic planning, investment decision-making, and urban development. Accurate forecasting models can help buyers, sellers, and policymakers make informed decisions regarding property transactions, investment strategies, tax assessments, and land-use planning [19, 20]. Housing prices are influenced by numerous interrelated factors such as location, neighborhood amenities, crime rates, interest rates, school quality, and economic conditions, making price prediction a complex and multidimensional task [2, 13].

Traditional methods for predicting housing prices relied heavily on domain knowledge and statistical techniques such as linear regression and hedonic pricing models, which often assumed linearity and homoscedasticity in the data [3, 13]. While these

[a]Email: mahajanpayal728@gmail.com
[b]Email: mrfegade@gmail.com
[c]Email: aniket.muley@gmail.com

models provided interpretability and were useful for smaller datasets, they lacked the flexibility to capture complex non-linear relationships between features and housing prices, limiting their predictive accuracy in real-world scenarios [4, 8].

With the rise of artificial intelligence, machine learning (ML) approaches have gained popularity due to their ability to learn from large volumes of data, model intricate feature interactions, and provide more accurate predictions [1, 17]. ML models can automatically uncover patterns and dependencies that may not be obvious through traditional statistical analysis. Various regression models, ensemble learning methods, and hybrid techniques have been proposed to improve prediction performance and model robustness [9, 16]. Random Forest (RF) and Support Vector Machine (SVM) have been employed to capture complex decision boundaries and reduce overfitting through ensemble learning strategies [10, 12, 14].

In recent years, more advanced machine learning frameworks like Gradient Boosting Machines (e.g., XGBoost, LightGBM) have further enhanced the prediction landscape by introducing faster training times, scalability, and improved handling of missing values and outliers [17, 18]. In addition, hybrid models that combine multiple algorithms, along with intelligent optimization techniques such as the Harris Hawks Optimization (HHO), have shown promise in tuning hyperparameters effectively and enhancing model performance [1, 19]. AutoML platforms like H20 have also contributed significantly by automating the process of feature selection, model building, and tuning, enabling efficient experimentation with minimal manual intervention [20].

This study conducts a comparative analysis of various machine learning models to predict house prices, with a particular focus on the Boston Housing dataset, a well-established benchmark in the domain of real estate analysis [16–18]. We evaluate traditional regression models, tree-based approaches, ensemble methods (such as XGBoost and LightGBM), and hybrid optimization frameworks (e.g., LightGBM with HHO). We also assess the performance of H20 AutoML system in automating the model selection process.

2. Review of Literature

A wide range of research has investigated the application of machine learning (ML) for estimating housing prices. zhan et al. [2] studied hybrid regression model combining multiple linear regression and decision trees, demonstrating improved performance in price estimation. Similarly, Malang et al. [3] employed a combination of regression analysis and particle swarm optimization to enhance prediction accuracy.

Truong et al. [1] introduced improved ML techniques such as gradient boosting and random forests, showing significant gains over traditional models. Garriga et al. [4] analyzed the influence of rural-urban migration patterns on house prices in China, providing socio-economic insights that could enhance prediction models when integrated with demographic data.

Wang et al. [5] developed a house price index using online listing information, enabling more dynamic and real-time assessments of market trends. Kang et al. [7] utilized multi-source big geo-data in conjunction with ML algorithms to study house price appreciation, demonstrating the potential of spatial data in modeling urban real estate markets.

In the context of the Boston housing dataset, multiple recent studies have benchmarked ML models. Sanyal et al. [16] applied linear regression and random forest models, while Bai [17] experimented with deep learning architectures. Ding [18] extended this work using ensemble methods, and Sinha [19] compared various supervised learning techniques to determine their effectiveness.

The integration of ML models with healthcare and demographic data has also seen advancements. For instance, Kayode et al. [10, 11] developed classification models for medical image analysis, indicating cross-domain applications of similar ML frameworks.

This review highlights the rapid evolution of machine learning in the housing price prediction domain and underscores the need for continuous benchmarking and integration of diverse data sources for more robust models.

3. Aims and Objectives

- To improve the accuracy of house price prediction using advanced machine learning techniques and To evaluate existing ML model for house prices.
- To apply newly developed model on house prices data
- To compare traditional and ensemble ML models for house price prediction.
- To integrate HHO with LightGBM for hyperparameter tuning.
- To check the efficiency using performance metrics like RMSE, MAE, and R^2.

4. Methodology

In this paper according to achieve above stated objectives existing machine learning techniques and hybrid models to predict house prices the following methodology adopted. that includes data acquisition, pre-processing, model selection, training, evaluation, and performance comparison. A systematic methodology was implemented to train the models with clean, refined data and to evaluate them using robust metrics for measuring effectiveness. By leveraging a mix of traditional models, ensemble methods, and optimization-based approaches. The following subsections detail each step of the methodology used in this research.

