Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 63(2): 123-132 (2026)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(63-2)708

Pakistan Academy of Sciences

Research Article

A Robust Stacked Ensemble Approach for House Price Prediction

Muhammad Khateeb Khan[1]*, Asad Ullah Khan[2], Salman Ahmed[3], and Wajid Ali[4]

[1]Faculty of Engineering & Computing, Software Engineering, NUML, Islamabad, Pakistan

[2]Faculty of Computing & Technology, Computer Science, IQRA, Islamabad, Pakistan

[3]Faculty of Engineering & Computing, Computer Science, NUML, Islamabad, Pakistan

[4]School of Computer Science & Artificial Intelligence, Computer Science,
Duy Tan University, Danang, Vietnam

Abstract: Real estate price prediction plays a fundamental role in decision-making and ensuring market stability for buyers, sellers, and property owners. Since the real estate sector has a significant impact on the global economy, accurate prediction models are crucial. In this study, we used the Ames Housing dataset and performed thorough preprocessing to handle missing values, reduce multicollinearity, and enhance feature quality. We introduced three new features, total square footage, total porch area and the age of the house to capture the better property characteristics. After these steps, we evaluated a range of machine learning (ML) models, including Linear Regression, Ridge, Lasso, Decision Trees, Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost and explored hybrid combinations with weighed ensemble and stacked ensemble strategies. We proposed CXL Stacked Ensemble (CatBoost + XGBoost + LightGBM) model with ridge regression as meta-learner and the model outperformed with R^2 of 0.9333 and a normalized RMSLE of 0.1132, demonstrating the superior model predictive capability. To validate its robustness fold-wise RMSE comparisons and significance tests (paired t-test and Wilcoxon signed-rank test) were conducted. Additionally, cross-validation and residual error analysis validated model reliability. This work depicts the effectiveness of stacked ensemble methods for real state price predication and provides practical insights for stakeholders to make informed decision.

Keywords: Real Estate, Stacked Ensemble Methods, Price Prediction, Residuals Analysis, Multimodal.

1. INTRODUCTION

Real estate is vital to the world economy and overall economic growth. In fact, real estate prices are intricately linked to the dynamics of the real estate market and play an important role in the development of nations, both domestically and internationally. Some of the factors that affect property pricing are geographic, size of the property, economic, and demographic changes. It is, therefore, essential that real estate buyers, sellers, investors, and realtors be able to predict the price of a house accurately when dealing in real estate investment. In the past few years, machine learning has proven to be an invaluable asset in extracting insightful patterns from intricate datasets, whether they are large or small, in various regions and industries of real estate. The Boston Housing dataset from the UCI ML Repository has been used in many studies and has several features related to the property and real estate market. Data mining algorithms were applied by Buyrukoglu and Uzut [1] in the field of real estate price prediction and they identified great potential of such methods in the market. A recent research study by Ding [2] has used three classifiers to test a bunch of features on the Boston housing dataset and XGBoost outperformed while comparing with others. Crime rates, property tax rates, and other local amenities are among the potential variables that could impact property value and are often tested and refined to develop a predictive model. For instance, Harrison and Rubinfeld [3] conducted a

Received: September 2025; Revised: May 2026; Accepted: June 2026
* Corresponding Author: Muhammad Khateeb Khan <m.khateeb231@gmail.com>

basic study on the Boston Housing Dataset and used multiple regression analysis to explore the effect of air pollution and other variables on housing prices. Other research used supervised ML algorithms including the Decision Trees and Support Vector Machines (SVM) to improve accuracy in prediction, Xibin *et al*. [4]. In addition, the use of robust and powerful ensemble models such as Random Forests (RF) has been highlighted as useful for high data variability contexts Lundberg *et al*. [5]. While the feature engineering and ensemble modeling techniques have led to significant improvements, there are still several issues that need to be addressed, including further fine-tuning the performance of the models and gaining insight into the conditions and limitations of each model. Aniobi *et al*. [6] made a comparison between six regression models LR, LASSO, Ridge, KNN-R, DTR, and ETR. Linear Regression, and had achieved an R^2 of 0.57 with moderate predictive capability.

Researchers have concentrated mainly on making improvements to prediction accuracy, rather than considering the deployment of models for real estate stakeholders. Further, the implementation of user-friendly interfaces for non-technical professionals is still a challenge and is a factor that is still constraining the actual use of these models, Tekouabou *et al*. [7]. In the study of prediction of house prices in Boston, Chen [8] used LR, RF, XGBoost and SVM. Random Forest had the highest R^2 value of 0.8641 while SVM had the lowest R^2 value of 0.5900. With the economy's current state, discussing property values has grown in importance over the years. Muralidharan *et al*. [9] employed Decision Trees and Artificial Neural Networks to forecast the price of houses in residential areas based on the housing and crime statistics. The study emphasizes the significance of housing markets in the U.S. economy, and companies such as Zillow use massive amounts of data to forecast the market and understand consumer behavior.

