INTERNATIONAL REAL ESTATE REVIEW

2025 Vol. 28 No. 1: pp. 53 - 69

Using an Asymmetric Loss Function to Alleviate the Risk of Loan Collateral Overvaluation

Changro Lee
Kangwon National University, Chuncheon, Republic of Korea,
Email: spatialstat@naver.com

Financial institutions are increasingly adopting machine learning-based valuation models to evaluate loan collaterals. However, most machine learning algorithms do not differentiate between the risks associated with the undervaluation and overvaluation of such assets. From the perspective of a lender, the risks of overvaluing loan collateral are more critical than those that arise from undervaluing them. In this study, we alleviate this risk of overvaluation by explicitly considering an asymmetric loss function. We customize a gradient boosting machine (GBM) by specifying an asymmetric loss function, and assigning a higher penalty for overvaluation. This customized GBM is then used to predict house prices in Gimhae, South Korea. The results show that the GBM effectively reduces overvaluation while maintaining prediction accuracy. Researchers and practitioners need to intentionally bias their machine learning algorithms to incorporate the asymmetric risks associated with their businesses. The approach proposed in this study can help stakeholders make informed decisions in the lending process, thereby mitigating the risk of default by borrowers, and ultimately promoting sustainable lending practices.

Keywords

Loss function, Asymmetric risk, Overvaluation, Gradient boosting machine, House valuation

1. Introduction

Managing the risks associated with lending activities has always been a critical issue for stakeholders in the financial industry. One critical aspect of this risk management is the appropriate valuation of properties that are pledged as security (Kruger and Maturana, 2021). Today, many banks and financial institutions are increasingly adopting machine learning-based valuation models to estimate the value of loan collaterals. Most machine learning algorithms treat underestimates and overestimates of loan collateral in the same manner, even though their impacts on financial institutions differ significantly. Compared to undervaluation, the overvaluation of assets is significantly costlier for financial institutions, due to the potential for serious repercussions like borrower defaults.

Financial institutions need to deal with the undervaluation and overvaluation of loan collaterals differently. In this study, we customize a gradient boosting machine (GBM), which is a popular machine learning algorithm, by incorporating an asymmetric loss function into its design. The GBM used in this study imposes a ten-times higher penalty for overvaluation. We apply this GBM to estimate house prices in Gimhae, South Korea, and the results are promising: the GBM effectively alleviates overvaluation cases while upholding performance quality.

In many business contexts, the risks associated with underestimation and overestimation are asymmetrical. Customizing machine learning algorithms by introducing asymmetric loss functions has been increasingly done across various domains, such as medical diagnoses (Hashemi et al., 2018; Bokhari and Bansal, 2020) and computer vision (Vogels et al., 2018; Zhou et al., 2023). However, similar efforts to tailor property valuation models by incorporating the asymmetric risk associated with inaccurate valuations have not been observed for real estate. In this study, we address this research gap. The approach adopted in this study is expected to promote sustainable lending practices and safeguard the real estate and financial industries.

The remainder of this paper is organized as follows. Section 2 reviews the asymmetric risk of businesses, and asymmetric loss functions in machine learning algorithms. In Section 3, the dataset, study area, and customized GBM are presented. Section 4 presents the findings and their implications. Finally, Section 5 provides a summary and proposes avenues for future research work.

2. Literature Review

2.1 Asymmetric Business Risk

Loss functions play a crucial role in machine learning algorithms. They measure the difference between the observed and predicted values, and give a penalty when predictions deviate from observed values, thereby leading to the convergence of the algorithm (Domingos, 2012). Most loss functions assign the same penalty to underestimation and overestimation if the absolute errors are the same. In most contexts, these symmetric loss functions are considered desirable because they can enhance training stability and algorithm robustness (Wang et al., 2020).

However, in many business environments, it is common for the consequences of underestimation and overestimation to be asymmetrical. For example, a retail store that underpredicts demand and runs out of stock of a particular product may suffer from drawbacks, like sales revenue losses and increasing complaints from customers. A store that overpredicts demand may have excessive inventory and product obsolescence. Depending on the retail context, the cost of incorrect predictions varies; if customer complaints cause serious damage to the brand reputation of a store, the cost of underestimation is greater than that of overestimation. In contrast, if the cost of storing inventory is high, the cost of overestimation is greater than that of underestimation.

