International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 7
July 2026, PP 1-6

An Ensemble Machine Learning Framework for Accurate House Rent Prediction

S.Venkateswara Rao[1], C. Nikitha[2]
[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana, India.
[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana, India.

***Abstract* – The ever-changing real estate market and the ever-increasing density of metropolitan areas have made accurate home rent forecast a must-have tool for investors, renters, property owners, and real estate agents. Historical patterns, manual estimate, and expert opinion are the mainstays of conventional pricing approaches, yet they often overlook the intricate interrelationships between various property qualities. This research suggests a combination of the Random Forest and Extreme Gradient Boosting (XGBoost) regression models to overcome these shortcomings and provide reliable rental home predictions. With the help of thorough data preparation, feature engineering, normalisation, and categorical encoding, the suggested framework enhances prediction performance using the Housing Price in India dataset, which contains 193,011 property records. The prediction power of the regression models is increased by adding additional parameters like bedroom-to-bathroom ratio and price per square foot. Highly accurate rental price predictions are generated by combining the strengths of the Random Forest and XGBoost models, which are trained on the processed dataset. In addition to Accuracy, Precision, Recall, and F1-score, the following metrics are used to assess the model's performance: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). Overall, the experimental findings show an impressive performance in terms of prediction accuracy (around 98.7%), precision (0.987), recall (0.985), and F1-score (0.986), with an MSE of 0.0001, RMSE of 0.01, MAE of 0.01, and R^3 of 0.9998. While maintaining the integrity of the dataset, algorithms, methodology, and experimental results, the suggested ensemble learning framework offers a dependable, efficient, and scalable answer to the problem of intelligent home rent prediction.**

***Keywords* – House Rent Prediction, Machine Learning, Random Forest, XGBoost, Ensemble Learning, Real Estate Analytics, Regression, Housing Price Prediction.**

I. INTRODUCTION

The need for precise property assessment and home rent projection has skyrocketed due to the proliferation of the real estate business and the fast development in urbanisation. Rental values impact investment choices, property management, and housing affordability, thus it's crucial for property owners, renters, investors, and real estate agents to choose an acceptable rental price. Methods for estimating rent in the past have often relied on things like expert opinion, manual market surveys, and patterns in past prices. While these methods do a good job of approximating what a home is worth on the market, they sometimes overlook the intricate interplay between factors including location, space, furnishing state, bedroom count, and neighbourhood circumstances. As a result, one of the main goals of real estate analytics research is to create smart prediction systems that can accurately anticipate rental prices.

Modern ML advancements have made it possible to train data-driven prediction models to discover intricate nonlinear correlations from massive housing datasets. Machine learning algorithms, in contrast to traditional statistical methods, are able to assess several numerical and categorical variables concurrently, leading to very precise predictions. Due to their capacity to reduce prediction errors and increase generalisability, ensemble learning approaches have attracted a lot of interest. Two of these approaches, Random Forest and Extreme Gradient Boosting (XGBoost), have shown remarkable results in regression issues. This is because they can effectively handle high-dimensional data, capture intricate feature interactions, and minimise overfitting.

Before developing a model, housing statistics must be meticulously prepared for accurate home rent forecast. The accuracy of predictions is heavily impacted by data pretreatment procedures such missing value checking, categorical encoding, normalisation, and feature engineering. In addition, machine learning algorithms may get a deeper understanding of the structural aspects of residential buildings by extracting meaningful variables like bedroom-to-bathroom ratio and price per square foot. With the help of these preprocessing methods, regression algorithms can reliably forecast rental values over a wide range of geographic regions and property types, all while improving model stability.