The recent studies were carried out on the Ames Housing dataset to show the effectiveness of machine learning techniques to predict house prices. Han [10] explored advanced regression models and emphasized the importance of validation and model selection in ensuring good performance. Kumar *et al*. [11] proposed an improved XGBoost model, which has high predictive accuracy with an R^2 of around 0.923 and low RMSE, which is better than the traditional approaches. Likewise, Zhao [12] tested a variety of models and found that ensemble methods, especially Random Forest, exhibit very good performance, with an R^2 of 0.85 and an RMSE of 3536. Though these are exciting developments, there are still open research questions concerning feature redundancy, model interpretability and optimal integration of heterogeneous learners.

This research aims of the present study are: (1) implement and test different algorithms of ML models to select the most accurate model for predicting property prices, and (2) use a stacked ensemble model to improve prediction accuracy, so that the stakeholders can determine the optimal value of properties by selecting the relevant features.

2. MATERIALS AND METHODS

In our methodology, we utilized the Ames Housing dataset and applied pre-processing techniques to remove noise and handle missing values. Feature engineering techniques were applied to construct additional informative variables. After pre-processing the dataset, we applied various ML and neural network models to achieve the desired predictions. A comprehensive evaluation was conducted using 10-fold cross-validation across multiple ML models, including linear, tree-based, and ensemble methods. The models were assessed using performance metrics such as R^2, RMSE, MAE, MAPE, and RMSLE. Furthermore, weighted averaging and stacking ensemble techniques were applied to improve predictive accuracy, and the best-performing model was identified based on comparative results, as illustrated in Figure1.

2.1. Dataset Description

The Ames Housing Database consists of a detailed information of housing properties in Ames, Iowa, recorded between 2006 and 2010. This particular housing dataset is comprised of 2,930 records, each containing 69 variables that cover a variety of aspects of each property. The features can be organized into various categories, which include information related to general housing attributes such as the type of structure, year of construction, and the overall condition of the house; information about the lot on which the house stands, including the lot size and neighborhood location; size of the

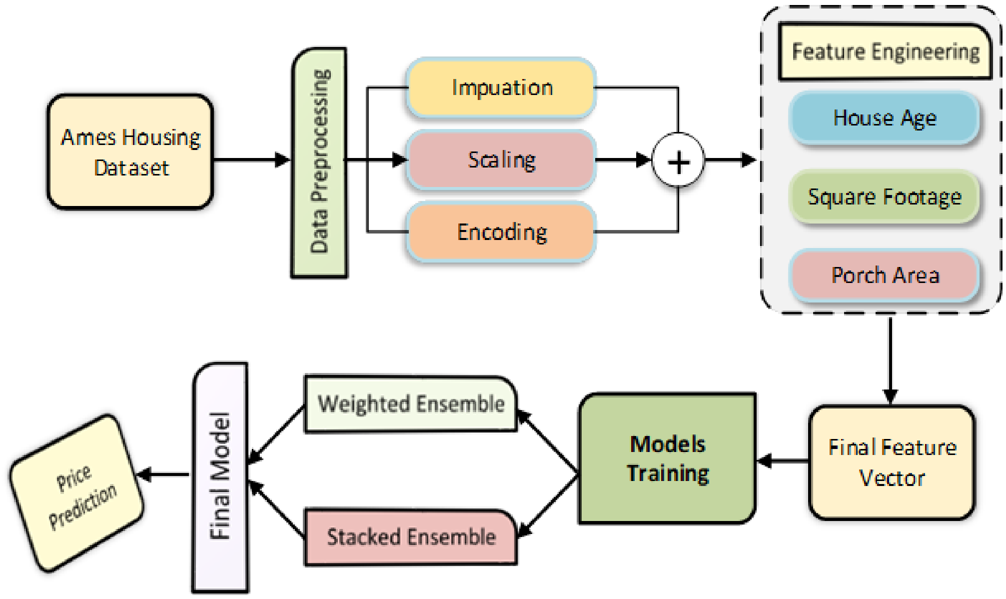

Fig. 1. Proposed Research Methodology.

house and room information; information about structural and exterior material composition of the house; information regarding the garage and basement attributes; amenity information including heating, central air, fireplace, and pool attributes; and time-related information including the month and year of sale.