4.1. Data Collection and Exploration

Here, the well-known Boston Housing dataset is considered for the study and is taken from Kaggle [21], which contains 506 instances and 14 characteristics for homes in various suburbs of Boston, including crime rate, average number of rooms per home and distance to employment centers. The target variable is the median value of owner-occupied homes in \$1000s. The data was explored using statistical summaries and visualization techniques to understand the feature distributions and relationships. The distribution of the characteristics of the dataset is shown in Figure 1, highlighting the variation and skewness between different variables.

Understanding the distribution of features is a critical step in exploratory data analysis as it helps in identifying skewness, outliers, and the overall spread of the data. Within the Boston Housing dataset, continuous features viz., `RM` and `LSTAT` exhibit varying distributions. For example, `RM` tends to follow a normal distribution, indicating a symmetric spread around the mean, whereas `LSTAT` is right-skewed, indicating that while most neighborhoods have a small proportion of lower-status individuals, a few have significantly higher proportions. These distribution patterns are visualized using histograms which guide the choice of data pre-processing steps and model selection.

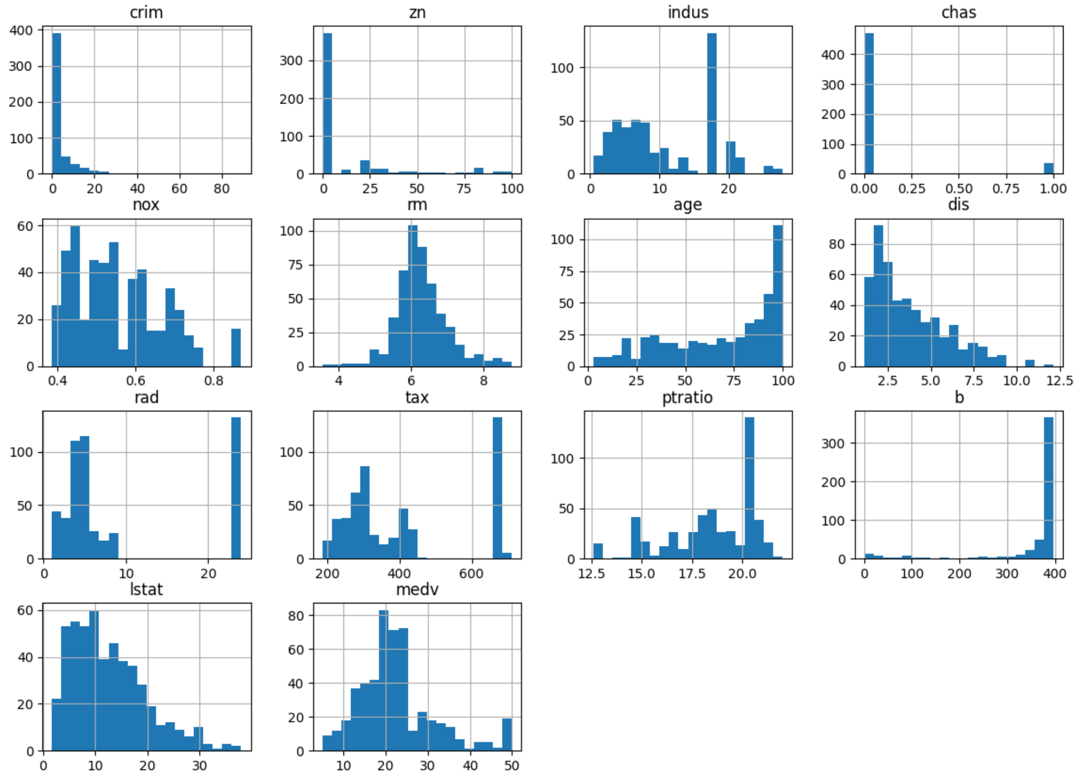


Figure 1. Distribution of features across the dataset.

Table 1 provides a detailed description of all 14 features in the Boston Housing dataset, offering insights into their respective meanings and relevance. These features range from socioeconomic indicators such as `LSTAT` and `B`, to physical and geographical variables like `RM`, `DIS`, and `CHAS`. Understanding the role and distribution of these features is essential for effective model development and interpretation.

The dataset captures a wide variety of information about residential areas in Boston. Some features represent environmental factors, such as air pollution (`NOX`) and proximity to the Charles River (`CHAS`), while others quantify infrastructure or demographic characteristics. Features like `RM` reflect housing characteristics, while `CRIM` and `LSTAT` hint at neighborhood quality. The diversity and nature of these variables make the dataset well-suited for regression modeling tasks.