In property management, the overestimation of the remaining useful life of properties often proves costlier than underestimation; in extreme cases, the overestimation of the remaining life of a property can have severe financial implications if it is damaged or reaches the stage of collapse early. In the case of machine and plant maintenance, the overestimation of their durability is also costlier than underestimation: it is well known that delayed repair generally costs more than preventive maintenance.

Property lending is the most representative case that demonstrates the asymmetric risk of incorrect decisions. Overvaluation of loan collaterals may result in several detrimental consequences: in the event of a borrower default, the lender may be unable to retrieve the full loan amount with the sale of the property pledged as security. On a broader scale, the overvaluation of loan collaterals can lead to overall market distortion, and thus threaten the health of the property market. In contrast, when loan collaterals are undervalued, the lender may lose the business opportunity for an additional loan, but the cost may be trivial compared to that of a borrower default.

Therefore, incorporating the asymmetric risk of property valuation into the real estate lending process is essential. Many valuation models adopted by banks and financial institutions have been developed based on machine learning algorithms (Steurer et al., 2021; Gao et al., 2022), and these algorithms are generally optimized by using conventional symmetric loss functions (Ecker et al., 2020). This study explicitly incorporates the asymmetric risk of overvaluation into a machine learning-based valuation model by customizing a loss function.

2.2 Customizing a Loss Function

Common loss functions for regression tasks include mean squared error (MSE), mean absolute error (MAE), and Huber loss (Kotta et al., 2021). The standard loss functions for classification problems include cross-entropy and hinge losses (Fei et al., 2020). These loss functions are all symmetric: that is, they assign the same penalty to underestimation and overestimation if the absolute error is the same.

Patton and Timmermann (2007) theoretically show that the standard MSE loss would not work effectively with even a slight deviation of an economic time series from the conventional assumptions made for economic forecasting. Instead, they suggest more general loss functions that could deal with asymmetric loss and nonlinearity. Since their study, a few empirical studies have applied asymmetric loss functions to their own domains. Gkillas et al. (2002) forecast price volatility in the oil market by using an asymmetric loss function. Berk (2011) applies an asymmetric loss function in criminal justice settings, and Gupta et al. (2020) customize a loss function to include an asymmetric feature for predicting sediment load. Ridnik et al. (2021) introduce asymmetric loss to image classification problems and show excellent results for multiple image datasets. More recently, Luo et al. (2023) design a class-adaptive asymmetric loss function to predict ingredients in food images.

As shown by previous studies, asymmetric loss functions have been applied in various domains, including the oil and gas industry, criminology, hydrology, and computer vision. However, to the best of the author's knowledge, there has been no attempt to apply an asymmetric loss function to a property valuation domain to alleviate the asymmetric risk involved in the real estate lending business. This study fills this gap by customizing a standard MSE loss function and applying this function to house valuation. As overvaluation can prove much costlier for a lender, so this study customizes a standard MSE loss function by assigning a higher penalty for overestimated house prices.

3. Data and Method

3.1 Dataset and Study Area

The transaction records of houses sold between January 2019 and April 2023 are used in this study. This dataset is available on the Korean government website in the form of a comma-separated-value file.[1] The city of Gimhae located in the southeastern part of the Korean Peninsula is selected as the study area. In 2022, its population was 317,000 (KOSIS, 2023). The economy of Gimhae is diverse, with a mix of industries that include the manufacturing, agriculture, and service sectors. The city is part of the broader Southeastern Economic Zone, which includes other major cities like Busan. Compared to major cities like Busan and Seoul, the housing market in Gimhae is generally more affordable. The growing economy, improving infrastructure, and relatively affordable prices of Gimhae make it an attractive option for real estate investment (Kim and Yoon, 2020).

Gimhae is notable for its dynamic housing market, where houses are readily bought and sold, and transactions are systematically documented by the local authority. The city is chosen for analysis due to its active housing market and accessibility of a substantial dataset.

Table 1 presents the summary statistics for the houses used in this study. This dataset is randomly split into samples of 2,257 (55%), 614 (15%), and 1,216 (30%) for training, hyperparameter validation, and testing, respectively. The summary statistics of the training data indicate that an average house in Gimhae has a site area of 228.0 m^2 and a building area of 163.0 m^2, with a value of 264 million KRW (approximately 200,000 USD). Typically, these houses are constructed with a reinforced-concrete framework and around 30 years old.

