Transparency in House Rent of Dhaka: Explainable AI Based Predictive Framework

Taeef Najib
Sidetrek
Dhaka, Bangladesh.
taeefnajib@gmail.com

Fahim Muntasir
BRAC University
Dhaka, Bangladesh
muntasirfahim.niloy@gmail.com

Wasif Al Wazed Wasi
Rajshahi University of Engineering & Technology
Rajshahi, Bangladesh
wwazed@live.com

Abstract

House rent is a crucial factor in any country's socio-economic scenario. This can act as an indicator of the financial and developmental situations of the stakeholders of the real estate business. So far, there have been approaches to predict house rent prices in several regions of Bangladesh, but without any clear explanation of the association of different factors to the prediction and how they are affecting the prices. This study touches on this lack of understanding and proposes an Explainable AI based framework. This framework produces accurate predictions on house rents with small margin of error with regression algorithm and can accurately depict the connection of various demographic features of the dataset by visualizing SHAP values. Our study finds that tree-based algorithms such as Decision Tree, Random Forest, XGBoost, Gradient Boost and Light Gradient Boost performed better at regression analysis on the nonlinear dataset of ours. The final model was a voting ensemble of all the tree-based algorithms, which encompasses the strengths of all the base models. We achieved an MAPE of 11% and R2 Score of 86%. The RMSE and MAE of the ensemble were 4398.23 and 2536.77 respectively. Finally, the SHAP Explainable AI determined how the features were correlated to the prediction and overall rent prices. The research introduces a novel framework for predicting house rents in Bangladesh, offering valuable insights through data preparation, model selection, performance assessments, and interpretability analyses, benefiting both scholars and stakeholders.

Keywords
House Rent, Prediction, Shapley Additive Explanations (SHAP), Explainable AI, Dhaka, Bangladesh

1. Introduction
The current urban population of Bangladesh is estimated to be around 38% to 40%, and projections indicate that this figure will increase to at least 48% by the year 2030 (Hasina 2019). The urban real estate market experiences significant pressure, with a special focus on the rental housing sector. The situation in Dhaka, the capital city of Bangladesh, has been worse by unplanned urbanization and inadequate infrastructure. Given the limited availability of adequate housing options, the current task of finding an affordably priced rental house that meets the desired criteria is particularly arduous. Furthermore, the cost of renting an apartment is subject to fluctuation due to a multitude of factors, including but not limited to the quantity of bedrooms and baths, the overall size of the unit, its geographical placement, and other relevant considerations. According to a study, this further complicates the procedure (Begum 2007).

One potential strategy for addressing the issue is enabling renters and tenants to make informed projections regarding apartment rental prices in advance, so alleviating the arduous process of locating an inexpensive dwelling. The achievement of this objective can be facilitated by the utilization of machine learning models, which possess the capability to forecast apartment rental prices by considering their distinct attributes. The utilization of machine learning in this situation yields three distinct advantages (Geethamani and Karthika 2023). This feature would enable tenants to ascertain affordable pricing in accordance with the anticipated geographical location and the quality of the residence. This would aid property owners in establishing competitive rental prices to attract potential renters. Furthermore, the provision of rent projections would offer urban economists with significant and essential data. Indeed, the utilization of machine learning for the prediction of apartment rents is not a novel concept. There exist various approaches for utilizing machine learning algorithms to make predictions on apartment rental prices. Several researchers have already employed several machine learning models in this task, yielding diverse outcomes (Geethamani and Karthika 2023; Khosravi et al. 2022; Kumar 2019; Mora-Garcia et al. 2022; Singh et al. 2021). Several research have been conducted to build models for the prediction of apartment prices and rents in Dhaka city (Jui et al. 2020; Kamruzzaman and Ogura 2007; Neloy et al. 2019; Samsuddin Ahmed and Islam 2014). However, the studies utilized distinct machine learning algorithms while utilizing an identical dataset, and subsequently evaluated the efficacy of each approach in predicting apartment rental prices in Dhaka.

The aim of this study was to utilize various machine learning algorithms to forecast apartment rental prices in Dhaka city, considering the characteristics of the flats. The study aimed to compare the performance of these algorithms and determine the most effective one. The proposed approach involves the utilization of various algorithms for the purpose of training models to predict apartment rents. These algorithms include SVR, MLP, AdaBoost, Ridge Regression, Linear Regression, KNN, Decision Tree, LGBM, GB, Random Forest, XgBoost, and Voting Ensemble. The evaluation of each model's performance is conducted by employing the four different metrics. Lastly, an explainable AI was used for interpreting model outcome and understanding overall impact of dataset.

1.1 Objectives

This study focuses its objectives on:

1. Creating a dataset of recent house rents in Dhaka city.
2. Develop and implement a comprehensive machine learning framework tailored for predicting house rent based on key variables and factors influencing rental prices.
3. Investigate and implement an ensemble learning approach to enhance the predictive performance of the developed model, exploring the synergies of combining multiple algorithms to improve accuracy, stability, and robustness.
4. Conduct an in-depth analysis to understand the behavior and nature of the selected machine learning algorithms during the prediction process.
5. Examining with AI, factors such as feature importance, model interpretability, and sensitivity to different input parameters.

