International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I10P106

Volume 73 Issue 10, 79-94, October 2025

Original Article

Ensemble Machine Learning-Based Real Estate Price Prediction with Explainable Artificial Intelligence Methods for Determinant Analysis

Matthew C. Okoronkwo[1] Ugochukwu E. Orji[2], Chikodili H. Ugwuishiwu[3], Caroline N. Asogwa[4], Emenike C. Ugwuagbo[5], Bande S. Ponsak[6]

[1,2,3,4,6]*Department of Computer Science, University of Nigeria, Nsukka, Enugu State, Nigeria.*
[5]*Centre for Space Transport and Propulsion (CSTP), Epe, 1 LASU Epe drive, Lagos, Nigeria.*

[3]*Corresponding Author : chikodili.ugwuishiwu@unn.edu.ng*

Received: 19 June 2025 Revised: 16 September 2025 Accepted: 25 September 2025 Published: 31 October 2025

Abstract *- The Real Estate (RE) industry is an essential part of many countries' economies, and accurately forecasting housing prices is beneficial to buyers, real estate agents, and the government. However, multiple factors influence the prices of RE properties, which are difficult to measure; the relationship between housing prices and housing characteristics is complex and nonlinear, requiring a flexible algorithm and tools. Three regression-based models were developed using Neural Network (NN), Random Forest (RF), and Extreme Gradient Boosting (XGB) algorithms to predict house prices. Explainable Artificial Intelligence (XAI) methods were deployed to explain the key factors influencing RE prices. The dataset used has 923,159 records, available on Kaggle. The models were evaluated using four zip codes, and the house size influenced the price prediction for the RF model. For efficient RE price performance evaluation, the following metrics were computed: squared (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics were applied to evaluate the machine learning models adopted in the research, and the results show that the XGB model performed better with R^2 = 0.817011. The XGB model, SHapley Additive exPlanations (SHAP) plot showed that acre lot and bath are the most influential determinants in predicting the price of houses, while the Individual Conditional Expectation (ICE) plots showed that bath, by the traditional evaluation methods. The result promises a better decision support to potential RE buyers in selecting houses that meet their specific needs.*

Keywords *- Explainable AI, Ensemble Learning, Housing Price Prediction, Random Forest, Gradient Boosting Analysis.*

1. Introduction

The real estate sector represents a substantial share of most national economies, and housing price prediction remains a longstanding challenge. Buyers, sellers, agents, and policymakers all benefit from reliable forecasts, yet housing prices are shaped by multiple interacting factors, including property features, location, and market demand. Classical valuation approaches-such as cost analysis, sales comparison, or income capitalization-have been widely used (Pagourtzi et al., 2003). However, these methods often fail to capture nonlinearities and high-dimensional feature interactions (Zurada et al., 2006), leading to inconsistent or inaccurate appraisals. Quantitative measurement of the benefits and liabilities accruing from the ownership of RE is part of the valuation process (Selim, 2009). The increasing availability of large housing datasets has accelerated the adoption of Machine Learning (ML) for real estate valuation. ML techniques, especially ensemble learning, are better suited to handle nonlinear relationships and uncover hidden patterns. Recent work has demonstrated that ensemble approaches, RF, and GBoosting achieve higher predictive accuracy than traditional approaches (Varma et al., 2018; Teoh et al., 2022). At the same time, interpretability remains critical. Without transparency, predictive models risk becoming "black boxes" that hinder trust and adoption (Gunning et al., 2019). As a result, it becomes increasingly difficult for sellers and buyers to have an efficient platform to determine the optimal price for a property. Hence, there is a need for the adoption of automated house appraisal methods. Many studies show that Artificial Intelligence (AI), especially ML, provides a more suitable method that approaches market values more (Naci, 2021). Explainable AI (XAI) addresses this challenge by offering interpretability tools such as SHAP values and ICE plots, which clarify how features contribute to predictions. While prior studies have used these methods individually, fewer have applied them in combination for housing price prediction. This study, therefore, develops and compares RF and XGB models for house price prediction while employing

SHAP and ICE to identify the most influential features. The goal is not only accurate prediction but also actionable explanations that can inform decision-making. Today, organizations and businesses are adopting intelligent data-driven processes using AI techniques to enhance business profitability (Fuster et al., 2020). Machine Learning (ML), as a subset of AI, is intended to improve its performance from previous experience. So far, ML techniques have been successfully deployed in virtually all sectors, including decision support systems, dynamic pricing, the banking sector, etc. (Pratt, 2020; Ugochukwu and Elochukwu, 2022). Recently, the RE industry has explored ML-driven house appraisal, which surpasses conventional house appraisal methods (Teoh et al., 2022). ML for predicting house prices has been the subject of a lot of research (Park & Bae, 2015; Manasa Gupta & Narahari, 2020; Varma et. al, 2018), but currently, a few studies have used both the SHAP and ICE plot XAi methods to explain the determinant features and identify the determinant factors that influence housing price prediction. XAi methods now enable users to understand, trust, and better manage AI systems (Gunning et al., 2021), making it possible to identify and describe factors that influence house prices via ML techniques.

Thus, this research aims to develop and compare two regression-based Ensemble Machine Learning (EML) models (RF and XGB) to predict housing prices based on the chosen metrics. The significant determinant factors that influence housing prices using two XAI methods (SHAP analysis and ICE plots) will be identified and interpreted. Intelligent house price predictions aid prospective home buyers in making decisions, guide developers on setting house prices, and enhance tax valuation, which boosts the overall nation's economy. Regression testing is essential in Machine Learning (ML) to guarantee that changes to the model, data, or environment do not negatively impact existing functionality (Owoc & Stambulski, 2025). The paper is organized in the following way: The theoretical analysis provides the description of the ML models deployed in this work and the XAi framework used. It is followed by the methodology, describing the dataset used and model evaluation methods. The next is the results, discussions, and summary of findings, lastly, the conclusion and recommendations.

2. Theoretical Analysis

The prediction of housing prices has been widely studied across economics, finance, and computer science. Earlier approaches commonly used hedonic regression, where housing prices are expressed as a function of property features, geographic context, and neighbourhood conditions. While these models provided interpretable insights, their linear functional forms often limited their ability to capture the complex, nonlinear dynamics of housing markets. The growth of machine learning has introduced more powerful alternatives. Decision-tree-based algorithms, particularly ensemble methods, have demonstrated superior predictive accuracy in real estate valuation tasks. RF is widely recognized for handling noisy, high-dimensional data effectively, while boosting methods such as GBM-and more recently XGBoost-have consistently ranked among the top performers in structured prediction tasks. Interpretability has also become a central concern. As machine learning models grow in complexity, stakeholders demand transparency in how predictions are formed. SHAP, grounded in Shapley values from cooperative game theory, has become a standard interpretability tool because it attributes portions of a prediction to each feature in a consistent, additive way. Similarly, ICE plots allow researchers to visualize heterogeneous effects across different observations, complementing global explanations with local nuance.

Recent studies applying these techniques in real estate have shown that location, property size, and neighbourhood characteristics consistently rank among the most influential determinants of price. At the same time, interpretability tools highlight that these effects are not uniform, underscoring the importance of context-specific modelling. This body of work provides the foundation for the present study, which combines RF and XGB for prediction while employing SHAP and ICE to enhance model transparency.

2.1. Understanding EML Models

EML is a particular form of machine learning paradigm that uses a group of base learners, sometimes referred to as weak learners, who are integrated and trained to evaluate and resolve real-world problems. In contrast, conventional learning methods construct only one learner from the training data. An ensemble's capacity for generalization is typically significantly greater than that of its component learners. EML techniques are also referred to as "meta-learning techniques" because of their capacity to learn from base learners (Zhou, 2009; Zhang & Ma, 2012). EML may transform weak learners into strong learners, who can predict outcomes far more accurately than random guesses (Zhou, 2011). The EML algorithms have the ability to capture nonlinear relationships, high-order interactions between inputs, and can offer higher prediction accuracy.

Within the past decade, EML methods have increasingly drawn the attention of researchers and analysts who have carried out a great number of ML experiments in research and competitions on platforms like Kaggle, KDD-Cups, etc. EML has achieved exceptionally satisfactory performance (Dong et al., 2020). The determination of house prices is complex, mostly nonlinear, and features dependent attributes. Dependent. EML algorithms allow more flexibility by combining individual models and averaging the results (Chen *et al.,* 2020). Therefore, EML algorithms are well-suited for modelling housing prices. Figure 1 shows the EML process: the ensemble formation and base learners, followed by the pruning to exclude some functions from the previous step. Lastly, the e integration to combine the chosen functions.