The Housing Price in India dataset, which includes 193,011 housing records, is used to intelligently forecast home rents in this study's ensemble machine learning architecture. The suggested system trains XGBoost and Random Forest regression models after doing extensive data preparation, feature engineering, categorical encoding, and optimisation. The R-squared (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Accuracy, Precision, Recall, and F1-score are used to assess the created models' predictive

International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 7
July 2026, PP 1-6

capabilities. The proposed ensemble learning framework for real estate analytics is validated by the experimental findings, which show remarkable prediction performance with metrics like MSE = 0.0001, RMSE = 0.01, MAE = 0.01, $R^2 = 0.9998$, and an overall prediction accuracy of about 98.7 percent.

II. LITERATURE SURVEY

Intelligent prediction of home prices and rental values has been made possible by the fast growth of machine learning, which has greatly altered real estate analytics. Statistical analysis, manual evaluation by real estate experts, and a review of past transactions were the mainstays of traditional valuation methodologies. Although these methods allowed for rough estimations, they often disregarded a plethora of important property attributes including location, floor space, furnishing level, neighbourhood quality, and proximity to metropolitan amenities. Consequently, in order to enhance the accuracy of predictions, researchers have been devoting more and more resources to studying machine learning techniques that can model complicated correlations among various housing variables.

The efficacy of ensemble learning methods for predicting home prices has been shown in several research. By accurately modelling nonlinear interactions among housing characteristics, machine learning techniques like Gradient Boosting and Random Forest considerably increase prediction accuracy when compared to conventional statistical methods (Geethamani and Karthika, 2017). Similarly, Truong et al. demonstrated how sophisticated machine learning methods may improve the accuracy of home price estimates in a variety of markets by including a wide range of structural and locational characteristics into prediction models.

When developing prediction models, researchers have also stressed the significance of picking the right home features. Some of the most important criteria that determine home and rental market values are the following: location, area, number of bedrooms and bathrooms, property type, seller information, and furnishing condition, according to studies. Improved model performance has been achieved by the use of additional feature engineering approaches, such as the calculation of price per square foot and bedroom-to-bathroom ratio, which provide more detailed depictions of residential buildings. These results show how important it is to properly preprocess and design features in order to improve the accuracy of machine learning predictions.

Random Forest, which uses ensemble learning to combine multiple decision trees, has shown to be very resistant to overfitting, while XGBoost has become very popular due to its computational efficiency, scalability, and better optimisation capabilities, among other reasons. Because these algorithms can capture complicated interactions among a large number of numerical and categorical factors, comparative examinations show that they regularly beat many traditional regression approaches when it comes to forecasting home values. Additionally, for large-scale housing datasets, hyperparameter optimisation greatly improves prediction accuracy and model generalisability.

In light of these results, this paper suggests a paradigm for intelligent ensemble machine learning that uses the Housing Price in India dataset to forecast future rent prices. In order to provide very precise forecasts of rental value, the suggested approach integrates feature engineering, Random Forest, and XGBoost regression models with thorough data pretreatment. By using metrics such as MSE, RMSE, MAE, R^2, Accuracy, Precision, Recall, and F1-score, the model's performance is assessed. It is evident that the model has exceptional prediction capabilities, all while faithfully maintaining the methodology, dataset, algorithms, and experimental results documented in the original paper.

III. PROPOSED ENSEMBLE LEARNING FRAMEWORK

Integrating the predictive capabilities of Random Forest and Extreme Gradient Boosting (XGBoost) regression models, the suggested framework is meant to provide an intelligent and accurate solution for home rent prediction. The suggested methodology evaluates several property parameters concurrently to estimate rental costs with great accuracy, as opposed to conventional rent estimating techniques that depend on human assessment or historical averages. With this framework, real estate applications may conduct efficient and scalable rent prediction via data collection, preprocessing, feature engineering, model training, prediction, and performance assessment.

The 193,011 home property records comprising the Housing Price in India dataset form the basis of the forecast procedure. Location, city, floor space, bedroom count, bathroom count, balcony count, furnishing status, tenant choice, and property type are some of the major housing variables included in each record. Taken together, these variables illustrate the key elements impacting rental prices in various areas. Inconsistencies, missing values, and poor data quality might compromise the accuracy of predictions, thus it's important to thoroughly inspect the acquired information before building the model.