2. Literature Review

The study conducted by the authors aimed to investigate the comparative accuracy of various machine learning models in predicting housing values. A sample including 5359 row dwellings located in Virginia, USA was utilized. The findings of the study indicated that the implementation of the RIPPER algorithm resulted in significant enhancements in price forecasting accuracy (Park and Bae 2015). In their study, Mora-Garcia et al. (2022) employed Ensemble learning techniques, specifically boosting algorithms such as Gradient Boosting Regressor, Extreme Gradient Boosting, and Light Gradient Boosting Machine, as well as bagging algorithms including random forest and extra-trees regressor. These models were subsequently compared to a linear regression model to forecast apartment prices in Spanish cities amidst the COVID-19 pandemic. The predicted accuracy of supervised learning systems for housing rents was assessed by some authors in Shenzhen, China. The researchers employed a set of machine learning techniques, including Random Forest (RF), Extra Tree Regression (ETR), Gradient Boosting Regression (GBR), Support Vector Regression (SVR), Multilayer Perceptron Neural Networks (MLP-NN), and Kohonen Neural Network (k-NN). The results showed superior prediction performance in the Random Forest (RF) and Extra Trees Regression (ETR) methods (Hu et al. 2019). In another study, the author employed a variety of machine algorithms based on multiple regression to make predictions regarding the prices of homes in the Netherlands. He obtained the most precise outcomes (90%) by employing LGBM, XGBM, CatBoost, and RF algorithms, with CatBoost being particularly effective (Voutas Chatzidis 2019). In another study, the Kaggle's "Ames housing" dataset was utilized to employ various regression approaches, including Ridge, Lasso, and ElasticNet regression, to forecast house rental prices. These techniques were applied by considering criteria such as the physical condition of the property, its location, and the size of the land (Kumar 2019).

In another work, the researchers developed an Artificial Neural Network (ANN) model to predict the median price of properties in various neighborhoods of Dhaka City, including Dhanmondi, Baridhara, Gulshan, Mirpur, Uttara, and Old Dhaka. The model was trained using forty demographic attributes specific to each neighborhood (Samsuddin Ahmed and Islam 2014). Neloy et al. (2019) employed ensemble learning techniques to forecast the rental apartment prices in Dhaka, Bangladesh. The researchers employed many algorithms in their analysis, including Advanced Linear Regression, Random Forest (RF), Support Vector Machine (SVM), Decision Tree Regressor, Ensemble Gradient Boosting Regressor, Ensemble AdaBoost Regressor, and Ensemble XGBoost. Furthermore, to integrate advanced regression approaches, the researchers employed Ridge Regression, Elastic Net Regression, and Lasso Regression. The results indicated that the RF algorithms had a reduced mean square error. In a separate study conducted in Dhaka, Jui et al. (2020) employed both linear regression and regression tree/random forest regression techniques to construct a predictive model for flat prices. The comparative analysis of the two techniques demonstrated that the random forest regression model exhibited superior predictive performance.

In their study, Yoshida et al. (2019) conducted a comparative analysis of regression and machine learning-based models to forecast apartment rent prices. The researchers utilized a substantial dataset and considered the spatial dependence of observations. The findings indicated that with an increase in the sample size, both XGBoost and RF models had superior performance compared to the nearest neighbor Gaussian processes (NNGP) model in terms of accuracy in out-of-sample predictions. XGBoost demonstrated superior predictive accuracy across all sample sizes and error metrics. Rajkumar et al. (2023) conducted a study wherein they utilized data obtained through web scraping from an Indian rental website. The primary objective of their research was to construct and assess predictive models for rental apartment pricing using a range of regression and boosting methods (Rajkumar et al. 2023; Yoshida et al. 2022).

The findings of the study indicate that the utilization of a stacking strategy in the implementation of an ensemble machine learning algorithm resulted in superior performance compared to other algorithms, as evidenced by higher accuracy rates and a decrease in errors. Moreover, no work in our demographic focused on model interpretability, so this study was conducted keeping these in mind.

3. Methodology

The total workflow of the paper has been summarized in the following diagram. Firstly, the data collection was performed, which is elaborated in section 4. The colored boxes represent the significant steps of our workflow.