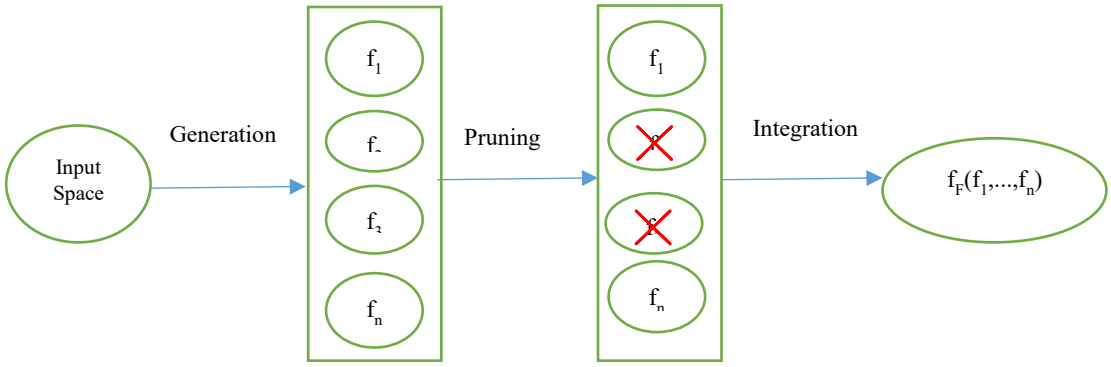

Fig. 1 An EML process adapted from Zounemat-Kermani et al., (2021)

2.2. Details of EML Models Deployed

According to Zounemat-Kermani et al. (2021), bagging and boosting are the most commonly used EML methods. The major difference is that in the bagging, a number of models are randomly trained in parallel using a subset of the data, while in the boosting approach, models are sequentially trained to learn from the mistakes made by the previous model. Brief descriptions of both methods are provided below.

The Bagging algorithms have been applied to both regression and classification problems to improve precision in ML approaches and aid in reducing variance and increasing robustness (Polikar, 2012; Wu et al., 2020). Notable bagging algorithms include RF, Random Subspace Methods (RSM), and Extremely Randomized Trees (ERT).

The Boosting algorithm is designed to improve the accuracy and performance of ML algorithms. By sequentially boosting weak learners (base learners) into strong ones, it can reduce the overfitting of decision trees and decrease variance and bias in an EML for greater prediction accuracy.

This technique also increases diversity within the primary training data set through random sampling and replacement.

Here, the first model is created from the training dataset, with subsequent models based on the performance of the previous one (Schapire, 2003; Zounemat-Kermani et al., 2021). Examples include Stochastic Gradient Boosting (SGB), AdaBoost, and eXtreme Gradient Boosting (XGB).

In this study, two regression-based EML algorithms (RF and XGB) were deployed to model housing prices; below is a brief description of both models.

2.3. Extreme Gradient Boosting (XGB) Model

XGB is a flexible and extensible implementation of the GBBoost framework by Friedman et al. (2000). It is a supervised learning algorithm that implements the boosting approach to yield more accurate models (Mitchell & Frank, 2017).

Table 1. XGB model description

SYMBOL	MEANING
n	Total number of samples
m	Total number of features
x_i	Features$_i$ information of the i-th sample, xi ∈ R^m
y_i	The actual label (or value) of the i-th sample
$\hat{y}_i$	The predicted label (or value) of the i-th sample
$\hat{y}_i^{(t)}$	The predicted value up to the t-th tree
$l(y_i, \hat{y}_i)$	The loss function of the i-th sample
$L(y, \hat{y})$	The loss function of the total sample
$\Omega(f_k)$	The regular term of the objective function to prevent overfitting, where f_k represents the k-th decision tree.

Given a data set containing n samples and m features, $D = \{(x_i, y_i) \mid x_i \in R^m, y_i \in R\}$ and $x_i = \{x_{i1}, x_{i2}, \cdots, x_{im} \mid i = 1,2, \cdots, n\}$. The XGB model is tasked with building t trees so that the predicted value. $\hat{y}_i^{(t)}$Up to the t^{-th} tree satisfies formula (1).

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (1)$$

In each iteration of the gradient boosting algorithm, a weak classifier $f_k(x_i)$ (i.e., a decision tree) is generated, and the predicted value $\hat{y}_i^{(t)}$ The sum of the predicted value of the previous iteration $\hat{y}_i^{(t-1)}$ and the decision tree result of this round $f_t(x_i)$ (Li et al., 2020).

Shrinkage and column subsampling are two important techniques introduced by the XGB method. In Shrinkage, the impact of a tree decreases, and overfitting is addressed by scaling new weights at each step of boosting. In column subsampling, only a randomly chosen set of input features is used to build a tree, in an effort to speed up the training process (Meng et al., 2020; Sheridan et al., 2016; Yaman & Subasi, 2019). XGB makes better predictions than the RF model and is comparable to deep neural nets (Sheridan et al., 2016),

In contrast to simple gradient boosting algorithms, XGB algorithms, unlike simple boosting, do not add weak learners at each stage, but use a multi-threaded approach that better uses the machine's CPU core, thus improving performance and speed.

2.4. Random Forest (RF) Model

According to Breiman (2001), RF is an EML algorithm that uses decision trees as weak learners. It is possible to employ RFs for both classification and regression problems (Liaw & Wiener, 2002). RFs are seen as one of the leading EML approaches that reduce over-fitting by computing the outcome mean. The RF technique is composed of a number of steps: First, random samples are selected from the dataset. Next is the construction of a decision tree using each sample to get a result, followed by voting to decide the most efficient model as the final forecast (Breiman, 2001; Cutler et al., 2012; Heddam, 2021). The RF forecast is the unweighted mean of the collection (Segal, 2004). RF prediction is the unweighted average over the collection, according to Segal (2004), as given below:

$$h(\mathrm{x}) = (1/K)\sum_{k=1}^{K} h(\mathrm{x};\theta_k) \quad (2)$$

As k → ∞ the Law of Large Numbers ensures,

$$E_{\mathrm{X},Y}(Y - h(\mathrm{X}))^2 \rightarrow E_{\mathrm{X},Y}(Y - E_{\theta}h(\mathrm{X};\theta))^2 \quad (3)$$

Where $h(\mathrm{x};\theta_k)$, $k = 1,\ldots,K$ is a collection of tree predictors, x represents the observed input (covariate) vector of length p with associated random vector X, and the θ_k are independent and identically distributed random vectors.

2.5. Neural Network (NN)

A NN is a model tailored according to how the human brain processes information and is widely used in AI and ML, eg, predicting house prices, medical diagnosis, etc.

It is a system composed of connected nodes (neurons) ordered into layers to learn patterns and relationships from data, as shown in Figure 2.

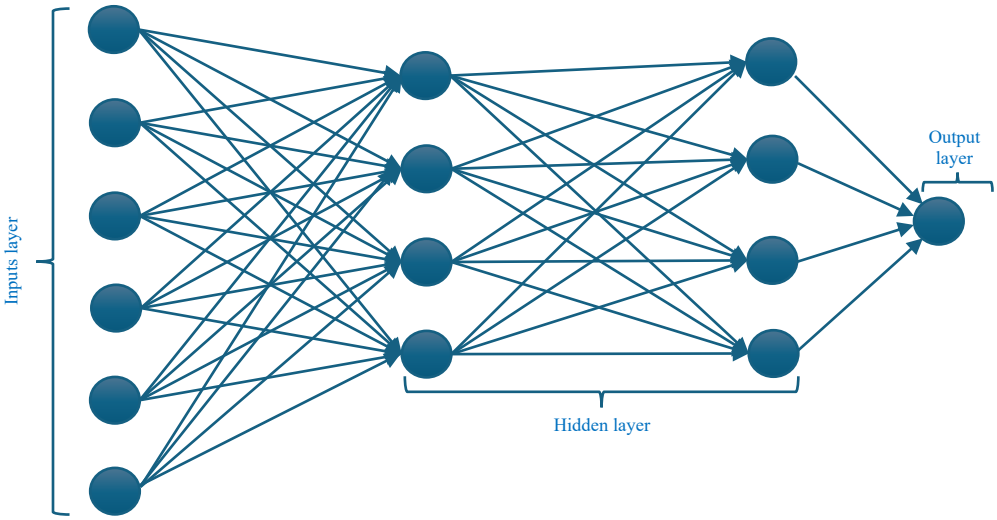

Fig. 2 Artificial neural netwok architecture (Hounmenou et.al 2021)

2.6. The Mathematics of Neural Networks

2.6.1. Weighted Sum

The Neurons compute a weighted sum of inputs and the bias term, which, when passed into the activation function, introduces nonlinearity and output of the neuron. Then, a cost function measures the difference between the predicted and target value as the error. Then, applying the chain rule to compute the gradient of error and bias. The forward pass, error computation, and backward pass are repeated to minimize error and increase prediction accuracy.