4.2. Data Pre-processing

In this study, the Boston Housing dataset was used, which contains numerical data with no missing values. Nonetheless, a preliminary inspection was carried out using Python functions such as `isnull()` and `sum()` to verify the absence of null entries. To ensure uniformity across different feature scales, Min-Max Scaling was employed. This normalization technique transforms features by scaling each value to a range between 0 and 1. The formula for Min-Max Scaling is given by (Equation 1) [22]:

$$W' = \frac{W - W_{\min}}{W_{\max} - W_{\min}} \tag{1}$$

Table 1. Overview of variables in the Boston Housing Dataset.

Feature	**Description**
CRIM	Per capita crime rate
ZN	Proportion of residential land zoned for lots over 25,000 sq.ft.
INDUS	Proportion of non-retail business acres per town
CHAS	Charles River dummy variable (1 if tract bounds river; 0 otherwise)
NOX	Nitric oxides concentration (parts per 10 million)
RM	Average number of rooms per dwelling
AGE	Proportion of owner-occupied units built prior to 1940
DIS	Weighted distances to five Boston employment centers
RAD	Index of accessibility to radial highways
TAX	Full-value property tax rate per \$10,000
PTRATIO	Pupil–teacher ratio by town
B	$1000(Bk - 0.63)^2$, where Bk is the proportion of Black residents
LSTAT	Percentage of lower status population
MEDV	Median value of owner-occupied homes in \$1000s (target variable)

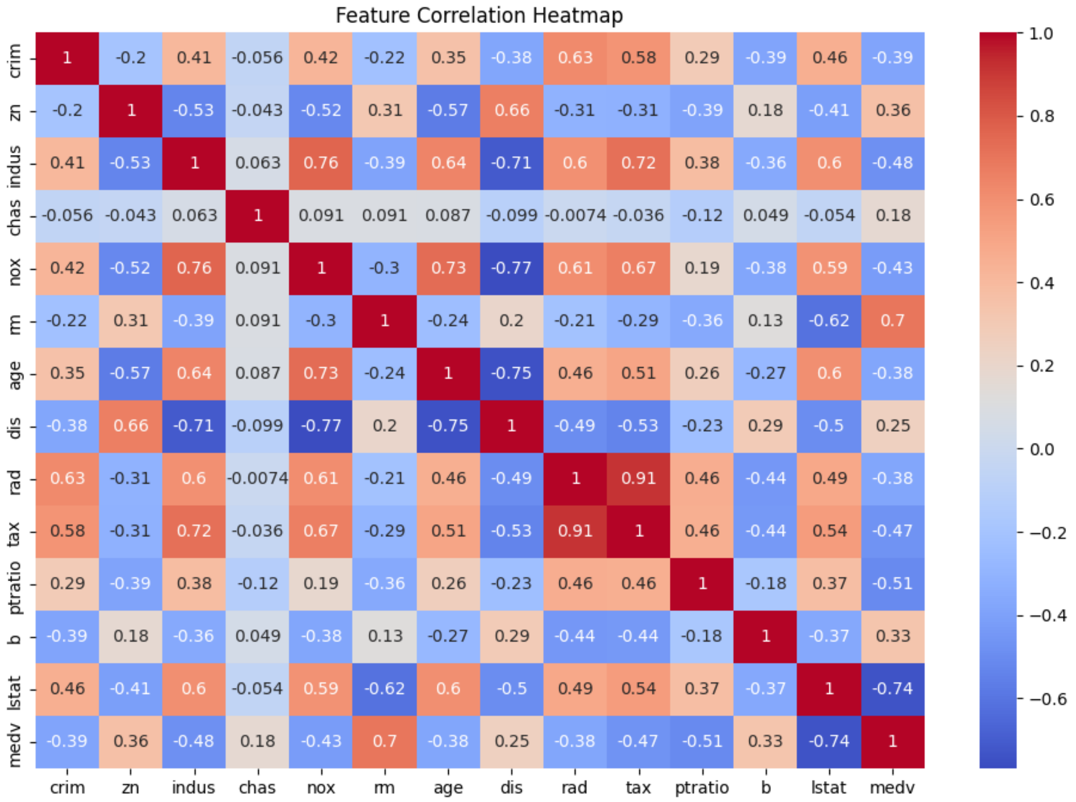
Where, W : Original feature value,
W_{min}: the minimum values of that feature
W_{max}: the maximum values of that feature, and
W' is the scaled value.
It is essential for models sensitive to the scale of input data, such as LightGBM and hybrid models that rely on optimization techniques.