2.2. Pre-processing

In pre-processing stage, first we examined the missing values and used the Simple Imputer module for handling the missing values by replacing them by median and removed the features like Alley, fence with large proration of missing values. One hot encoding techniques was used to cope the nominal categories such as (Neighborhood, MSZoning) and ordinal encoding technique was used to handle the ordinal variables like (ExterQual, OverallQual) to preserve their rank. Numerical features were standardized to have zero mean and unit variance, and outliers in key features, such as GrLivArea and LotArea, were detected and handled to reduce skewness. The preprocessed data were used for feature engineering to finalize the feature vector which was fed into the ML models for house price prediction.

2.3. Feature Engineering

The feature engineering stage was used to construct composite variables that maintain the overall information regarding property size, usability, and age-related characteristics. In this work, three key features are derived. First, the total square footage is computed by aggregating basement, first-floor, and second-floor areas, from existing feature space and then the total porch area is calculated by aggregating all porch-related spaces, reflecting outdoor living area, which has significant role in property value. A temporal feature is introduced representing age of the house and is being calculated by the difference between the year of sale and the year of construction, which helps model depreciation or appreciation effects over time. These engineered features significantly enhance the model's ability to capture structural and temporal variations in housing prices.

2.4. Final Feature Vector Space

The features were preprocessed, and then the engineered features were added to the feature vector space. This high dimensional feature space is drawn in Figure 2, where housing samples are distributed according to their corresponding price levels, ranging from low to high. To facilitate interpretation, the entire feature space was projected into a two-dimensional embedding using t-SNE, represented by Dimension 1 and Dimension 2. Furthermore, the top 10% high-priced houses were identified and highlighted based on the dataset features, particularly housing size and related structural attributes.

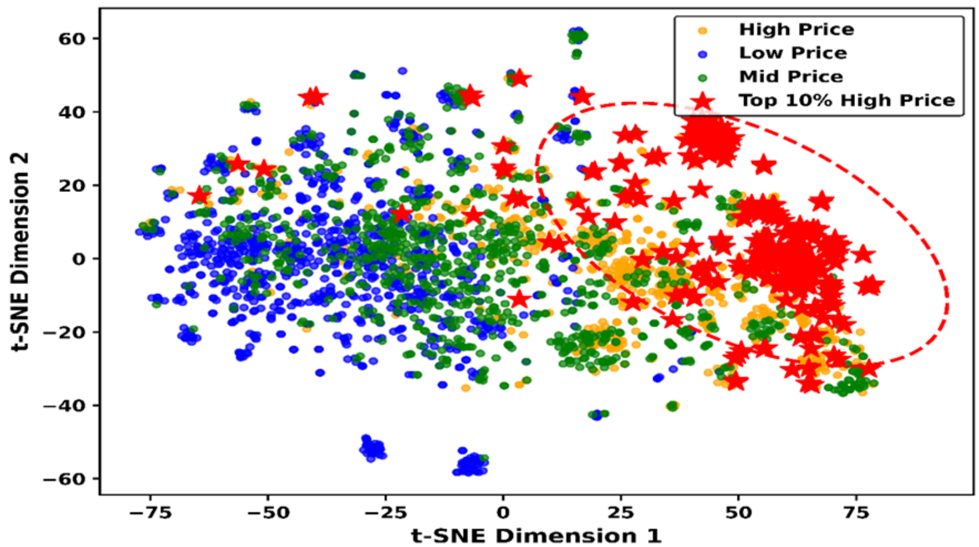

Fig. 2. Final Feature Vector Space Visualization.

2.5. Model Training and Testing

In the model selection phase, we conducted extensive experiments with a variety of classifiers, including Linear Regression, James *et al*. [13] and Shbool *et al*. [14]; Ridge, Lasso, Decision Tree, Random Forest, Breiman [15]; hybrid model porcupine and BP neural network, Hong and Li [16], Gradient Boosting and XGBoost, Hong and LightGBM, Himanshu [17], and neural network, Wijaya [18]. Additionally, we investigated the various fusion approaches such as weighted and stacked ensemble methods, to find and select the suitable model for real estate price prediction.

Model evaluation is conducted using a 10-fold cross-validation strategy to ensure generalization and avoid overfitting. Out-of-fold predictions are generated using cross_val_predict, allowing unbiased performance estimation on unseen data without requiring a separate hold-out test set. Multiple evaluation metrics are employed, including R^2, adjusted R^2, RMSE, normalized RMSE, MAE, MAPE, and RMSLE, providing a holistic assessment of model performance both in original and log-transformed scales.

2.6. Proposed Stacking Ensemble

The three gradient boosting algorithms of XGBoost, CatBoost and LightGBM are used as our base learners. The two-level stacking ensemble of a Ridge regression with meta-learner (with α = 1.0) to enhance the stability of the predictions. The target variable is log-transformed ( ) to reduce the skewness. Preprocessing: Median imputation for numeric features, most frequent imputation for categorical, and feature engineering (TotalSF, TotalPorchSF, and HouseAge). To avoid data leakage, out-of-fold predictions from stratified 10-fold cross validation are used as meta-features. Base models are then retrained using the complete training set, and predictions are then inverse-transformed. The entire process is summarized in Algorithm 1.

