Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
VOLUME:9, ISSUE:1, JANUARY 2026, PAPER NUMBER:12

Predicting Future House Price Using Machine Learning and Web-Based Deployment for Price Estimation

Dr.C.Radhika[1], Eswari.P[2],Ganga Devi.J[3],A.S.Jaya Sarala[4],R.Keerthana[5],S.Lakshya[6]

[1]Faculty, [2-6]Students
Department of Computer Science -SFS ,
Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for women,
University of Madras, India. email:radhika.c@sdnbvc.edu.in[1] , gangadevigangsj@gmail.com

Abstract: *The research involves creating a predictive model by using advanced machine learning approaches for future house price forecasting. The housing market is influenced by various factors, including location, property characteristics, and economic indicators, making accurate price prediction is complex, yet essential task for buyers, sellers, and investors. The process begins with the collecting large comprehensive dataset that includes historical house prices and relevant features such as square footage, number of bedrooms, and local amenities. The dataset undergoes thorough preprocessing to clean and normalize the features. Used boosting algorithms, such as AdaBoost and CatBoost, are then trained and evaluated based on their predictive accuracy. The best-performing model is selected for future price predictions. Additionally, the model is deployed on a user-friendly webpage, enabling stakeholders to input property details and receive accurate price estimates for prospective sales, thereby facilitating informed decision-making in the real estate market.*
Keywords: *Machine learning, preprocessing, CatBoost, AdaBoost.*

1. INTRODUCTION:

The housing market is a dynamic and complex environment influenced by a multitude of factors, including location, property characteristics, and broader economic indicators. For buyers, sellers, and investors, understanding future house prices is crucial for making informed decisions. Accurate price predictions can help stakeholders navigate the market effectively, whether they are looking to buy their dream home, sell a property at the right time, or invest in real estate for profit. However, the intricacies of the housing market make this task challenging, necessitating the use of advanced analytical techniques to derive meaningful insights PonnagantiArjun et al[1].

The goal of the work intends to address this difficulty by creating a prediction model using machine learning methods to project future property prices. Start with compiling a thorough dataset including past housing prices together with the appropriate details including square footage, number of bedrooms, nearby facilities, and economic data. The study is based on this dataset, which lets us find trends and connections impacting property values Miguel Alvarez de LineraAlperi et al[2].

By employing robust data preprocessing methods, we ensure that the dataset is clean and normalized, setting the stage for effective model training. Once the data is prepared, we explore various machine learning algorithms, particularly boosting techniques like AdaBoost and CatBoost, to train our predictive model. These algorithms are evaluated based on their accuracy in predicting house prices, and the best-performing model is selected for deployment AurelienGeron[3].
The design is a user-friendly platform where stakeholders may enter property details and get reliable pricing predictions for next sales in order to make our results easily available. Apart

from improving knowledge of the housing market, this initiative offers an efficient way for stakeholders to make informed decisions based on facts.

2. OBJECTIVES:

The aim of the "Predicting Future House Prices Using Machine Learning with Python" is to create a reliable predictive model that effectively estimates house prices by leveraging historical data and a range of significant factors. This initiative employs machine learning techniques, including AdaBoost and CatBoost, to scrutinize essential attributes such as property features and geographical location, thereby uncovering trends that influence pricing. The model is developed using an extensive dataset, with its efficacy assessed through performance indicators like the R^2 score and Mean Squared Log Error.

The research aspires to furnish stakeholders with dependable price forecasts, facilitating informed choices in real estate dealings and investments. Furthermore, it aims to illustrate the connections between various features and the target variable, thereby improving the clarity of the model's predictions. By providing insights into the impact of different elements on housing prices, the project will offer valuable guidance for buyers, sellers, and investors. Ultimately, the objective is to produce a robust predictive instrument that not only anticipates housing prices but also aids in comprehending market dynamics, thus enhancing strategic decision-making in real estate transactions.

3. LITERATURE SURVEY :

Table1 summarizes a selection of studies that explore various tools and algorithms employed in this domain.Krishna Rao N. V. and DhanaLaxmi B. [4]utilized PYOT and PYUIC for automated graphical user interface (GUI) development, which streamlines the process of creating user interfaces. However, their approach is limited when it comes to handling complex relationships within the data. ZonaKostic[5]focused on image processing tools, which offer the advantage of integrating visual data into analyses. This integration can enhance the interpretability of results; however, it also introduces challenges related to computational complexity, making it resource-intensive.NimaRizun and Anna Baj-Rogowska[6] employed cross-correlation analysis, which allows for innovative utilization of data by identifying relationships between different variables. Despite its strengths, this method faces limitations concerning data availability and the need for log transformations, which can complicate the analysis.