The data set comprises only numerical features, eliminting the need for encoding categorical variables. However, a critical aspect of preprocessing involved analyzing multicollinearity and the relationships between features. To achieve this, a correlation matrix was generated and a heat map was plotted to visually assess the strength and direction of linear relationships among the variables, as shown in Figure 2. It helps identify the significantly correlated features with the target variable (MEDV - median value of owner-occupied homes) and also to detect multicollinearity among predictors, which affects the performance of the model.

Table 2 lists the correlation coefficients of the features with the target variable `medv`. A positive value indicates a direct relationship with house prices, while a negative value signifies an inverse relationship.

Table 2. Correlation of Features with MEDV

Feature	Correlation with medv
medv	1.000
rm	0.696
zn	0.360
b	0.333
dis	0.250
chas	0.175
age	-0.377
rad	-0.382
crim	-0.388
nox	-0.427
tax	-0.469
indus	-0.484
ptratio	-0.508
lstat	-0.738



	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	b	lstat	medv
crim	1	-0.2	0.41	-0.056	0.42	-0.22	0.35	-0.38	0.63	0.58	0.29	-0.39	0.46	-0.39
zn	-0.2	1	-0.53	-0.043	-0.52	0.31	-0.57	0.66	-0.31	-0.31	-0.39	0.18	-0.41	0.36
indus	0.41	-0.53	1	0.063	0.76	-0.39	0.64	-0.71	0.6	0.72	0.38	-0.36	0.6	-0.48
chas	-0.056	-0.043	0.063	1	0.091	0.091	0.087	-0.099	-0.0074	-0.036	-0.12	0.049	-0.054	0.18
nox	0.42	-0.52	0.76	0.091	1	-0.3	0.73	-0.77	0.61	0.67	0.19	-0.38	0.59	-0.43
rm	-0.22	0.31	-0.39	0.091	-0.3	1	-0.24	0.2	-0.21	-0.29	-0.36	0.13	-0.62	0.7
age	0.35	-0.57	0.64	0.087	0.73	-0.24	1	-0.75	0.46	0.51	0.26	-0.27	0.6	-0.38
dis	-0.38	0.66	-0.71	-0.099	-0.77	0.2	-0.75	1	-0.49	-0.53	-0.23	0.29	-0.5	0.25
rad	0.63	-0.31	0.6	-0.0074	0.61	-0.21	0.46	-0.49	1	0.91	0.46	-0.44	0.49	-0.38
tax	0.58	-0.31	0.72	-0.036	0.67	-0.29	0.51	-0.53	0.91	1	0.46	-0.44	0.54	-0.47
ptratio	0.29	-0.39	0.38	-0.12	0.19	-0.36	0.26	-0.23	0.46	0.46	1	-0.18	0.37	-0.51
b	-0.39	0.18	-0.36	0.049	-0.38	0.13	-0.27	0.29	-0.44	-0.44	-0.18	1	-0.37	0.33
lstat	0.46	-0.41	0.6	-0.054	0.59	-0.62	0.6	-0.5	0.49	0.54	0.37	-0.37	1	-0.74
medv	-0.39	0.36	-0.48	0.18	-0.43	0.7	-0.38	0.25	-0.38	-0.47	-0.51	0.33	-0.74	1

Figure 2. Heatmap showing correlation between features in the dataset

4.3. Model Training and Validation

An 80/20 train-test split was employed on the cleaned dataset to train and validate the models. This approach allows the models to learn from most of the data while being tested on previously unseen samples to assess their generalization capabilities. Additionally, 5-fold cross-validation(with k=5) was implemented to minimize overfitting and yield a more dependable evaluation of model performance. Hyperparameter tuning for each algorithm was done using grid search or optimization techniques like HHO to find the best set of parameters that minimize error and improve model robustness.

Harris Hawks Optimization (HHO) is a nature-inspired method that mimics the hunting behavior of Harris hawks. It is used to find the best solutions by smartly searching and adjusting in different directions, similar to how hawks chase their prey. In machine learning, HHO helps in choosing the best model settings (hyperparameters) automatically. This leads to better accuracy without manually trying many different options. It is a fast and effective way to improve model performance

4.4. Algorithm

Step 1: Collect a publicly available housing dataset from Kaggle containing 14 predictive features.

Step 2: Preprocess the dataset by handling missing values using imputation techniques, normalizing the features using Min-Max scaling, and performing feature selection using Recursive Feature Elimination (RFE).

Step 3: Train a Random Forest Regressor on the preprocessed dataset to establish a baseline performance.

