

## Article

# Intelligent Feature Selection Ensemble Model for Price Prediction in Real Estate Markets

Daniel Cristóbal Andrade-Girón <sup>1</sup>, William Joel Marin-Rodriguez <sup>2,\*</sup>  and Marcelo Gumercindo Zuñiga-Rojas <sup>3</sup>

<sup>1</sup> Department of Formal and Natural Sciences, Universidad Nacional José Faustino Sánchez Carrión, Lima 15136, Peru; dandrade@unjfsc.edu.pe

<sup>2</sup> Department of Engineering Systems, Computer and Electronics, Universidad Nacional José Faustino Sánchez Carrión, Lima 15136, Peru

<sup>3</sup> Department of Social Sciences and Communication, Universidad Nacional José Faustino Sánchez Carrión, Lima 15136, Peru; mzuniga@unjfsc.edu.pe

\* Correspondence: wmarin@unjfsc.edu.pe; Tel.: +51-990455214

**Abstract:** Real estate is crucial to the global economy, propelling economic and social development. This study examines the effects of dimensionality reduction through Recursive Feature Elimination (RFE), Random Forest (RF), and Boruta on real estate price prediction, assessing ensemble models like Bagging, Random Forest, Gradient Boosting, AdaBoost, Stacking, Voting, and Extra Trees. The results indicate that the Stacking model achieved the best performance with an MAE (mean absolute error) of 14,090, MSE (mean squared error) of  $5.338 \times 10^8$ , RMSE (root mean square error) of 23,100,  $R^2$  of 0.924, and a Concordance Correlation Coefficient (CCC) of 0.960, also demonstrating notable computational efficiency with a time of 67.23 s. Gradient Boosting closely followed, with an MAE of 14,540,  $R^2$  of 0.920, and a CCC of 0.958, requiring 1.76 s for computation. Variable reduction through RFE in both Gradient Boosting and Stacking led to an increase in MAE by 16.9% and 14.6%, respectively, along with slight reductions in  $R^2$  and CCC. The application of Boruta reduced the variables to 16, maintaining performance in Stacking, with an increase in MAE of 9.8% and a  $R^2$  of 0.908. These dimensionality reduction techniques enhanced computational efficiency and proved effective for practical applications without significantly compromising accuracy. Future research should explore automatic hyperparameter optimization and hybrid approaches to improve the adaptability and robustness of models in complex contexts.



Academic Editor: Olga Kurasova

Received: 18 March 2025

Revised: 29 April 2025

Accepted: 15 May 2025

Published: 20 May 2025

**Citation:** Andrade-Girón, D.C.; Marin-Rodriguez, W.J.; Zuñiga-Rojas, M.G. Intelligent Feature Selection Ensemble Model for Price Prediction in Real Estate Markets. *Informatics* **2025**, *12*, 52. <https://doi.org/10.3390/informatics12020052>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** ensemble model; machine learning; prediction; real estate; regression; dimensionality reduction; feature selection; model optimization

## 1. Introduction

The real estate market represents an essential component of the global economy [1], substantially impacting macroeconomic and microeconomic dynamics [2,3]. The importance is evidenced by its ability to influence economic growth [4,5], mainly through changes in real estate prices [2]. A paradigmatic example of the inherent vulnerability of this market was the mortgage crisis of 2007 and 2008, which triggered a global recession [6]. That crisis revealed the structural fragility of the real estate sector [7] and its deep interconnection with the international financial system [8,9], underscoring the necessity for more accurate price forecasting models.

There is a clear need to refine predictive tools in the real estate market [10,11], particularly in contexts of accelerated growth. Accurate prediction of real estate prices would

not only reduce the risk of forming speculative bubbles and financial collapses [12,13] and promote local economic development by facilitating more efficient planning [14]. This predictive capability would drive socioeconomic growth locally and globally and support informed decision-making by governments, real estate agents, financial institutions, and market analysts [15–17].

To solve the problem, traditional models for price prediction have been applied, such as the Comparative Market Method [8], the Income Capitalization Method (Income or Rent Method) [18], the Replacement Cost or Replacement Value Method, the Automatic Valuation Method (AVM) [19], and the Dynamic Residual Value Method (Discounted Cash Flow) [20]. These models have proven effective in stable markets, where price trends typically follow predictable patterns based on historical data [21]. However, they have significant limitations, particularly in environments with high volatility or when prices react to more complex economic dynamics. Traditional models often underestimate or overestimate prices in highly speculative markets or during crises, as they fail to adequately consider macroeconomic factors, regulatory changes, large data volumes, or sudden fluctuations in supply and demand [22,23].

In recent years, the increasing availability and storage of large volumes of data have created new opportunities to tackle the complex issue of price prediction in the real estate market [24–27]. Traditional valuation models, primarily based on statistical methods, have been outperformed in accuracy and predictive capability by machine learning algorithms [28,29], which are proving to be a highly effective alternative for enhancing real estate price estimates [30,31].

Numerous empirical studies have investigated the predictive capability of various machine learning algorithms, yielding promising results in real estate price prediction [24]. Among the most notable methods are Random Forest (RF) [32,33], Support Vector Machines (SVMs) [34], Multiple Linear Regression (MLR) [35], and regularization techniques such as LASSO (Least Absolute Shrinkage and Selection Operator) [36]. Additionally, algorithms like K-Nearest Neighbors (KNNs) [37] and decision trees (DTs) [38] have demonstrated remarkable performance across different dataset configurations.

Research on machine learning algorithms applied to regression problems has gained significant importance in the real estate sector in recent years [31,39]. However, regression issues in real-world contexts often involve highly complex internal and external factors [40]. Furthermore, different machine learning algorithms exhibit considerable variations in terms of scalability and predictive performance [41,42], which presents additional challenges for their effective application in practical settings by tuning their parameters to select the best model [40,43]. Although this strategy is commonly used, it faces three critical challenges [44–46]. First, defining the “best” model is complicated when multiple candidates show comparable performance, complicating the final choice [47]. This issue is exacerbated when the algorithm is sensitive to local optima and the amount of training data are limited [48]. Second, excluding less successful models may lead to the loss of valuable information that could enhance prediction accuracy [49]. Third, variable selection is essential in mitigating the curse of dimensionality, as using too many features may introduce noise and redundancy, impacting model generalization [50].

As a result, research in the real estate market encounters methodological and computational challenges that demand a rigorous and systematic approach. The lack of studies examining these issues highlights the necessity of developing innovative strategies that effectively enhance models’ predictive accuracy and generalizability, ensuring their relevance in complex and dynamic scenarios within the real estate sector [51,52].

In this context, this study aimed to develop and compare ensemble models with optimized feature selection for price prediction in real estate markets. The proposed

methodology integrates multiple base algorithms and employs advanced dimensionality reduction strategies, including RF, Recursive Feature Elimination (RFE), and Boruta, to identify and retain only contributing variables. This approach aims to maximize predictive accuracy, model robustness, and generalizability, ensuring optimal performance in highly complex and structurally variable environments.

## 2. Literature Review

An extensive literature review on price prediction in real estate markets was conducted, emphasizing machine learning algorithms. The analysis identified and assessed the most pertinent scientific articles in this field. In this context, Park et al. [53] developed a housing price prediction model utilizing C4.5, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), Naïve Bayes, and AdaBoost, comparing their classification accuracy. The results indicate that RIPPER surpasses the other algorithms in predictive performance.

In the same line of research, Varma et al. [54] implemented a weighted average approach using multiple regression techniques to enhance the accuracy of predicting real estate values. This method reduces error and outperforms individual models in stability and accuracy, optimizing valuations by incorporating environmental contextual information. Conversely, Rafiei et al. [55] developed an innovative approach based on a Restricted Boltzmann machine (RBM) combined with a genetic algorithm that does not involve mating. Their model evaluates the viability of the real estate market by incorporating economic variables, seasonality, and temporal effects, achieving computationally efficient optimization on standard workstations. The validity and applicability of the model were verified through a case study, demonstrating its accuracy in predicting real estate prices.

An effective strategy to tackle this problem is to implement ensemble methods, which combine multiple base models to enhance the stability and generalizability of the final model [56]. This approach addresses the individual limitations of each algorithm, reducing both variance and bias, thereby minimizing the risk of overfitting and optimizing predictive accuracy in real estate price estimation [57]. In this context, models based on ensemble methods, such as RF, have demonstrated superior performance in predicting prices in real estate markets due to their ability to decrease variance, improve stability, and capture nonlinear relationships in the data [32].

In a related study, Phan [45] evaluated the performance of SVM, RF, and Gradient Boosting Machines (GBMs) in real estate valuation in Hong Kong, discovering that RF and GBMs surpassed SVM regarding accuracy. However, while SVM exhibited lower performance, it excelled in efficiency for rapid predictions. The authors concluded that machine learning presents a promising alternative for property valuation.

Arabameri et al. [57] employed data mining techniques and regression models, including LASSO, RF, and GBMs, to predict prices in the Wroclaw real estate market, achieving 90% accuracy and demonstrating these approaches' effectiveness in modeling real estate prices. Regarding the enhancement in capturing nonlinear relationships with high accuracy, Ribeiro and dos Santos [58] evaluated several machine learning techniques for real estate price prediction, highlighting the superiority of RF over traditional models such as linear regression and SVM. Furthermore, the authors discussed potential future advancements in real estate estimation.

Adetunji et al. [33] obtained concordant results by employing RF to predict housing prices using the Boston dataset (UCI). Their model achieved an accuracy margin of  $\pm 5\%$ , demonstrating its effectiveness in estimating individual real estate values. Ensemble methods like Bagging, GBMs, and Extreme Gradient Boosting (XGBoost) have shown promising outcomes in predicting real estate prices. They are noted for their capacity to enhance predictive accuracy and model robustness across various contexts [42,58,59].