TABLE 1 :Different tools used for prediction of future house price.

Author	Tools	Advantages	Disadvantages
[4]Krishna Rao N. V. and DhanaLaxmi B.	PYOT, PYUIC	Automated GUI Development	Not suitable for complex Relationship
[5]Aleksandar, Jevremovic	Image processing Tools	Integration of Visual Data	Computational Complexity
[6]NimaRizun, Anna Baj-Rogowska	Cross correlation analysis	Innovative Data Utilization	Data Limitation, Log Consideration

Table2 summarizes a selection of studies that explore various algorithms employed in this domain.J.J. Wang et al[7]. utilized the back-propagation algorithm, which is known for its accuracy in predicting house prices. However, this method is also characterized by a high

susceptibility to errors, which can affect the reliability of the predictions.Zongyan Yang et al[8]. implemented Graph Convolutional Neural Networks (GCNNs), which demonstrate robustness across diverse datasets, making them versatile for various applications. Nonetheless, this approach incurs high computational costs, which may limit its practicality in resource-constrained environments.HeminVasani et al[9] focused on advanced regression techniques, achieving enhanced predictive accuracy. However, the complexity of these advanced models can pose challenges in terms of interpretability and implementation.ShashiBhushanJha[10] employed XGBoost, a method that leverages real data to compare different models, providing valuable insights. However, its application is limited to a specific location, which may restrict the generalizability of the findings.Yuying Wu and Youshan Zhang [11]utilized Convolutional Neural Networks (CNN) in conjunction with Natural Language Processing (NLP) techniques, resulting in more accurate predictions. However, this approach requires powerful computing resources, which may not be accessible to all researchers.

TABLE 2:Different Algorithms used for prediction of future house price.

Author	Algorithms Used	Advantages	Disadvantages
[7]J.J.Wang et al.	Back-propagation	Accurate house price prediction	High error-susceptibility
[8]Zongyan Yang et al.	GCNNS	Robustness across diverse dataset	High computational lost
[9]Heminvasani et al.	Advanced Regression Techniques	Enhanced predictive Accuracy	Complexity of advanced regression models
[10]ShashiBhushan Jha et al.	XGBoost	Uses Real data compares different model	Limitation to one location
[11]Yuyingwu, Youshan Zhang	CNN, NLP	More Accurate prediction	Require powerful computers

4. METHODOLOGIES:

Machine Learning:

Machine learning is a form of artificial intelligence that facilitates the self-learning of data and then the implementation of that learning without the necessity of human involvement.Y.Liu[12]. Machine learning forecasts an output by aggregating data with statistical instruments. Corporate uses this output then to provide practical insights. The machine uses an algorithm to create responses from data input M.Swamynathan[13].

Boosting Algorithms:

- Boosting is an ensemble modeling technique that attempts to build a strong model from the number of weak model.
- Boosting can improve the accuracy of the model by combining several weak models accuracies and averaging them for regression to increase the accuracy of the model.

Types of boosting algorithm:

1. AdaBoost
2. CatBoost

Adaboost:

AdaBoost is a boosting algorithm that creates a strong classifier by combining weak classifiers. The algorithm works by weighting the weak classifiers so that they vote with more importance.

AdaBoostRegressor is a machine learning algorithm that is used for regression problems in python. It is an ensemble method that combines multiple weak models (i.e., models with low predictive power) to create a strong model with high predictive power T. Hastie et al[14].

Catboost:

CatBoost uses a method called ordered encoding to encode categorical features. Ordered encoding considers the target statistics from all the rows prior to a data point to calculate a value to replace the categorical feature. Another unique characteristic of CatBoost is that it uses symmetric trees.CatBoost converts categorical values into numbers using various statistics on combinations of categorical features and combinations of categorical and numerical featuresProkhorenkova et al[15].

