A model for Real Estate Price Prediction using Multi-Level Stacking Ensemble Technique

Lesi Nnadozie
Department of Computer Science, Rivers State University, Nigeria

Dr. Daniel Matthias
Department of Computer Science, Rivers State University, Nigeria.

Dr. E.O Bennett
Department of Computer Science, Rivers State University, Nigeria.

Citation: Nnadozie L, Matthias D., and Bennett E.O. (2022) A model for Real Estate Price Prediction using Multi-Level Stacking Ensemble Technique, *European Journal of Computer Science and Information Technology*, Vol.10, No.3, pp.33-45

ABSTRACT: *Recent research and economic publications have shown the impact of real estate investment on the over economy of Nigeria. It is therefore crucial to employ machine learning technique to predict the price for real estate properties. Real estate price analysis and prediction will assist in establishment of real estate policies and can also be used to aid real estate property stakeholders to come up with informative decisions without bias or prejudice. Thus, it is imperative to develop a model to improve the accuracy of real estate price prediction. The goal of this research is to develop a model using a multi-level stacking ensemble technique to predict price of real estate property. The dataset utilized for the study was collected from transactions done by real estate firms in Port Harcourt and it consist of a total of 1053 rows with twelve features. The base model used includes Random Forest(RF), Extreme Gradient Boosting Algorithm(XGBoost), Light Gradient Boosting Machine(LightGBM), Decision Tree regression and ElasticNet Regression. Various combinations of the base models were stacked using StackingCVRegressor. The final model was developed by combining the best performing stacked models and evaluated using R-Square, Mean Absolute Error(MAE), Root Mean Square Error(RMSE), Mean Square Error(MSE) and Training time. The proposed model outperformed the various individual base model with R-square of 0.985203, MSE of 0.013438, RMSE of 0.115923, MAE of 0.063411 and training time of 0.599398. The result show that multi-level stacking significant improve the accuracy of a model. Again, it was observed stacking improve the performance accuracy of a model at the cost of computational time. Stacking by using blending function for the proposed model significantly reduced the computational time for training the model to 0.599398 second when compared to using StackingCVRegressor with training time of 107.054931 seconds. Therefore, multi-level stacking ensemble technique can be employed to improve the predictive accuracy of a prediction model. Future work can be done by increasing the dataset and also increasing the number of features.*

KEYWORDS: Real Estate, Machine Learning, Multi-level Stacking Ensemble Technique, Random Forest, Extreme Gradient Boosting Algorithm(XGBoost), Light Gradient Boosting Machine(LightGBM), Decision Tree regression and ElasticNet Regression, Evaluation Matrix.

INTRODUCTION

Real estate are assets that comprises of land, structures such as building on it, including its natural resources namely crops, minerals resources etc. although much interest is vested on real property, mostly buildings or housing in general. Real estate business has significant impact on the economic situation in a nation. The market trend in the field of real estate business poses great concern in the stakeholders involved such as the buyer and seller. It is therefore of outmost importance to device a means of forecast or make prediction in this sector to aid buyer and sellers make timely and right judgement without any bias or prejudice. The importance of real estate property was greatly overlooked in the past since it was only viewed as a means parent handover inheritance to their children. With changing times, understanding has evolved since real estate properties have a worthy advantage against inflation and it is also one of the most important means of capital increase thereby increasing its popularity in the financial and investment market. Igbinosa stated that "Nigeria stands on the threshold of establishing a secondary mortgage market to mobilize capital market finance for the primary market, the need to assist real estate professionals with information on the influence of property features/characteristics on residential property values cannot be over stressed" (Igbinosa, 2011).

Of recent times, machine learning methods are being used by different individual or organization to investigate the real estate sector of various town and cities around the globe. In order to improve or increase the accuracy of machine learning prediction, other techniques have been devised, one of which is ensemble learning techniques. Ensemble learning method can be defined as a machine learning method which joins together more than one individual machine learning models (weak learners) in order to produce a hybrid model which is likely to yield more accurate prediction with reduced variance and less deviation as the case maybe. Ensemble learning is classified into three major categories or method such as bagging, boosting and stacking. Developing a model that can combine the result of more than one machine algorithms in order to produce a better or improved prediction will provide the greater technical potential to intensely help real estate stakeholders to make good and sound business decision. This research seeks to employ one ensemble learning method; stacking to forecast the cost of real-estate property in Nigeria. It will involve the implementation of an ensemble learning method to provide improved accuracy and precise outcome.