Sibindi et al. [59] evaluated the effectiveness of XGBoost in predicting real estate prices, achieving an accuracy of 84.1% compared to the 42% obtained by hedonic regression. Their study, which analyzed 13 variables, reaffirms the superiority of machine learning approaches over traditional models in this field. Similarly, Gonzales [60] examined land price prediction in Seoul from 2017 to 2020 using Random Forest (RF) and XGBoost. This study considered 21 variables and assessed their impact on predictive accuracy. The results indicated that XGBoost outperformed RF, demonstrating greater generalizability and accuracy in estimating real estate values.

Meanwhile, Sankar et al. [61] applied regression techniques and assembly methods to predict real estate prices, considering key variables such as location and demographics. The models developed achieved an accuracy of 94%, evidencing their usefulness in real estate investment decision-making. Consistent with these findings, Kumkar et al. [62] conducted a comprehensive comparative analysis of ensemble methods, including Bagging, RF, Gradient Boosting, and XGBoost, in Mumbai real estate valuation, focusing on hyperparameter optimization. To improve the accuracy of the models, they implemented advanced data preprocessing techniques, which mitigated biases and reduced variance [63]. The results confirm that ensemble methods are robust and efficient tools for price estimation in real estate markets characterized by high complexity and volatility [59].

In a bibliometric analysis, Takouabou et al. [10] examined 70 articles indexed in Scopus, noting that scientific production in this field primarily comes from the USA, China, India, Japan, and Hong Kong—regions known for their high levels of digitization and well-established research ecosystems. However, the review revealed recurring methodological limitations, such as dependence on small datasets and a preference for straightforward machine learning models, complicating the issue of the curse of dimensionality. These limitations highlight the need to integrate more advanced methodologies that enhance predictive performance and improve the interpretability of models applied to the real estate market.

In recent years, machine learning techniques for predicting real estate prices have significantly advanced, enhancing the models' accuracy and efficiency. However, critical challenges remain that restrict these methods' interpretability, generalizability, and robustness, particularly in situations with vast data volumes and high market heterogeneity. One major gap in the literature is the identification, selection, and optimization of relevant variables for predictive models in real estate. Although advancements in machine learning have enabled the development of more sophisticated approaches, challenges remain in integrating multiple data sources and addressing biases stemming from the structural complexity of real estate markets.

In recent years, machine learning techniques for real estate price prediction have advanced significantly, improving the accuracy and efficiency of predictive models. Despite the increasing adoption of machine learning methods in forecasting prices and behaviors within the real estate market, critical challenges persist that limit the interpretability, generalizability, and robustness of these models [26]. These challenges are especially pronounced in contexts characterized by large data volumes, high market heterogeneity, and structural complexity in the factors influencing property value [20].

One of the main knowledge gaps identified in the current literature is selecting and optimizing relevant features [21]. This issue not only compromises model accuracy but also constrains the replicability and transferability of models to other geographical or socioeconomic contexts. The literature indicates that, although there are algorithms capable of ranking variables to some extent, these approaches still face limitations when addressing nonlinear interactions, latent variables, and multiscale contextual effects [38].

Another critical challenge is integrating multiple data sources, such as urban cadastres, public records, satellite imagery, socioeconomic indicators, and transactional databases [44]. This integration requires advanced data fusion, normalization, and hierarchical modeling techniques, which have not yet been standardized or empirically validated in the real estate domain. Consequently, structural biases and multicollinearity issues persist, undermining the quality of inferences [54].

This gap in the literature underscores the need for research to develop integrated methodological frameworks that concurrently address feature selection, heterogeneous data integration, and bias mitigation. Such frameworks are essential to ensure more transparent, replicable, and effective predictive models.

In this context, this study tackles this gap by creating a model based on ensemble methods while optimizing the selection of key variables with advanced machine learning techniques. The goal is to enhance predictive accuracy and reduce potential biases, thereby contributing to the development of more robust and generalizable models in real estate valuation.

### 3. Methodology

This study follows the positivist paradigm, employs a quantitative approach, uses experimental design, and functions at a predictive level. As detailed below, the methodology for developing the machine learning model describes the design of machine learning systems.

#### 3.1. Data Acquisition

The Ames Housing dataset, utilized in various competitions organized by Kaggle ([https://www.kaggle.com/datasets/ahmedmohameddawoud/ames-housing-data?select=Ames\\_Housing\\_Data.csv](https://www.kaggle.com/datasets/ahmedmohameddawoud/ames-housing-data?select=Ames_Housing_Data.csv), accessed on 20 October 2024), was employed. This dataset contains structured information on 2930 real estate transactions recorded in Ames, Iowa. Eighty-one variables are included, 79 of which pertain to structural, spatial, and qualitative attributes of the properties, while the dependent variable (SalePrice) represents the sales price in U.S. dollars. The included variables described essential aspects of the real estate market, such as the location and zoning of the lot, structural characteristics of the building, distribution of living space, and availability of additional amenities, including garages, swimming pools, and fireplaces. Details on building materials and remodeling were also captured. It was noted that the SalePrice variable exhibited an asymmetric distribution, necessitating the application of mathematical transformations to enhance the stability and fit of the regression models. Methodologically, we dealt with a high-dimensional dataset, which facilitated the implementation of advanced feature selection and dimensionality reduction techniques.

#### 3.2. Preprocessing

An exploratory analysis was carried out with the functions `pandas.info()` and `pandas.describe()`. These tools allowed the identification of null values, data type inconsistencies, and possible distribution anomalies. For data integration, merging and concatenation techniques were implemented, ensuring structural coherence through standardizing variable names and types [62]. Missing values were managed through elimination and imputation strategies with `SimpleImputer`, applying methods based on the mean, median, or mode, depending on the nature of each variable. Outlier detection was performed using Z-score and interquartile range, while redundant records were eliminated using `drop_duplicates()`, ensuring the integrity of the dataset [63].



Categorical variables were transformed using one-hot encoding through OneHotEncoder and pd.get\_dummies(), optimizing their representation for integration into machine learning models. Normalization and scaling techniques were utilized to enhance data homogeneity: StandardScaler for standardization. These strategies enhanced the model's stability and reduced the impacts of data heterogeneity, ensuring more accurate and robust predictions for estimating real estate prices.

### 3.3. Selection Methods

The importance of the features was evaluated using the varImp() function in the Python language (Version: 3.11.12), which employs the RF algorithm to estimate the relative contribution of each variable in the model prediction. This method allowed us to identify influential characteristics by randomly permuting each variable and measuring the decrease in accuracy or increase in impurity. The RFE technique was implemented to optimize the variable selection using sklearn.feature\_selection.RFE, which iteratively eliminates those variables with lower relevance until an optimal subset [64,65] is reached. Additionally, the Boruta algorithm was incorporated using BorutaPy, an extension of RF based on statistical hypothesis testing for feature selection. Boruta compares each variable with randomized synthetic attributes (shadow features), providing a robust criterion for their relevance. Its ability to capture nonlinear interactions and relationships makes it a key tool for improving the interpretability and generalization of models in high-dimensional environments [66]. In constructing decision trees within the RF model, the quality of each partition was evaluated using impurity metrics, which quantify the reduction of heterogeneity after each split. This criterion ensured that the groups formed were as homogeneous as possible, optimizing the model's predictive capacity. In this context, the impurity function  $i(t)$  measures the heterogeneity of the node  $t$ , and the optimal partition is selected by maximizing the impurity reduction, given by the following equation:

$$\Phi(s, t) = i(t) + P_R i(t_R) - P_L i(t_L)$$

where

- $\Phi(s, t)$  is the goodness of partitioning at node  $t$  using the partitioning criterion  $s$ .
- $i(t)$  is the impurity of the parent node.
- $i(t_R)$  and  $i(t_L)$  are the impurities of the child nodes after partitioning.
- $P_R$  and  $P_L$  represent the proportions of data assigned to the right and left nodes, respectively.

The most commonly used impurity functions in decision trees are as follows:

Gini index ( $Gini(t)$ ):

The Gini index measures the probability that an element is incorrectly classified if it is randomly chosen according to the distribution of classes in the node.

It is calculated as follows:

$$Gini(t) = 1 - \sum_{i=1}^c p_i^2$$

where  $p_i$  is the proportion of elements of the class  $t$  in the node.

Entropy ( $H(t)$ ):

Entropy measures the amount of information held in the node.

It is defined as follows:

$$H(t) = - \sum_{i=1}^c p_i \log_2 p_i$$

Evaluating the goodness of partitioning and impurity in RF allows optimal variable selection using well-founded mathematical criteria. Methods such as RFE and Boruta use this structure to select the most relevant variables, optimize predictive models, and reduce the problem's dimensionality.

### 3.4. Learning Algorithm Selection

Ensemble models are chosen for land price prediction because they can improve the accuracy and robustness of predictions by combining the results of multiple base models. This is especially advantageous in applications such as real estate appraisal, where data can have high variability and nonlinear characteristics.

#### 3.4.1. AdaBoost

AdaBoost for regression, known as AdaBoost.R, extends the boosting methodology by minimizing a continuous error instead of a binary ranking function [67]. At each iteration, a weak regressor is trained on a dynamically adjusted weight distribution according to the absolute or quadratic error of the prediction [68]. The error is normalized and used to compute a weighting coefficient, where regressors with lower error have a more significant influence on the final combination [69]. AdaBoost (Adaptive Boosting) optimizes the error by weighting instances that are difficult to classify. In regression, an exponential cost function is minimized by iteratively updating weights:

$$F_m(x) = \sum_{m=1}^M \alpha_m h_m(x),$$

where  $h_m(x)$  are base estimators (commonly weak regressors such as decision trees), and  $\alpha_m$  represents their weighting coefficients.

#### 3.4.2. Gradient Boosting

Gradient Boosting is a boosting-based machine learning method that optimizes an additive model by minimizing the loss function using downward gradients in the functional space [70]. It aims to build a strong model as a sequential combination of weak models, where each new model is designed to correct the errors of the previous models as follows [71]:

$$F_m(x) = F_{m-1} + \gamma_m h_m(x)$$

where  $\gamma_m$  is the learning coefficient obtained by minimizing the following:

$$\gamma_m = \operatorname{argmin}_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i)) + \gamma h_m(x_i)$$

Unlike AdaBoost, which adjusts sample weights according to their error, Gradient Boosting builds a sequential model by fitting each new estimator to the residual gradients. This formulation allows flexibility using various loss functions, such as mean square error for regression or classification [59].