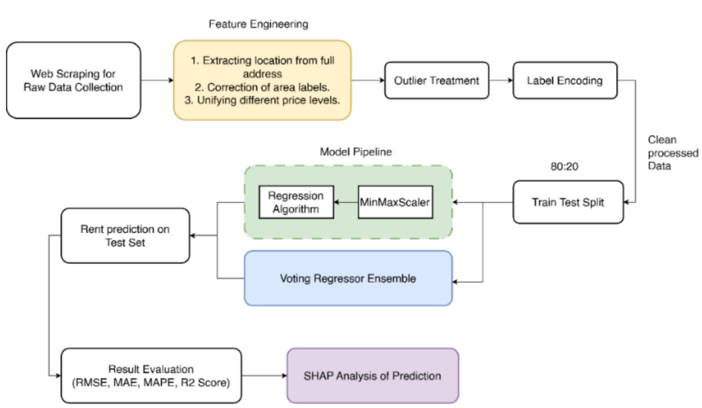

Figure 1. The Workflow Diagram

3.1 Feature Scaling

Prior to being inputted into the models, the data underwent scaling using the MinMaxScaler function from the scikit-learn preprocessing library (Pedregosa et al. 2011). The process of scaling was of utmost importance in this context, given that the various attributes exhibited varying degrees of data. The process of scaling was employed to transform the data into a standardized range of values, specifically within the interval of 0 to 1. This would facilitate quicker convergence for all algorithms, particularly the ensemble method, while also mitigating the risk of overfitting. Indeed, in the absence of appropriate scaling, the random forest algorithm exhibited a complete inability to provide any meaningful predictions.

3.2 Algorithms

Based on the findings obtained from the literature review, the selection of XgBoost, Random Forest, Linear Regression, Ridge, Decision Tree, Gradient Boost, AdaBoost, MLP, LGBM, SVR, and k-NN was done. The decision was made to employ XGBoost and Random Forest as tree-based algorithms due to their superior performance compared to statistical methods. Later voting regressor, an ensemble approach was tried and tested.

The open-source package **XGBoost** was employed in this study. This choice facilitated the explicit programming process and placed greater emphasis on the elucidation and analysis of results (Chen and Guestrin 2016). The gradient boosting method (GBM) typically constructs an ensemble of regression trees (Friedman 2001). XGBoost extends the concept of gradient boosting by incorporating explicit regularization techniques on individual trees. The algorithm in question exhibits minimal bias and volatility and possesses the ability to accommodate outliers. The scalability of XGBoost and its suitability for our dataset are enhanced by this feature.

On the contrary, the **random forest** algorithm involves the selection of a random subset of decision trees and applies the bagging technique, coupled with the inclusion of feature randomness (Breiman 2001). As the quantity of trees within a forest increases significantly, the generalization error tends to approach a finite value. The occurrence of this inaccuracy is contingent upon the individual strength of each tree within the forest, as well as the interconnections among them. This guarantees that the composition of the forest consists of trees that are not associated with each other. The primary distinction between conventional decision trees and Random Forests is in their respective characteristics.

The term "**ridge**" refers to a long, narrow elevated landform that is characterized by linear regression is a statistical approach commonly employed to address the issue of multicollinearity in each dataset. Multicollinearity is a phenomenon that arises when there is a high correlation between two or more independent variables in a regression model. This correlation can result in unstable and incorrect estimations of the regression coefficients. The inclusion of a regularization element in the least squares objective function serves the goal of imposing a penalty on coefficients with large values (McDonald 2009).

Gradient Boosting (GB) and **AdaBoost** are two popular ensemble learning methodologies extensively employed in the field of machine learning. Gradient Boosting is an influential and adaptable boosting method that constructs a robust predictive model by the sequential aggregation of weak learners, often in the form of decision trees. Each subsequent tree in the sequence aims to rectify the errors made by its predecessors. The training procedure involves minimizing a pre-established loss function to optimize the model for enhanced predictive accuracy (Natekin and Knoll 2013). AdaBoost, also known as Adaptive Boosting, is a boosting algorithm that aims to enhance the efficacy of weak learners by giving weight to training instances (Schapire 2013). The proposed approach involves the iterative training of weak models, wherein the weights of misclassified examples are adjusted to prioritize their correct classification in later rounds.

The methods, Decision Tree, Multi-Layer Perceptron (MLP), Support Vector Regression (SVR), and k-Nearest Neighbors (K-NN) exhibit a range of properties that differentiate them from one another. The Decision Tree model is a flexible and easily interpretable algorithm that divides data into subsets using feature splits to generate predictions. The MLP, which stands for Multilayer Perceptron, is a specific sort of artificial neural network. It is composed of interconnected nodes that are arranged in layers. This architectural arrangement allows the MLP to effectively capture intricate relationships within data by means of nonlinear transformations. The Support Vector Regression (SVR) technique utilizes support vectors to transform data into high-dimensional spaces, facilitating the accurate representation of intricate associations, especially in regression scenarios. In contrast, the K-Nearest Neighbors (K-NN) algorithm operates by leveraging the resemblance between examples inside the feature space, hence making predictions based on the predominant class among its k-nearest neighboring instances. Every algorithm possesses unique characteristics and is well-suited for our proposed model, rendering them invaluable tools to our house rent prediction applications.

Lastly, The Voting Regressor is an ensemble machine learning technique specifically developed for regression tasks, which involve the prediction of continuous outcomes. The proposed approach involves aggregating the predictions generated by multiple base regressor models and computing their average, therefore harnessing the collective knowledge of diverse algorithms. In the realm of forecasting apartment rental prices, the Voting Regressor emerges as a formidable instrument for augmenting the precision and dependability of the predictive framework. The utilization of an ensemble strategy demonstrates its advantageous nature in effectively managing the intricate interplay of several aspects that impact apartment rental prices. The feature enhances the precision and comprehensiveness of the decision-support system, catering to the needs of landlords and tenants in the ever-changing and heterogeneous real estate industry.