Stacking method also known as Stacking generalization involves using an algorithm called a meta-learner to learn how to efficiently combine the prediction of two or more machine learning algorithms called based learners to produce better prediction. Stacking Generalization was introduced by David H. Wolpert in 1992 whose aim was to reduce the generalization error of different machine learning models (Wolpert, 1992). The general purpose of the Stacking Generalization method is to generate a Meta-Model. And it has been confirmed to be better in terms of intrusion detection (Syarif *et al.*, 2012) The stacking method has the capacity of combining or harnessing the capability and advantages of multiple base learners and make prediction that has improved performance than any single model in the ensemble.

Multi-level stacking ensemble involves using more than two levels; the first level consists of the base learners and the second consist of the meta-learner. Multi-level stacking goes further by creating more than two meta-learners at the second level before feeding their predictions to the third level meta-learner for the final prediction. The goal of this research is to build a multi-level stacking model with the aim of determining if it can actually improve the level of predictions made by the second level meta-learner.

RELATED WORKS

Several works have been carried out in the area of real estate price prediction. Most of the algorithms adopted to be used to build base model for the proposed model have been proven to good prediction power in real estate forecast in terms of classification and regression. XGBoost algorithm has been utilize by different researcher for real estate price prediction and has it given high prediction accuracy. In 2021, a research was carried out on house price prediction using machine learning algorithms such as XGBoost, Random Forest, Decision Tree Regression and linear regression with dataset from Boston housing. Their result revealed that XGBoost outperform others with a mean cross validation score of 0.88, followed by Random Forest with score of 0.85. (Dabreo et al., 2021). Although XGBoost performed better than Random forest in this case, the difference in their score is not high. Random Forest has also been used and have shown that it has a high predictive power. (Ja'afar et al., 2021)

Li. (2018) used LightGBM to forecast monthly house rent using 196539 training dataset with 108 features. His choice of algorithm was based on its ability to train large volume of data very fast when compared to its counterpart XGBoost and that LightGBM is a distributed and high performance algorithm that uses decision tree framework. The model built was evaluated using R-Square with the result of 96.13%. Masisas *et al.* (2016) used machine learning algorithms random forest, neural network and SVM to analyze and made prediction using input-data from the new housing market of Santiago, Chile. The objective of their study was to compare their performance in terms of predictive accuracy which was done using ordinary least square regression. The dataset used consist 16,472 samples of price records of the new housing unit in the focused area. The outcome of the analysis indicates that Random Forest has better performance than the other models. They were able to come to a conclusion that machine learning models can be utilize as a valuable tool for obtaining relevant information on housing markets**.**

In Nigeria, several work have been done on real estate price prediction. Abidoye & Chan. (2018) adopted Hedonic Pricing Model (HPM) approach for real estate valuation in Nigeria. Their analysis demonstrated that using HPM method can give rise to erroneous prediction in property valuation. They came to this conclusion by using evaluation criteria such as mean absolute percentage error, root means square error and mean absolute error (MAE). HPM has a high MAPE, MAE, and RMSE value which stockholders in the real estate sector may not accept in a real practice. Therefore, Hedonic Price Model is completely dependable to yield accurate price prediction in real estate property market. Consequently, further research is needed the able to produce a model that produce high level accuracy that can be accepted by real estate investors. Again, Oyedeji *et al.* (2018) carried out a research on Property Rental Value Classification adopting artificial neural network, logistic regression and support vector machine as models using Osogbo, Osun State, Nigeria for their case study. In their study, over 80% of the properties forecasted were been classified by the three models adopted, thus the model used is dependable for making forecast in residential property in terms of rental. Again, they were able to identify negative impact of feature such as distance from cultural site as the topmost on the cost of renting in the study area. This work was able to only classify the properties and did not go into the actual price prediction of each property. Igbinosa (2011). carried out a study to investigate the property features or attributes that determines the value of real estate properties in Nigeria using Artificial neural network with Lagos and Benin for his case study. The study revealed that while residential properties in real estate market have numerous features, but the ones that have high level impact of the market price or value of the properties are not many. Nine (9) residential property features such as neighborhood attractiveness, land size, years sold, number of bath rooms, property type, neighborhood category, number bedrooms en-suite, property category were identified to have comparatively robust control on market price and how such characteristics impact the level of selling and buying decision of property stakeholders these attributes can therefore be used by stakeholders in real estate business to make guided and unbiased decision. Again, with the understanding of the impact of these features, real estate stakeholders, amongst others, should be able to improve on these attributes in their properties in order to increase sales level thereby increasing their profitability. This research seeks to these features identified can there be used to further predict the price of real estate properties.