Mathematically, it is expressed as (García, 2022):

$$z = \sum_{i=1}^{n} w_i x_i + b$$

a=f(z)

Where:

z	=	weighted sum(linear part),
wi	=	xi feature weight
xi	=	neuron input features,
b	=	bias term
n	=	next layer inputs
I	=	counter
f(z)	=	activation function
a	=	the neuron output

2.6.2. Activation Functions

There are many ways neurons make regarding f(z). This includes:

Rectified Linear Units (ReLU) to ensure the output is not less than zero. This is given as: *f(z) = max(0, z), the* Tanh: given as: *f(z) = tanh(z)* and the Sigmoid activation: *f(x) = 1 / (1 + e^(-1*z)). T*he range is (0, 1).

2.6.3. Neural Network for Real Estate Price Prediction

Real estate prices depend on many factors, such as:

Location (urban vs. rural, proximity to schools, roads, hospitals, etc.)

Size (length & width, number of bedrooms, bathrooms), condition, and maintenance of the property,

Prevailing demand (consumer behaviour)

Amenities (swimming pool, garage, garden, etc.)

A Neural Network (NN) can capture nonlinear relationships between these features, so it's well-suited for this task in property price.

Input Layer: Each feature (location score, size, number of rooms, etc.) is represented as an input neuron. Example:

x_1=size (m^2), x_2=number of bedrooms and x_3=location index

A neural network has 3 main types of layers:

Hidden Layers: The network processes these inputs through one or more layers of neurons that pass through an activation function (like ReLU, sigmoid, or tanh). The Output Layer then produces a single value as the predicted house price.

2.7. Explainable AI (XAi)

Even though ML models variously proved reliable, efficient, and most importantly, accurate, their dominance comes with the cost of complexity (Guidotti et al., 2018).

However, accuracy against interpretation is a major issue (Gunning et al., 2019). The contribution of various determinants is difficult to measure. The XAi, however, was formulated to increase understanding and transparency (Gunning et al., 2021; Adadi & Berrada, 2018).

The concept of explainability sits at the forefront of AI, focusing, according to Hagras (2018) on:

- Transparency: end-users of ML models deserve to follow a model operation (Weller, 2017).
- Determinants: explore the possibilities of obtaining additional information from ML models, like the explanations for some underlying phenomena.
- Bias: ensures balanced models in the training data or objective function.
- Fairness: ensures objectivity.
- Safety: certifies the dependability and veracity of ML models.

In this paper, we used two different XAI-based Graphic Interpretation Tools: SHAP analysis and ICE plots for our determinant analysis of the models.

2.8. SHAP

SHAP as a method for explaining the predictions of a specific instance (Lundberg et. al., 2017). It has become a popular tool for interpreting natural and social phenomena (Stojić, 2019; Janizek, 2018), finding solid theoretical ground in game theory. SHAP attempts to interpret a model at each point *x.* The function thus defines the expectation value for a conditional distribution in a subset of S, and is given as:

$$e_S = E[f(x) \mid x_S = x_S^*] \tag{4}$$

As described by Holzinger et al. (2022), the contribution of a variable *j* is denoted by ϕ_j and calculated as the weighted average over all possible subsets S:

$$\phi_j(\text{val}) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \setminus \{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} \left(\text{val}\left(S \cup \{x_j\}\right) - \text{val}\,(S)\right) \tag{5}$$

Where p = the number of features, S = a subset of features, x = an instance of feature values in the model being explained, and val(S) = feature value predicted in the set S.

The Shapley approach accounts for the contribution of each determinant using various combinations of explanatory variables to evaluate each contribution (Teoh et al., 2022).

The value of absolute Shapley per characteristic is computed as global importance given below:

$$I_j = \sum_{i=1}^{n} \left|\Phi_j^{(i)}\right| \tag{6}$$

Where $\Phi_j^{(i)}$= the SHAP value of the j-th feature for instance i.

SHAP was implemented in Python using the XGBoost and SHAP packages.

2.9. ICE Plots

An ICE plot of the actual prediction functions $f^{(i)}$, rather than the means. It provides a visualization disaggregation of

typical PDPs. The plot shows a range of independent variables on the x-axis and forecasts on the y-axis. Every line represents one expectation (Jordan, Paul, & Philips, 2022).

"This study considered a set of observations $\{(X_{si}, X_{ci})\}_{i=1}^{N}$, and studied the response function f.

There are two types of ICE plots: a "centred" ICE plot, or c-ICE (Goldstein et al. 2015, Jordan, Paul, & Philips, 2022), that chooses some location x^* along x_s and forces such lines to run along the point. The c-ICE plots for mapping each explanatory variable are:

$$f_{cICE}^{(i)} = f^{(i)} - 1\, f\left(x^*, x_{Ci}\right) \quad \ldots\ldots\ldots (7)$$

Where $f^{(i)}$ = the given ICE curve and $f(x_\cdot, x_{ci})$represent the forecast value x^* for the i^{th} observation. An ICE plot is best when computing a heterogeneous relationship between features (Jordan, Paul, & Philips, 2022).

2.10. The Hyperparameters of RF and XGBoost

RF and XGBoost are both ensemble methods based on decision trees, but they differ in their approach and thus have distinct sets of important hyperparameters.

2.11. Random Forest Hyperparameters

RF builds multiple independent decision trees and combines their predictions. Key hyperparameters include: the n_estimators, which specify the number of decision trees; a bigger number improves performance but increases processing cost. Other parameters include: x_depth; depth reduction to prevent overfitting, min_samples_split, min_samples_leaf, max_features for randomness and decorrelation, and bootstrap; usually set to True in RF.

2.12. XGBoost Hyperparameters

XGBoost forms trees sequentially; previous errors are corrected by the new tree. It has an extensive set of hyperparameters, categorized into general, booster, and learning task parameters, which are usually set to refine/tune the model performance:

3. Materials and Methods

3.1. Research Flow

Figure 3 provides a high-level display of the research process. The first phase is data pre-processing to clean and get the data ready for analysis (Jadhav et al., 2019). This is followed by the Exploratory Data Analysis (EDA), which explains data patterns and provides understanding via statistical and visual displays. Thirdly, we transform the data to remove noise and reduce the skewness. In the second phase, we train, evaluate, and compare the results of the models. Then, finally, in the third phase, we use SHAP to explain a particular prediction by quantifying each feature's contribution and ICE to explain a specific feature's value influence on house prices for each mode.

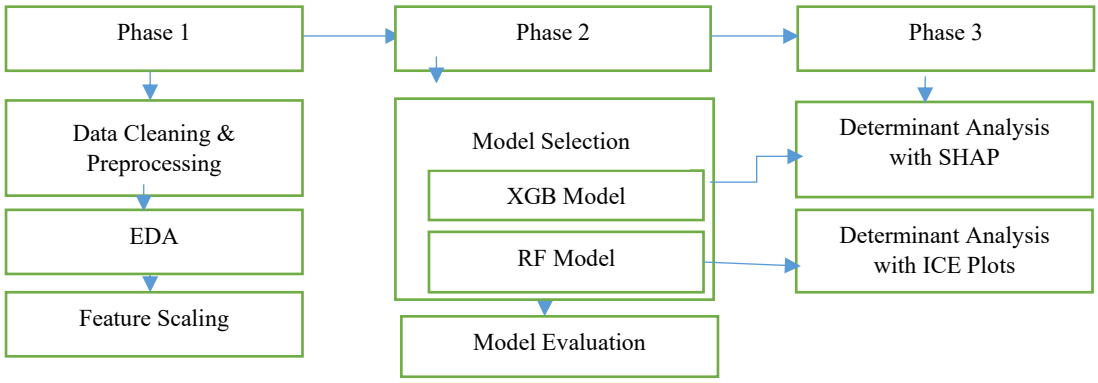

Fig. 3 Flow of the experiment

3.2. Details of the Dataset

3.2.1. Data Source

We use a large-scale dataset obtained from Kaggle, which contains more than 900,000 property entries described by 12 variables. These variables include property-specific features (e.g., size, number of bedrooms, and bathrooms), lot characteristics, geographic identifiers, and financial details such as sale price. The dataset covers a wide range of properties across different U.S. regions, offering both scale and diversity for robust modelling. It is available at: https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset/) as described in Table 2.