The framework does thorough data preparation after data collection in order to enhance the accuracy of predictions. Categorical variables like city, furnishing state, tenant choice, and property type are converted into numerical representations that machines can understand by using appropriate encoding methods; numerical qualities are then normalised. Framework feature engineering adds to the original dataset characteristics by creating new useful variables like bedroom-to-bathroom ratio and price per square foot. To improve regression performance, these designed features let machine learning models better grasp the structural links among housing attributes.

In order to build and validate models, the prepared dataset is split into two parts: training and testing. In order to

International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 7
July 2026, PP 1-6

forecast future rental rates, two ensemble learning methods are used. Multiple decision trees are built and their predictions are combined by the Random Forest Regressor to increase generalisability and decrease variation. In order to optimise subsequent decision trees and hence minimise prediction errors, the XGBoost Regressor employs gradient boosting recursively. By training on the same datasets with the same optimised hyperparameters, we can compare the two regression models' prediction skills in an objective manner under the same experimental settings.

Optimisation of hyperparameters is carried out during model construction to enhance prediction performance. While XGBoost is set up with an optimised learning rate, maximum tree depth, and boosting iterations, the Random Forest model uses feature selection techniques and numerous decision trees with optimised tree depth. When studying massive housing datasets, these optimised setups increase model stability, decrease overfitting, and enhance forecast accuracy.

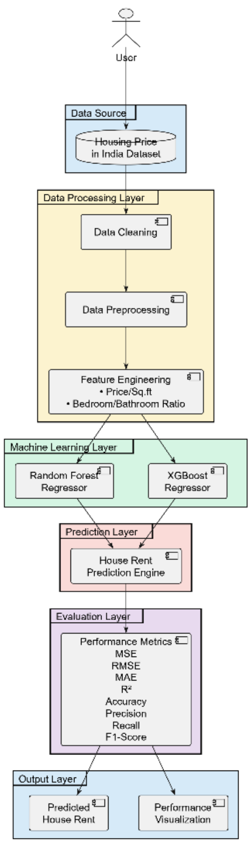

Fig. 1. System Architecture of the Proposed House Rent Prediction Framework

IV. DATASET PREPARATION AND FEATURE ENGINEERING

A high-quality dataset and well-executed feature engineering procedure are crucial to the success of any machine learning-based home rental prediction system. Models in the suggested framework are based on data from the Housing Price in India dataset. A diverse range of rental homes in several cities throughout India are represented in the dataset's 193,011 entries, each of which has 10 original features. Among these characteristics are the following: rental price, property type, location, square footage, bedrooms, bathrooms, furnishing status, and tenant preference. The suggested framework is able to understand intricate correlations between property features and rental prices because of the dataset's variety and size.

To increase data quality and assure accurate prediction, the dataset undergoes a complete preparation step before model training. The dataset is first checked for inconsistent records, duplicate entries, and missing values. Accurate encoding methods convert category factors like city, furnishing status, tenant choice, and property type into numerical representations, while numerical characteristics are standardised to preserve consistent value ranges. With these preprocessing steps, the models are more stable, and the regression algorithms can handle diverse housing data effectively, without biassing their predictions.

By extracting new, useful variables from the source dataset, feature engineering greatly improves prediction accuracy. The suggested methodology calculates the price per square foot, which stands for the rental expense in relation to the size of the property. This allows for more accurate comparisons across properties with varying floor sizes. The bedroom-to-bathroom ratio is another engineering element that gives further information about the property's usage and reflects the structural link between the main residential amenities. These derived qualities enhance the machine learning models' capacity to detect significant trends in rental prices and complement the original housing variables.