The quest to improve the level of accuracy of prediction in machine learning gave rise to using the ensemble techniques. Several algorithm has been developed based on ensemble technique, example are random forest, XGboost, LightGBM etc., all of which have been discovered to improve accuracy of prediction. Stacking which is also one of type of ensemble has been used in the property price prediction. Implementation of stacking ensemble technique in House Price Prediction research was carried out by Maharshi Modi, Ayush Sharma & Madhavan P., in their work -- Applied Research On House Price Prediction Using Diverse Machine Learning Techniques. Machine learning algorithms such as Extra Tree, Support Vector Machine, K Nearest Neighbor, Naive Bayes, Logistic Regression, Stochastic Gradient Descent were combined using stacking ensemble method for create a new model which outperformed single conventional models used for the ensemble (Modi *et al.*, 2020). Jiajia Wang, Cheng Cheng, and Wei Xu in 2012, implemented an ensemble technique in business application. It was a comparative study of ensemble methods such as stacking, boosting and random spacing in Financial Market Prediction. They concluded that bagging performed best because the iteration of ensemble technique is modified and therefore the higher the iteration can give rise to an incompetent and underperforming model that has overfitting on data (Wei *et al.*, 2012). Shivdutt Vishwakarma & Swasti Singhal in 2020, carried out a research on House Price Forcasting based on Hybrid Multi-Regression Model. Four models model (ElasticNet, GradientBoost, KernelRidge and Lasso) were mix and averaged to produce a hybrid. Using statistical error analysis tool, it is concluded that the hybrid model performed better than other conventional models (Vishwakarma & Singhal, 2020).

Most of the stacking ensemble techniques implemented are based of two level stacking. Wu *et al*., (2021) in their work -- Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. They suggested that further study can be carried out on stacking by implementing a multi-level (more than two) to enhance the accuracy of the ensemble. Based on their suggestion, this research seeks to implement a multi-level stacking for prediction of real estate price.

METHODOLOGY

Data Description

This research uses a dataset collected from different real estate firms in Port Harcourt. The dataset consists of 12 explanatory features based on Igbinosa (2011) research that employed Artificial Neural Network to determine property features that determines the cost of real estate properties in Nigeria. It has a total 1053 entries of houses rented in Port Harcourt Nigeria. The table below gives the property attributes names and their descriptions. Price is the dependent variable to be predicted which is numerical datatype. The independent variable description.

Table 1: Description of Dataset

S/NO	VARIABLES NAMES	VARIABLES DESCRIPTION	DATATYPE
1	Neigh_Attractiveness	Neigbourhood attractiveness	String
2	Floor_Area	Floor Area (M^2)	Numeric
3	Year	Year	Numeric
4	No_BDR	Number. of Bedroom	Numeric
5	No_BTR	Number of Bathrooms	Numeric
6	Property_Type	Property Type	String
7	Property_Cat	Property Category	String
8	No_R_Ensuite	Number of Room Ensuite	Numeric
9	No_BQ	No of Boys Quarters	Numeric
10	Location	Property Location	String
11	Light	Quality of Light Availability	String
12	Price	Cost of Renting	Numeric

European Journal of Computer Science and Information Technology
Vol.10, No.3, pp.33-45, 2022
Print ISSN: 2054-0957 (Print),
Online ISSN: 2054-0965 (Online)

ALGORITHMS USED

Random Forests Algorithm

Random forests algorithm is an algorithm built of the principle of ensemble learning technique and it works for both classification and regression problem. It works by building several decision trees when fitted on training data and returns the highest class for a classification task or mean of different trees for a regression task. Random forests improve on the drawback of decision tree algorithm which is over fitting in the training data-set. Leo Breiman in 2001 created and later improve on by Adele Cutler in 2012. (Breiman, 2001; Andy, 2012). The random forest method harnesses the idea of Breiman on bagging and the selection of random features that was pioneered by Ho independently and Amit and Geman to be able to construct a group of decision trees with controlled variance. (Ho, 1995; Amit & Geman, 1997). Viewing computationally, Random Forests algorithm are attractive because it can be used for multiple class classification problem and regression problem, it also takes less computational time for both training and prediction though that depends on the parameter tuning, it also has an in-built mechanism for handling generalization error and can be used directly on problems with high-dimensional features etc. From statistical standpoint, Random Forests are alluring since they provide supplementary attributes like measurement of variable importance, means of visualization and detection of outlier etc.