Figure 1 shows the study area and location of the 4,087 houses. A few clusters in the figure represent densely populated residential areas. The areas where houses are sparsely distributed are either farmland or mountainous areas.

[1] https://rtdown.molit.go.kr/. The Ministry of Land, Infrastructure and Transport maintains this website.

Table 1 Summary Statistics of 4,087 Houses sold between January 2019 and April 2023

Training data (n = 2,257)

	Min.	Median	Mean	Max.
Price (million KRW)*	15.6	264.0	354.1	5,112.0
Property age (year)	10	30	32	118
Site area (m^2)	27.8	228.0	282.6	9,477.0
Building area (m^2)	18.5	163.0	206.2	1,741.6
Transaction year	2019 (511 houses), 2020 (689 houses), 2021 (405 houses), 2022 (443 houses), 2023 (209 houses)			
Slope	flat (1,518 houses), sloping (739 houses)			
Bearing	east (415 houses), west (330 houses), north (651 houses), south (861 houses)			
Building structure	reinforced-concrete (789), steel frame (39), brick masonry (581), block masonry (531), wooden frame (175), mobile homes (112), other (30)			

Validation data (n = 614)

	Min.	Median	Mean	Max.
Price (million KRW)*	16.8	249.6	345.1	2,682.0
Property age (year)	10	30	33	122
Site area (m^2)	30.0	229.5	306.7	8,456.0
Building area (m^2)	23.0	163.0	198.1	1,512.8
Transaction year	2019 (122 houses), 2020 (167 houses), 2021 (128 houses), 2022 (138 houses), 2023 (59 houses)			
Slope	flat (409 houses), sloping (205 houses)			
Bearing	east (117 houses), west (96 houses), north (161 houses), south (240 houses)			
Building structure	reinforced-concrete (201), steel frame (11), brick masonry (142), block masonry (150), wooden frame (65), mobile homes (36), other (9)			

Test data (n = 1,216)

	Min.	Median	Mean	Max.
Price (million KRW)*	16.2	257.4	353.1	4,890.0
Property age (year)	10	30	32	116
Site area (m^2)	30.5	226.0	273.0	8,622.0
Building area (m^2)	23.1	164.0	204.0	1,599.4
Transaction year	2019 (298 houses), 2020 (342 houses), 2021 (200 houses), 2022 (251 houses), 2023 (125 houses)			
Slope	flat (824 houses), sloping (392 houses)			
Bearing	east (212 houses), west (185 houses), north (361 houses), south (458 houses)			
Building structure	reinforced-concrete (445), steel frame (23), brick masonry (282), block masonry (288), wooden frame (95), mobile homes (75), other (8)			

Notes: * million KRW = 690 USD (as of March 2025)

Figure 1 Study Area and Location of Houses

Notes: Study Area is highlighted in Gray on Left Map, while Dots denote Location of Houses on Right Map.

3.2 GBM with an Asymmetric Loss Function

The GBM is a machine learning technique that combines the predictions of multiple basic learners, such as decision trees, through sequential training (Natekin and Knoll, 2013).[2] The GBM is adopted as the main valuation model in this study because: (1) it shows excellent performance in many competitions, such as the Kaggle competitions (Taieb and Hyndman, 2014), and (2) it is relatively easy to customize a loss function manually, for example, deriving the first and second derivatives of a loss function.

The residual is calculated by subtracting the predicted value from the observed value as follows:

$$residual = y - \hat{y} \tag{1}$$

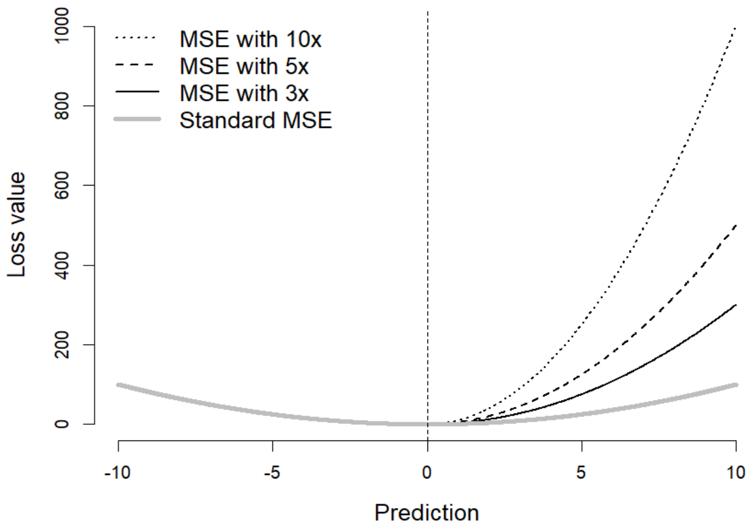
A negative residual indicates that the predicted value is higher than the observed value, or in other words, there is an overestimation. As overvaluation needs to be avoided as much as possible in the business of lending, a custom loss function is designed to assign a higher penalty to a negative residual. In contrast, a residual with a positive sign indicates that the house price is underestimated, which is not the focus of business practitioners.

---

2 This technique is often referred to as an ensemble model in the literature. Details on the GBM are succinctly explained in Friedman (2002).

As the output of a house valuation is a numerical value (price), an MSE loss function is adopted. Figure 2 shows the behavior of the MSE loss function as the severity of the penalty for overvaluation varies. True values are set to zero, and predictions are generated to have values that range from -10 to 10. MSE with "10 x " means MSE with a ten-times higher penalty for overvaluation. The MSE loss function is customized such that it assigns a ten-times higher penalty to negative residuals than positive ones. The degree of penalty given for the overvaluation of loan collateral varies according to the policies of financial institutions. Our consultations with bank managers reveal that they are willing to prevent overvaluation at all cost; thus, this study adopts the MSE loss function and assigns a ten-times higher penalty for overvaluation.[3]

Figure 2 MSE Loss Functions with Varying Penalties for Overvaluation



The input variables used to predict house prices are as follows: site area, floor area, property age, road width on which a lot abuts, geographical coordinates (longitude and latitude), transaction year, zone, slope, lot shape, bearing, building structure, type of roof, and neighborhood characteristics. The house sales prices (the target variable) are log-transformed and scaled to improve the training process. Table 2 lists the input variables employed to train the GBM.

---

3 The MSE loss function customized in this way is used for both training and validation losses. In the GBM, an important hyperparameter that needs to be optimized through validation loss is the number of boosting iterations.

Table 2 Input Variables Employed to train the GBM

Name (type)	Remarks	Name (type)	Remarks
Site area (numerical)	Min-max scaled	Zone (categorical)	Residential, commercial, etc. (14 levels)
Floor area (numerical)	Min-max scaled	Slope (categorical)	Flat, sloped (2 levels)
Property age (numerical)	Min-max scaled	Lot shape (categorical)	Rectangular, trapezoidal, etc. (4 levels)
Road width (numerical)	Min-max scaled	Bearing (categorical)	East, north, south, west (4 levels)
Longitude (numerical)	Min-max scaled	Building structure (categorical)	Reinforced concrete, wooden frame, etc. (7 levels)
Latitude (numerical)	Min-max scaled	Type of roof (categorical)	Slab, shingle, other (3 levels)
Transaction year (categorical)	2019, 2020, 2021, 2022, 2023 (5 levels)	Neighborhood (categorical)	Downtown, suburbs, etc. (5 levels)

4. Results

4.1 Model Performance

The ordinary least-squares (OLS) model is used as the baseline model. The GBM employed in this study uses decision trees as basic learners. Two types of GBMs are fitted to the dataset. The first GBM, which employs a standard MSE loss function, is optimized after 87 boosting iterations. The second GBM, equipped with an asymmetric MSE loss function that assigns a ten-times higher penalty for overvaluation, is optimized after 26 boosting rounds.[4]

Table 3 shows the model performance on the test dataset (1,216 houses). As expected, the two GBMs in general perform better than the OLS. Among the three models, the ordinary GBM performs best with 0.004 when the standard MSE is used as the metric. However, the most important metric in this study is the asymmetric MSE, and the customized GBM shows the best performance