#### 3.4.3. Random Forest Regressor

Random Forest Regressor is a decision tree-based ensemble method that improves prediction accuracy and stability by combining multiple trees trained on random subsets of the data [57]. The central idea of RF is to reduce the variance of individual models by taking advantage of the diversity of multiple predictors, which makes it less prone to overfitting compared to a single decision tree as follows [56]:

$$F(x) = \frac{1}{T} \sum_{t=1}^T h_t(x)$$

where  $h_t(x)$  are individual regression trees trained with bootstrap sampling.

#### 3.4.4. Extra Trees Regressor

Extra Trees Regressor is a variant of RF that introduces more randomization into the construction of the trees, reducing the variance of the model and improving its robustness to noise in the data [72]. Unlike RF, where the split thresholds in each node are selected by optimization, Extra Trees assigns these thresholds completely randomly within the subset of selected characteristics.

$$F(x) = \frac{1}{B} \sum_{t=1}^B h_b(x)$$

The function  $h_b(x)$  denotes the prediction made by the  $b$ -th tree. This strategy introduces a higher bias compared to RF but reduces the variance and correlation between the trees, improving generalizability.

#### 3.4.5. Bagging Regressor

Bagging Regressor is an ensemble method that decreases the variance of the base models by training them on multiple subsets of data generated by sampling with replacement and combining their predictions [73]. Each base model  $h_b(x)$  is fitted independently, and the final ensemble prediction is obtained as the average of the individual predictions as follows:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B h_b(x)$$

where each  $h_b(x)$  is a base model trained on a random subset of the data.

The relationship explains the impact of Bagging on variance reduction:

$$Var(\hat{y}) = \frac{1}{B} Var(h_b) + \left(1 - \frac{1}{B}\right) Cov(h_b, h_{b'})$$

where the decrease in variance is more significant when the model's  $h_b$  values are less correlated. Bagging improves model stability and generalizability, reducing overfitting without significantly increasing bias [58].

#### 3.4.6. Stacking Regressor

Stacking Regressor is an ensemble method that combines multiple regression models in a two-level hierarchical architecture to improve predictive ability [74]. Its theoretical foundation lies in the optimal combination of base estimators using a meta-model that learns to correct its biases and exploit its strengths:

$$\hat{y} = g(h_1(x), h_2(x), \dots, h_M(x)),$$

where  $g$  is a meta-model trained with the outputs of the base regressor  $sh_m(x)$ . Stacking can be seen as an optimal predictor combination problem, where the meta-model learns to minimize the loss function  $L(y, \hat{y})$  more efficiently than a simple aggregation by averaging (as in Bagging) or weighting (as in Voting). Its flexibility allows the integration of heterogeneous models with different biases and variances, achieving a more robust ensemble [75].

#### 3.4.7. Voting Regressor

Voting Regressor is an ensemble method that combines the predictions of multiple regression models to improve stability and generalizability [76,77]. Its underlying principle is that by merging several estimators with different biases and variances, a more robust



prediction is obtained that is less sensitive to fluctuations in the training data. Voting by simple averaging, one obtains the following:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B h_b(x)$$

where all contributions have the same weight.

Weighted voting:

$$\hat{y} = \sum_{b=1}^B w_b h_b(x), \quad \sum_{b=1}^B w_b = 1,$$

Each model  $h_b$  receives a weight  $w_b$  proportional to its performance, which is usually determined by validation metrics such as the coefficient of determination  $R^2$  [78].

### 3.5. Model Training

The program was developed using the *Scikit-learn* (Version: 1.6.1) Python package, employing a rigorous approach centered on ensemble models and *k-fold* cross-validation ( $k = 10$ ) to ensure optimal training and evaluation. This alternating procedure offered robust evaluation, minimizing bias and variance related to specific data partitions while reducing the risk of overfitting [79]. A sample split of 70% for training and 30% for testing was employed, ensuring final model evaluation on fully independent data. During training, ensemble models iteratively optimize errors by combining predictions from multiple base models to enhance accuracy and reduce bias. In addition, a systematic hyperparameter optimization process was implemented using techniques such as Grid Search and Randomized Search, integrated into the cross-validation framework. This allowed the identification of optimal parameter combinations that maximize model performance while balancing complexity and generalization capacity.

### 3.6. Model Evaluation

To evaluate the accuracy, robustness, and computational efficiency of the predictive models for land prices, we employed a set of standard regression performance metrics, including mean squared error (MSE), mean absolute error (MAE), and the coefficient of determination ( $R^2$ ). Additionally, we incorporated Lin's Concordance Correlation Coefficient (CCC) to assess the degree of agreement between predicted and observed values, capturing both precision and accuracy. To complement the performance analysis, we also reported the execution time of each model as a proxy for computational efficiency, providing a more holistic comparison across different model configurations and feature selection strategies.

#### 3.6.1. Coefficient of Determination ( $R^2$ )

Statistical metric to evaluate the quality of a regression model. It indicates what proportion of the variability in the dependent variable ( $y$ ) is explained by the model as a function of the independent variables ( $X$ ). The formula for  $R^2$  is as follows:

$$R^2 = 1 - \text{SSE}/\text{SST}$$

where SSE is the sum of squared errors ( $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ ), which represents the variability not explained by the model; SST is the total sum of squares ( $\sum_{i=1}^n (y_i - \bar{y})^2$ ), representing the total variability in the real data; and  $\bar{y}$  is the mean of the real values ( $y_i$ ).

### 3.6.2. Mean Squared Error (MSE)

Mean Squared Error is an evaluation metric that measures the average squared difference between the actual values ( $y_i$ ) and the model predictions ( $\hat{y}_i$ ); the formula is as follows:

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|^2$$

where  $n$  is the total number of observations;  $y_i$  is the actual value of the  $i$ -th observation; and  $\hat{y}_i$  is the value predicted by the model for the  $i$ -th observation.

### 3.6.3. Root Mean Squared Error (RMSE)

Root mean squared error is an evaluation metric used mainly in regression models. It represents the square root of the average of the squared errors between the actual values  $y_i$  and the predictions of the model  $\hat{y}_i$ ; the formula is as follows:

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

### 3.6.4. Mean Absolute Error (MAE)

Mean absolute error is a metric used to evaluate the accuracy of a model in regression tasks. It represents the average of the absolute differences between the actual values ( $y_i$ ) and the predictions of the model ( $\hat{y}_i$ ). The formula is as follows:

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

## 4. Results and Discussion

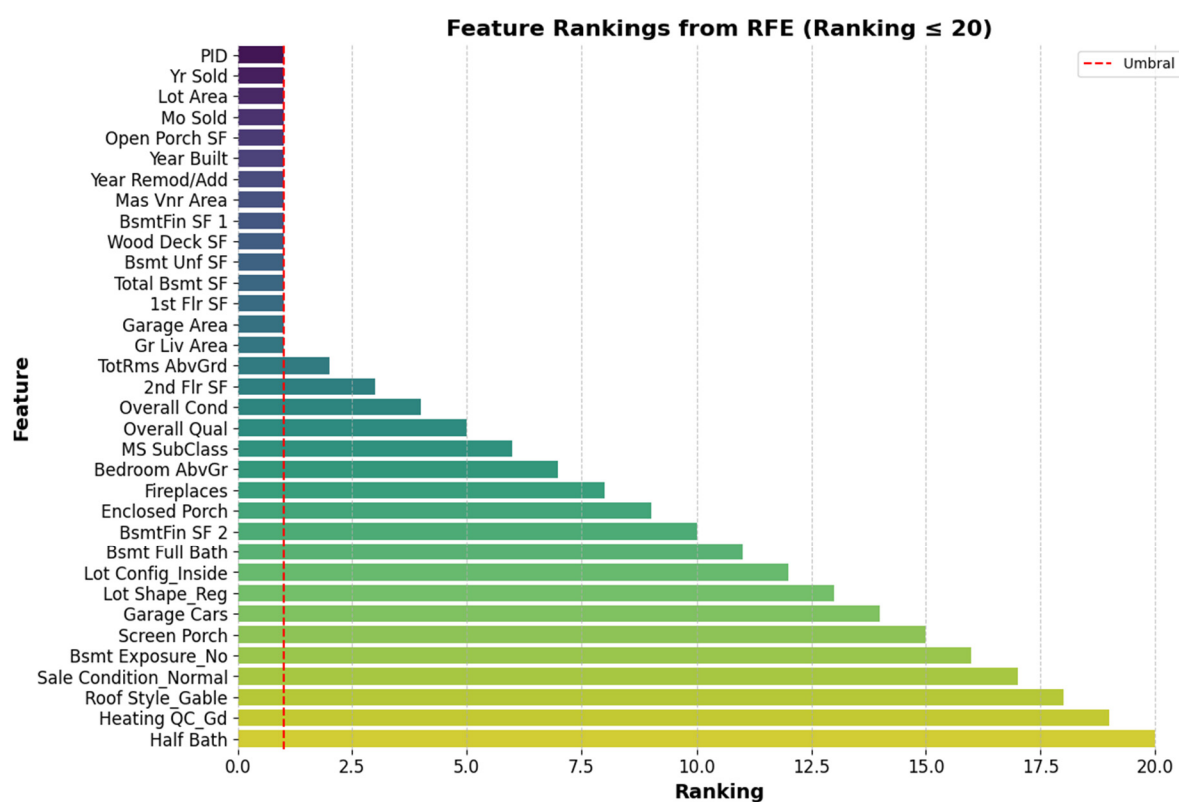
Accurate estimation of real estate prices is crucial for investment and urban planning [16,17], given the sector's significance in the global economy [1,15]. Dimensionality reduction improves computational efficiency, enhances model interpretability, and reduces overfitting, resulting in more robust predictions [49,50]. The comparative analysis of ensemble models in machine learning and dimensionality reduction techniques, such as RFE, RF, and Boruta [65,66], shows that effective feature selection maximizes the predictive accuracy and stability of the model, confirming its efficacy in predicting real estate prices.