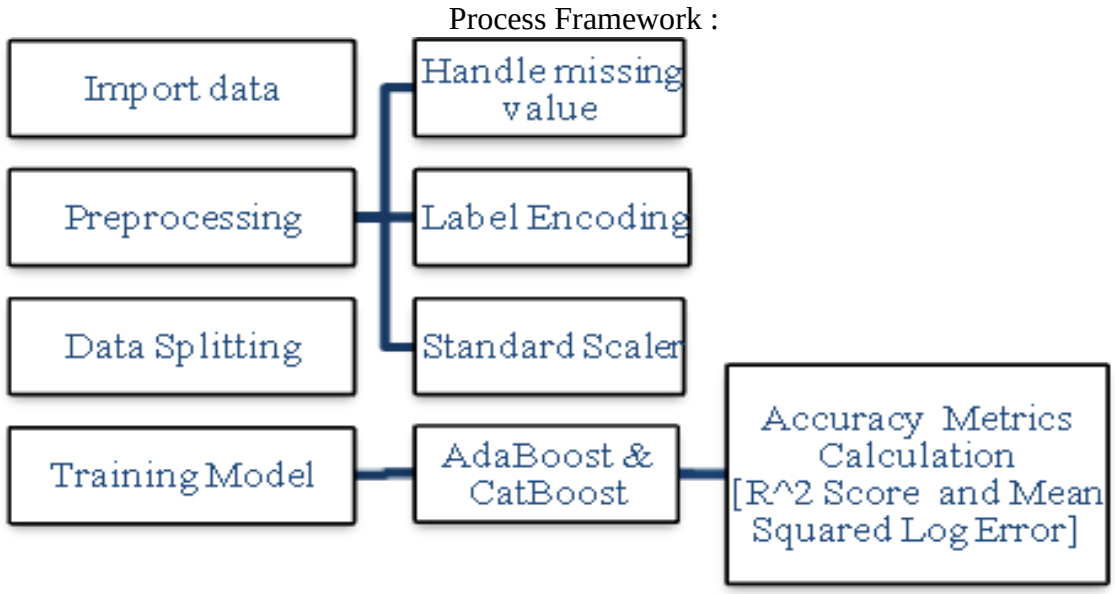


Figure 1:Detailed Analysis of Machine Learning Models and deployment of webpage.

The Process framework outlines the entire workflow,first import the libraries that are important for the computer to be able to perform any data manipulations as required, handle any machine learning functions, and also for performing model evaluation. These libraries include NumPy for numerical functions, Pandas for table data, Pickle for Python object serialization, Scikit-learn containing many machine learning functions, and Warnings which will disable any warning messages during the program run. Subsequently, the house prices data set is brought in from the CSV by loading it up into a Data Frame which is a way of representing data that can be easily examined and manipulated afterwards. The attributes utilized in the house price prediction dataset include various features that users can input through a user-friendly interface. These features encompass:

House Type: A dropdown menu for selecting the type of house (e.g., apartment, bungalow, condominium).
Location: A text input field for entering the location of the house.
Size of the House: A numeric input field for specifying the size of the house in square feet.
Number of Bathrooms: A numeric input field for indicating the number of bathrooms in the house.
Balcony Area: A numeric input field for entering the size of the balcony area in square feet.
Submit Button: A button that, when clicked, submits the input features to the backend server for processing.

By incorporating these attributes, the model can effectively learn from the training data and make accurate predictions based on user inputs.

The dataset was then prepared for creating and evaluating a predictive model by segmenting these given features into a training portion from which the model will learn and a testing portion comprising 20% of the total which will be utilized to test the models efficacy after training completes.Afterwards, the AdaBoost regression model is constructed with the proper parameters and after that it is fit into the training dataset with set learning rate and number of estimators. In addition, the code also uses the CatBoost model, which is specialized for categorical features, and this model is as well fit to the same training data. In this regard, a function is specified to show differences in the performance of both the models AdaBoost and CatBoost by making predictions and measuringperformance by calculating efficiency and printing the formatted outcomes to decide which model has smarter prediction capability by receiving the results based upon the R^2 scores.

5. PERFORMANCE EVALUATION METRIC OF THE MODELS

The performance of AdaBoost and CatBoost was evaluated using Mean Squared Log Error (MSLE) and R^2 score. CatBoost outperformed AdaBoost with an R^2 of 0.6139 compared to 0.2519, indicating better pattern recognition. Its lower MSLE of 0.1695 further demonstrates its effectiveness in capturing relative variations. The performance analysis of different machine learning models are depict in Table3.These results confirm CatBoost's superiority, making it the primary model for house price predictions in the research.The last step in implementing the code is testing the trained model with an input vector of a fictional house to predict its price depending on its specific parameters. In addition, the trained model does not require retraining, as it is saved to a file through Pickle. A plotting tool is also used to demonstrate the results and show the accuracy rates of both models in Figure 2, thus allowing everyone to easily compare their performance.

TABLE 3: The performance analysis of different machine learning models.