XGBoost Algorithm

Extreme gradient boosting also known as XGBoost is one of the famous and widely used gradient boosting algorithm that is based on ensemble technique having improved speed and performance in machine learning algorithms using the principle decision tree (sequential decision trees). XGBoost was beget or created by Tianqi Chen (Chen & Guestrin, 2016). XGBoost works by generating decision trees sequentially. Then, independent variables are allocated with weights to be used by the decision tree to make prediction. Weight given to the wrongly predicted variables in the decision tree is improved and then the variables are given to the second decision tree. These distinct classifiers or predictors then ensemble to produce a robust and more accurate model. It can be used for regression, ranking, classification, and user-defined prediction tasks. Recently, XGboost has been recognized for its great ability in providing solution to machine learning problems in diverse areas where it is applied. It is known for its speed as it has the ability of processing process large volumes of input dataset efficiently in a parallel manner. It implemented upon the gradient boosting. It uses greedy method and it is thus greedy in operation. It is high in performance and speed. Extreme Gradient Boost(XGBoost) can easily be upgraded or improved on when used for learning in tree ensemble technique and this the brain behind its high performance.

Decision Tree Regression

Decision tree Algorithm is form based on tree structure. It can be used for both regression or classification task. It divides a dataset into subsets and concurrently developing related decision tree in incremental manner with the expected result of a tree with decision nodes (which have two or more branches and leaf nodes (which is the decision on the numerical target). Decision trees can be implemented for both categorical and numerical datasets. The decision tree algorithm works

@ECRTD-UK: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

by using a data point and goes through the whole tree by asking true or false questions. It starts from the root node where questions are asked and creates different branches based on each answer. It uses recursive partitioning to create the tree and continues until the leaf node is gotten. A decision tree algorithm is a supervised machine learning model, and thus, learns to chart data to the outputs at the training phase of building the model. The decision to split for decision tree regression trees is based Mean squared error (MSE).

ElasticNet Regression

Elasticnet Regression is form by employing the drawbacks found in Lasso and Ridge regression. This regression model fuses the lasso regression and ridge regression modus operandi by using the knowledge gained from their deficiencies to make better the regularization of models. The elasticnet method boosts the Lasso's shortcomings, especially where lasso uses few numbers of dataset with high dimension, Elasticnet regression offers the additional number (N) of variables till satiety. Again, it also improves Lasso by including a quadratic function ($||\beta||^2$) in calculating the penalty and when that is used only, it turns to ridge regression. This quadratic function used when defining the penalty uplifts the loss function into convex curve. The elasticnet regression algorithm utilizes the strength of both ridge and lasso egression. The procedure for getting the estimator in elasticnet methods is done by it finding the coefficients of ridge regressions before taking the next move by making use a lasso regression to shrink the coefficients. The elasticnet method simultaneously does variable selection and regularization. It is appropriate to use elastic net where dimensional data is bigger than the numbers of sample used.

Gradient Boosting Regression (GBM)

Gradient boosting is a machine learning built on boosting ensemble paradigm that is employed for both classification and regression task. It works by using decision tree to form it model for prediction that performs prediction using boosting ensemble technique to improve the prediction of weak models. Gradient Boosting Regression algorithm is mostly employed to train prediction model that involves continuous value. Though boosting can cause increase to the accuracy of a base or weak learner, example a decision tree, linear regression etc., it does that at the expense of interpretability and intelligibility. The major aim GBM framework is geared towards building the new model or weak learners that has high associativity with loss function with negative gradients.