A dataset from the Kaggle machine learning repository (<https://www.kaggle.com/datasets/ahmedmohameddawoud/ames-housing-data>) (accessed on 20 October 2024) was utilized. The Ames Housing dataset, which is well known in the scientific literature and machine learning competitions, contains detailed information on 2930 homes in Ames, Iowa, described by 81 variables, including structural, spatial, and contextual characteristics. A thorough process of preprocessing and data cleaning was implemented to predict the properties' sale prices. First, the dataset's structure was examined to evaluate the types of variables and the existence of missing values. A threshold of 5% was set for the removal of variables with a high proportion of missing data, ensuring a balance between reducing bias and preserving information. Consequently, the database was narrowed down to 69 variables.

To mitigate the impact of null values, we applied differentiated imputation strategies. We used the median to reduce sensitivity to outliers for numerical variables, while for categorical variables, we employed the mode to maintain class coherence. Additionally, we identified and removed duplicate records to minimize potential redundancies in the analysis. We transformed categorical variables using one-hot encoding, ensuring appropriate numerical representation and avoiding collinearity issues by excluding the first category

of each variable. Next, we performed a normalization process for the numerical variables using StandardScaler, ensuring each variable had a distribution with a zero mean and a unit standard deviation. Data preprocessing resulted in a final dataset of 2930 records and 227 variables derived from the transformation of categorical variables through one-hot encoding. This strategy provided a structured and fitting representation for subsequent analysis, facilitating the application of feature selection techniques and machine learning models with an optimized and uniform database.

The research evaluated the impact of the RFE algorithm in conjunction with an RF classifier to identify the most relevant variables in a structured dataset. The results indicated that the progressive elimination of irrelevant features reduced the complexity of the model without significantly affecting its accuracy. The selected features had the highest contribution to predicting the target variable. Additionally, an importance ranking was generated, allowing visualization of the relative influence of each variable in the model, as shown in Figure 1.



**Figure 1.** Application of the RFE method for the selection of variables. The bar colors indicate the relative importance of features according to the ranking obtained by the Recursive Feature Elimination (RFE) algorithm, where darker tones correspond to higher importance (lower ranking) and lighter tones to lower importance within the top 20.

A database was reduced to fifteen variables after conducting attribute engineering (feature selection). For the comparative analysis, seven widely recognized ensemble models were evaluated: AdaBoost, Gradient Boosting, Random Forest, Extra Trees, Bagging, Stacking, and Voting. These were chosen for their ability to enhance prediction accuracy by combining multiple base models, thus optimizing the robustness and generalization of the results.

The predictive performance of each model was quantified using standard metrics for regression problems. Among these metrics, the coefficient of determination ( $R^2$ ) indicated the proportion of variability in the data explained by the model, serving as a crucial

indicator of its fit. The root mean square error (RMSE) directly measures the magnitude of error and is expressed in the same units as the dependent variable, allowing for intuitive interpretation of precision. The mean absolute error (MAE) assesses the average absolute deviation between predictions and actual values, providing a complementary metric that does not excessively penalize extreme errors. Lastly, the mean squared error (MSE) was used to impose stricter penalties on significant errors, offering additional insight into model quality by identifying highly discrepant predictions.

The analysis of these metrics enabled a thorough comparison of the models, revealing their strengths and limitations in predicting real estate prices. The observed differences underscore the importance of choosing the correct algorithm based on the problem's specific characteristics and the analysis goals.

Table 1 presents the results for each model, along with a critical analysis that evaluates their relative performance in terms of accuracy, robustness, and generalizability. This analysis establishes a solid foundation for identifying the optimal configurations for practical applications and future research.

**Table 1.** Performance of machine learning ensemble models without variable selection and with variable selection through RFE.

Model	Features	MAE	MSE	RMSE	R <sup>2</sup>	CCC	Time (seg)
AdaBoost	227	23,630	$1.044 \times 10^9$	32,320	0.851	0.918	1.709
RFE + AdaBoost	15	25,390	$1.174 \times 10^9$	34,260	0.833	0.904	0.757
Gradient Boosting	227	14,540	$5.643 \times 10^8$	23,750	0.920	0.958	1.756
RFE + Gradient Boosting	15	17,450	$6.754 \times 10^8$	25,990	0.904	0.947	1.231
Random Forest	227	15,190	$6.066 \times 10^8$	24,630	0.914	0.953	6.445
RFE + Random Forest	15	17,540	$7.455 \times 10^8$	27,300	0.894	0.943	2.992
Extra Trees	227	15,480	$6.574 \times 10^8$	25,640	0.906	0.945	5.409
RFE + Extra Trees	15	16,760	$6.783 \times 10^8$	26,040	0.904	0.946	1.100
Bagging	227	15,130	$6.391 \times 10^8$	25,280	0.909	0.950	1.964
RFE + Bagging	15	17,430	$7.575 \times 10^8$	27,520	0.892	0.939	0.761
Stacking	227	14,090	$5.338 \times 10^8$	23,100	0.924	0.960	67.231
RFE + Stacking	15	16,510	$6.473 \times 10^8$	25,440	0.908	0.951	20.946
Voting	227	15,720	$5.971 \times 10^8$	24,440	0.915	0.953	13.433
RFE + Voting	15	17,510	$6.940 \times 10^8$	26,340	0.901	0.944	4.752

This study evaluated the performance of various machine learning models in predicting land prices and analyzed the impact of dimensionality reduction using Recursive Feature Elimination (RFE). Baseline models employing 227 features were compared with their optimized counterparts, in which 15 features were selected through RFE.

The results indicate that the Stacking model achieved the highest overall performance, with a Mean Absolute Error (MAE) of 14,090, a Mean Squared Error (MSE) of  $5.338 \times 10^8$ , a Root Mean Squared Error (RMSE) of 23,100, an R<sup>2</sup> of 0.924, and a Lin's Concordance Correlation Coefficient (CCC) of 0.960. It was closely followed by Gradient Boosting, which recorded an MAE of 14,540 and an R<sup>2</sup> of 0.920, highlighting strong predictive accuracy and generalization capabilities. In contrast, the Adaboost and RFE + Adaboost models demonstrated the weakest performance, with MAE values exceeding 23,000 and R<sup>2</sup> scores below 0.85, suggesting difficulties in adequately fitting the training data.

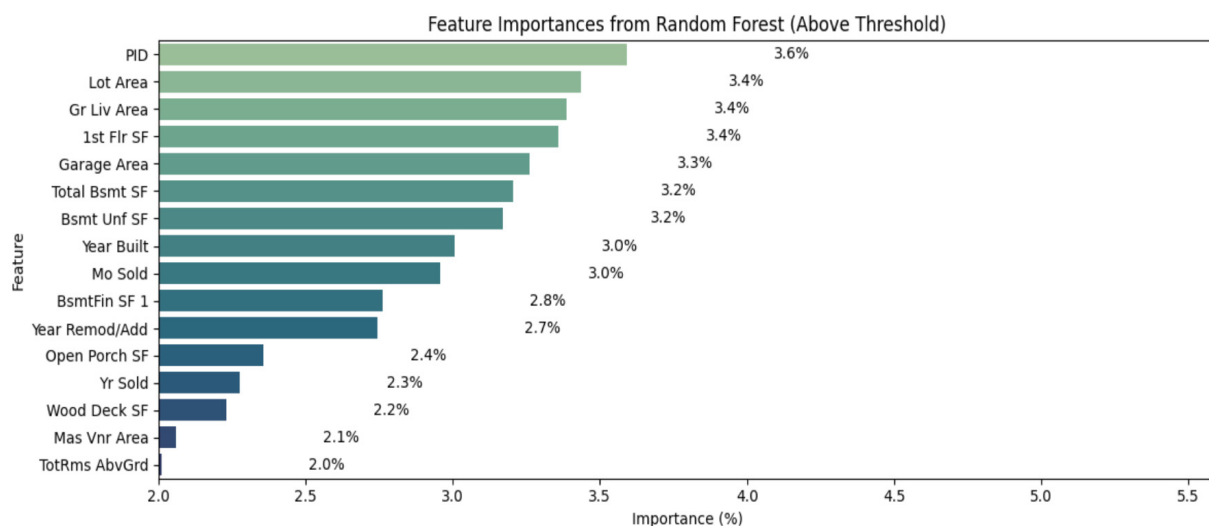
Including the CCC metric provided a comprehensive assessment of both the precision and accuracy of the models' predictions, while execution time served as a robust indicator of computational efficiency. In this regard, models such as RFE + Bagging (0.76 s) and RFE + AdaBoost (0.76 s) demonstrated significantly reduced computation times compared to more complex models such as Stacking (67.23 s) and Voting (13.43 s).

Although dimensionality reduction enhances efficiency and interpretability, the findings suggest that feature elimination through RFE did not consistently improve predictive accuracy. Specifically, the Gradient Boosting model without RFE significantly outperformed its reduced counterpart, achieving a 16.9% reduction in MAE and a 1.6% increase in  $R^2$ . A similar pattern was observed in the Stacking model, where the full-feature version attained a 14.6% lower MAE than its RFE-optimized version, implying that the removed features may have contained relevant information for the prediction task.

This trend was also observed in models such as Random Forest, Extra Trees, Bagging, and Voting, where RFE increased prediction errors (MAE, MSE, and RMSE) and led to a decline in  $R^2$ . This confirms that these algorithms benefit from access to more input features. Nevertheless, dimensionality reduction resulted in significant improvements in execution time, which is a critical advantage in resource-constrained environments.

From a computational perspective, RFE-based feature selection can be an effective strategy for balancing accuracy and efficiency. In scenarios where model interpretability is crucial—such as in regulatory environments or explanation-driven decision-making—limiting the input to 15 features enhances the understanding of each variable's impact on the model without significantly compromising predictive performance.

The method outlined involves calculating the significance of each variable using RF. Figure 2 illustrates the results of the variable selection.