Table 2. Dataset description

Type	Name	Description	Class
	Status	House Status (on sale, sold, or other options)	Categorical
	Bed	No of Bedrooms	Numeric

Predictor Variables	Bath	No of baths	Numeric
	Acre_lot	The lot size of the house	Numeric
	Full_address	House location	Categorical
	Street	Street name	Categorical
	City	City name	Categorical
	State	City State	Categorical
	Zip_code	The house location Zip code	Numeric
	House_size	House in square feet	Numeric
	Sold_date	Sold Date (If sold)	Date
Target Variable	Price	Price of a house in USD	Numeric

3.3. Data Preprocessing

Data cleaning was performed to handle missing entries and remove extreme outliers that could bias the models. Categorical variables such as location were encoded into numeric form, and continuous variables were standardized where necessary. Randomly, the dataset was divided into training and test subsets to ensure unbiased evaluation. Data preparation, according to Famili et al. (1997), is crucial to any knowledge discovery project. This aims at handling missing information and cleaning out unnecessary or noisy aspects of the data (Jadhav et al., 2019).

On initial analysis, we discovered that the dataset contained 809,370 duplicate records out of the 923,159 available records, i.e., 87.67% of the data were duplicates, and after removal, we were left with 113,789 records, i.e., 12.33% of the initial data. Further analysis showed that the data still contained lots of null/missing values. At this stage, we used a data imputation approach to replace the missing values with the sample median or mode based on the distribution of the data. Furthermore, we dropped some columns like sold date, street, status, and full address since they would have no impact on our prediction model. Finally, we converted our categorical data into numerical data using Scikit-learn LabelEncoder.

3.4 Exploratory Data Analysis (EDA)

EDA is an important stage in data preprocessing that must be carried out before building an ML model. The goal of EDA is to explore the dataset to uncover anomalies, test hypotheses, and verify assumptions.

The information gained from EDA helps researchers choose appropriate ML approaches to solving the needed problem (Patil, 2018). Table 3 is the summary of the features; one can observe that it is right-skewed, with the bulk of the outliers lying in the final quartile. The Correlation Matrix Heatmap Figure 4 shows that no strong correlation exists between the various independent variables.

Note: Each record shows the Pearson correlation coefficient: green colours represent a positive correlation, while other colours indicate a mild or negative correlation. Furthermore, the house location is important in determining the price of the house.

Using a boxplot in Figure 5, we explored the dataset's range of house prices in various states. The data shows that South Carolina is the least expensive locality, while New York is the most expensive.

Table 3. Statistical summary of the features

Features	Mean	STD.	Min.	Q1 (25%)	Median (50%)	Q3 (75%)	Max.
price	9.095336^{e+05}	3.418652^{e+06}	0.000000^{e+00}	2.500000^{e+05}	4.499000^{e+05}	8.000000^{e+05}	8.750000^{e+08}
bed	3.261255	1.712655	1.000000	2.000000	3.000000	4.000000	123.000000
bath	2.446537	1.612504	1.000000	2.000000	2.000000	3.000000	198.000000
acre_lot	12.958765	836.871860	0.000000	0.150000	0.260000	0.570000	100000.000000
zip_code	8267.006670	4580.863932	601.000000	6010.000000	8005.000000	10301.000000	99999.000000
house_size	2.002986^{e+03}	4.824839^{e+03}	1.000000^{e+02}	1.376000^{e+03}	1.664000^{e+03}	2.035000^{e+03}	1.450112^{e+06}

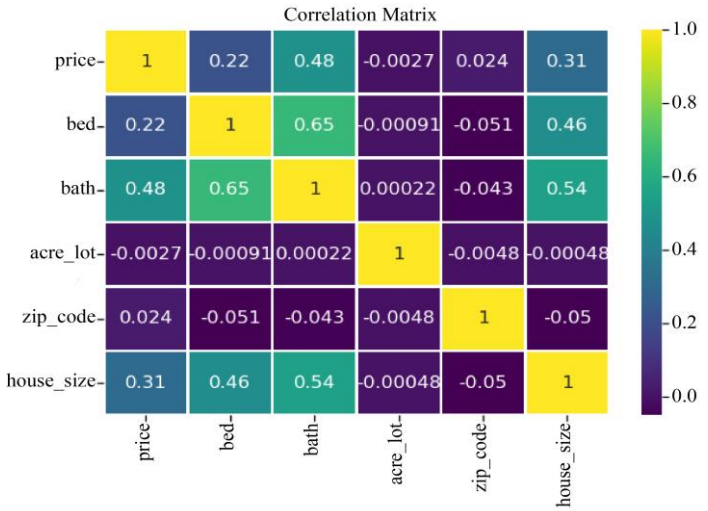

Fig. 4 Correlation plot of the dataset

3.5. Experiment Setup

Two ensemble machine learning models were implemented:

Random Forest (RF): In RF, numerous trees are grown from resampled training subsets, with each split restricted to a randomly chosen set of predictors, producing diverse trees whose results are later aggregated.

Extreme Gradient Boosting (XGB): XGBoost builds trees in a forward, stage-wise manner, where a new tree is optimized to reduce residual errors left by earlier ones. Regularization terms, column sampling, and learning-rate shrinkage are incorporated to prevent overfitting and enhance generalization.

Both models were tuned using grid search and cross-validation to identify optimal hyperparameters. The Python libraries on Jupyter Notebook on an Intel(R) HD Graphics 5500 GPU, a 2.60-GHz Intel(R) Core(TM) i7-5600U CPU, and 8GB DDR3 RAM were used to perform our experiments. First, feature scaling was applied to the data via Power Transformer scaling to make the distribution more Gaussian-like.

This technique finds an optimal scaling factor that stabilizes variance as well as minimizes skewness through maximum likelihood estimation (Roy, 2022). Figure 6 shows a scatter plot of the normalized data. The next step was to divide the normalized data into training (70%) and testing (30%) subsets for model training and validation, respectively.