3. RESULTS AND DISCUSSION

This section presents the detailed experimental results obtained from various ML models on Ames housing datasets. The model's performance was evaluated based on 10-fold cross validation strategy and assessed through various performance parameters including coefficient of determination (R^2), root mean squared error (RMSE), root mean squared logarithmic error (RMSLE), mean absolute error (MAE), mean absolute percentage error (MAPE). We also performed various statistical test including paired t-test and Wilcoxon signed-rank test were conducted on cross-validated predictions to validate the top model's performance.

3.1. Individual Model Performance with Base Dataset

To determine the contribution of individual models for real state price predication, all nine models were trained and tested with same cross validation

Algorithm 1. Pseudocode for Stacking Ensemble with Ridge Meta-Learner.

```
Input: X, y = log1p(SalePrice), K = 10
Output: y_pred
Xp ← preprocess(X)
base_models ← [XGBoost, CatBoost, LightGBM]
Z_train ← hstack([cross_val_predict(m, Xp, y, K) for m in base_models])
ridge_model ← Ridge.fit(Z_train, y, alpha=1.0)
fitted_models ← [fit(m, Xp, y) for m in base_models]
Xp_new ← preprocess(X_new)
Z_test ← hstack([m.predict(Xp_new) for m in fitted_models])
y_pred ← expm1(ridge_model.predict(Z_test))
return y_pred
```

technique and the results are presented in Table 1. In this experiment we only used preprocessed dataset without including engineering features as total square footage, total porch area and the age of the house.

We compared the performance of basic ML and ensemble models as illustrated in Table 1 for house price predication and evaluated based on R^2, adjusted R^2, RMSE, NRMSE, and MAPE. A significant variation was observed between traditional ML models and ensemble models. When comparing the R^2 values, XGBoost showed the best performance with a value of 0.9298, followed closely by CatBoost with 0.9293, LightGBM of 0.9225 and Gradient Boosting at 0.9188. On the other hand, random forest outperformed with R^2 of 0.8897 from traditional ML models and the linear regression achieved minimum R^2 of 0.3331. The same behavior was observed for all models from other evaluating parameters including NRMSE, adjusted R^2 and MAPE. The results demonstrate that ensemble-based machine learning models consistently outperform traditional machine learning methods for real estate price prediction tasks.

3.2. Individual Models Performance Based on Engineered Features

To investigate the impact of feature engineering on real state price prediction, we introduced three new features including total square footage area, porch area and the age of the house and append these features with our base dataset. We trained our base models again with the updated dataset with same data spilt strategy and used same evaluation parameters. The performance of each model both

Table 1. Models Comparison based R^2, Ajd.R^2, RMSE ($), NRMSE and MAPE (%) with Base Dataset.

Model	R^2	Adj. R^2	RMSE ($)	NRMSE	MAPE (%)
XGBoost	0.9298	0.9211	21164.54	0.1171	7.78
CatBoost	0.9293	0.9206	21244.21	0.1175	7.66
LightGBM	0.9225	0.9130	22233.90	0.1230	8.10
Gradient Boosting	0.9188	0.9088	22755.76	0.1259	8.16
Random Forest	0.8897	0.8761	26524.85	0.1467	9.63
Decision Tree	0.8175	0.7951	34119.27	0.1887	13.12
Ridge	0.7086	0.6728	43113.93	0.2385	8.93
Lasso	0.6296	0.5840	48613.34	0.2689	9.16
Linear Regression	0.3331	0.2511	65225.98	0.3608	9.51

from traditional ML models and ensemble models is improved and the CatBoost achieved the maximum R^2 of 0.9312 as shown in Table 2 which was improved from base dataset as presented in Table 1. Similar pattern was observed from others models and again the ensemble models performed better while comparing with the traditional ML models.

3.3. Hybrid Models Performance Based on Final Feature Vector

We employed hybrid modeling strategies to assess global predictive performance using two distinct ensemble approaches. In the first approach, the top three high-performing models CatBoost, XGBoost, and LightGBM were combined, as they share similar boosting-based architectures and demonstrated superior individual performance. In the second approach, we introduced architectural diversity by integrating a traditional ensemble model, Random Forest, with CatBoost. This combination was designed to balance bias variance, where Random Forest contributes to variance reduction while CatBoost enhances predictive accuracy through gradient boosting.