TRAINING MODAL	**R^2 SCORE**	**MEAN SQUARED LOG ERROR(MSLE)**
AdaBoost	0.2519	0.2670
Catboost	0.6139	0.1695

After providing accurate predictions for house prices, "Room Hive"(webpage) presents this information on a user-friendly website designed to enhance the overall experience. The application predicts house prices based on various input features, with the frontend crafted using HTML, CSS, and JavaScript to ensure a visually appealing and interactive user interface. The backend is powered by Python Flask, a micro web framework that facilitates the rapid development of web applications. Figure 3 and Figure 4 represent the web page of prediction of house price. Flask is recognized for its versatility and user-friendly nature, allowing developers to create applications with ease. Its minimalist design encourages experimentation and customization, enabling developers to focus on feature development without the burden of unnecessary complexity, making it suitable for both beginners and experienced developers alike.

Figure 2 :Visual Representation of Predicted House Prices.

Figure 3: Selecting the features of the house according to user's need.

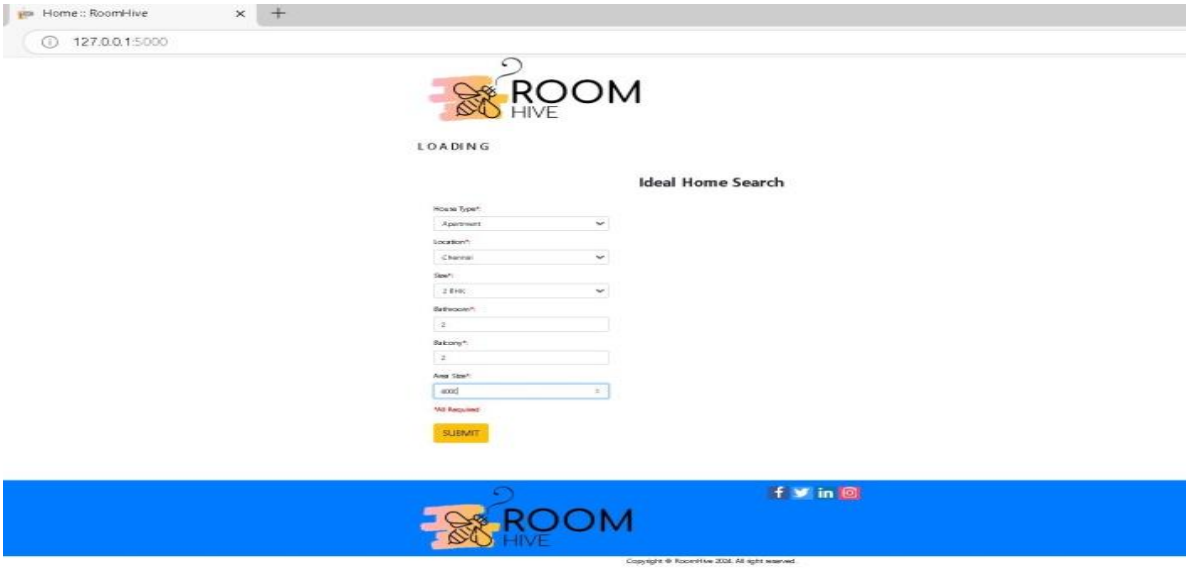
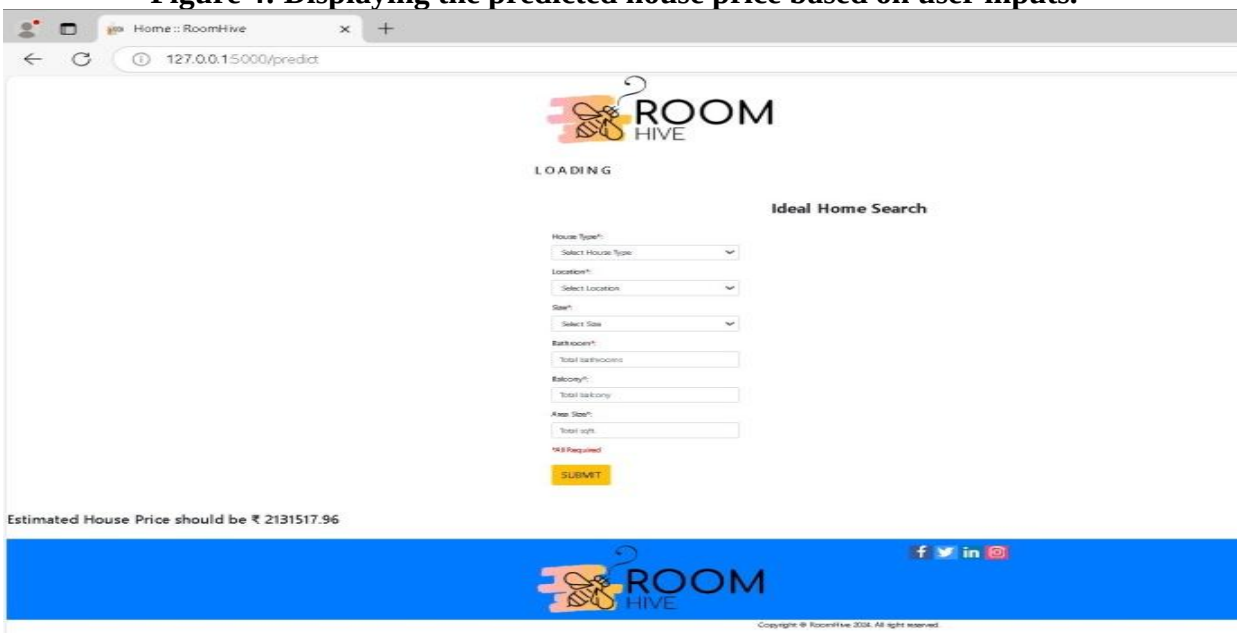