**Figure 2.** Application of the RF method for the selection of variables. The color intensity of the bars reflects the relative importance of each feature as determined by the Random Forest model. Lighter shades indicate higher predictive importance, whereas darker tones represent lower importance, though still above the 2% threshold. This color encoding enhances the visual interpretation of the most influential predictors in the model.

The results obtained in the evaluation of different ensemble algorithms for price prediction show apparent differences in their performance, according to the indicators of root mean square error (RMSE), explained variance (EV), mean absolute error (MAE), and coefficient of determination ( $R^2$ ). The results are presented in Table 2.

**Table 2.** Performance of machine learning ensemble models without variable selection and with variable selection by RF.

Model	Features	MAE	MSE	RMSE	R <sup>2</sup>	CCC	Time (seg)
AdaBoost	227	23,630	$1.044 \times 10^9$	32,320	0.851	0.918	1.709
RF + AdaBoost	16	24,370	$1.124 \times 10^9$	33,530	0.840	0.901	0.705
Gradient Boosting	227	14,540	$5.643 \times 10^8$	23,750	0.920	0.958	1.756
RF + Gradient Boosting	16	17,310	$6.725 \times 10^8$	25,930	0.904	0.947	1.261
Random Forest	227	15,190	$6.066 \times 10^8$	24,630	0.914	0.953	6.445
RF + Random Forest	16	17,170	$7.265 \times 10^8$	26,950	0.897	0.944	2.057
Extra Trees	227	15,480	$6.574 \times 10^8$	25,640	0.906	0.945	5.409
RF + Extra Trees	16	16,780	$6.684 \times 10^8$	25,850	0.905	0.945	1.109
Bagging	227	15,130	$6.391 \times 10^8$	25,280	0.909	0.950	1.964
RF + Bagging	16	17,650	$7.408 \times 10^8$	27,220	0.895	0.939	0.742
Stacking	227	14,090	$5.338 \times 10^8$	23,100	0.924	0.960	67.231
RF + Stacking	16	16,590	$6.388 \times 10^8$	25,280	0.909	0.951	22.465
Voting	227	15,720	$5.971 \times 10^8$	24,440	0.915	0.953	13.433
RF + Voting	16	17,390	$6.870 \times 10^8$	26,210	0.902	0.944	4.005

This study evaluated the performance of various machine learning models in predicting real estate prices. It compared full-feature versions (227 features) with their respective reduced versions obtained through Random Forest-based feature selection (RF-FS), which limited the input space to 16 variables. The analysis focused on the trade-off between predictive accuracy and computational efficiency resulting from dimensionality reduction.

Among all models, Stacking achieved the best overall performance, with a Mean Absolute Error (MAE) of 14,090, a Mean Squared Error (MSE) of  $5.34 \times 10^8$ , a Root Mean Squared Error (RMSE) of 23,100, and a coefficient of determination ( $R^2$ ) of 0.924, along with Lin's Concordance Correlation Coefficient (CCC) of 0.960, confirming its robustness and excellent generalization capacity. The Gradient Boosting model followed closely, with an MAE of 14,540 and an  $R^2$  of 0.920, further supporting its high predictive power.

The impact of RF-based dimensionality reduction varied significantly across models. In the case of Gradient Boosting, the RF-optimized version exhibited a 19.1% increase in MAE and a 1.6% reduction in  $R^2$  compared to its full-feature counterpart. A similar pattern was observed in the Stacking model, where the reduced version resulted in a 17.7% increase in MAE, suggesting that key predictive variables may have been excluded during feature selection.

Other ensemble models, such as Random Forest and Extra Trees, also experienced decreased accuracy after dimensionality reduction. Specifically, RF + Random Forest increased MAE by 13.1%, while RF + Extra Trees increased MAE by 8.3%, indicating that reducing the feature space negatively affected these models' predictive capabilities. Conversely, RF + AdaBoost slightly outperformed the full-feature AdaBoost version, yielding a 5.3% reduction in MAE, which suggests that RF-FS may help mitigate overfitting and enhance performance in models prone to overparameterization.

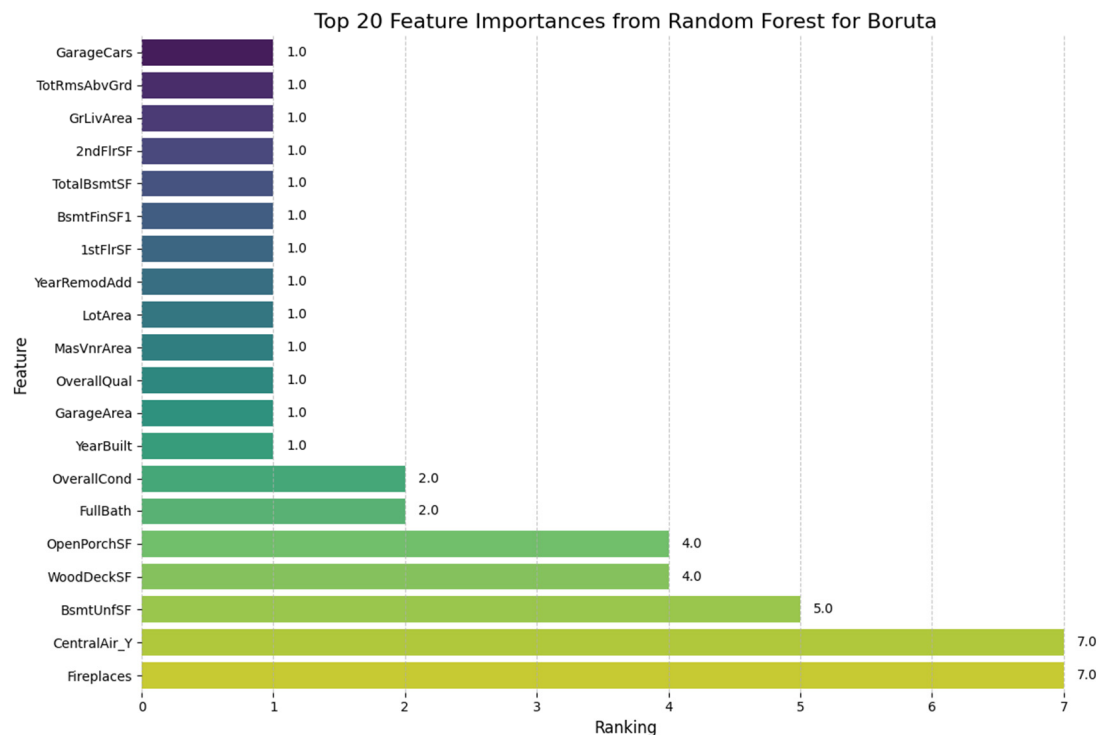
From a computational perspective, dimensionality reduction through RF-FS led to significant reductions in execution time across all models. For instance, RF + Bagging and RF + AdaBoost required only 0.74 s and 0.70 s, respectively, compared to 1.96 s for full Bagging and 1.71 s for full AdaBoost. The computational gains were even more pronounced in complex models such as Stacking, where runtime decreased from 67.23 s to 22.47 s after feature selection, and Voting, from 13.43 s to 4.00 s.

Despite occasional losses in predictive accuracy, the RF-FS optimized models maintained acceptable performance with only 16 features, compared to their original versions with 227 features. These findings highlight the value of RF-based feature selection as a



method to reduce computational complexity and enhance interpretability, particularly in resource-constrained environments or fields where explainability is crucial.

Concerning the Boruta model, 16 variables were selected based on the ranking established, as shown in Figure 3.



**Figure 3.** Application of the Boruta method for the selection of variables. Color intensity indicates relative importance: darker shades represent more important features (lower ranking), while lighter shades indicate less important ones (higher ranking).

Similar to the prior models, the performance of several machine learning models in predicting land prices was assessed using metrics such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and the coefficient of determination ( $R^2$ ), as shown in Table 3.

**Table 3.** Performance of machine learning ensemble models without variable selection and with variable selection using Boruta.

Model	Features	MAE	MSE	RMSE	$R^2$	CCC	Time (seg)
AdaBoost	227	23,630	$1.04 \times 10^9$	32,317	0.851	0.918	1.710
Boruta + AdaBoost	16	23,529	$1.04 \times 10^9$	32,277	0.852	0.920	0.560
Gradient Boosting	227	14,537	$5.64 \times 10^8$	23,755	0.920	0.958	1.760
Boruta + Gradient Boosting	16	16,073	$6.53 \times 10^8$	25,555	0.907	0.950	0.830
Random Forest	227	15,186	$6.07 \times 10^8$	24,629	0.914	0.953	6.440
Boruta + Random Forest	16	15,841	$6.86 \times 10^8$	26,195	0.902	0.949	1.750
Extra Trees	227	15,476	$6.57 \times 10^8$	25,639	0.907	0.945	5.410
Boruta + Extra Trees	16	15,899	$7.23 \times 10^8$	26,893	0.897	0.946	1.030
Bagging	227	15,134	$6.39 \times 10^8$	25,281	0.909	0.950	1.960
Boruta + Bagging	16	15,576	$6.87 \times 10^8$	26,204	0.902	0.946	1.120
Stacking	227	14,092	$5.34 \times 10^8$	23,104	0.924	0.960	67.230
Boruta + Stacking	16	15,472	$6.45 \times 10^8$	25,401	0.908	0.951	21.970
Voting	227	15,724	$5.97 \times 10^8$	24,436	0.915	0.953	13.430
Boruta + Voting	16	16,228	$6.58 \times 10^8$	25,658	0.906	0.949	3.760

This study systematically evaluates the influence of applying the Boruta algorithm for feature selection across various ensemble-based machine learning models. The original models were trained using 227 predictors, while their optimized versions utilized a reduced subset of 16 variables selected by Boruta as the most relevant. The assessment focused on predictive accuracy—measured through MAE, MSE, RMSE,  $R^2$ , and the Concordance Correlation Coefficient (CCC)—and computational efficiency, represented by training time.

The findings indicate that Boruta achieved a 92.9% reduction in the number of variables without significantly compromising predictive performance. In most models, both  $R^2$  and CCC remained close to the values obtained with the full feature set, thereby highlighting the effectiveness of the feature selection process in maintaining predictive quality with a more compact dataset.