The results from the experiments presented in Figure 3 indicate that the stacking ensemble method with a Ridge meta-learner outperforms while comparing with other approaches. The best predictive performance comes from a weighted averaging ensemble (R^2 = 0.9334) followed by the best average (R^2 = 0.9330). The Random Forest and CatBoost hybrid combination has a comparative score of 0.9313, which is lower than the other models. The advantages of stacking can be explained by the fact that this system can learn optimal combinations. Uses a meta-learner to adjust base model predictions instead of a fixed or manually assigned weights. This enables stacking to be very effective at capturing complementary strengths and minimizing individual model biases with better generalization ability. In contrast, weighted averaging assumes a static contribution from each model, while the Random Forest–CatBoost combination may suffer from reduced diversity due to overlapping boosting-based learning behavior, limiting its overall variance reduction effectiveness.

The results based on RMSLE as illustrated in Figure 4, indicate that the stacking ensemble with a Ridge meta-learner achieves the lowest prediction error 0.1132, followed closely by the weighted averaging ensemble 0.1138, while the Random Forest with CatBoost hybrid combination shows comparatively higher error 0.1211. This demonstrates that stacking provides better generalization by effectively learning an optimal combination of base model predictions, whereas

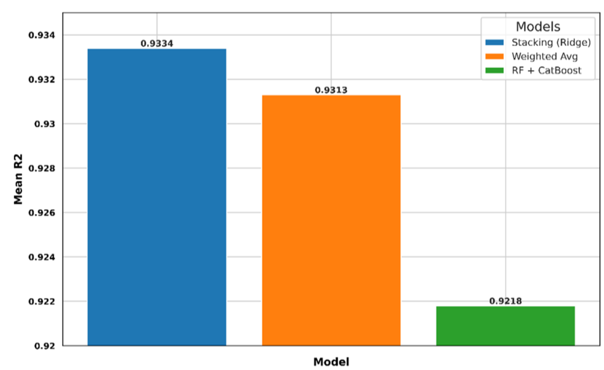

Fig. 3. R^2 performance comparison for hybrid (stacking and weighted averaging) ensemble methods.

Table 2. Models Comparison based R^2, Ajd.R^2, RMSE ($), NRMSE and MAPE (%) with new Features.

Model	R²	Adj. R²	RMSE ($)	NRMSE	MAPE (%)	RMSLE
CatBoost	0.931209	0.922653	20,949.09	0.11587	7.6383	0.11393
XGBoost	0.929067	0.920244	21,272.86	0.11766	7.7443	0.11519
LightGBM	0.922117	0.912430	22,290.61	0.12329	8.0257	0.11951
Gradient Boosting	0.921036	0.911215	22,444.73	0.12414	8.1217	0.11992
Random Forest	0.898268	0.885615	25,475.89	0.14091	9.5320	0.13735
Decision Tree	0.804418	0.780092	35,323.62	0.19538	13.0818	0.18681
Ridge	0.707936	0.671611	43,165.72	0.23875	8.9255	0.13566
Lasso	0.625607	0.579041	48,872.49	0.27032	9.1611	0.13831
Linear Regression	0.334176	0.251364	65,174.83	0.36049	9.5098	0.15828

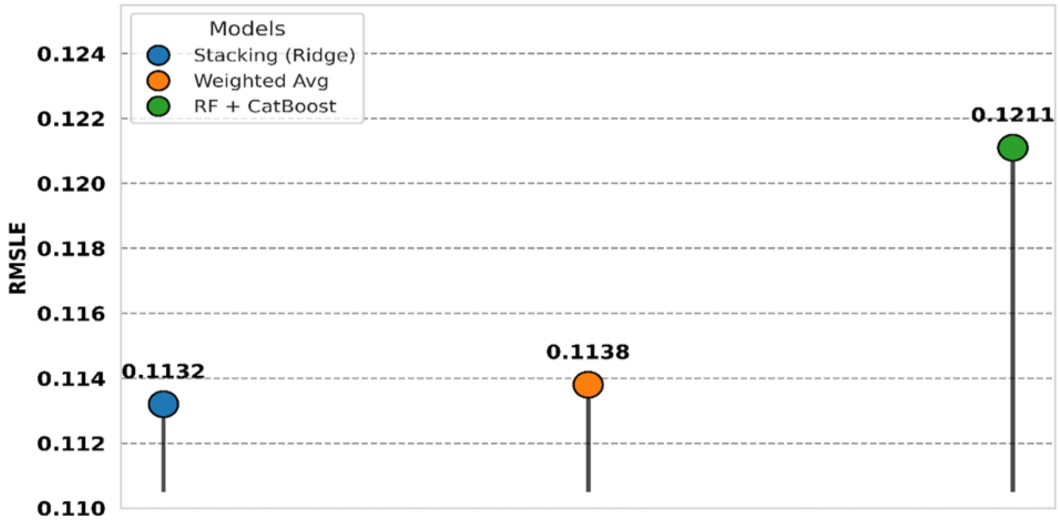

Fig. 4. RMSLE performance comparison for hybrid (stacking and weighted averaging) ensemble methods.

weighted averaging applies fixed contributions and may not capture model complementarities. The relatively higher error in the Random Forest and CatBoost ensemble suggests limited diversity in data.