4 The asymmetric MSE loss function for OLS can be specified as follows when imposing a ten-times higher penalty for overvaluation: $loss = \frac{1}{n}\sum_{i=1}^{n}\begin{cases} 10 \times (y_i - \hat{y}_i)^2 & \text{if } y_i < \hat{y}_i \\ (y_i - \hat{y}_i)^2 & \text{otherwise} \end{cases}$, where y_i represents the actual value and $\hat{y}_i$ indicates the predicted value. Given that the focus of this study is on improving machine learning models, OLS with an asymmetric MSE loss function is not investigated.

with 0.011 according to this criterion. The customized GBM also yields the lowest ratio of overvaluation (17.8%) among the three models.[5]

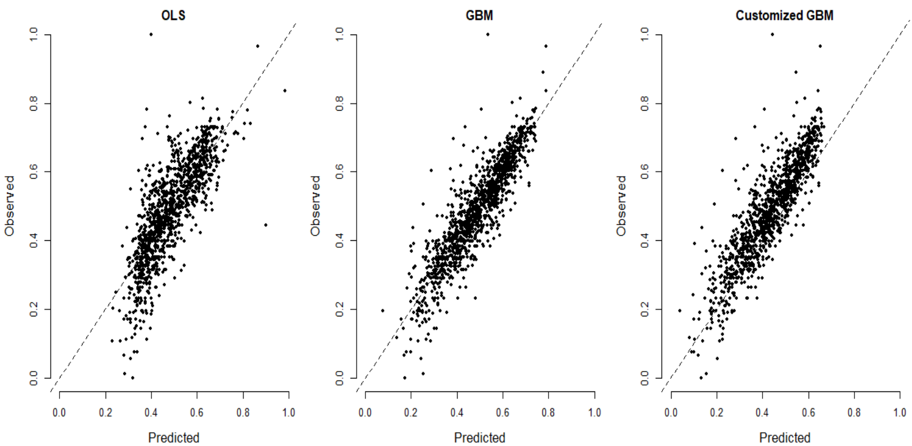
Figure 3 graphically presents the results from Table 3. In the figure, the horizontal axis shows the predicted prices from the models, while the vertical axis shows the observed prices. Data points on the dotted diagonal line represent accurate predictions. The predicted values deviate insignificantly from the observed values in the OLS. In the GBM, the degree of deviation is significantly decreased. In the customized GBM, the degree of deviation between the predicted and observed values remains the same as in the ordinary GBM; however, the overall mass of data points shifts upward, thus indicating that the estimated house prices are adjusted downward while keeping the deviation between the predicted and observed values as small as possible.

Table 3 Model Performance on the Test Dataset

	Standard MSE	Asymmetric MSE	Ratio of overvaluation*
OLS	0.007	0.040	48.9%
GBM	0.004	0.021	49.9%
Customized GBM	0.006	0.011	17.8%

Notes: * Number of cases where predicted values are higher than observed values ÷ total number of cases × 100

Figure 3 Goodness-of-Fit of Models



5 In this study, a ten-times higher penalty for overvaluation is employed. The results of simulations with varying penalty values are presented in Appendix 1.

4.2 Implications

Figure 4 shows the distribution of residuals from the OLS (pink), GBM (green), and customized GBM (blue). All of the residual distributions are centered near zero, but with notable differences in shape, spread, and symmetry. The OLS residuals are approximately symmetric around zero, which is expected due to the mathematical properties of OLS. They show the widest spread among the three models, which indicates lower prediction accuracy compared to the other models. The standard GBM residuals show a notably narrower distribution compared to the OLS and maintain symmetry around zero due to the MSE loss function. Their higher density near zero indicates more predictions with small errors. In the customized GBM, the residuals show a narrower distribution compared to OLS and maintain intentional asymmetry with a right-side skew. The higher concentration of positive residuals (observed price > predicted price) shows a systematic tendency to undervalue. In short, the bias of the customized GBM toward undervaluation represents a conservative valuation strategy. This approach shows how machine learning models can be adapted to domain-specific requirements.