Light Gradient Boosting Machine (LightGBM)

Light Gradient Boosting Machine (LightGBM) is a machine learning algorithm that employs gradient boosting method to make a model more efficient with added advantage of decreasing the memory consume in the process or decreasing consumption of memory. LightGBM employs Exclusive Feature Bundling and Gradient-based One Side Sampling technique which eliminates shortfalls of algorithms that are based on histogram which are mostly utilize for Gradient Boosting Decision Tree methods. Both these methods (GOSS and EFB) makes LightGBM algorithm to be more effective and gives it a superiority over the other Gradient Boost Decision Tree methods. Gradient-based One Side Sampling Technique for LightGBM uses different data instances to compute information gain. Data instances with bigger gradient contributes more to information gain and therefore retains those instances with bigger gradients that is gradients that exceeds the predefined threshold so as to preserve the correctness of the estimate information gain. Exclusive Feature Bundling Technique in LightGBM aims at dropping the number of features or attributes or high-dimensional data by regrouping mutually exclusive features into bundles or group thereby, looking at them as a single attribute. The LightGBM has advantages such as Faster speed for training with higher accuracy, lesser memory usage. It also treats overfitting well when working with a small dataset.

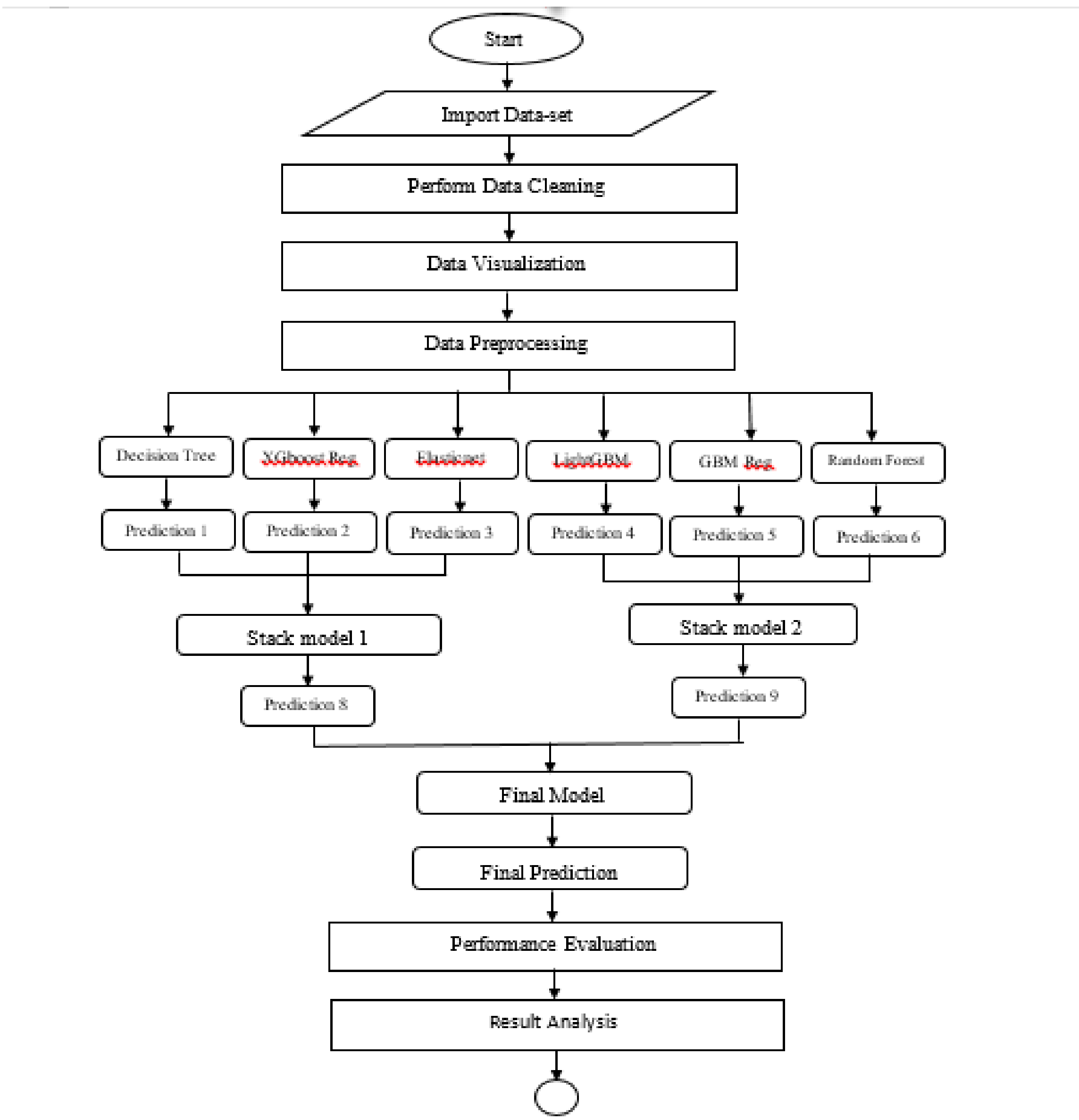

Fig 1: Simple flowchat of the proposed model

Implementation

Stacking generalization can be implemented Python 3 in many ways for example; using StackingCVRegressor in MLxtend Library, blending function. These two were used in the research. StackingCVRegressor StackingCVRegressor uses the concept of out-of-fold predictions. Input data gotten from the out-of-fold are used as the input dataset for the next level meta-regressor. K-fold cross-validation technique is employed to split the dataset then use k-1 fold for training of the first level (base learner) and the remaining data is used for prediction. The outcome of their various prediction is then given to meta-learner as input. The framework used by StackingCVRegressor ensures the model does not over-fit. Blending ensemble is a type of stacking where by the meta-model is trained using predictions on a holdout validation dataset in place of out-of-fold predictions.