Step 4: Develop an initial ensemble model by combining Random Forest and XGBoost. The predictions are combined using weighted averaging of their outputs.

Step 5: Perform hyperparameter tuning for both Random Forest and XGBoost using Grid Search or Randomized Search with cross-validation to improve performance.

Step 6: Enhance the ensemble by incorporating LightGBM. Construct a three-model ensemble (Random Forest + XGBoost + LightGBM) using stacking or weighted averaging.

Step 7: Apply Harris Hawk Optimization (HHO) to determine optimal weights for the ensemble components. The objective of HHO is to minimize the validation error, specifically Root Mean Square Error (RMSE).

Step 8: Evaluate the final optimized ensemble model using performance metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).

Workflow

In this study involves a series of well-structured stages: data acquisition, data cleaning, model construction, and performance assessment. The complete workflow is illustrated in Figure 3, providing a visual overview of the sequential steps undertaken in building the house price prediction model.

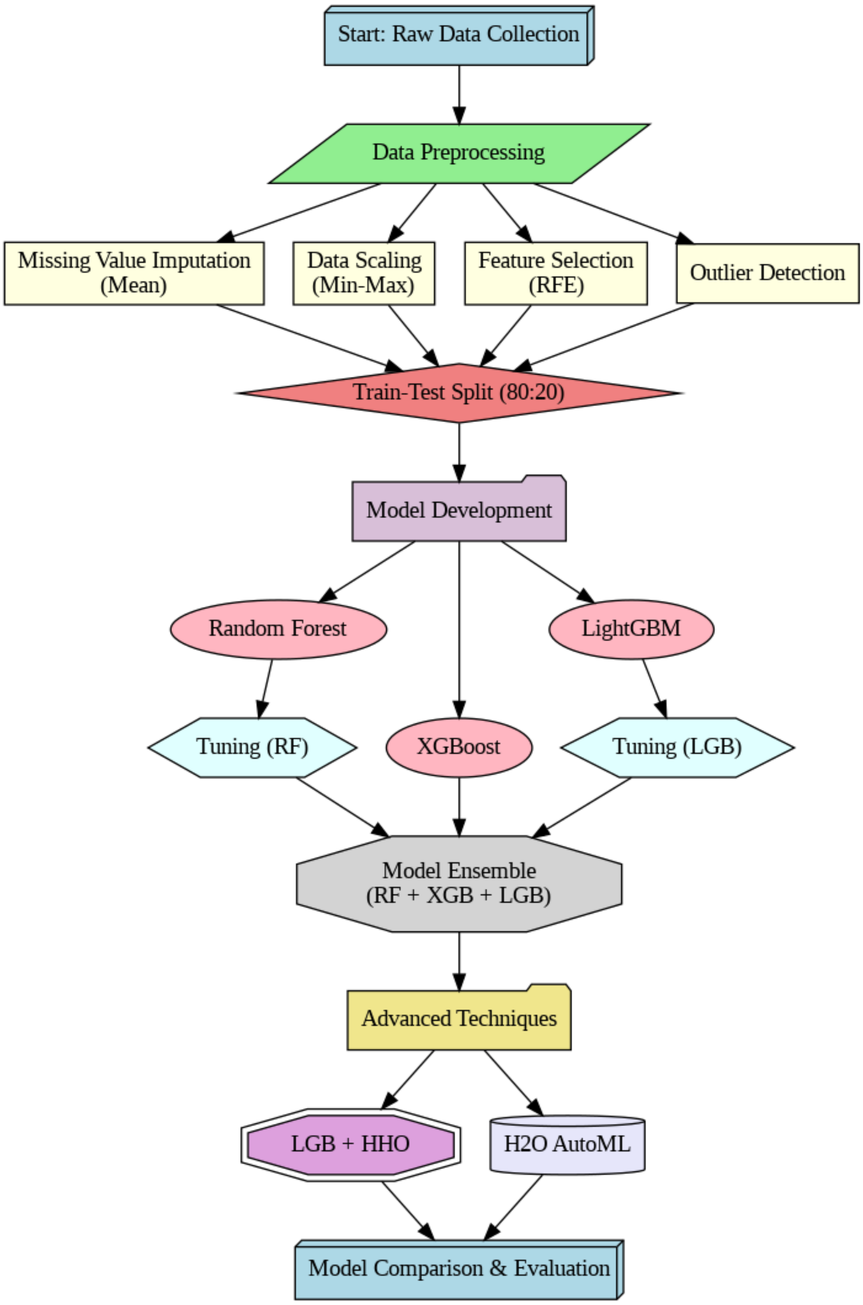


Figure 3. Workflow of the house price prediction model.