3.3 Evaluation Metrics

RMSE, R2 Score, MAE, and MAPE were used to analyze performance. These measurements can help explain algorithm superiority. Residuals are the differences between observed and expected values. MAE can also be expressed as a percentage, called MAPE. We always aim for the lowest RMSE, MAE, and MAPE values in our analysis. On the other hand, a higher R2 value means the model can better account for dataset variability.

3.4 SHAP Analysis

"SHapley Additive exPlanations (SHAP)" is an abbreviation used to refer to a machine learning (ML) technique that aims to provide explanations for individual predictions, hence enhancing the interpretability of conventional black box ML models. The fundamental concept that underlies the development of machine learning models based on Shapley values is to distribute the credit for the model's output among its input characteristics by employing equitable allocation results derived from cooperative game theory (Lundberg and Lee 2017). The method facilitates the comprehension of the influential aspects in a prediction and the manner in which they exert their impact (Mokhtari et al. 2019). The Shapley value of a feature value refers to its contribution to the overall payout. In this study, we selected SHAP analysis in conjunction with XGBoost as it offered greater flexibility and interpretability (Lundberg et al. 2020). In numerous instances, tree-based models have demonstrated superior performance compared to neural networks (McDonald 2009).

4. Data Collection

The dataset used in this paper was web scrapped from www.bproperty.com using Selenium library. It has 5 columns and 28800 rows. We have scrapped the location, area in square feet, number of bedrooms, number of bathrooms and monthly rent of apartment listings in Dhaka posted from 8 February 2022 to 8 April 2022 on the mentioned website.

Table 1. Column Description

Column	Data Type	Description
Location	‘object’	Location where the apartment is
Area	‘object’	Total area in square feet
Bed	‘int64’	Total number of bedrooms
Bath	‘int64’	Total number of bathrooms
Rent	‘object’	The monthly rent of the apartment

4.1 Feature Engineering

To preprocess the dataset, several steps were taken. There were 13 data points where the location column only had the city name, which was “Dhaka”. Those rows were dropped. The city name was removed from all the values in the location column, keeping only the thana/PO/area names. The string “square feet” and “,” were removed from the area column and the data type was changed to ‘‘int64’’ from ‘object’. The string values in the rent column were converted to numeric values. For example, 65 thousand was changed to 65000 and 1.5 Lakh was changed to 150000. The data type of the rent column was changed to ‘int64’ from ‘object’. Outliers are removed as required. The dataset preprocessing features are summarized in table 1.

4.2 Data Cleaning

The investigation revealed a subset of apartments with rental costs beyond 100,000. This subset comprised 0.98% of the dataset. We used two methods to eliminate these rare situations. First, all data points above 1% were removed for data cleansing. This technique tried to eliminate likely outliers. After accepting that bedrooms and bathrooms affect rental pricing, we selectively eliminated outliers based on their relationship with these major qualities. This rigorous technique reduced outliers to increase our dataset's dependability and better represent most rental circumstances in the analysis.

After outlier treatment, noteworthy dataset outcomes were observed. First, outliers like six-bedroom apartments were removed to handle extreme values. The dataset also omitted three- and four-bedroom contemporary apartments. Our focus on mid-range flats in our analysis was a conscious choice to narrow the investigation to a specific rental market segment. We carefully removed these outliers to better align the dataset with our analysis and better represent the mid-range flat rental scenario in the results.

5. Results and Discussion

This section presents a discussion of the findings derived from our research. We choose to utilize both numerical findings and pictorial representations to enhance comprehension. The graphical results have incorporated and elucidated the trend analysis. Subsequently, a depiction of shape analysis was presented.

5.1 Numerical Results

The results presented illustrate the performance indicators of diverse models. The ensemble model has demonstrated robust performance in accurately forecasting load data, exhibiting a commendable ability to capture 86% of the variance and exhibiting a negligible average prediction error. Both Random Forest and Decision Tree algorithms have demonstrated strong performance, but with variable degrees of precision. The performance of the Support Vector Regression (SVR) model is the least satisfactory.

Table 2. Result analysis from evaluation metrics

Models	RMSE	MAE	MAPE	R2
SVR	11042.92	5763.88	0.24	0.07
MLP	8627.14	5075.56	0.26	0.43
Ada	6867.59	4461.65	0.23	0.64
Ridge	6576.25	4109	0.21	0.67
LR	6576.15	4116.61	0.21	0.67
KNN	4857.43	2760.77	0.13	0.82
DT	4703.71	2710.13	0.13	0.83
LGBM	4574.4	2659.4	0.13	0.84
GB	4542.07	2524.2	0.12	0.84
RF	4526.47	2649.45	0.13	0.84
XGBoost	4462.72	2559.85	0.13	0.85
Ensemble	**4398.23**	**2536.77**	**0.11**	**0.86**

5.2 Graphical Results

At first, we check the plots of RMSE and MAE errors. We see the Ensemble model has the lowest RMSE and MAE errors. XgBoost, Random Forest, GB and LGBM have also performed well.