The Stacking model emerged as the top performer in both configurations. The full model achieved an MAE of 14,092, an RMSE of 23,104, an  $R^2$  of 0.9241, and a CCC of 0.960. Its reduced counterpart (Boruta + Stacking) preserved a high level of agreement (CCC = 0.951), despite a modest decline in  $R^2$  (0.908) and an increase in MAE (15,472). Notably, computational time decreased from 67.23 to 21.97 s—a reduction of 67.3%—demonstrating that the model retains strong predictive performance even after a significant reduction in dimensionality.

Similarly, Gradient Boosting demonstrated resilience to feature reduction. The CCC slightly decreased from 0.958 to 0.9502, with a marginal increase in MAE (from 14,537 to 16,073) and a controlled decline in  $R^2$  (from 0.920 to 0.907). This stability in  $R^2$  and CCC underscores Boruta's ability to preserve key structural relationships between predictors and the target variable.

Conversely, tree-based models such as Random Forest and ExtraTrees exhibited greater sensitivity to feature reduction. In the case of Random Forest, CCC decreased from 0.953 to 0.949, while ExtraTrees saw a drop from 0.945 to 0.946. Although these changes are modest, they indicate that such models benefit from broader feature sets to maintain the diversity of decision trees. The dimensionality reduction may compromise the robustness and consistency of their predictions.

A particularly noteworthy case is the Adaboost model. The optimized version performed nearly identically to its full-feature counterpart, with CCC values of 0.920 and 0.918, MAE values of 23,529 and 23,630, and  $R^2$  values of 0.851 and 0.851, respectively. This result aligns with existing literature suggesting that Adaboost is sensitive to data noise and prone to overfitting, thereby benefiting from more selective and parsimonious input spaces. Computational time was also substantially reduced—from 1.71 to 0.56 s.

Similarly, the Bagging model demonstrated minimal performance changes following dimensionality reduction. Its CCC slightly decreased (from 0.9496 to 0.9464), accompanied by only moderate increases in MAE and RMSE. However, the training time was nearly halved (from 1.96 to 1.12 s), highlighting the advantages of feature selection in scenarios where time constraints and interpretability are critical.

Overall, incorporating the Concordance Correlation Coefficient (CCC) as an additional evaluation metric provides a more comprehensive understanding of each model's predictive capacity. Unlike  $R^2$ , which only assesses the proportion of variance explained, CCC captures both precision and accuracy, enabling a more thorough identification of potential biases or systematic misalignments in predictions.

The computational advantages of dimensionality reduction were significant across all models. This increase in efficiency is particularly valuable in situations where rapid model iteration is necessary or where computational resources are limited. Moreover, by concentrating the analysis on a smaller and more informative subset of variables, model interpretability is enhanced, which is a desirable characteristic in high-stakes applications

such as real estate valuation or medical diagnosis, where the rationale behind predictions must be clearly justified.

Dimensionality reduction significantly improved computational efficiency. Decreased variables optimized training time and model responsiveness, which is crucial in prediction applications in massive data environments [50]. Dimensionality reduction also effectively mitigated the adverse effects of the exponential increase in complexity as the number of variables in the dataset increased [47].

Feature selection enhances model accuracy by reducing dimensionality and removing irrelevant or redundant variables that introduce noise and increase the risk of overfitting. This process optimizes model variance and boosts generalization capacity, consistent with the bias–variance tradeoff. From a statistical perspective, it improves the signal-to-noise ratio and lowers model complexity, resulting in decreased estimator variance.

However, accuracy may decline if predictively relevant variables are eliminated, introducing bias and resulting in underfitting. Additionally, some selection methods struggle to capture nonlinear relationships or interactions between variables, and if they are not properly integrated within a cross-validation framework, they may lead to overfitting in the training set.

The findings indicate that no single model is universally superior for predicting real estate prices. While Gradient Boosting and Stacking stand out for their accuracy and generalizability, models like Random Forest and Extra Trees offer robustness and stability, especially when utilizing variable selection techniques. Future research could focus on automated hyperparameter optimization and incorporate hybrid methods to improve predictive performance in highly complex situations.

Accurate land value prediction using ensemble machine learning models, supported by feature selection and dimensionality reduction processes, significantly enhances credit risk assessment by providing more precise valuations of mortgage-backed assets. At the same time, it strengthens data-driven urban planning, enabling informed decision-making regarding land use and infrastructure allocation. Integrating dimensionality reduction techniques captures complex interactions among spatial, structural, and socioeconomic variables, thereby increasing model robustness and generalization capacity. This approach positions price prediction not merely as a technical task but as a strategic tool for territorial management and evidence-based decision-making in both public and private sectors.

Dimensionality reduction in real estate price prediction models enhances interpretability and computational efficiency. By reducing the number of features, the most relevant variables, such as location, size, and structural characteristics of properties, become more easily identifiable, thereby improving the model's transparency and understanding. This is especially valuable for professionals in the real estate sector, such as appraisers and urban planners, who can make more informed decisions based on clear and comprehensible results.

Regarding computational efficiency, dimensionality reduction optimizes training time and resource usage, resulting in faster and more cost-effective models that facilitate implementation in financial institutions and real estate companies. This approach enables quicker and more accurate appraisals, enhancing credit risk assessment and decision-making in urban development. Collectively, these benefits position dimensionality reduction as a crucial factor in improving operational effectiveness and the practical adoption of predictive models in the real estate sector.

## 5. Conclusions

The results of this study underscore the remarkable effectiveness of ensemble learning models in predicting land prices, with Stacking and Gradient Boosting emerging as the most accurate and robust methods. These models consistently achieved the highest coefficients of

determination ( $R^2$ ) and Concordance Correlation Coefficients (CCCs), along with the lowest prediction errors (MAE and RMSE), establishing themselves as reliable tools for capturing complex, high-dimensional patterns. Stacking demonstrated outstanding stability across all scenarios, even after dimensionality reduction using RFE, Random Forest, and Boruta, while maintaining an optimal balance between predictive accuracy and computational efficiency.

The findings also highlight the strategic relevance of dimensionality reduction techniques for enhancing the performance of ensemble models. Methods such as RFE and Boruta reduced the number of predictive features by over 90%, significantly lowering computational complexity while maintaining satisfactory levels of predictive performance. This efficiency was particularly evident in Gradient Boosting and Stacking, where variable reduction led to only marginal losses in accuracy, making these approaches highly suitable for applications that require model scalability, interpretability, and operational speed.

Conversely, models such as AdaBoost, Random Forest, and Extra Trees demonstrated greater sensitivity to feature reduction, highlighting the need to carefully implement these strategies in contexts where information loss could impair model performance. Interestingly, the combination of Boruta with AdaBoost resulted in minimal performance differences between the full and reduced models, suggesting that this pairing may offer valuable advantages in resource-constrained environments.

The practical implications of this research are significant, particularly in real estate valuation, financial risk assessment, and spatial planning, where predictive accuracy and computational efficiency are crucial. Nevertheless, several directions for future work are identified. These include evaluating the generalizability of ensemble models across more diverse datasets spanning different geographic regions and temporal settings, and incorporating advanced hyperparameter optimization strategies to further enhance model performance in a fully automated manner.

Finally, exploring hybrid approaches that combine the strengths of multiple ensemble algorithms and variable selection methods, such as adaptive combinations of Stacking, Voting, and Boosting with Boruta or RFE, represents a promising avenue for developing more robust and flexible methodological frameworks. Such advancements would contribute to the design of high-impact predictive solutions that are particularly suited to real-world contexts characterized by high dimensionality, uncertainty, and the demand for interpretable and reproducible outcomes.