Figure 4 Distribution of Residuals from the Models

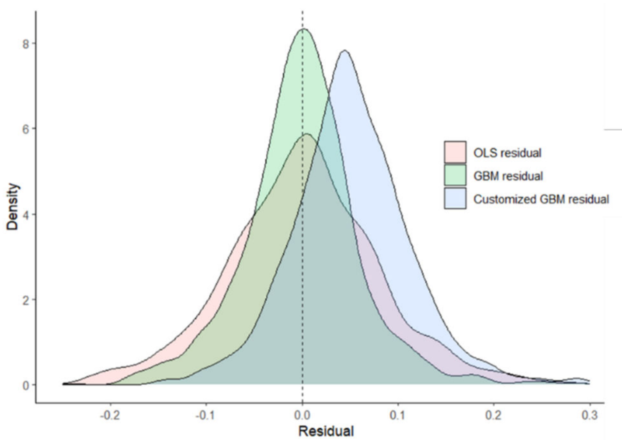
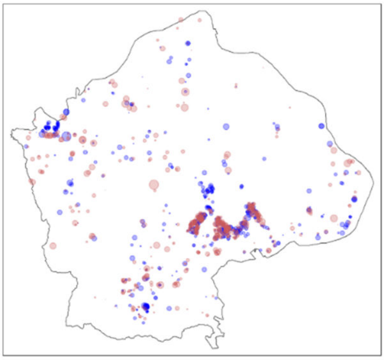


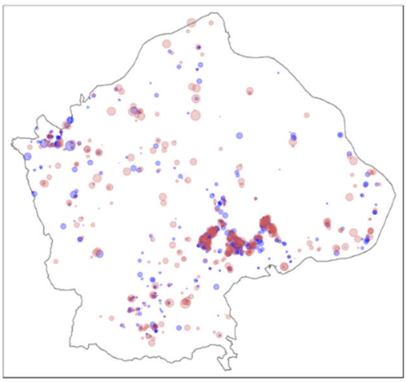
Figure 5 shows the spatial distribution of the residuals, where the red circles indicate overvaluation (negative residuals), and blue circles indicate undervaluation (positive residuals). The size of the plotted circles is in proportion to the absolute residual value. In the OLS and GBM, overvaluation can be observed throughout the study area. In the customized GBM, many overvaluation cases are converted into undervaluation cases by using the asymmetric loss function. In particular, two clusters in the southeastern and northwestern regions of the study area show dramatic changes in color. Based on this change, it can be inferred that most overvaluation cases occur in densely populated residential areas.

Figure 5 Spatial Distribution of Residuals

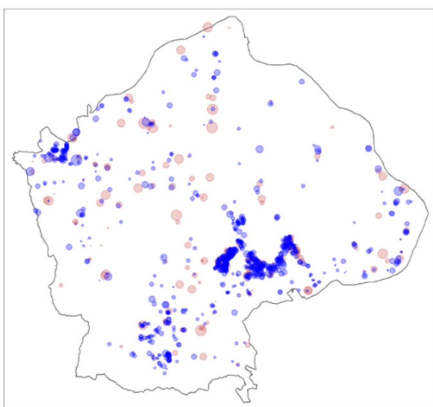
a. OLS Residuals



b. GBM Residuals



c. Customized GBM Residuals



Overvaluation can stem from various sources.[6] Borrowers may have better information about the true value of the asset than the lender, which is likely to lead to overvaluation. The use of outdated valuation models can also lead to inaccurate valuations. In some cases, deliberate misrepresentation or fraud can result in overvaluation. Banks and financial institutions invest significant resources to mitigate the overvaluation of loan collaterals. First, they perform due diligence on houses to reduce information asymmetry. Second, they keep automated valuation models (AVMs) up to date, incorporate advanced algorithms, and regularly review collateral values. In addition, they implement strict risk management policies and internal controls to prevent fraudulent activities.

The implications of the customized GBM with an asymmetric loss function are significant for the financial industry, particularly in real estate lending. By penalizing overestimation more heavily, the model reduces the risk of overvaluing properties that serve as loan collateral. This is crucial for financial institutions as overvaluation can lead to increased risk of defaults and financial losses, thereby threatening market stability. By adopting this approach, lenders can fine-tune their AVMs, ensure more accurate and conservative property valuations, and contribute to the overall sustainability of the real estate lending market. Additionally, the spatial analysis of residuals can help institutions target specific geographical areas prone to overvaluation, thus enabling more focused auditing and monitoring efforts.

5. Conclusion

Machine learning algorithms are used across various business sectors, and one of their essential elements is the loss function. Most practically used loss functions assume that the costs of underestimation and overestimation are the same. However, the consequences of these two scenarios differ in real business settings. In this study, we customize a GBM by specifying an asymmetric loss function that assigns a ten-times higher penalty for overestimation, and then apply the GBM to the prediction of house prices in Gimhae, South Korea. The customized GBM successfully reduces overvaluation cases while maintaining performance quality.