3.4. Statistical Evaluation of Best Performing Models

To evaluate the robustness and stability of the best performing models (CatBoost, XGBoost, and LightGBM) we performed the statistical evaluation and used two strategies paired t-test and Wilcoxon signed ranked test. To investigate the statistical significance of best performing model, paired test was employed on their predictive results. As presented in Figure 5, the predictive results are comparable for the models XGBoost and CatBoost (0.9470); and XGBoost and LightGBM (0.0970), as these models are not statistically significant with minimum margin of difference in their predication. However, the comparison between CatBoost and LightGBM ($p = 0.0363$) reveals a statistically significant difference, suggesting a measurable performance gap between these two models as presented in Figure 5.

Similarly, the Wilcoxon signed-rank test was applied to the best-performing models to assess the median differences in their predictive results without assuming normality. The statistical comparison as shown in Figure 6, varying levels of difference among the evaluated models. The highest p-value corresponds to the comparison between XGBoost

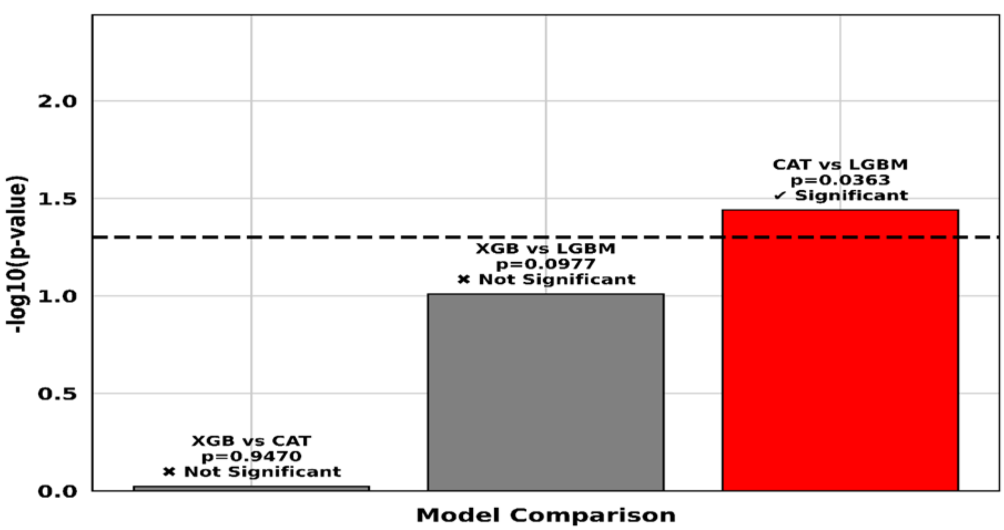

Fig. 5. Paired T-Test analysis of best performing ensemble models.

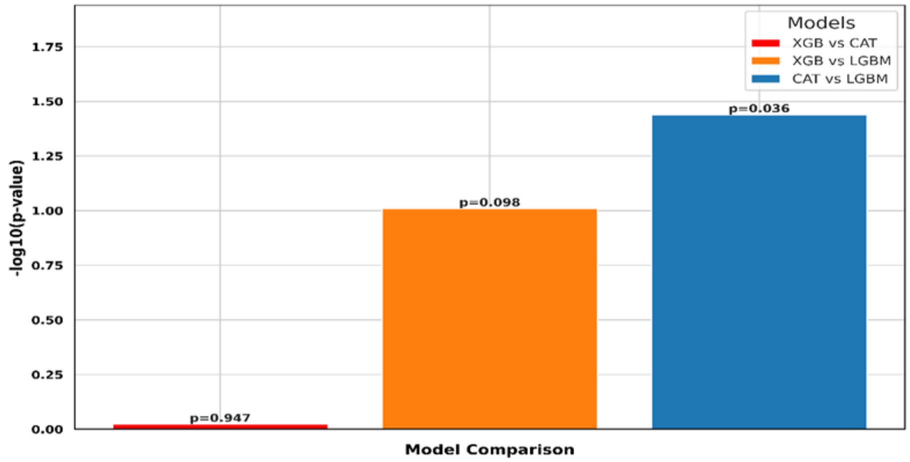

Fig. 6. Wilcoxon singed ranked analysis of best performing ensemble models.

and CatBoost, indicating minimal difference and highly similar performance between the two models.