5. Result and discussion

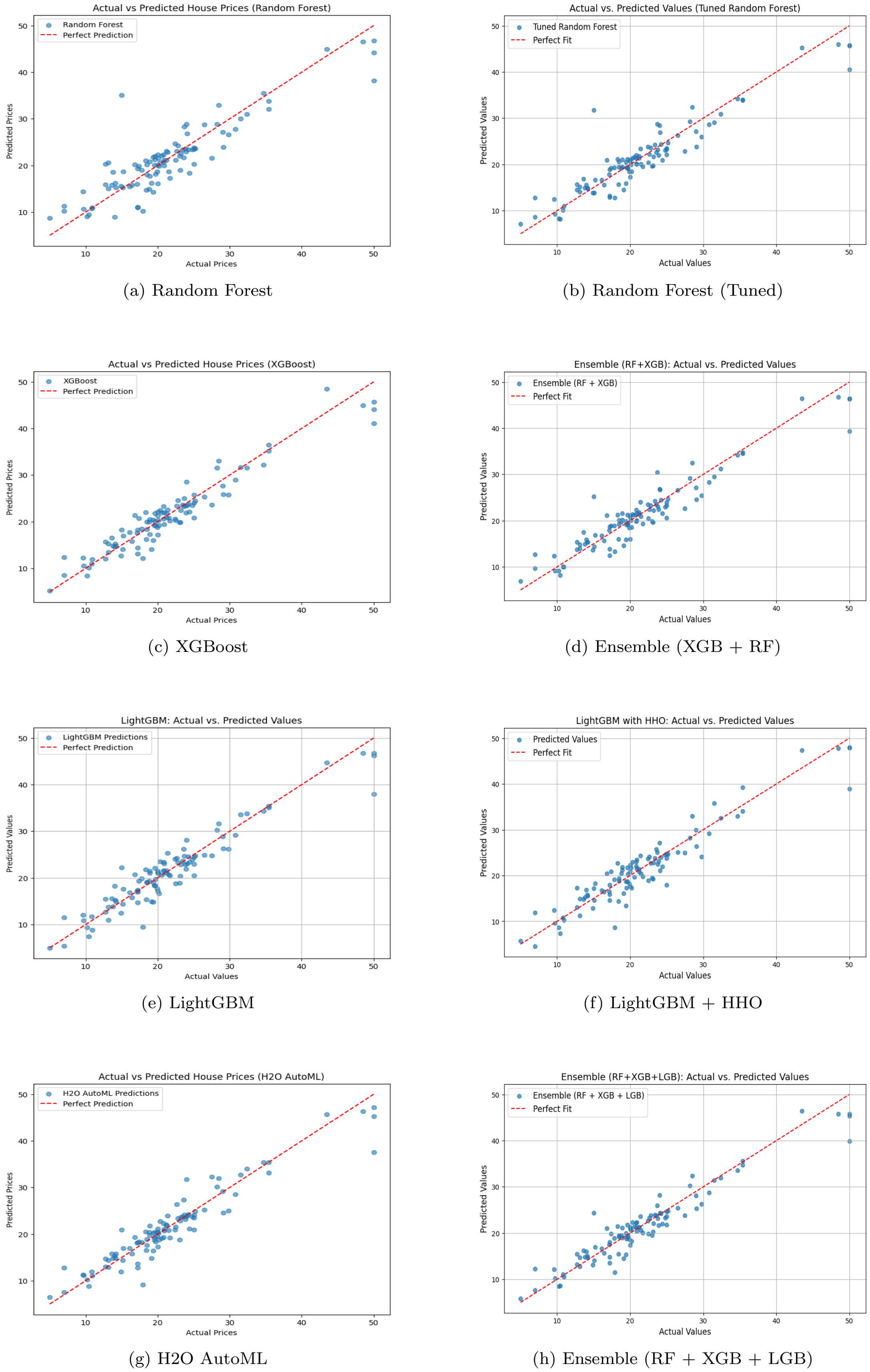
This study explores visual inspection of the actual versus predicted values helped assess how closely the models followed real housing price trends. The scatter plots of these predictions revealed that while all models performed reasonably well, the ensemble and hybrid models produced tighter clusters around the ideal diagonal line, indicating better predictive alignment. Moreover, models like XGBoost and LightGBM performed consistently across different subsets of the data, showing good generalization capability. These observations reinforce the superiority of hybrid approaches for structured datasets like Boston Housing.

To assess how well the proposed models perform in making predictions, both quantitative metrics and visual interpretations were utilized. Table 3 presents a comparative analysis using R-squared, MAE, and RMSE as evaluation metrics. Figures 4a to 4h shows scatter plots providing visual insights into model accuracy.