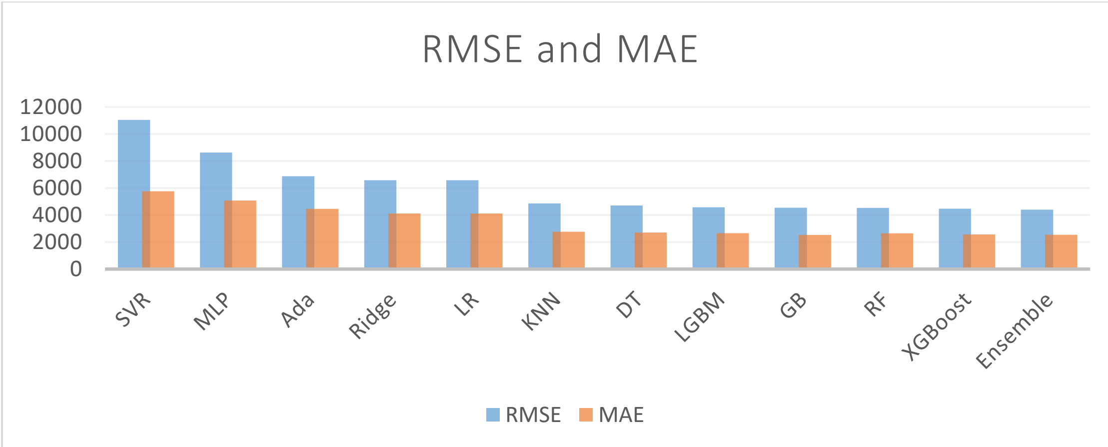

Figure 2. RMSE and MAE Error Graph

Next, we check the MAPE plot. Here, we see GB performs better than XgBoost, while LGBM, Random Forest, Decision Tree & k-NN have equal score as XgBoost. Ensemble model holds the top position. Lastly, we go for R^2 Score. Here, ensemble model achieved the best score, while XgBoost, Decision Tree, Random Forest, LGBM, GB, and k-NN had achieved similar results.

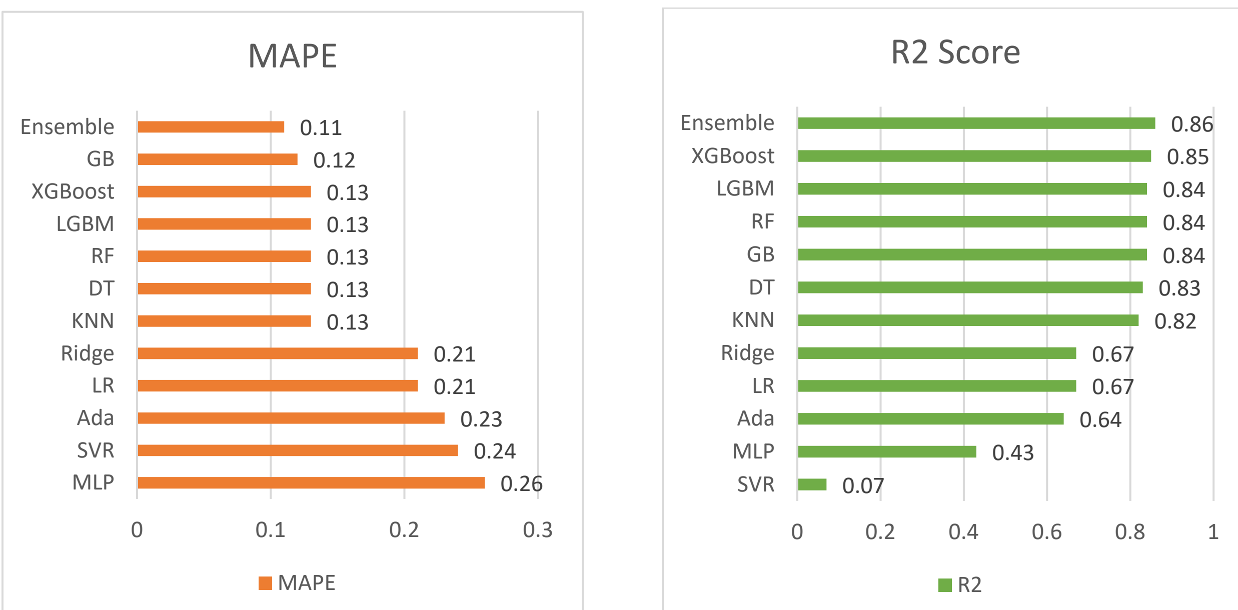

Figure 3. MAPE & R^2 Plot of Different Algorithms

5.3 Performance Analysis

A residual plot is a visual tool used to depict the discrepancies between the observed values and the anticipated values (known as residuals) within a regression study. Residual plots play a crucial role in evaluating the adequacy of a regression model's fit and detecting any discernible patterns or trends within the residuals. The presence of a consistent pattern in the residual plot may suggest that the model lacks some information or exhibits a systematic mistake (Larsen and McCleary 1972).