RESULT AND DISCUSSION

A total of 1053 dataset with twelve attributes were used. All attributes were checked to ensure there was no null value (see fig 2 & 3). The attributes consist of seven (7) numeric and five (5) categorical variables. One hot encoding was used to convert the categorical variable to ensure our models can interpret them. Log1P was used is transforming the dependent variable Y(Price) in order to normalize and reduce skewedness in the data. The dataset was divided into training set and test set. All base models were trained and tested using same dataset. To obtain the various stacked models, Model1 combined LightGBM, Gradient Boosting Machine, and Random Forest using Random Forest as the meta-learner. Model2 combined Extreme Gradient Boosting, Decision Tree Regression, and Elasticnet Regression with decision tree regression and meta-learner. Our proposed models were implemented in python using StackingCVRegressor function imported from MLxtend Library. The final model combining stacked Model 1 and 2 built with both StackingCVRegressor and blending function.

```
In [9]: #Check for missing data
        df1.isnull().sum().sort_values(ascending=False)

Out[9]: NEIGH_ATTRACTIVENESS    0
        FLOOR_AREA              0
        YEAR                    0
        NO_BDR                  0
        NO_BTR                  0
        PROPERTY_TYPE           0
        PROPERTY_CAT            0
        NO_R_ENSUITE            0
        NO_BQ                   0
        LOCATION                0
        LIGHT                   0
        PRICE                   0
        dtype: int64
```

Fig 2: Missing Data Count

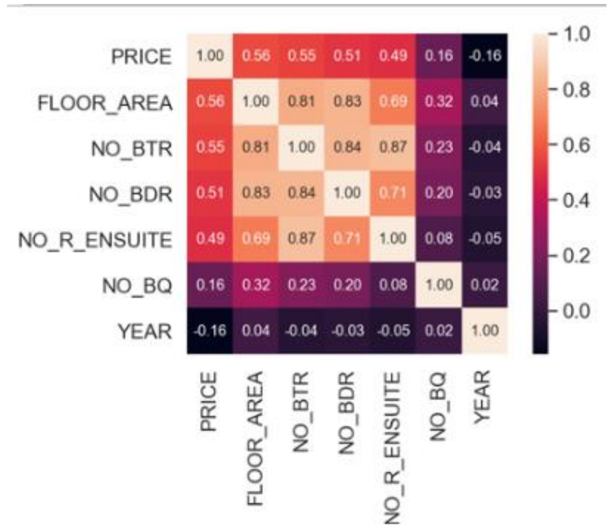

Fig 3: Correlation Matrix Between Different features

```
Test set evaluation for Proposed Model using StackCVRegressor:
_____________________________________
MAE: 0.070408689624476352
MSE: 0.013044767370933609
RMSE: 0.11421369169645822
R2 Square: 0.9856363367957113
Training Time 213.91266918182373
__________________________________

Out[73]: Text(0.5, 1.0, 'Actual Rent vs Predicted Rent of Proposed Model using StackCVRegressor')
```