In contrast, the comparison between CatBoost and LightGBM shows a statistically significant difference, suggesting a consistent performance gap. The intermediate comparison between XGBoost and LightGBM reflects a moderate difference that is not statistically significant. Overall, these findings indicate that while most models exhibit comparable predictive behavior, certain model pairs demonstrate meaningful variations that are supported by statistical evidence.

Based on our extensive experiments, we investigated that tree-based ensemble models perform better as compared to traditional machine learning models. Moreover, it was also observed that model performance is enhanced by adding suitable features in the base dataset. The predictive results improved further through model fusion, and by using stacking techniques, the proposed model outperformed and achieved the best results. Furthermore, the statistical analysis confirms that most ensemble model comparisons are not significantly different, indicating minimal variation in predictive performance, with only specific model pairs showing statistically significant differences.

3.5. Discussion

The results indicate that proposed stacked ensemble model achieved the maximum accuracy on the Ames Housing dataset, with an R^2 score of 0.9334, RMSLE of 0.1132, NRMSE: 0.1140 and MAPE: 7.58% and RMSE (20606.16 $). The performance of our stacking model was improved by introducing new features in the base dataset. This performance significantly outperforms several existing models reported in the literature. The XGBoost model proposed by Kumar *et al*. [11] and model achieved the R^2 of approximately 0.923 RMSE of (24630.91 $) and our model outperformed while comparing it as shown in Table 3.

Similarly, the Random Forest model was employed by Zhao [12] on Ames dataset and the model achieved the performance with an R^2 of 0.85and RMSE of 24630.91$. The performance of the model is comparably low as our proposed model. Another research study proposed by

Table 3. Proposed model comparison with other models.

Papers	DATASETS	Models	R^2	RMSLE	RMSE ($)
Proposed Method	Ames Housing	CXL Stacked	0.9334	0.1132	20606.16
Kumar *et al*. [11]	Ames Housing	XGBoost	0.923	----	24630.91
Zhao [12]	Ames Housing	Random Forest	0.85	-----	3536
Han [10]	Ames Housing	Stacked Model	-----	0.1189	-----
Ye [19]	Ames Housing	XGBoost	0.88	0.1249	30718.06

Han [10] utilized a stacked ensemble model and achieved an RMSLE of 0.1249, demonstrating a competitive performance compared to our model as illustrated in Table 3. The research based on XGBoost proposed by Ye [19] on the Ames dataset achieved an R^2 of 0.88, an RMSLE of 0.1249, and an RMSE of 30718.06 \$. Based on this comparison, our model outperformed the existing approach, as we introduced enhanced feature engineering along with a stacking strategy for effective model fusion. The proposed CXL ensemble method, which combines the predication of CatBoost, XGBoost, and LightGBM and used ridge regression as meta-learner and model achieved the best an RMSLE of 0.1132 and an R^2 score of 0.9334. This improvement highlights the effectiveness of advanced feature engineering combined with a stacked ensemble approach in enhancing model performance. By transforming and enriching the original features, and leveraging multiple models through stacking, the proposed method captures complex patterns more effectively while reducing over fitting.

4. CONCLUSIONS

The proposed study demonstrates the integration of new features with CXL ensemble model significantly improve the performance of house price prediction The CXL ensemble model combing the (CatBoost, XGBoost, and LightGBM) with ridge regression as meta-learner improved the performance and the model achieved the best RMSLE of 0.1132 and an R^2 score of 0.9334 as compared to existing approaches. The robustness and reliability of our model were ensured through Wilcoxon singed ranked and T-paired test and it the ensemble based model shown minimum variation for the predication performance, demonstrating its stability and consistency in house price prediction. For future work, the framework can be extended to larger and more diverse datasets while integrating explainable AI (XAI) techniques to enhance interpretability and provide actionable insights.

5. CONFLICT OF INTEREST

Authors declare no conflict of interest.

6. FUNDING

No funding was received to conduct this research.

7. AUTHORSHIP CONTRIBUTION STATEMENT

Muhammad Khateeb Khan was responsible for the conception of the methodology, model development, experimental analysis, and presentation of results; Asad Ullah Khan originated the initial study idea and performed data collection; Salman Ahmed and Wajid Ali jointly conducted the literature review and led the manuscript writing, review, and proofreading. All results, analyses, and interpretations were conducted by the authors; Grammarly and Google Translate were used for English language refinement.

8. ETHICAL STATEMENT

In the present study no human or animal subjects are involved; therefore, ethical approval is not required.