**Author Contributions:** D.C.A.-G.: writing—review and editing, writing—original draft, visualization, resources, methodology, investigation, formal analysis, conceptualization. W.J.M.-R.: writing—original draft, methodology, investigation, formal analysis, supervision. M.G.Z.-R.: writing—review and editing, investigation, supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Universidad Nacional José Faustino Sánchez Carrión under RCU No. 0696-2024-CU-UNJFSC.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Shi, X.; Cheng, Q.; Xia, M. The Industrial Linkages of the Real Estate Industry and Its Impact on the Economy Caused by the COVID-19 Pandemic. In Proceedings of the 25th International Symposium on Advancement of Construction Management and Real Estate; Lu, X., Zhang, Z., Lu, W., Peng, Y., Eds.; Springer: Singapore, 2021; pp. 1015–1028. [\[CrossRef\]](#)
- Karanasos, M.; Yfanti, S. On the Economic Fundamentals behind the Dynamic Equicorrelations among Asset Classes: Global Evidence from Equities, Real Estate, and Commodities. *J. Int. Financ. Mark. Inst. Money* **2021**, *74*, 101292. [\[CrossRef\]](#)
- Abdul Rahman, M.S.; Awang, M.; Jagun, Z.T. Polycrisis: Factors, Impacts, and Responses in the Housing Market. *Renew. Sustain. Energy Rev.* **2024**, *202*, 114713. [\[CrossRef\]](#)
- Newell, G.; McGreal, S. The Significance of Development Sites in Global Real Estate Transactions. *Habitat Int.* **2017**, *66*, 117–124. [\[CrossRef\]](#)
- El Bied, S.; Ros McDonnell, L.; de-la-Fuente-Aragón, M.V.; Ros McDonnell, D. A Comprehensive Bibliometric Analysis of Real Estate Research Trends. *Int. J. Financ. Stud.* **2024**, *12*, 95. [\[CrossRef\]](#)
- Schoen, E.J. The 2007–2009 Financial Crisis: An Erosion of Ethics: A Case Study. *J. Bus. Ethics* **2017**, *146*, 805–830. [\[CrossRef\]](#)
- Trovato, M.R.; Clienti, C.; Giuffrida, S. People and the City: Urban Fragility and the Real Estate-Scape in a Neighborhood of Catania, Italy. *Sustainability* **2020**, *12*, 5409. [\[CrossRef\]](#)
- Zhang, P.; Hu, S.; Li, W.; Zhang, C.; Yang, S.; Qu, S. Modeling Fine-Scale Residential Land Price Distribution: An Experimental Study Using Open Data and Machine Learning. *Appl. Geogr.* **2021**, *129*, 102442. [\[CrossRef\]](#)
- De la Luz Juárez, G.; Daza, A.S.; González, J.Z. La crisis financiera internacional de 2008 y algunos de sus efectos económicos sobre México. *Contaduría Adm.* **2015**, *60*, 128–146. [\[CrossRef\]](#)
- Tekouabou, S.C.K.; Gherghina, Ş.C.; Kameni, E.D.; Filali, Y.; Idrissi Gartoumi, K. AI-Based on Machine Learning Methods for Urban Real Estate Prediction: A Systematic Survey. *Arch. Comput. Methods Eng.* **2024**, *31*, 1079–1095. [\[CrossRef\]](#)
- Porter, L.; Fields, D.; Landau-Ward, A.; Rogers, D.; Sadowski, J.; Maalsen, S.; Kitchin, R.; Dawkins, O.; Young, G.; Bates, L.K. Planning, Land and Housing in the Digital Data Revolution/The Politics of Digital Transformations of Housing/Digital Innovations, PropTech and Housing—the View from Melbourne/Digital Housing and Renters: Disrupting the Australian Rental Bond System and Tenant Advocacy/Prospects for an Intelligent Planning System/What Are the Prospects for a Politically Intelligent Planning System? *Plan. Theory Pract.* **2019**, *20*, 575–603. [\[CrossRef\]](#)
- Fabozzi, F.J.; Kynigakis, I.; Panopoulou, E.; Tunaru, R.S. Detecting Bubbles in the US and UK Real Estate Markets. *J. Real Estate Financ. Econ.* **2020**, *60*, 469–513. [\[CrossRef\]](#)
- Pollestad, A.J.; Næss, A.B.; Oust, A. Towards a Better Uncertainty Quantification in Automated Valuation Models. *J. Real Estate Financ. Econ.* **2024**. [\[CrossRef\]](#)
- Munanga, Y.; Musakwa, W.; Chirisa, I. The Urban Planning-Real Estate Development Nexus. In *The Palgrave Encyclopedia of Urban and Regional Futures*; Brears, R.C., Ed.; Springer International Publishing: Cham, Switzerland, 2022; pp. 2046–2054. ISBN 978-3-030-87745-3. [\[CrossRef\]](#)
- Kwangwama, N.A.; Mafuku, S.H.; Munanga, Y. A Review of the Contribution of the Real Estate Sector Towards the Attainment of the New Urban Agenda in Zimbabwe. In *New Urban Agenda in Zimbabwe: Built Environment Sciences and Practices*; Chavunduka, C., Chirisa, I., Eds.; Springer Nature: Singapore, 2024; pp. 151–168. ISBN 978-981-97-3199-2. [\[CrossRef\]](#)
- Behera, I.; Nanda, P.; Mitra, S.; Kumari, S. Machine Learning Approaches for Forecasting Financial Market Volatility. In *Machine Learning Approaches in Financial Analytics*; Maglaras, L.A., Das, S., Tripathy, N., Patnaik, S., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 431–451. ISBN 978-3-031-61037-0. [\[CrossRef\]](#)
- Goldfarb, A.; Greenstein, S.M.; Tucker, C.E. Economic Analysis of the Digital Economy 2015. Available online: <http://www.nber.org/books/gree13-1> (accessed on 12 February 2025).
- Bond, S.A.; Shilling, J.D.; Wurtzbaach, C.H. Commercial Real Estate Market Property Level Capital Expenditures: An Options Analysis. *J. Real Estate Finance Econ.* **2019**, *59*, 372–390. [\[CrossRef\]](#)
- Arcuri, N.; De Ruggiero, M.; Salvo, F.; Zinno, R. Automated Valuation Methods through the Cost Approach in a BIM and GIS Integration Framework for Smart City Appraisals. *Sustainability* **2020**, *12*, 7546. [\[CrossRef\]](#)
- Moorhead, M.; Armitage, L.; Skitmore, M. Real Estate Development Feasibility and Hurdle Rate Selection. *Buildings* **2024**, *14*, 1045. [\[CrossRef\]](#)
- Htun, H.H.; Biehl, M.; Petkov, N. Survey of feature selection and extraction techniques for stock market prediction. *Financ. Innov.* **2023**, *9*, 26. [\[CrossRef\]](#)
- Glaeser, E.L.; Nathanson, C.G. An Extrapolative Model of House Price Dynamics. *J. Financ. Econ.* **2017**, *126*, 147–170. [\[CrossRef\]](#)
- Zheng, M.; Wang, H.; Wang, C.; Wang, S. Speculative Behavior in a Housing Market: Boom and Bust. *Econ. Model.* **2017**, *61*, 50–64. [\[CrossRef\]](#)
- Fotheringham, A.S.; Crespo, R.; Yao, J. Exploring, Modelling and Predicting Spatiotemporal Variations in House Prices. *Ann. Reg. Sci.* **2015**, *54*, 417–436. [\[CrossRef\]](#)

25. Kumar, S.; Talasila, V.; Pasumarthy, R. A Novel Architecture to Identify Locations for Real Estate Investment. *Int. J. Inf. Manag.* **2021**, *56*, 102012. [\[CrossRef\]](#)
26. Barkham, R.; Bokhari, S.; Saiz, A. Urban Big Data: City Management and Real Estate Markets. In *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities: Designing for Sustainability*; Pardalos, P.M., Rassia, S.T., Tsokas, A., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 177–209. ISBN 978-3-030-84459-2. [\[CrossRef\]](#)
27. Moghaddam, S.N.M.; Cao, H. Housing, Affordability, and Real Estate Market Analysis. In *Artificial Intelligence-Driven Geographies: Revolutionizing Urban Studies*; Moghaddam, S.N.M., Cao, H., Eds.; Springer Nature: Singapore, 2024; pp. 361–393. ISBN 978-981-97-5116-7. [\[CrossRef\]](#)
28. Parmezan, A.R.S.; Souza, V.M.A.; Batista, G.E.A.P.A. Evaluation of Statistical and Machine Learning Models for Time Series Prediction: Identifying the State-of-the-Art and the Best Conditions for the Use of Each Model. *Inf. Sci.* **2019**, *484*, 302–337. [\[CrossRef\]](#)
29. Rane, N.; Choudhary, S.P.; Rane, J. Ensemble Deep Learning and Machine Learning: Applications, Opportunities, Challenges, and Future Directions. *Stud. Med. Health Sci.* **2024**, *1*, 18–41. [\[CrossRef\]](#)
30. Naz, R.; Jamil, B.; Ijaz, H. Machine Learning, Deep Learning, and Hybrid Approaches in Real Estate Price Prediction: A Comprehensive Systematic Literature Review. *Proc. Pak. Acad. Sci. A. Phys. Comput. Sci.* **2024**, *61*, 129–144. [\[CrossRef\]](#)
31. Breuer, W.; Steininger, B.I. Recent trends in real estate research: A comparison of recent working papers and publications using machine learning algorithms. *J. Bus. Econ.* **2020**, *90*, 963–974. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Adetunji, A.B.; Akande, O.N.; Ajala, F.A.; Oyewo, O.; Akande, Y.F.; Oluwadara, G. House Price Prediction Using Random Forest Machine Learning Technique. *Procedia Comput. Sci.* **2022**, *199*, 806–813. [\[CrossRef\]](#)
33. Jui, J.J.; Imran Molla, M.M.; Bari, B.S.; Rashid, M.; Hasan, M.J. Flat Price Prediction Using Linear and Random Forest Regression Based on Machine Learning Techniques. In *Embracing Industry 4.0*; Mohd Razman, M.A., Mat Jizat, J.A., Mat Yahya, N., Myung, H., Zainal Abidin, A.F., Abdul Karim, M.S., Eds.; Springer: Singapore, 2020; pp. 205–217. [\[CrossRef\]](#)
34. Wang, X.; Wen, J.; Zhang, Y.; Wang, Y. Real Estate Price Forecasting Based on SVM Optimized by PSO. *Optik* **2014**, *125*, 1439–1443. [\[CrossRef\]](#)
35. Liu, G. Research on Prediction and Analysis of Real Estate Market Based on the Multiple Linear Regression Model. *Sci. Program.* **2022**, *2022*, 5750354. [\[CrossRef\]](#)
36. Sharma, M.; Chauhan, R.; Devliyal, S.; Chythanya, K.R. House Price Prediction Using Linear and Lasso Regression. In Proceedings of the 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 1–3 March 2024; pp. 1–5. [\[CrossRef\]](#)
37. Harahap, T.H. K-Nearest Neighbor Algorithm for Predicting Land Sales Price. *Al'adzkiya Int. Comput. Sci. Inf. Technol. J.* **2022**, *3*, 58–67.
38. Mohd, T.; Jamil, N.S.; Johari, N.; Abdullah, L.; Masrom, S. An Overview of Real Estate Modelling Techniques for House Price Prediction. In *Charting a Sustainable Future of ASEAN in Business and Social Sciences*; Kaur, N., Ahmad, M., Eds.; Springer: Singapore, 2020; pp. 321–338. [\[CrossRef\]](#)
39. Zhu, D.; Cai, C.; Yang, T.; Zhou, X. A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization. *Big Data Cogn. Comput.* **2018**, *2*, 5. [\[CrossRef\]](#)
40. Huang, J.-C.; Ko, K.-M.; Shu, M.-H.; Hsu, B.-M. Application and Comparison of Several Machine Learning Algorithms and Their Integration Models in Regression Problems. *Neural Comput. Appl.* **2020**, *32*, 5461–5469. [\[CrossRef\]](#)
41. Ahsan, M.M.; Mahmud, M.A.P.; Saha, P.K.; Gupta, K.D.; Siddique, Z. Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies* **2021**, *9*, 52. [\[CrossRef\]](#)
42. Rane, N.L.; Rane, J.; Mallick, S.K.; Kaya, Ö. *Scalable and Adaptive Deep Learning Algorithms for Large-Scale Machine Learning Systems*; Deep Science Publishing: San Francisco, CA, USA, 2024.
43. Jiaji, S. Machine Learning-Driven Dynamic Scaling Strategies for High Availability Systems. *Int. IT J. Res.* **2024**, *2*, 35–42. [\[CrossRef\]](#)
44. Ho, W.K.O.; Tang, B.-S.; Wong, S.W. Predicting Property Prices with Machine Learning Algorithms. *J. Prop. Res.* **2021**, *38*, 48–70. [\[CrossRef\]](#)
45. Phan, T.D. Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia. In Proceedings of the 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, NSW, Australia, 3–7 December 2018; pp. 35–42. [\[CrossRef\]](#)
46. Pai, P.-F.; Wang, W.-C. Using Machine Learning Models and Actual Transaction Data for Predicting Real Estate Prices. *Appl. Sci.* **2020**, *10*, 5832. [\[CrossRef\]](#)
47. Dhal, P.; Azad, C. A Comprehensive Survey on Feature Selection in the Various Fields of Machine Learning. *Appl. Intell.* **2022**, *52*, 4543–4581. [\[CrossRef\]](#)