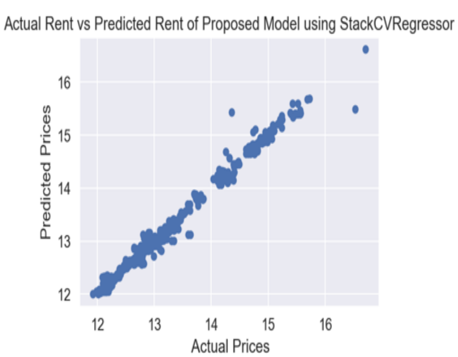

Fig 4: Proposed Model using StackingCVRegressor

```
Test set evaluation for Proposed Model using Blending function:
_____________________________________
MAE: 0.06341144852785793
MSE: 0.013438192486019625
RMSE: 0.11592321806273162
R2 Square: 0.9852031342947762
Training Time 0.5993983745574951
__________________________________

Out[71]: Text(0.5, 1.0, 'Actual Rent vs Predicted Rent of Proposed using Blending function')
```

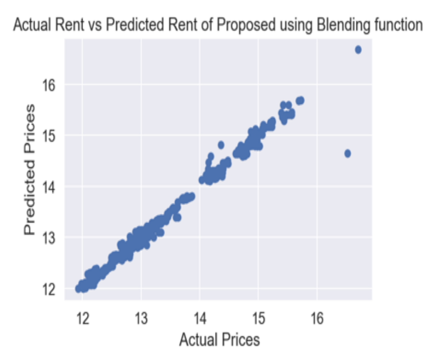

Fig 5 Proposed Model Using Blending

Table 2: Table Summary of Results

MODEL	R-SQUARE	MSE	RMSE	MAE	TRAINING TIME(s)
PROPOSED MODEL (MODEL 1, 2 with Blending function)	0.985203	0.013438	0.115923	0.063411	0.599398
PROPOSED MODEL (MODEL 1, 2 with StackCVRegressor)	0.985636	0.013045	0.114214	0.070409	107.054931
MODEL_1(LGBM, GBR, RF)	0.983561	0.014930	0.122187	0.066697	55.655737
MODEL_2 (DTR, XGBR,ELNR)	0.979328	0.018774	0.137019	0.073886	17.233479
MODEL_3(DTR, XGBR,ELNR,RF,LGBM,GBR)	0.978689	0.019354	0.139118	0.078666	64.861121
XGBOOST REGRESSION(XGBR)	0.978476	0.019547	0.139812	0.085151	0.749679
DECISION TREE REGRESSION(DTR)	0.971153	0.026199	0.161860	0.064675	0.011881
RANDOM FOREST(RF)	0.969657	0.027557	0.166002	0.071801	5.664087
GBOOST REGRESSION(GBR)	0.967432	0.029577	0.171980	0.105374	0.148140
LIGHTGBM REGRESSION(LGBM)	0.892574	0.097562	0.312349	0.197025	1.991120
ELASTICNET REGRESSION(ELNR)	0.861404	0.125870	0.354781	0.244597	0.504250

Performance Analysis

The models were analyzed using evaluation criteria such as R-Square, Root Mean Square Error(RMSE), Mean Square (MSE), Mean Absolute Error(MAE) and Training time (seconds) as shown in Table 2 above. The base models Random Forest, XGboost, Elasticnet, LightGBM, Decision Tree and Gradient Boosting Machine when evaluated using R-Square yield the following output 0.969657, 0.978476, 0.861404, 0.892574, 0.971153 and 0.967432 with training time (seconds) of 5.664087, 0.749679, 0.504250, 1.991120, 0.011881 and 0.148140 respectively. The second level stacked model Model1(LightGBM, Gradient Boosting Machine, Random Forest) and Model2 (Extreme Gradient Boosting, Decision Tree Regression, Elasticnet Regression) have R-Square result of 0.983561and 0.979328 with training time 55.655737 and 17.233479 respectively. R-Square for the final model using StackingCVRegressor is 0.985636 and blending is 0.985203 with training time of 107.054931 and 0.599398 respectively.

From the R-Square obtained from various models, stacking ensemble improves the prediction accuracy of a machine learning model. Although the difference is not high but for prediction any increase is significant with the overall performance of a model. The result for different combinations (see table 2) indicates that the greater the number of base model combined, the higher the training time involved. Again, multi-level stacking ensemble can be used to increase the accuracy in the prediction performance of a model.

An efficient model should not only improve the accuracy but the time-taken for training should be considered especially in situation where resources are limited. Computational time is very vital in computing. In order to reduce the computational time for training for our proposed, blending function was used to combine the second level model.