9. REFERENCES

1. S. Buyrukoglu and O.G. Uzut. Prediction of real estate prices with data mining algorithms. *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences* 7(9): 77-84 (2020). https://www.researchgate.net/publication/342782159_Prediction_of_real_estate_prices_with_data_mining_algorithms
2. H. Ding. Predicting Boston housing prices using machine learning models. *Proceedings of the 2024 2nd International Conference on Management Innovation and Economy Development (MIED 2024), Advances in Economics, Business and Management Research*. *Atlantis Press* (2024). https://doi.org/10.2991/978-94-6463-542-3_51
3. D. Harrison and D.L. Rubinfeld. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* 5(1): 81-102 (1978). https://doi.org/10.1016/0095-0696(78)90006-2
4. W. Xibin, J. Wen, Y. Zhang, and Y. Wang. Real estate price forecasting based on SVM optimized by PSO. *Optik* 125(3): 1439-1443(2014). https://doi.org/10.1016/j.ijleo.2013.09.017
5. S.M. Lundberg, G.G. Erion, and S.I. Lee. Consistent individualized feature attribution for tree ensembles. *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics* 2018: 13-27 (2018). https://arxiv.org/abs/1802.03888
6. D.E. Aniobi, C.O. Ochuba, and S.B. Nguideen. House price prediction: *Comparative analysis of regression-based machine learning algorithms.*

International Journal for Research in Applied Science & Engineering Technology 11(X): 1550 (2023). https://doi.org/10.22214/ijraset.2023.56232

7. S.C.K. Tekouabou, Ş.C. Gherghina, E.D. Kameni, Y. Filali, and K.I. Gartoumi. AI-Based on Machine Learning Methods for Urban Real Estate Prediction: A Systematic Survey. *Archives of Computational Methods in Engineering* 31(2): 1079-1095 (2024). 10.1007/s11831-023-10010-5
8. Y. Chen. Research on the prediction of Boston house price using various models. *Highlights in Business, Economics and Management* 21: 27-37 (2023). https://doi.org/10.54097/hbem.v21i.13600
9. S. Muralidharan, K. Phiri, S.K. Sinha, and B. Kim. Analysis and prediction of real estate prices: A case of the Boston housing market. *Issues in Information Systems* 19(2): 109-118(2018). https://doi.org/10.48009/2_iis_2018_109-118
10. Y. Han. Price Prediction of Ames Housing Through Advanced Regression Techniques. *BCP Business & Management* 38: 1965-1974 (2023). https://doi.org/10.54691/bcpbm.v38i.4014
11. M.S. Kumar, M. Saraswathi, B.Y. Singh, D. Divya, and J.S. Shanker. An Efficient and Interpretable XGBoost Framework for Housing Price Prediction: Enhancements in Accuracy and Computational Performance. *International Journal of Computer Engineering Research and Trends* 11(1s): 23-30 (2024). https://www.ijcert.org/index.php/ijcert/article/view/1041/900
12. T. Zhao. Predicting House Prices Using Machine Learning Models. *Transactions on Computer Science and Intelligent Systems Research* 9: 629-633 (2025). https://doi.org/10.62051/e9x5cc73
13. G. James, D. Witten, T. Hastie, and R. Tibshirani (Eds.). An introduction to statistical learning: With applications in R. *Springer* (2013). https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611259312432/ISLR+Seventh+Printing.pdf
14. M.A. Shbool, R. Al-Dmour, B.A. Al-Shboul, N.T. Albashabsheh, and N. Almasarwah. Real estate decision-making: precision in price prediction through advanced machine learning algorithms. *International Journal of Housing Markets and Analysis* 19(3): 771-795 (2026). https://doi.org/10.1108/IJHMA-01-2025-0004
15. L. Breiman. Random forests. *Machine Learning* 45(1): 5-32 (2001). https://doi.org/10.1023/A:1010933404324
16. Q. Hong and H. Li. Prediction of house price in Boston based on the fusion of crested porcupine optimizer algorithm and BP neural network. In: *Proceedings of the 2025 International Conference on Digital Economy and Information Systems (DEIS '25), Association for Computing Machinery, New York, NY, USA* pp. 247-252 (2025). https://doi.org/10.1145/3745133.3745175
17. S. Himanshu. Benchmarking Predictive Performance of Machine Learning Approaches for Accurate Prediction of Boston House Prices: An In-Depth Analysis. *International Journal of Advanced Research and Interdisciplinary Scientific Endeavours* 1(2): 78-91 (2024). DOI: 10.5281/zenodo.13841369
18. R. Wijaya. Multi-level dense layer neural network model for housing price prediction. *arXiv preprint arXiv:*2310.08133v1 (2023). https://arxiv.org/pdf/2310.08133
19. Q. Ye. House price prediction using machine learning for Ames, Iowa. *Applied and Computational Engineering* 55: 44-54 (2024). https://doi.org/10.54254/2755-2721/55/20241483