48. Charles, Z.; Papailiopoulos, D. Stability and Generalization of Learning Algorithms That Converge to Global Optima. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 745–754. Available online: <https://tinyurl.com/2p8mj8u9> (accessed on 25 February 2025).
49. Gilpin, L.H.; Bau, D.; Yuan, B.Z.; Bajwa, A.; Specter, M.; Kagal, L. Explaining Explanations: An Overview of Interpretability of Machine Learning. In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018; pp. 80–89. [\[CrossRef\]](#)
50. Aremu, O.O.; Hyland-Wood, D.; McAree, P.R. A Machine Learning Approach to Circumventing the Curse of Dimensionality in Discontinuous Time Series Machine Data. *Reliab. Eng. Syst. Saf.* **2020**, *195*, 106706. [\[CrossRef\]](#)
51. Khan, W.A.; Chung, S.H.; Awan, M.U.; Wen, X. Machine Learning Facilitated Business Intelligence (Part II): Neural Networks Optimization Techniques and Applications. *Ind. Manag. Data Syst.* **2020**, *120*, 128–163. [\[CrossRef\]](#)
52. Qin, H. Athletic Skill Assessment and Personalized Training Programming for Athletes Based on Machine Learning. *J. Electr. Syst.* **2024**, *20*, 1379–1387. [\[CrossRef\]](#)
53. Park, B.; Bae, J.K. Using Machine Learning Algorithms for Housing Price Prediction: The Case of Fairfax County, Virginia Housing Data. *Expert Syst. Appl.* **2015**, *42*, 2928–2934. [\[CrossRef\]](#)
54. Varma, A.; Sarma, A.; Doshi, S.; Nair, R. House Price Prediction Using Machine Learning and Neural Networks. In Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 20–21 April 2018; pp. 1936–1939. [\[CrossRef\]](#)
55. Rafiei, M.H.; Adeli, H. A Novel Machine Learning Model for Estimation of Sale Prices of Real Estate Units. *J. Constr. Eng. Manag.* **2016**, *142*, 04015066. [\[CrossRef\]](#)
56. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Tree-Based Ensemble Methods for Predicting PV Power Generation and Their Comparison with Support Vector Regression. *Energy* **2018**, *164*, 465–474. [\[CrossRef\]](#)
57. Arabameri, A.; Pal, S.C.; Rezaie, F.; Chakraborty, R.; Saha, A.; Blaschke, T.; Di Napoli, M.; Ghorbanzadeh, O.; Ngo, P.T.T. Decision Tree Based Ensemble Machine Learning Approaches for Landslide Susceptibility Mapping. *Geocarto Int.* **2022**, *37*, 4594–4627. [\[CrossRef\]](#)
58. Ribeiro, M.H.D.M.; dos Santos Coelho, L. Ensemble Approach Based on Bagging, Boosting and Stacking for Short-Term Prediction in Agribusiness Time Series. *Appl. Soft Comput.* **2020**, *86*, 105837. [\[CrossRef\]](#)
59. Sibindi, R.; Mwangi, R.W.; Waititu, A.G. A Boosting Ensemble Learning Based Hybrid Light Gradient Boosting Machine and Extreme Gradient Boosting Model for Predicting House Prices. *Eng. Rep.* **2023**, *5*, e12599. [\[CrossRef\]](#)
60. Kim, J.; Won, J.; Kim, H.; Heo, J. Machine-Learning-Based Prediction of Land Prices in Seoul, South Korea. *Sustainability* **2021**, *13*, 13088. [\[CrossRef\]](#)
61. Sankar, M.; Chithambaramani, R.; Sivaprakash, P.; Ithayan, V.; Dilip Charaan, R.M.; Marichamy, D. Analysis of Landlord's Land Price Prediction Using Machine Learning. In Proceedings of the 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 18–20 September 2024; pp. 1514–1518. [\[CrossRef\]](#)
62. Zelaya, C.V.G. Towards Explaining the Effects of Data Preprocessing on Machine Learning. In Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE), Macao, China, 8–11 April 2019; pp. 2086–2090. [\[CrossRef\]](#)
63. Nair, P.; Kashyap, I. Hybrid Pre-Processing Technique for Handling Imbalanced Data and Detecting Outliers for KNN Classifier. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; pp. 460–464. [\[CrossRef\]](#)
64. Mallikharjuna Rao, K.; Saikrishna, G.; Supriya, K. Data Preprocessing Techniques: Emergence and Selection towards Machine Learning Models-a Practical Review Using HPA Dataset. *Multimed. Tools Appl.* **2023**, *82*, 37177–37196. [\[CrossRef\]](#)
65. Habibi, A.; Delavar, M.R.; Sadeghian, M.S.; Nazari, B.; Pirasteh, S. A Hybrid of Ensemble Machine Learning Models with RFE and Boruta Wrapper-Based Algorithms for Flash Flood Susceptibility Assessment. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *122*, 103401. [\[CrossRef\]](#)
66. Kursu, M.B. Robustness of Random Forest-Based Gene Selection Methods. *BMC Bioinform.* **2014**, *15*, 8. [\[CrossRef\]](#)
67. Solomatine, D.P.; Shrestha, D.L. AdaBoost.RT: A Boosting Algorithm for Regression Problems. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, Hungary, 25–29 July 2004; Volume 2, pp. 1163–1168. [\[CrossRef\]](#)
68. Zhang, P.; Yang, Z. A Robust AdaBoost.RT Based Ensemble Extreme Learning Machine. *Math. Probl. Eng.* **2015**, *2015*, 260970. [\[CrossRef\]](#)
69. Zharmagambetov, A.; Gabidolla, M.; Carreira-Perpiñán, M.Ê. Improved Boosted Regression Forests Through Non-Greedy Tree Optimization. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8. [\[CrossRef\]](#)
70. Hoang, N.-D.; Tran, V.-D.; Tran, X.-L. Predicting Compressive Strength of High-Performance Concrete Using Hybridization of Nature-Inspired Metaheuristic and Gradient Boosting Machine. *Mathematics* **2024**, *12*, 1267. [\[CrossRef\]](#)

71. Hoang, N.-D. Predicting Tensile Strength of Steel Fiber-Reinforced Concrete Based on a Novel Differential Evolution-Optimized Extreme Gradient Boosting Machine. *Neural Comput. Appl.* **2024**, *36*, 22653–22676. [\[CrossRef\]](#)
72. Borup, D.; Christensen, B.J.; Mühlbach, N.S.; Nielsen, M.S. Targeting Predictors in Random Forest Regression. *Int. J. Forecast.* **2023**, *39*, 841–868. [\[CrossRef\]](#)
73. González, S.; García, S.; Del Ser, J.; Rokach, L.; Herrera, F. A Practical Tutorial on Bagging and Boosting Based Ensembles for Machine Learning: Algorithms, Software Tools, Performance Study, Practical Perspectives and Opportunities. *Inf. Fusion* **2020**, *64*, 205–237. [\[CrossRef\]](#)
74. Hajihosseini, M.; Maghsoudi, A.; Ghezelbash, R. Stacking: A Novel Data-Driven Ensemble Machine Learning Strategy for Prediction and Mapping of Pb-Zn Prospectivity in Varcheh District, West Iran. *Expert Syst. Appl.* **2024**, *237*, 121668. [\[CrossRef\]](#)
75. Tummala, M.; Ajith, K.; Mamidibathula, S.K.; Kenchetty, P. Driving Sustainable Water Solutions: Optimizing Water Quality Prediction Using a Stacked Hybrid Model with Gradient Boosting and Ridge Regression. In Proceedings of the 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 4–6 December 2024; pp. 1584–1589. [\[CrossRef\]](#)
76. Revathy, P.; Manju Bhargavi, N.; Gunasekar, S.; Lohit, A. Voting Regressor Model for Timely Prediction of Sleep Disturbances Using NHANES Data. In *ICT Systems and Sustainability*; Tuba, M., Akashe, S., Joshi, A., Eds.; Springer Nature: Singapore, 2024; pp. 53–65. [\[CrossRef\]](#)
77. Chen, S.; Luc, N.M. RRMSE Voting Regressor: A Weighting Function Based Improvement to Ensemble Regression. *arXiv* **2022**, arXiv:2207.04837. [\[CrossRef\]](#)
78. Battineni, G.; Sagaro, G.G.; Nalini, C.; Amenta, F.; Tayebati, S.K. Comparative Machine-Learning Approach: A Follow-Up Study on Type 2 Diabetes Predictions by Cross-Validation Methods. *Machines* **2019**, *7*, 74. [\[CrossRef\]](#)
79. Kumkar, P.; Madan, I.; Kale, A.; Khanvilkar, O.; Khan, A. Comparison of Ensemble Methods for Real Estate Appraisal. In Proceedings of the 2018 3rd International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 15–16 November 2018; pp. 297–300. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.