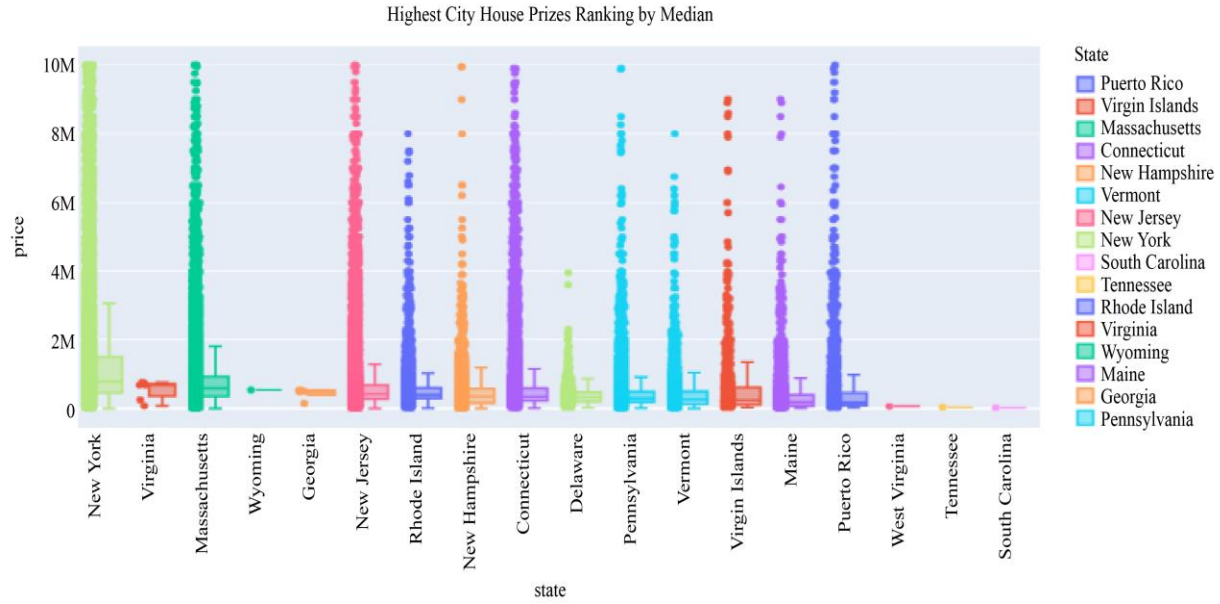

Fig. 5 House price range across various states in the dataset dataset

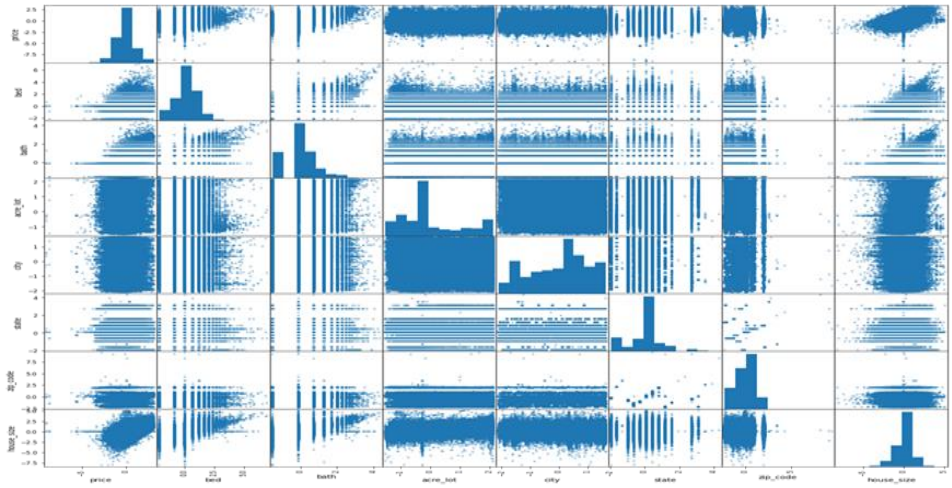

Fig. 6 Scatter plot of the data after scaling

3.6. Model Evaluation

Performance was assessed using the coefficient of determination (R^2), RMSE, and MAPE. These metrics capture complementary aspects of accuracy, ensuring that results reflect both overall fit and predictive ability. Still, it is difficult sometimes to predict the exact value of a regression model; hence, we aimed at showing the closeness of the forecast values to the actual values. The R^2 is a good metric to evaluate the fitting of the model; it examines the percentage of the predicted price explained by the features via a regression relationship.

The MSE calculates an absolute measure of goodness of fit; hence, it is often utilized for evaluating regression models. To accurately reflect prediction errors, MAE was calculated, being the average value of absolute errors.

Additionally, the MAPE calculates error in terms of percentage. Reading below 10% for MAPE reflects high-quality predictive modelling. Ideally, model MAE and RMSE scores of near 0, and an R^2 score of near 1/100% is adjudged the best possible outcome and informs performance accuracy of the model.

The metrics are illustrated as follows:

Given n samples, let/be the i prediction of the sample and y_i the actual value, and let r be the mean value of the sample.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \text{ where } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i^p\right)^2} \quad (9)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(y_i - y_i^p\right)\right| \quad (10)$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\left(\frac{y_i - y_i^p}{y_i}\right)\right| * 100 \quad (11)$$

4. Results and Discussion

4.1. Model Performance

Both RF and XGB achieved strong predictive accuracy, with RF performing slightly better with 0.817011, as shown in Table 4.

The RF, XGB, and Neural Network models' residual plots were plotted in Figures 7, 8, and 9, respectively.

The Residuals Plot and prediction error visualizations confirmed a balanced model behaviour, with errors distributed symmetrically and predictions aligning closely with actual values.

The Prediction error plot for RF and XGB was demonstrated in Figures 9 and 10, respectively.

Figure 11 presents the comparison of neural networks with other Models.

Table 4 shows the model's performance comparison (R-squared).

Table 4. Prediction result

Model	R^2 score	MAE	RMSE	MAPE %
XGB	0.81701	87739.69504	157851.4610	16.09018
RF	0.8150	87739.6950	157851.4610	16.09018
NN	0.53116	165454.32413	251331.94010	34.34042