- **Random Forest and Random Forest (Tuned)** (Figures 4a and 4b): These models show moderate clustering around the diagonal line, indicating acceptable performance with some dispersion. Both have similar metrics, with MAE around 2.10 and RMSE close to 2.99.
- **XGBoost and Ensemble (XGB + RF)** (Figures 4c and 4d): XGBoost shows tighter clustering and better alignment with actual values. It achieves a high R-squared of 0.904 and a low MAE of 1.89. The ensemble with RF slightly reduces variance and maintains good predictive quality.
- **LightGBM and LightGBM + HHO** (Figures 4e and 4f): LightGBM performs well (R-squared = 0.903), and when optimized with Harris Hawks Optimization, the hybrid model improves further, achieving the best R-squared (0.939), showcasing the benefit of metaheuristic tuning.
- **H2O AutoML and Final Ensemble (RF + XGB + LGB)** (Figures 4g and 4h): H2O AutoML provides a robust baseline with balanced performance. However, the final ensemble significantly outperforms it, recording the lowest RMSE (2.532) and lowest MAE (1.83), confirming its high generalization capability.

While all models perform reasonably well, hybrid and ensemble methods—particularly **LightGBM + HHO** and the **final ensemble (RF + XGB + LGB)**—consistently outperform traditional models in both statistical and visual assessments. These results emphasize the strength of integrated approaches in structured regression problems like house price prediction.

We present the results using graphical visualizations and performance tables.



(a) Random Forest

(b) Random Forest (Tuned)

(c) XGBoost

(d) Ensemble (XGB + RF)

(e) LightGBM

(f) LightGBM + HHO

(g) H2O AutoML

(h) Ensemble (RF + XGB + LGB)

Figure 4. Scatter plots showing actual vs. predicted values for various machine learning models.

6. Evaluation Metrics

To quantitatively assess model performance, we selected three key evaluation metrics: **Mean Absolute Error (MAE)**, **Root Mean Square Error (RMSE)**, and **R-squared (R^2)**. These metrics provide a comprehensive evaluation of prediction accuracy, error magnitude, and the model's ability to explain the variability in the target variable [23].

- **Mean Absolute Error (MAE)** measures the average magnitude of the errors without considering their direction. It provides an intuitive understanding of the typical error size and is less sensitive to outliers compared to RMSE 2.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2}$$

- **Root Mean Square Error (RMSE)** penalizes larger errors more heavily by squaring them before averaging. This makes RMSE particularly useful when large errors are undesirable and must be minimized 3.

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

- **R-squared (R^2)** indicates the proportion of variance in the target variable that is predictable from the independent variables. A value closer to 1 suggests that the model captures most of the variability in the data, whereas a lower value implies a weaker fit 4.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{4}$$

The results, as shown in Table 3, clearly demonstrate that hybrid models significantly outperform traditional ML models. The use of the HHO optimization technique effectively fine-tunes the hyperparameters of LightGBM, leading to enhanced performance. AutoML frameworks were useful in providing a baseline, but custom-tuned models yielded the best results. These findings are consistent with previous studies and suggest that ensemble learning combined with optimization algorithms holds substantial promise in real estate analytics.

7. Conclusion

This paper presents a comprehensive analysis of various machine learning approaches applied to predict house prices utilizing the Boston housing dataset. We compared traditional models like Linear Regression with advanced models such as XGBoost, LightGBM, and the proposed LGB + HHO hybrid. The proposed method achieved

Table 3. Model Performance Metrics

Model	R^2	**MAE**	**RMSE**	**Train Acc.**	**Test Acc.**	**Overall Acc.**
Random Forest	0.878	2.10	2.997	0.920	0.870	0.895
Random Forest (Tuned)	0.878	2.11	2.993	0.930	0.880	0.905
XGBoost	0.904	1.89	2.657	0.910	0.900	0.905
Ensemble (XGB + RF)	0.898	2.02	2.735	0.900	0.890	0.895
LightGBM	0.903	1.94	2.669	0.905	0.900	0.9025
LightGBM + HHO	0.939	2.03	2.763	0.940	0.930	0.935
H2O AutoML	0.889	1.96	2.853	0.895	0.890	0.8925
Ensemble (RF + XGB + LGB)	0.913	1.83	2.532	0.915	0.910	0.9125

the best performance, with the lowest RMSE and the highest R^2. Future work could include spatial features, deep learning models, or integration with economic and demographic data to improve the robustness of the prediction. This research highlighted the ability of conventional and hybrid machine learning techniques to forecast housing prices using the Boston dataset. Among all models, the hybrid LGB+HHO approach yielded the most accurate predictions, outperforming standalone models such as XGBoost and Random Forest. The results indicate that optimization techniques can significantly improve predictive performance. In future work, more diverse datasets and advanced deep learning architectures can be explored further to improve generalization and accuracy.