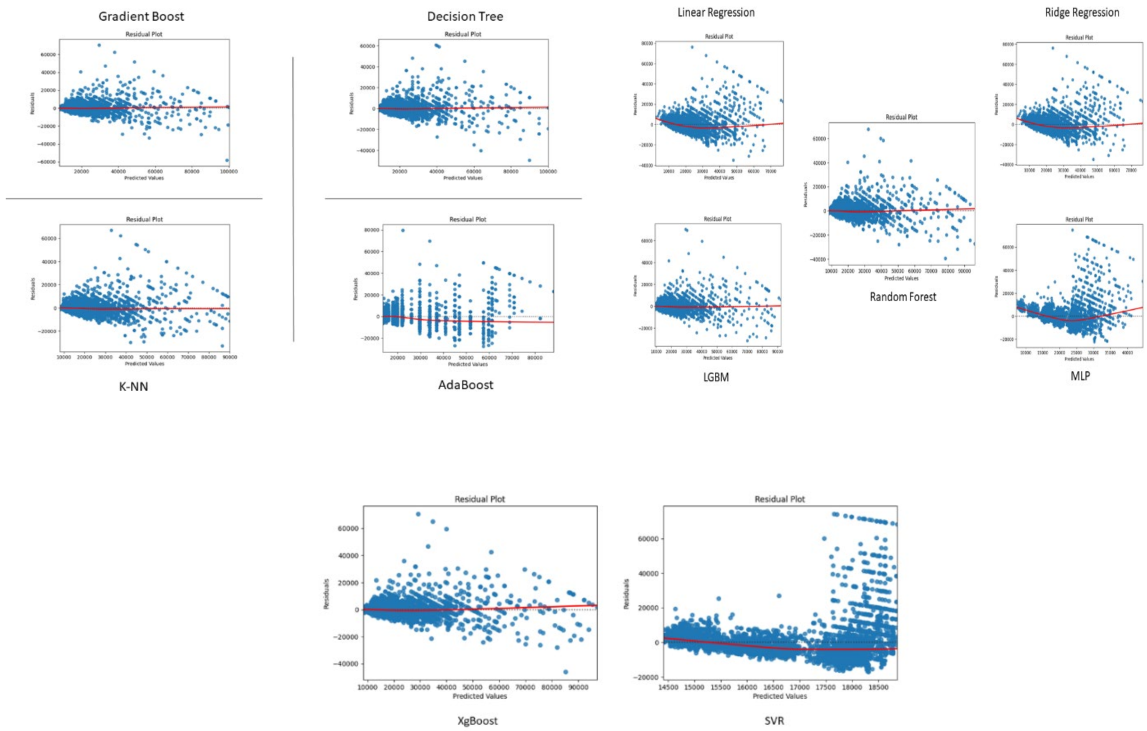

Figure 4. Residual Plots of Algorithms Used

From the plots, we found that the algorithm, GB, DT, XgBoost, LGBM and RF, had nearly linear zero line and the most values are close to the line. These algorithms did not have many outliers. On the contrary, SVR, MLP, Ridge Regression and LR had the worst non-linear zero line and most values were treated as outliers. So, the performance of these algorithms was also dropped.

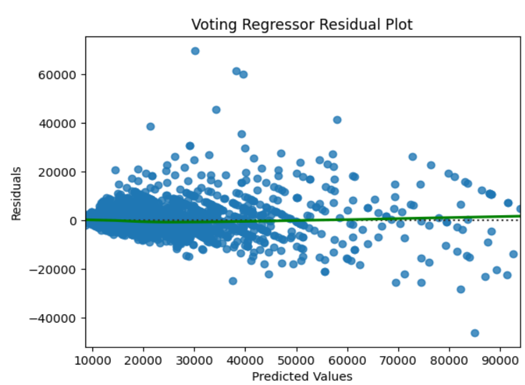

Figure 5. Residual Plot of Ensemble Method

As for the ensemble method, shown in figure 5, the zero line was nearly linear. The values were completely close to the line and almost zero outliers. This method showed the best outcome.

5.4 SHAP Analysis

The SHAP summary plot shows how each feature is connected to the prediction models. For the feature "Area", we can see that the feature value gets high when SHAP value is increasing. It means, the larger the apartment area, the higher the rent. It is true for real life applications. For "Location" feature, we see that both high and low values are overlapping. It means that apartments of both high and low values are in the same location. Increasing of bath and bed also increases the apartment fare, but their impact to the prediction is low. "Area" and "Location" are the main features while predicting house rent.