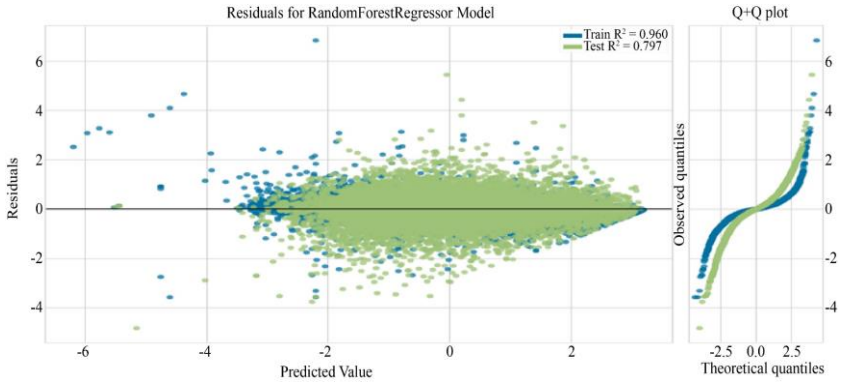

Fig. 7 RF model residual plot

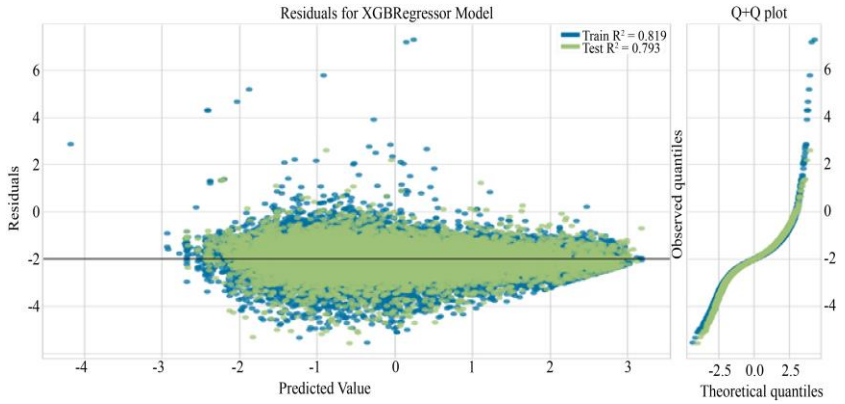

Fig. 8 XGB model residual plot

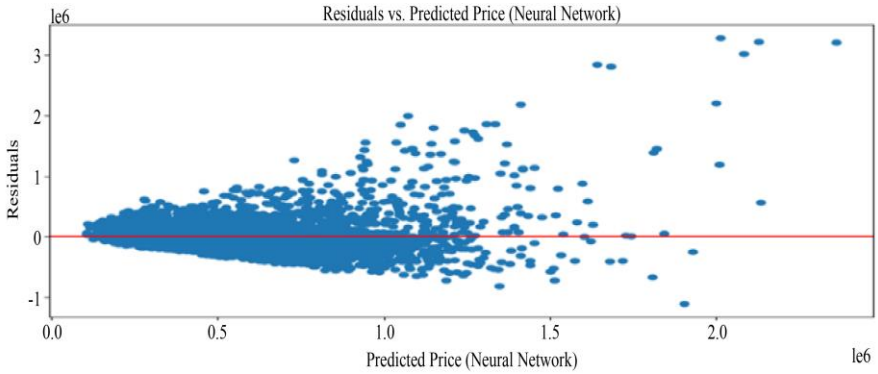

Fig. 9 Neural network residual plot model residual plot

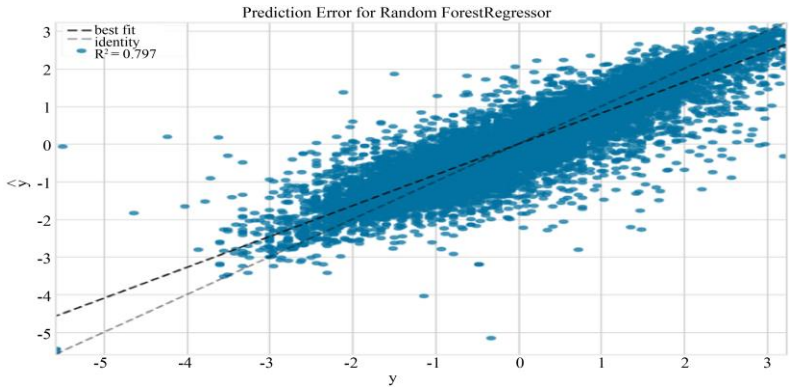

Fig. 10 Prediction error plot for RF model

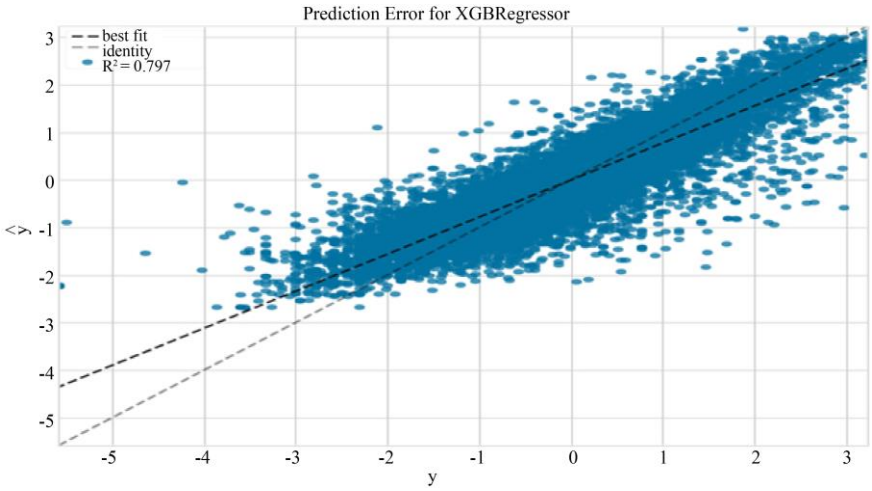

Fig. 11 Prediction error plot for XGB model

4.1.1. Comparison of Neural Networks and Other Models

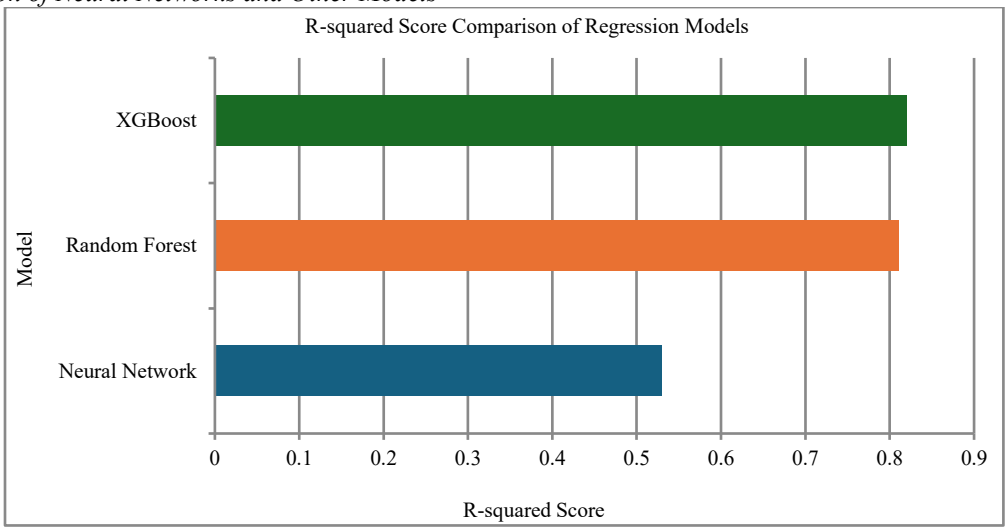

Fig. 12 The models comparison with neural networks model

4.2. Determinant Analysis

SHAP summary plots highlighted lot size (acre_lot) and number of bathrooms as the strongest predictors in the XGB model. For RF, ICE plots revealed that bathrooms, house size, and zip code played major roles in price determination. These interpretability tools reinforce that both structural features and geographic identifiers drive housing values.

Figure 13 provides an understanding of the nature of the relationships. The XGB model, for bath, house_size, and acre_lot, observes that an increase in feature value increases the SHAP values.

An indication that larger values for the features will lead to a higher predicted price for the RE property. Furthermore, it is observed that the zip code and city where the RE property is located also serve as major determinant factors for the house prices. Additionally, comparing the curves of distinct instances is made simpler by the centre ICE plots. When we want to see the difference between a prediction and a fixed point in the feature range rather than the absolute change of a predicted value, this can be useful.

Figure 14 reveals that, on average, the dataset features for bath, zip_code, and house_size were the major determining factors that influenced the price prediction. While on average, the bed, acre_lot, and city features remained constant. Finally, the SHAP summary plot and the ICE plot depict the impact of features on the outcome.

A house feature selection function can be provided by RE trading platforms to customers based on the ranking of feature importance and their impact on the RE price. This makes it easier for RE customers to select houses according to their needs.

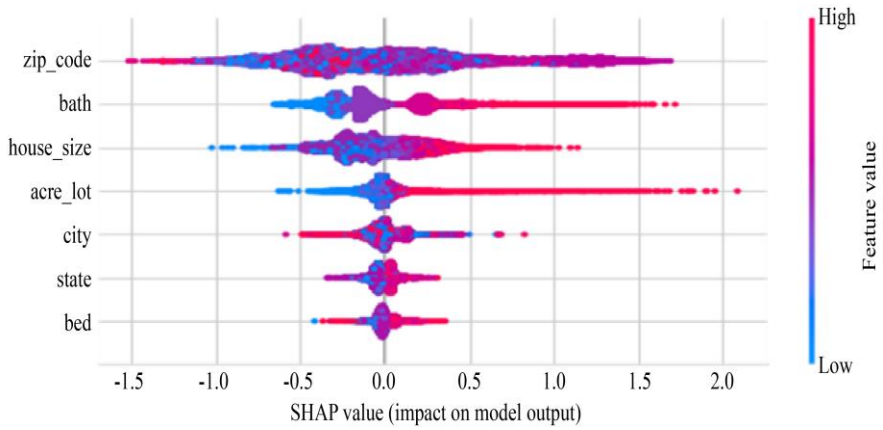

Fig. 13 SHAP summary plot for XGB model

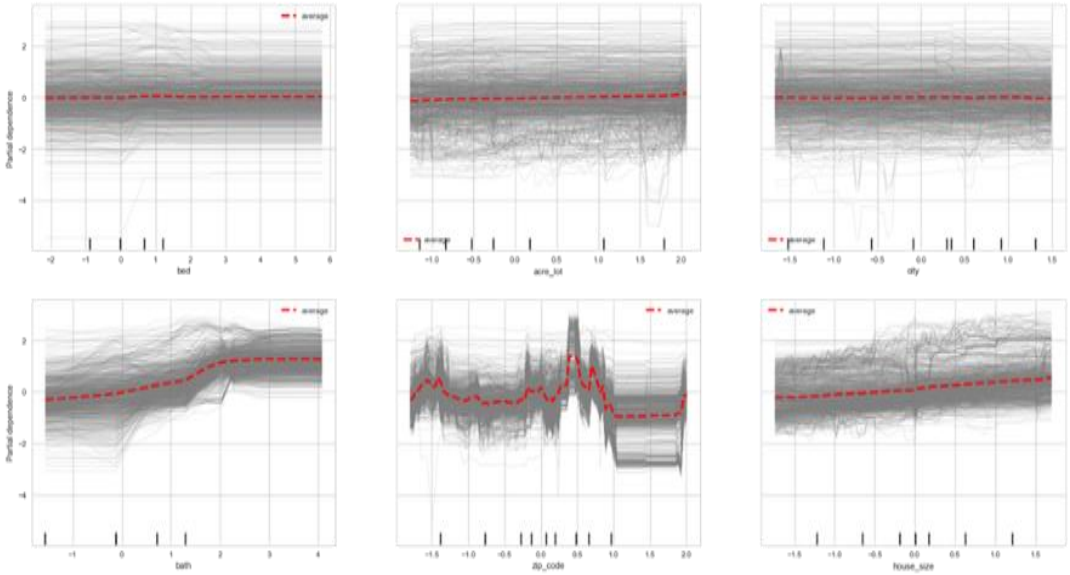

Fig. 14 ICE plot for RF model

4.3. Feature Importance

Figures 15, 16, and 17 present the top ten model feature importance from high (Grade) to low(bedrooms) order of contribution. This means that house grade is the highest determinant of the house price prediction, while bedroom is the tenth feature.