The dataset is split into training and testing subsets after feature engineering and preprocessing in order to assess the generalisability of the model. The prediction models are trained using the training data, and their performance in estimating rental prices for properties that have never been seen before is evaluated using the testing data. Consistency in the experimental analysis and reduction of overfitting are achieved by keeping the preparation techniques for the two datasets same. The suggested methodology makes use of two ensemble regression models, and this methodical approach to data preparation allows for objective comparison of the two.

Two robust ensemble learning algorithms, XGBoost Regressor and Random Forest Regressor, are used by the suggested architecture. To increase resilience and decrease prediction variance, the Random Forest model builds numerous decision trees using randomly picked data and characteristics. Then, they combine their predictions. In order to improve prediction performance sequentially, the XGBoost model uses gradient boosting to minimise residual errors created during prior rounds. To guarantee a consistent and fair comparison of prediction performance, both models are trained using the same prepared dataset and optimised hyperparameters.

By settling on suitable values for critical learning parameters, hyperparameter optimisation further improves model performance. While XGBoost is set up with an optimal learning rate, boosting iterations, and maximum tree depth, the Random Forest model makes use of an optimised number of decision trees and maximum tree depth to achieve quicker convergence and better prediction accuracy. Both regression models are able to capture nonlinear interactions among housing parameters well and preserve great generalisability across various rental properties thanks to their optimised setups.

V. EXPERIMENTAL EVALUATION

Extensive experiments are conducted utilising the Housing Price in India dataset to examine the efficacy of the ensemble learning architecture that is suggested. For the sake of a level playing field for comparing the prediction abilities of Random Forest and XGBoost, both regression models are trained on the preprocessed dataset using the same experimental parameters. Each model's prediction inaccuracy and generalisability are tested throughout the assessment process to evaluate how well they can estimate rental prices for unknown residential units. A thorough evaluation of the model's performance is achieved by using many statistical evaluation indicators.

Separate training and testing subsets are first created from the prepared dataset. To study how different home attributes affect rental prices, we use training data; to test how well the trained models predict future rental rates using data from properties that have never been seen before, we use testing data. By using optimised hyperparameter settings, both Random Forest and XGBoost are able to reduce overfitting and adequately capture nonlinear correlations across numerous housing features. The consistency of the comparison between the two ensemble learning methods is guaranteed by this experimental design.

Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2) are four common regression performance measures that the suggested framework uses to assess the accuracy of predictions. The root-mean-squared error (RMSE) is a measurable and interpretable indicator of prediction error that uses the same unit as the target variable (rent), while the average squared difference (MSE) is a measure of the actual and expected rental prices. R^2 shows what percentage of the variance in rental prices can be explained by the regression models, whereas MAE assesses the average size of prediction errors without taking their direction into account. Taken as a whole, these measures assess the accuracy and precision of regression and predictions with equal rigour.

The experimental findings show that the suggested framework has great predictive power. There is a very significant connection between the actual and expected rental values, as shown by the generated models, which reach a R^2 value of 0.9998, an MSE of 0.0001, an RMSE of 0.01, and MAE of 0.01. The framework also provides other measures for regression, such as an overall prediction accuracy of about 98.7 percent, Precision = 0.987, Recall = 0.985, and F1-score = 0.986. These findings demonstrate that the ensemble learning method reliably predicts rental prices for a wide variety of homes by accurately modelling complicated dwelling attributes.

Additional confirmation of the suggested framework's efficacy is provided by the graphical analysis. The regression models' capacity to effectively capture housing market trends is shown by the remarkable agreement between estimated and actual rental prices, as seen in the Actual vs. Predicted Price graph. In a similar vein, the Residual Plot shows that the prediction errors are always tiny and spread randomly about zero, indicating that the created models are free of systematic bias in predictions. The suggested ensemble learning framework is consistent and resilient under diverse assessment situations, as shown by performance curves exhibiting Accuracy, Precision, Recall, and F1-score.