The overvaluation of loan collaterals has always been a critical issue in the financial industry, which threatens the overall sustainability of the real estate

---

6 Kim (2017) reports that the average default rate for mortgage loans is approximately 0.55%. As of September 2024, the outstanding balance of mortgage loans in South Korea stands at 413 billion USD (Financial Supervisory Service, 2024). Consequently, the annual loss from bad loans can be roughly estimated at 2.3 billion USD. A portion of this loss can be attributed to the overvaluation of loan collaterals.

lending market. The approach adopted in this study is expected to be conveniently used in business practices and contribute to sustainable lending practices.

In this study, we chose the GBM to customize a loss function. In the GBM, it is relatively simple to specify an asymmetric loss function and implement this function on the training data. However, in other machine learning algorithms such as neural networks, this may not be the case: calculating derivatives of a loss function and applying the function to the training process may be difficult or even impossible. Therefore, asymmetric loss functions for different machine learning algorithms should be investigated through future research. Additionally, the empirical findings of this study are limited to Gimhae. In future research, customized machine learning algorithms should be applied to other geographical settings to generalize the results of this study.

Data Availability Statement

The data that support the findings of this study are available on the Korean government website in the form of a comma-separated-value file, https://rtdown.molit.go.kr/.

References

Berk, R. A. (2011). Asymmetric loss functions for forecasting in criminal justice settings. *Journal of Quantitative Criminology,* 27(1), 107-123. http://dx.doi.org/10.1007/s10940-010-9098-2

Bokhari, W., and Bansal, A. (2020). 'Asymmetric Error Control for Binary Classification in Medical Disease Diagnosis'. *2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE).* Laguna Hills, CA, USA, December 9-13. IEEE. https://doi.org/10.1109/AIKE48582.2020.00013

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87. https://doi.org/10.1145/2347736.2347755

Ecker, M., Isakson, H., and Kennedy, L. (2020). An Exposition of AVM Performance Metrics. *Journal of Real Estate Practice and Education*, 22(1), 22-39. https://doi.org/10.1080/15214842.2020.1757352

Fei, R., Yao, Q., Zhu, Y., Xu, Q., Li, A., Wu, H., and Hu, B. (2020). Deep Learning Structure for Cross-Domain Sentiment Classification Based on Improved Cross Entropy and Weight. *Scientific Programming*. https://doi.org/10.1155/2020/3810261

Financial Supervisory Service. (2024). Financial Statistics Information System. https://efisis.fss.or.kr/fss/fsiview/indexw.html

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis,* 38(4), 367-378. https://doi.org/10.1016/S0167-9473(01)00065-2

Gao, Q., Shi, V., Pettit, C.J., and Han, H. (2022). Property valuation using machine learning algorithms on statistical areas in Greater Sydney, Australia. *Land Use Policy*, 123(2), 106409. http://dx.doi.org/10.1016/j.landusepol.2022.106409

Gkillas, K., Gupta, R., and Pierdzioch, C. (2020). Forecasting realized oil-price volatility: The role of financial stress and asymmetric loss. *Journal of International Money and Finance*, 104, 102137. https://doi.org/10.1016/j.jimonfin.2020.102137

Gupta, D., Hazarika, B. B., and Berlin, M. (2020). Robust regularized extreme learning machine with asymmetric Huber loss function. *Neural Computing and Applications*, 32, 12971-12998. https://doi.org/10.1007/s00521-020-04741-w

Hashemi, S. R., Salehi, S. S. M., Erdogmus, D., Prabhu, S. P., Warfield, S. K., and Gholipour, A. (2018). Asymmetric Loss Functions and Deep Densely-Connected Networks for Highly-Imbalanced Medical Image Segmentation: Application to Multiple Sclerosis Lesion Detection. *IEEE Access,* 7, 1721-1735. https://doi.org/10.1109/ACCESS.2018.2886371

Kim, J. (2017). 주택담보대출의 대손율 추정과 결정요인에 대한 연구 [A Study on the Estimation of Expected Loss Rate of Mortgage Loan and its Determinants]. *Journal of Housing and Urban Finance*, 2(2), 57-91. In Korean. Available at: https://www.ejhuf.org/archive/view_article?pid=jhuf-2-2-57

Kim, M. and Yoon, K. (2020). 지역 정체성과 지자체 캐릭터 연구 - 김해시 캐릭터를 중심으로 [Character Study of Local Identity & Local Government - Focused on Gimhae City Character]. *Design Forum of Korea*, 25(3), 109-120.