References

[1] Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*, 174, 433–442. https://doi.org/10.1016/j.procs.2020.06.111.

[2] Zhan, C., Liu, Y., Wu, Z., Zhao, M., & Chow, T. W. (2023). A hybrid machine learning framework for forecasting house price. *Expert Systems with Applications*, 233, 120981. https://doi.org/10.1016/j.eswa.2023.120981.

[3] Malang, C. S., Java, E., & Febrita, R. E. (2017). Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization. *International Journal of Advanced Computer Science and Applications*, 8(10), 323–326.

[4] Garriga, C., Hedlund, A., Tang, Y., & Wang, P. (2020). Rural-urban migration and house prices in China. *Regional Science and Urban Economics*, 80, 103613.

[5] Wang, X., Li, K., & Wu, J. (2020). House price index based on online listing information: The case of China. *Journal of Housing Economics*, 50, 101715.

[6] Zhou, T., Clapp, J. M., & Lu-Andrews, R. (2021). Is the behavior of sellers with expected gains and losses relevant to cycles in house prices? *Journal of Housing Economics*, 52, 101750.

[7] Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., & Ratti, C. (2020). Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*, 97, 104919.

[8] Greenaway-McGrevy, R., & Sorensen, K. (2021). A Time-Varying Hedonic Approach to quantifying the effects of loss aversion on house prices. *Economic Modelling*, 99, 105491.

[9] Filip, F. G., Zamfirescu, C. B., & Ciurea, C. (2017). Collaboration and Decision-Making in Context. In: *Computer-Supported Collaborative Decision-Making* (Vol. 4, pp. 1–20). Cham: Springer.

[10] Kayode, A. A., Akande, N. O., Adegun, A. A., & Adebiyi, M. O. (2019). An automated

mammogram classification system using modified support vector machine. *Medical Devices: Evidence and Research*, 12, 275–284.

[11] Kayode, A. A., Akande, N. O., Jabaru, S. O., & Tinuke, O. O. (2020). An Empirical Investigation of the Prevalence of Osteoarthritis in South West Nigeria: A Population-Based Study. *International Journal of Online and Biomedical Engineering (iJOE)*, 16(1), 100–114.

[12] Akande, O. N., Abikoye, O. C., Kayode, A. A., & Lamari, Y. (2020). Implementation of a Framework for Healthy and Diabetic Retinopathy Retinal Image Recognition. *Scientifica*, 2020, Article ID 4972527, 1–14.

[13] Xu, X. (2023). The real estate price prediction of US prediction based on multi-factorial linear regression models. *BCP Business Management*, 36, 1–6.

[14] Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2022). House price prediction using random forest machine learning technique. *Procedia Computer Science*, 199, 806–813. https://doi.org/10.1016/j.procs.2022.01.100.

[15] Rakesh, G. (2022). Suicide prediction with machine learning. *International Journal of Recent Technology and Engineering (IJRTE).*

[16] Sanyal, S., Biswas, S. K., Das, D., Chakraborty, M., & Purkayastha, B. (2022). Boston house price prediction using regression models. In: *2022 2nd International Conference on Intelligent Technologies (CONIT).* IEEE, 1–6.

[17] Bai, S. (2022). Boston house price prediction: Machine Learning. In: *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP).* 1678–1684.

[18] Ding, H. (2024). Predicting Boston Housing Price Using Machine Learning Models. In: *2024 2nd International Conference on Management Innovation and Economy Development (MIED 2024).* Atlantis Press, 439–444.

[19] Sinha, H. (2024). Benchmarking ML Models for Boston House Price Prediction. *International Journal of Research and Analytical Reviews (IJRAR)*, 11(3).

[20] Zhang, B. (2024). Analysis and Forecast of Influencing Factors of House Price in Boston. *International Journal of Research and Analytical Reviews (IJRAR)*, 11(3), 123–132.

[21] Kaggle. *Boston Housing Dataset.* Available at:
urlhttps://www.kaggle.com/datasets/altavish/ boston-housing-dataset

[22] Jain, A., Singh, R., & Sharma, D. (2022). Data Preprocessing Techniques for Machine Learning: A Review. *International Journal of Computer Applications*, 184(43), 1–7.

[23] Tandon, R. (2024). The Machine Learning Based Regression Models Analysis For House Price Prediction. *International Journal of Research and Analytical Reviews (IJRAR)*, 11(3), 296–305.