Figure 4: Displaying the predicted house price based on user inputs.



6. CONCLUSION :

This research focuses on predicting house prices through machine learning. It examines multiple factors to deliver accurate predictions, assisting buyers, sellers, and real estate agents in making informed decisions. Advanced algorithms such as AdaBoost and CatBoost provide reliable, data-driven results. The proposed model designed to be user-friendly, even for individuals without technical expertise, featuring clear visuals and metrics. This tool enhances market comprehension and empowers users to navigate real estate transactions with confidence, adapting to market changes for ongoing support. The future of predicting house prices using machine learning is promising. By leveraging real-time data and advanced algorithms, while taking into account various factors such as economic conditions, we can achieve more accurate predictions. Developing user-friendly applications for both buyers and real estate agents can enhance the impact of this technology. As the market evolves, these machine learning tools can adapt, providing innovative solutions. Overall, this project aims to facilitate smarter investment decisions and improve understanding of the real estate market.

7. References:

[1] P. Arjun et al., "Enhancing House Price Prediction Accuracy and Precision: A Data Mining Approach with Python and Stacking Algorithm," 2024.

[2]M. A. de LineraAlperi, J. P. Fernandez, and F. G. Martinez, "AI Algorithms in Real Estate: A Roadmap to Precision Housing Price Predictions," Int. J. Artif. Intell.Research, vol. 12, no. 3, pp. 250-265, 2024.doi: 10.1109/IJAI.2024.00125.

[3]A. Geron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems," 2022.

[4]Krishna Rao N. V. and DhanaLaxmi B., "Regression Based Prediction of House Prices using Python," 2019.

[5]Z. Kostic and A. Jevremovic, "What image features boost housing market predictions," J. Real Estate Research, vol. 40, no. 2, pp. 87-102, 2021. doi: 10.1109/JRER.2021.1009342.

[6]N. Rizun and A. Baj-Rogowska, "Can web search queries predict price change on the real estate valuation," J. Comput. Econom., vol. 28, no. 5, pp. 1123-1138, 2021. doi: 10.1109/JCE.2021.0023245.

[7] J. J. Wang, Z. Xie, L. Zhang, and Y. Xu, "Predicting future house prices with a memristor-based artificial neural network," IEEE Access, vol. 6, pp. 5012-5022, 2018. doi: 10.1109/ACCESS.2018.2800652.

[8]Z. Yang, Z. et al., "Graph convolutional network-based model for megacity real estate valuation," IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 6, pp. 2298-2312, 2022.doi: 10.1109/TNNLS.2022.3150095.

[9] H.Vasani et al., "House Price Prediction Using advanced regression Techniques",2024.

[10] S. B. Jha, R. Mishra, and P. K. Sahu, "Machine learning approaches to real estate market prediction problem," Comput. Ind. Eng., vol. 92, pp. 97-107, 2020. doi: 10.1016/j.cie.2020.06.002.

[11]Y. Wu and Y. Zhang, "House price prediction based on deep learning," J. Real Estate Finance Econ., vol. 62, no. 3, pp. 297-312, 2022. doi: 10.1007/s11146-022-09824-9.

[12]Y. Liu, "Python Machine Learning By Example," 2017.

[13]M. Swamynathan, "Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python," 2019.

[14]T. Hastie et al.,"*The Elements of Statistical Learning: Data Mining, Inference, and Prediction"*, 2009.

[15]Prokhorenkova, L. et al., "CatBoost: Gradient boosting with categorical features," J. Mach. Learn. Res., vol. 20, no. 1, pp. 1-7, 2019.doi: 10.1007/JMLR.2019.00124.