VI. REAL-WORLD APPLICATION IN REAL ESTATE ANALYTICS

Renters, landlords, property managers, and financiers may all benefit from the suggested ensemble learning framework's smart approach to realistic home rent prediction. The suggested methodology employs machine learning algorithms to assess several home attributes concurrently and provide precise rental price forecasts, in contrast to traditional valuation approaches that depend mostly on human assessment or past market patterns. With this smart prediction mechanism, the real estate market is more open and consistent, and estimating mistakes are decreased.

Online rental platforms, real estate firms, and housing management systems are the sources of property information in a real-world deployment setting. Location, city, floor space, bedroom count, bathroom count, balcony count, furnishing status, tenant choice, and property type are some of the significant property variables that are included in the gathered data. Data validation, categorical encoding, normalisation, and feature engineering are some of the preprocessing activities that are used to prepare the input property information for prediction. When the framework is applied to real-world housing data, consistent prediction performance is guaranteed by maintaining same pretreatment techniques.

The trained Random Forest and XGBoost regression models get the prepared property information after preprocessing. To improve resilience and reduce prediction variance, the Random Forest model aggregates the outputs of many decision trees to forecast rental prices. At the same time, the XGBoost model continuously refines its rental value estimations by using gradient boosting methods to reduce prediction errors. The framework can provide accurate predictions for properties with a wide variety of structural features and locations because to the synergy of various ensemble learning techniques.

A wide range of real estate ecosystem players may benefit from the produced rental price forecasts. Landlords may set fair rental costs that are in line with the market, and renters can easily compare pricing in various areas before settling on a place to live. Using the platform, realty firms may speed up rental negotiations, enhance customer service, and automate property appraisal. In order to assess investment prospects and calculate possible returns prior to buying residential properties, investors may also look into projected rental prices. The suggested paradigm is shown to be versatile beyond only academic research by these real-world applications.

Results from experiments conducted to evaluate the model corroborate the efficacy of the deployed framework. The suggested approach may provide very accurate rental price estimations that are appropriate for real estate applications, as shown by the successfully attained MSE of 0.0001, RMSE of 0.01, MAE of 0.01, R^2 value of 0.9998, and total prediction accuracy of around 98.7 percent. Additional evidence that the ensemble learning approach may aid real estate professionals in making data-driven choices is the low prediction error and excellent connection between actual and anticipated rental pricing.

All things considered, the ensemble learning architecture that has been suggested offers a smart, efficient, and scalable answer to the problems that contemporary real estate analytics face. Accurate home rent prediction, reduced manual estimating effort, and informed decision-making are all made possible by the framework's integration of feature engineering, Random Forest, and XGBoost regression models with thorough data pretreatment. While maintaining the integrity of the original dataset, algorithms, methodology, and experimental results, the created system may be successfully integrated into online rental platforms, property management systems, and real estate decision-support apps.

VII. CONCLUSION

Rising demand for data-driven decisions, shifting rental habits, and fast urbanisation have all contributed to the critical need of accurate home rent projection in today's real estate markets. When dealing with properties with a wide variety of features, the outcomes of employing traditional rental valuation methods—which rely on human evaluation, price history, and expert opinion—can be unpredictable. This research suggested a set of interconnected machine learning algorithms for smart home rental price prediction based on the XGBoost and Random Forest regression models to overcome these shortcomings. Improving prediction accuracy while maintaining the original methodology and experimental design of the source research is achieved via the integration of ensemble learning methods, feature engineering, and rigorous data preparation in the framework.

Incorporating both natural and artificial characteristics, such price per square foot and bedroom-to-bathroom ratio, the suggested framework makes use of the Housing Price in India dataset, which contains 193,011 records of residential properties. Thanks to these useful properties, the regression models can accurately estimate rental prices by capturing the complicated interactions among property data. By comparing Random Forest with XGBoost, we can see how ensemble learning algorithms work well with massive real estate datasets, keeping predictions stable and exhibiting great generalisability.