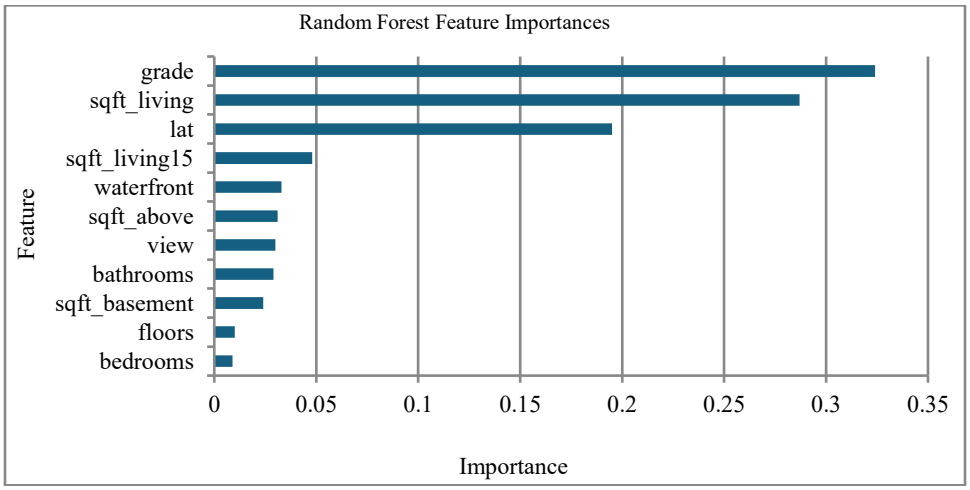

Fig. 15 Random Forest feature importance

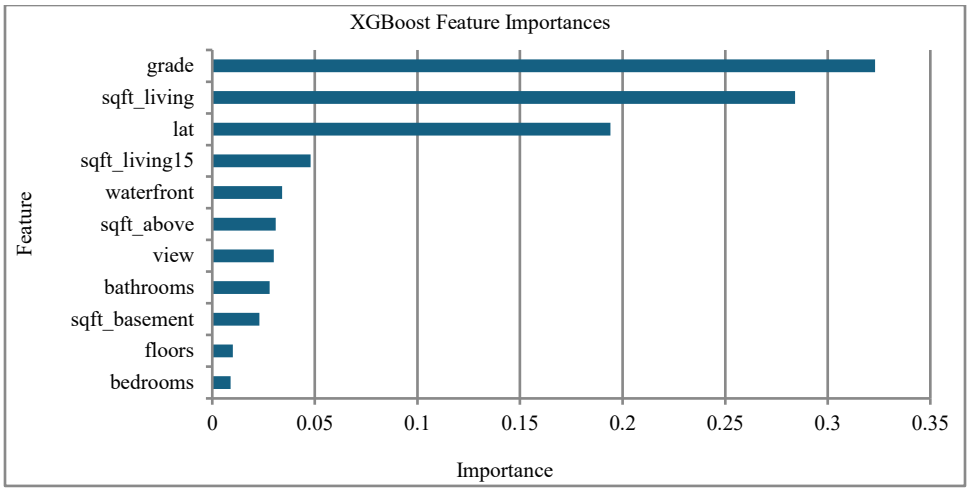

Fig. 16 XGB feature importance

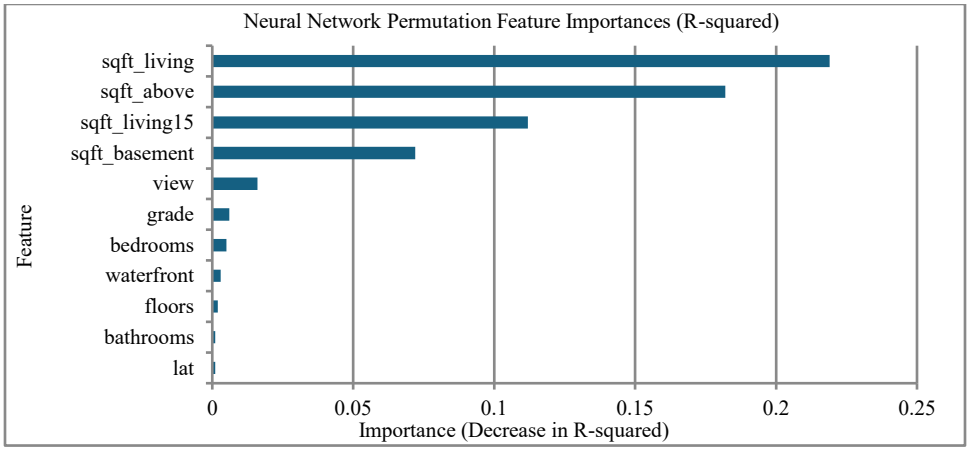

Fig. 17 Neural network feature importance

4.4. How Metrics Impact Real-World Decisions on Estate Price Prediction

The study adopted some common metrics for house price prediction, including MAE**,** MSE /RMSE**,** R^2 (coefficient of determination), and ror (MAPE.

Metrics affect real-world decisions as follows:

Mortgage Lending & Risk Management: Banks use predictions to decide loan amounts. If MAE is too high**,** borrowers may be over-lent or under-lent. If RMSE is minimized**,** the bank avoids catastrophic losses on luxury properties.

Property Taxation: Governments often rely on predicted market values for setting property taxes. High bias in predictions could lead to unfair taxation (over-taxing poorer neighborhoods, under-taxing wealthy ones).

Real Estate Investment: Investors look at predictions to decide when/where to buy properties. Low R^2 means the model does not capture market drivers (risky investments). MAPE helps investors compare errors across cities or countries.

Urban Planning & Policy: City planners use models to forecast housing affordability and plan infrastructure. If residuals show systemic bias (e.g., always undervaluing rural houses), policies may favour cities over rural communities.

The results demonstrate that ensemble models, particularly RF, offer reliable predictions of housing prices in large, diverse datasets. Integrating SHAP and ICE allowed us to move beyond accuracy metrics to understand why predictions were made. For instance, the positive influence of bathrooms and lot size aligns with conventional expectations of property valuation but provides a quantified, data-driven confirmation.

This combination of predictive accuracy and interpretability has practical implications. Buyers can use such models to evaluate properties against their budgets, developers can benchmark pricing strategies, and policymakers can better assess regional housing market

dynamics. Moreover, the results show that explainable ML can bridge the gap between high-performing algorithms and stakeholder trust

5. Conclusion and Future Works

We applied RF, XGB, and NN models to predict housing prices using a large U.S. real estate dataset, which was evaluated with standard regression metrics. RF slightly outperformed XGB, achieving an R^2 of 0.817011. SHAP and ICE analyses revealed that bathrooms, lot size, and house size are among the most significant determinants of price. The findings contribute a dual benefit: robust predictive performance and transparent explanations of model behaviour. This framework can inform decision-making for buyers, sellers, and regulators, helping to improve trust and efficiency in real estate markets.

Future research should expand to multimodal data sources (e.g., images, socio-economic indicators) and explore deep learning approaches such as ANN, ANFIS, and MANFIS. These methods may enhance predictive power further, particularly for high-value properties or emerging markets. In many developing economies, lack of transparency in the RE market, legal infrastructure, expert personnel, knowledge and experience in valuation, data bank, and deficiencies in the economy are factors that cause RE valuation to be poorly conducted. The use of more advanced methods will enhance the evaluation accuracy of the valuation reports to various stakeholders.

5.1. Limitations

The Kaggle housing datasets used in the study (e.g., USA real estate dataset) are location-specific, capturing features like neighborhood, proximity to schools, lot size, or zoning regulations in one city/region. The dataset is not good for the generalizability of findings because the geographic insights from that dataset may not transfer to other regions or contexts. Generalizability of findings may also be affected by regional economic factors, cultural and social preferences, and policy and regulation.