CONCLUSION

Machine learning models have been used to improve accuracy in predictions based on human error. The need to improve the accuracy of prediction of machine learning models that can provide stakeholders in market place (Real Estate) high technical insight gave rise to different ensemble techniques. In our research, a multi-level stacking model was built using dataset collected from Port Harcourt metropolis in Rivers State Nigeria. From the results obtained from the various base model and the proposed model built using multi-level stacking ensemble, the proposed model outperforms the base models in terms of prediction accuracy. Although the prediction accuracy was improved, the computational time for training the proposed model also increased. Blending function was used to combine the second level model to reduce the training time of the proposed model. It is therefore worthy of note that increasing the number of level or layers in stacking can give rise to models with improve accuracy of performance.

RECOMMENDATION

In order to improve the accuracy of a model, multiple level stacking model can be utilized. The features used in this research are restricted to some features proposed by Igbinosa that determines the price of real estate properties in Nigeria (Igbinosa, 2011). Further work can be done by increasing the number of features and also the number of dataset since there is a likelihood of overfitting with smaller dataset. Again, stacking by averaging or blending can be done at the second level before using StackingCVRegressor or blending on the final level to see its impact on not just the prediction accuracy but also the computational time.

REFERENCES

Abidoye, R. B., & Chan, A. P. (2018). Hedonic Valuation of Real Estate Properties in Nigeria. *Journal of African Real Estate Research, 1*(1), 122-140. doi:10.15641/jarer.v1i1.452.

Amit, Y., & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Computation., 9*(7), 1545-1588. doi:10.1162/neco.1997.9.7.1545.

Andy, L. (2012). *Documentation for R package random Forest.*

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi:https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining.* San Francisco, CA, USA: ACM Sigkdd. doi: http://dx.doi.org/10.1145/2939672.2939785

Dabreo, S., Rodrigues, S., Rodrigues, V., & Shah, P. (2021). Real Estate Price Prediction. *International Journal of Engineering Research & Technology (IJERT), 10*(4), 644-649.

Ho, T. K. (1995). Random Decision Forest. *Proceedings of the 3rd International Conference on Doc-ument Analysis and Recognition,* (pp. 278–282). Montreal: QC.

Igbinosa, S. O. (2011). Determinants of Residential Property Value in Nigeria – A Neural Network Approach. *International Multidisciplinary Journal, Ethiopia, 5*(5), 152-168. doi:http://dx.doi.org/10.4314/afrrev.v5i5.13

Ja'afar, N. S., Mohamad, J., & Ismail, S. (2021). Machine learning for property price prediction and price valuation: a systematic literature review. *Journal of the Malaysian Institute of Planners, 19*(3), 411-422.

Li, J. (2018). Monthly Housing Rent Forecast based on LightGBM (Light Gradient Boosting) Model. *International Journal of Intelligent Information and Management Science, 7*(6), 58-65.

Masías, V. H., Valle, M., Crespo, F., Crespo, R., Schüler, V. A., & Laengle, S. (2016). Property Valuation using Machine Learning Algorithms: A Study in a Metropolitan-Area of Chile. *AMSE Conference.* Santiago.

Modi, M., Sharma, A., & Madhavan, P. (2020). Applied Research On House Price Prediction Using Diverse Machine Learning Techniques. *International Journal of Scientific & Technology Research, 9*(4), 372-376.

Oyedeji, J. O., Oshodi, O. S., & Oloke, O. C. (2018). Property Rental Value Classification Model: A Case of Osogbo, Osun State, Nigeria. *Covenant Journal of Research in the Built Environment (CJRBE),, 6*(1), 52-61.

Syarif, I., Zaluska, E., Prugel-Bennett, A., & Wills, G. (2012). Application of Bagging, Boosting and Stacking to Intrusion Detection. *Perner P. (eds) Machine Learning and Data Mining in Pattern Recognition. MLDM 2012. Lecture Notes in Computer Science,. 7376*, pp. 593-602. Berlin, Heidelberg: Springer. doi:https://doi.org/10.1007/978-3-642-31537-4_46

Vishwakarma, S., & Singhal, S. (2020). House Price Forecasting Based on Hybrid multi. *International Conference on Innovative Computing and Communication.*

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*(2), 241-259.

Wu, W. Z., Jiao, X., Guo, W., & Hamoud, Y. A. (2021). Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. *Computers and Electronics in Agriculture*. doi:https://doi.org/10.1016/j.compag.2021.106039