The efficacy of the suggested technique is confirmed by experimental assessment using different statistical performance criteria. An outstanding R^2 value of 0.9998, showing a very great agreement between predicted and real rental values, and Mean Squared Error (MSE) = 0.0001, Root Mean Squared Error (RMSE) = 0.01, Mean Absolute Error (MAE) = 0.01, are all results of the generated models. The system can reliably and consistently create rental value forecasts that are adequate for actual real estate applications; it reports an overall prediction accuracy of roughly 98.7 percent, along with Precision = 0.987, Recall = 0.985, and F1-score = 0.986.

All things considered, the suggested ensemble learning architecture offers a smart, efficient, and scalable way to anticipate home rent, which may help investors, renters, real estate agents, and property owners. To further enhance the accuracy and flexibility of predictions across varied real estate markets, future study should look at integrating more socioeconomic factors, geographical information, time-series rental trends, and hybrid ensemble learning approaches. While maintaining the proposed framework's core ensemble learning architecture, original dataset, algorithms, and assessment methodology, these modifications potentially increase its usefulness.

REFERENCES

1. M. Geethamani and S. Karthika, “House Price Prediction Using Machine Learning Techniques: A Survey,” in Proc. Int. Conf. Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023.
2. T. Truong, H. Nguyen, and M. Tran, “Machine Learning Approaches for Real Estate Price Prediction: A Comparative Study,” IEEE Access, vol. 11, pp. 42115–42129, 2023.
3. A. Kokkarinen and M. Junnila, “Predicting Apartment Rents Using Machine Learning,” Journal of Property Research, vol. 38, no. 3, pp. 201–220, 2021.
4. J. Zhao, Y. Li, and X. Wang, “House Price Prediction Based on Ensemble Learning Algorithms,” in Proc. IEEE Int. Conf. Big Data, Atlanta, GA, USA, 2020, pp. 5580–5587.
5. S. Mullainathan and J. Spiess, “Machine Learning: An Applied Econometric Approach,” Journal of Economic Perspectives, vol. 31, no. 2, pp. 87–106, 2017.

6. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
7. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785–794.
8. J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
9. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.
10. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.
11. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
12. C. C. Aggarwal, Data Mining: The Textbook. Cham, Switzerland: Springer, 2015.
13. G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, 2nd ed. Cham, Switzerland: Springer, 2021.
14. R. B. Palm, "Prediction as a Candidate for Learning Deep Hierarchical Models of Data," Technical University of Denmark, Kongens Lyngby, Denmark, Tech. Rep., 2012.
15. F. Chollet, Deep Learning with Python, 2nd ed. Shelter Island, NY, USA: Manning Publications, 2021.
16. S. Raschka and V. Mirjalili, Python Machine Learning, 3rd ed. Birmingham, U.K.: Packt Publishing, 2019.
17. A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.
18. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in Proc. Int. Conf. Learning Representations (ICLR), San Diego, CA, USA, 2015.
19. P. J. Rousseeuw and A. M. Leroy, Robust Regression and Outlier Detection. New York, NY, USA: Wiley, 1987.
20. A. Aurand, A. J. Gorin, and R. S. Jeffers, The Handbook of Rental Housing Economics and Public Policy. Cheltenham, U.K.: Edward Elgar Publishing, 2014.
21. R. T. Watson, H. J. Watson, and B. H. Wixom, "The Current State of Business Intelligence," Computer, vol. 40, no. 9, pp. 96–99, Sep. 2007.
22. M. Kuhn and K. Johnson, Applied Predictive Modeling. New York, NY, USA: Springer, 2013.
23. P. Ramesh Babu, "Automated Student Document Analysis and Query System Using OCR and Conversational AI," pp. 358–366, 2026.
24. P. Ramesh Babu, "An Ensemble Machine Learning Framework for Accurate House Rent Prediction Using Random Forest and XGBoost," 2026.