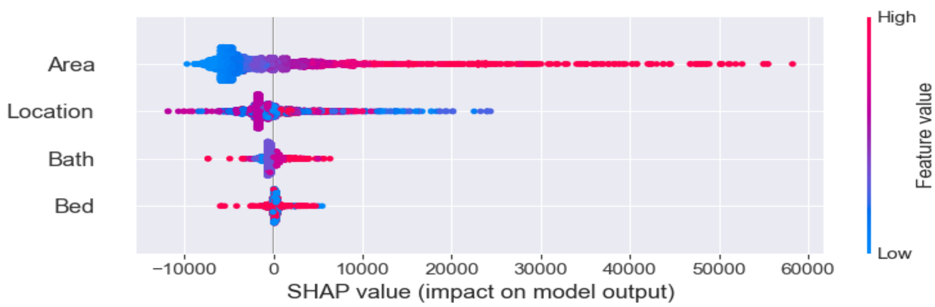

Figure 6. SHAP Summary Plot of the Dataset

To check how accurate our model is, we also viewed the SHAP force plot on single instance predictions. The voting regressor was used for creating SHAP values of force plot. Let's take a point, 3000. For the 3000th point, the predicted value is 18892.56 while the actual value is 16500. It shows how close the prediction is done by voting regressor. And the value is so close to the base value (19400). For the fare prediction, we see that the location and area of apartment is lowering the rent while amount of bed and bath in that apartment is pushing it higher.

Figure 7. SHAP Force Plot for Point 3000

For another point, let's say 634, we see that the predicted value is 14876 while actual value in the dataset is 15500. It shows the predicted value is almost close to the actual value. Low fare is predicted according to the area and bath while location and bed was pushing the fare high.

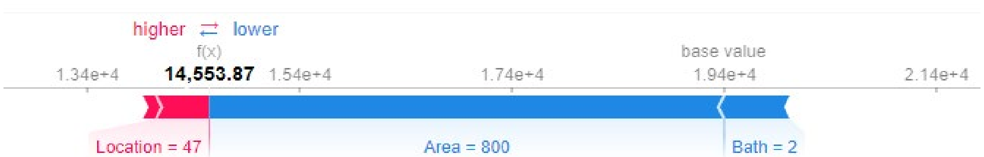

Figure 8. SHAP Force Plot for Point 634

6. Conclusion

In this article, a novel way to predict flat rent in Dhaka city is presented. This approach makes use of a comprehensive strategy that employs twelve different machine learning algorithms. The results reveal that the overall performance is solid, with the Voting Regressor Ensemble Model emerging as the most effective predictor. The significance of ensemble models in improving accuracy for real-world applications in the dynamic area of rent prediction is highlighted by this realization. Also, the black box nature of the models was interpreted here.

6.1 Proposed Improvement

Our algorithms performed exceptionally well. The ensemble model had a MAPE of 11%, lower than earlier research. We improved the existing works as follows:

1. No previous research has been done on the transparency of predicting apartment rent prices in Bangladesh. The recommended enhancement emphasizes the study's unique contribution. The study's distinctiveness and originality make it a local pioneering endeavor.
2. The proposal implies that real estate brokers and homeowners can benefit from the study. The analysis can help real estate decision-makers by projecting Dhaka City flat rents. This implies that the study has practical applications beyond academic curiosity.
3. The proposal offers a comparison analysis with research or works that have been undertaken in the same field in the past. From the previous works, we have discovered a great deal of room for development. First and foremost, we made use of the Voting Regressor ensemble model on the tree-based algorithms, which produced the most favorable results.
4. SHAP explainable AI made it possible to easily understand our findings and making it clear to all the stakeholders of the Bangladesh real estate community.

6.2 Future Research

Inherent to any research endeavor, it is imperative to acknowledge the limitations that may influence the scope, generalizability, and precision of the study. This section delineates key constraints encountered during the exploration of the study, shedding light on areas where the study may be refined or extended for future investigations:

1. Since the dataset does not contain comprehensive data on all components of a property, it leaves out important details, which restricts the level of depth that can be achieved in the research.
2. The study does not include any data from regions that have populations that are greater than one lakh, which may cause it to overlook significant patterns or variations in flat rents in regions that have a high population density.
3. The dataset used in the research is from 2022, which influences the accuracy of the model as well as its relevance to the dynamics of the current market. This is because the real estate and rental markets are both continuously evolving.

Although there were limitations, this study conducted a comparative analysis of various machine learning algorithms for estimating the monthly rental prices of flats in Dhaka city. By making suitable adjustments, it may be implemented on online flat rental search engines to assist users in seeking and forecasting flat rents in various locations of Dhaka city.