References

[1] Amina Adadi, and Mohammed Berrada, "Peeking inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138-52160, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[2] Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001. [CrossRef] [Google Scholar] [Publisher Link]

[3] Liujia Chen et al., "Measuring Impacts of Urban Environmental Elements on Housing Prices based on Multisource Data-A Case Study of Shanghai, China," *International Journal of Geo-Information*, vol. 9, no. 2, pp. 1-23, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Tianqi Chen, and Tong He, "Xgboost: Extreme Gradient Boosting," *R Package Version 0.4-2*, pp. 1-4, 2025. [Google Scholar]

[5] Adele Cutler, D. Richard Cutler, and John R. Stevens, *Random Forests*, Ensemble Machine Learning, Springer, pp. 157-175, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[6] Xibin Dong et al., "A Survey on Ensemble Learning," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 241-258, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] A. Fazel Famili et al., "Data Preprocessing and Intelligent Data Analysis," *Intelligent Data Analysis*, vol. 1, no. 1-4, pp. 3-23, 1997. [CrossRef] [Google Scholar] [Publisher Link]

[8] Jerome H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232, 2001. [Google Scholar] [Publisher Link]

[9] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, "Additive Logistic Regression: A Statistical view of Boosting (with Discussion and a Rejoinder by the Authors)," *The Annals of Statistics*, vol. 28, no. 2, pp. 337-407, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[10] Julia García Cabello, "Mathematical Neural Networks," *Axioms*, vol. 11, no. 2, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Alex Goldstein et al., "Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation," *Journal of Computational and Graphical Statistics*, vol. 24, no. 1, pp. 44-65, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[12] Riccardo Guidotti et al., "A Survey of Methods for Explaining Black Box Models," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1-42, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[13] David Gunning et al., "XAI-Explainable Artificial Intelligence," *Science Robotics*, vol. 4, no. 37, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Dave Gunning et al., "DARPA's Explainable AI (XAI) Program: A Retrospective," *Authorea Preprints*, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Hani Hagras, "Toward Human-Understandable, Explainable AI," *Computer*, vol. 51, no. 9, pp. 28-36, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[16] Salim Heddam, *Intelligent Data Analytics Approaches for Predicting Dissolved Oxygen Concentration in River: Extremely Randomized Tree Versus Random Forest, MLPNN and MLR*, Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation, Springer, pp. 89-107, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Andreas Holzinger et al., *Explainable AI Methods - A Brief Overview*, xxAI - Beyond Explainable AI, Springer, Cham, pp. 13-38, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Castro Gbememali Hounmenou, Kossi Essona Gneyou, and Romain Lucas Glele Kakaï, "A Formalism of the General Mathematical Expression of Multilayer Perceptron Neural Networks," *Preprints*, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Anil Jadhav, Dhanya Pramod, and Krishnan Ramanathan, "Comparison of Performance of data Imputation Methods for Numeric Dataset," *Applied Artificial Intelligence*, vol. 33, no. 10, pp. 913-933, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[20] Joseph D. Janizek, Safiye Celik, and Su-In Lee, "Explainable Machine Learning Prediction of Synergistic Drug Combinations for Precision Cancer Medicine," *BioRxiv Preprint*, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[21] Soren Jordan, Hannah L. Paul, and Andrew Q. Philips, "How to Cautiously Uncover the 'Black Box'of Machine Learning Models for Legislative Scholars," *Legislative Studies Quarterly*, vol. 48, no. 1, pp. 165-202, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Hua Li et al., "XGBoost Model and its Application to Personal Credit Evaluation," *IEEE Intelligent Systems*, vol. 35, no. 3, pp. 52-61, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[23] Andy Liaw, and Matthew Wiener, "Classification and Regression by Randomforest," *Scientific Research*, vol. 2, no. 3, pp. 18-22, 2002. [Google Scholar] [Publisher Link]

[24] Scott M. Lundberg, and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, pp. 4768 - 4777, 2017. [Google Scholar] [Publisher Link]

[25] CH. Raga Madhuri, G. Anuradha, and M. Vani Pujitha, "House Price Prediction using Regression Techniques: A Comparative Study," *2019 International Conference on Smart Structures and Systems (ICSSS)*, Chennai, India, pp. 1-5, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[26] J. Manasa, Radha Gupta, and N.S. Narahari, "Machine Learning based Predicting House Prices using Regression Techniques," *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Bangalore, India, pp. 624-630, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Yuan Meng et al., "What Makes an Online Review more Helpful: An Interpretation Framework using XGBoost and SHAP Values," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, no. 3, pp. 466-490, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Rory Mitchell, and Eibe Frank, "Accelerating the XGBoost Algorithm using GPU Computing," *PeerJ Computer Science*, vol. 3, pp. 1-37, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[29] Naci Büyükkaraciğ, "Modern Methods Approach in Real Estate Valuation," *Ankara Iksad Publications House*, pp. 1-130, 2021. [Google Scholar] [Publisher Link]

[30] Mieczysław Lech Owoc, and Adam Stambulski, "Software Quality Management: Machine Learning for Recommendation of Regression Test Suites," *Journal of Economics and Management*, vol. 47, no. 1, pp 117-137, 2025. [Google Scholar] [Publisher Link]

[31] Elli Pagourtzi et al., "Real Estate Appraisal: A Review of Valuation Methods," *Journal of Property Investment and Finance*, vol. 21, no. 4, pp. 383-401, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[32] Byeonghwa Park, and Jae Kwon Bae, "Using Machine Learning Algorithms for Housing Price Prediction: The Case of Fairfax County, Virginia Housing Data," *Expert Systems with Applications*, vol. 42, no. 6, pp. 2928-2934, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[33] Prasad Patil, What is Exploratory Data Analysis?, Towards Data Science, 2018. [Online]. Available: https://medium.com/data-science/exploratory-data-analysis-8fc1cb20fd15

[34] Robi Polikar, *Ensemble Learning*, Ensemble Machine Learning, Springer, New York, NY, pp. 1-34, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[35] Luca Rampini, and Fulvio Re Cecconi, "Artificial Intelligence Algorithms to Predict Italian Real Estate Market Prices," *Journal of Property Investment and Finance*, vol. 40, no. 6, pp. 588-611, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[36] Baijayanta Roy, "All About Feature Scaling," *Towards Data Science*, pp. 3-19, 2021. [Google Scholar] [Publisher Link]

[37] Robert E. Schapire, *The Boosting Approach to Machine Learning: An Overview*, Nonlinear Estimation and Classification, pp. 149-171, Springer, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[38] Mark R. Segal, "Machine Learning Benchmarks and Random Forest Regression," *UCSF: Center for Bioinformatics and Molecular Biostatistics*, pp. 1-14, 2004. [Google Scholar] [Publisher Link]

[39] Hasan Selim, "Determinants of House Prices in Turkey: Hedonic Regression Versus Artificial Neural Network," *Expert systems with Applications*, vol. 36, no. 2, pp. 2843-2852, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[40] Robert P. Sheridan et al., "Extreme Gradient Boosting as a Method for Quantitative Structure-Activity Relationships," *Journal of Chemical Information and Modeling*, vol. 56, no. 12, pp. 2353-2360, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[41] Andreja Stojić et al., "Explainable Extreme Gradient Boosting Tree-Based Prediction of Toluene, Ethylbenzene and Xylene Wet Deposition," *Science of The Total Environment*, vol. 653, pp. 140-147, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[42] Ean Zou Teoh et al., "Explainable Housing Price Prediction with Determinant Analysis," *International Journal of Housing Markets and Analysis*, vol. 16, no. 5, pp. 1021-1045, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[43] Ayush Varma et al., "House Price Prediction using Machine Learning and Neural Networks," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, India, pp. 1936-1939, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[44] Adrian Weller, "Challenges for Transparency," *Open Review*, pp. 1-8, 2017. [Google Scholar] [Publisher Link]

[45] Yanli Wu et al., "Application of Alternating Decision Tree with Adaboost and Bagging Ensembles for Landslide Susceptibility Mapping," *Catena*, vol. 187, pp. 1-58, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[46] Emine Yaman, and Abdulhamit Subasi, "Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Automated EMG Signal Classification," *BioMed Research International*, vol. 2019, pp. 1-13, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[47] Cha Zhang, and Yunqian Ma, *Ensemble Machine Learning: Methods and Applications*, 1st ed., Springer New York, NY, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[48] Zhi-Hua Zhou, "When Semi-Supervised Learning Meets Ensemble Learning," *International Workshop on Multiple Classifier Systems*, Reykjavik, Iceland, pp. 529-538, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[49] Zhi-Hua Zhou, "When Semi-Supervised Learning Meets Ensemble Learning," *Frontiers of Electrical and Electronic Engineering*, vol. 6, no. 1, pp. 6-16, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[50] Mohammad Zounemat-Kermani et al., "Ensemble Machine Learning Paradigms in Hydrology: A Review," *Journal of Hydrology*, vol. 598, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[51] Jozef M. Zurada, Alan S. Levitan, and Jian Guan, "Non-Conventional Approaches to Property Value Assessment," *Journal of Applied Business Research*, vol. 22 no. 3, pp. 1-14, 2006. [CrossRef] [Google Scholar] [Publisher Link]