References

Begum, A., Urban housing as an issue of redistribution through planning? The case of Dhaka city, *Social Policy & Administration*, vol.41, no.4, pp.410-418, 2007

Breiman, L., Random forests, *Machine learning*, vol.45, pp.5-32, 2001

Chen, T. & Guestrin, C., Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp.785-794, 2016

Friedman, J.H., Greedy function approximation: a gradient boosting machine, *Annals of statistics*, pp.1189-1232, 2001

Geethamani, R. & Karthika, B., HOUSE RENT PREDICTION USING MACHINE LEARNING ALGORITHMS – A METHODICAL REVIEW, *International Journal of Emerging Technologies and Innovative Research (www.jetir.org)*, vol.10, no.3, pp.112-116, 2023

Hasina, S. Bangladesh is booming-and its future looks even brighter| World Economic Forum. In: (October.2019

Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M. & Cai, Z., Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies, *Land use policy*, vol.82, pp.657-673, 2019

Jui, J.J., Imran Molla, M., Bari, B.S., Rashid, M. & Hasan, M.J., Flat price prediction using linear and random forest regression based on machine learning techniques, *Embracing Industry 4.0: Selected Articles from MUCET 2019*, pp.205-217, 2020

Kamruzzaman, M. & Ogura, N., Apartment housing in Dhaka City: Past, present and characteristic outlook, *Building Stock Activation, Tokyo, Japan*, 2007

Khosravi, M., Arif, S.B., Ghaseminejad, A., Tohidi, H. & Shabanian, H., Performance Evaluation of Machine Learning Regressors for Estimating Real Estate House Prices, 2022

Kumar, A., House Rent Price Prediction, *International Research Journal of Engineering and Technology*, vol.6, no.04, pp.3188-3191, 2019

Larsen, W.A. & McCleary, S.J., The use of partial residual plots in regression analysis, *Technometrics*, vol.14, no.3, pp.781-790, 1972

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. & Lee, S.-I., From local explanations to global understanding with explainable AI for trees, *Nature machine intelligence*, vol.2, no.1, pp.56-67, 2020

Lundberg, S.M. & Lee, S.-I., A unified approach to interpreting model predictions, *Advances in neural information processing systems*, vol.30, 2017

McDonald, G.C., Ridge regression, *Wiley Interdisciplinary Reviews: Computational Statistics*, vol.1, no.1, pp.93-100, 2009

Mokhtari, K.E., Higdon, B.P. & Başar, A., Interpreting financial time series with SHAP values, *Proceedings of the 29th annual international conference on computer science and software engineering*, pp.166-172, 2019

Mora-Garcia, R.-T., Cespedes-Lopez, M.-F. & Perez-Sanchez, V.R., Housing Price Prediction Using Machine Learning Algorithms in COVID-19 Times, *Land*, vol.11, no.11, pp.2100, 2022

Natekin, A. & Knoll, A., Gradient boosting machines, a tutorial, *Frontiers in neurorobotics*, vol.7, pp.21, 2013

Neloy, A.A., Haque, H.S. & Ul Islam, M.M., Ensemble learning based rental apartment price prediction model by categorical features factoring, *Proceedings of the 2019 11th International conference on machine learning and computing*, pp.350-356, 2019

Park, B. & Bae, J.K., Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data, *Expert systems with applications*, vol.42, no.6, pp.2928-2934, 2015

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R. & Dubourg, V., Scikit-learn: Machine learning in Python, *the Journal of machine Learning research*, vol.12, pp.2825-2830, 2011

Rajkumar, S., Ramanathan, L., Ramalingam, R. & Jantwal, M. Cloud Hosted Ensemble Learning-Based Rental Apartment Price Prediction Model Using Stacking Technique. In: Deep Learning Research Applications for Natural Language Processing (pp. 229-238): IGI Global 2023

Samsuddin Ahmed, M.M.R. & Islam, S., House rent estimation in Dhaka city by multi layer perceptions neural network, *International Journal of u-and e-Service, Science and Technology*, vol.7, no.4, pp.287-300, 2014

Schapire, R.E. Explaining adaboost. In: Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik (pp. 37-52): Springer 2013

Singh, A.P., Rastogi, K. & Rajpoot, S., House Price Prediction Using Machine Learning, *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp.203-206, 2021

Voutas Chatzidis, I., Prediction of housing prices based on spatial & social parameters using regression & deep learning methods, *Aristotle University of Thessaloniki*, 2019

Yoshida, T., Murakami, D. & Seya, H., Spatial prediction of apartment rent using regression-based and machine learning-based approaches with a large dataset, *The Journal of Real Estate Finance and Economics*, pp.1-28, 2022

Biography

Taeef Najib is a self-taught programmer, working as a Software Development Engineer (Data Science) at Sidetrek. He did his bachelor's and master's in business administration with Marketing as his concentration. His passion lies in data analytics, machine learning, and computer vision.

Fahim Muntasir, a Research Engineer at AIMS Lab, UIU, is currently pursuing his M.Sc. in Computer Science and Engineering from BRAC University. He obtained his B.Sc. in Electrical and Electronics Engineering from Rajshahi University of Engineering and Technology. His area of interest is interdisciplinary applications of machine learning, computer vision, explainable AI, and deep neural networks.

Wasif Al Wazed Wasi, an IT Engineer at Telecontran Limited, holds a B.Sc. in Electrical and Electronic Engineering from RUET. With a passion for Natural Language Processing (NLP), his research interests encompass machine translation, lexical semantics, and generative pre-trained models (GPT). Wasif explores innovative approaches in NLP, aiming to enhance language understanding and communication in diverse applications. His dedication to advancing NLP technologies reflects a commitment to shaping the future of human-computer interaction and linguistic intelligence.