KOSIS. (2023). 국내통계 [Populations and households in 2022, Korean Statistical Information Service, Daejeon City]. Available at:

https://kosis.kr/statisticsList/statisticsListIndex.do?vwcd=MT_ZTITLE&menuId=M_01_01

Kotta, H., Pardasani, K., Pandya, M., and Ghosh, R. (2021). 'Optimization of Loss Functions for Predictive Soil Mapping'. In Hassanien, A., Bhatnagar, R., Darwish, A. (eds). *Advanced Machine Learning Technologies and Applications*, *AMLTA 2020. Advances in Intelligent Systems and Computing, February 13-15*, *Jaipur, India*. Singapore: Springer, pp.95-104 https://doi.org/10.1007/978-981-15-3383-9_9

Kruger, S., and Maturana, G. (2021). Collateral Misreporting in the Residential Mortgage-Backed Security Market. *Management Science*, 67(5), 2729-2750.

Luo, M., Min, W., Wang, Z., Song, J., and Jiang, S. (2023). Ingredient prediction via context learning network with class-adaptive asymmetric loss. *IEEE Transactions on Image Processing,* 32, 5509-5523. https://doi.org/10.1109/TIP.2023.3318958

Natekin, A., and Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7, 21. https://doi.org/10.3389/fnbot.2013.00021

Patton, A. J., and Timmermann, A. (2007). Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity. *Journal of Econometrics*, 140(2), 884-918. http://dx.doi.org/10.1016/j.jeconom.2006.07.018

Ridnik, T., Ben-Baruch, E., Zamir, N., Noy, A., Friedman, I., Protter, M., and Zelnik-Manor, L. (2021). 'Asymmetric Loss For Multi-Label Classification'. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, Montreal, Quebec, Canada.* United States: IEEE, pp. 82-91.

Steurer, M., Hill, R. J., and Pfeifer, N. (2021). Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research*, 38(2), 99-129. https://doi.org/10.1080/09599916.2020.1858937

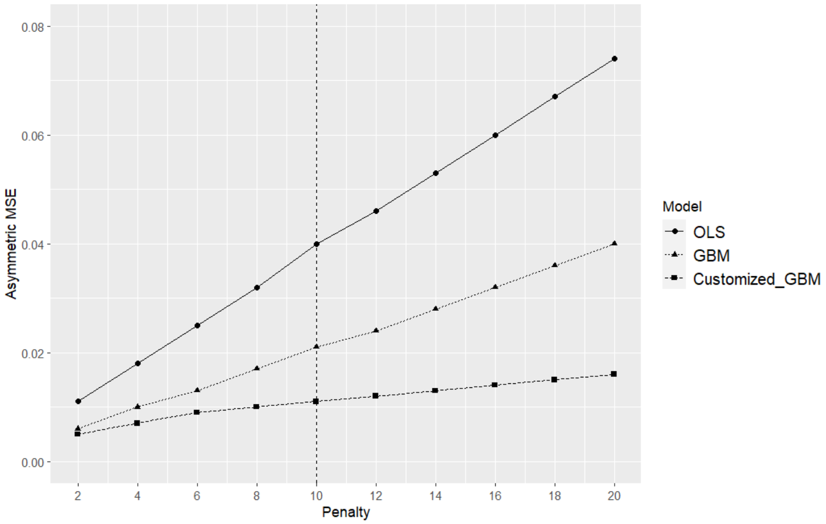
Taieb, S. B., and Hyndman, R. J. (2014). A Gradient Boosting Approach to the Kaggle Load Forecasting Competition. *International Journal of Forecasting*, 30(2), 382-394. http://dx.doi.org/10.1016/j.ijforecast.2013.07.005

Vogels, T., Rousselle, F., McWilliams, B., Röthlin, G., Harvill, A., Adler, D., and Novák, J. (2018). Denoising with kernel prediction and asymmetric loss functions. *ACM Transactions on Graphics,* 37(4), 1-15. https://doi.org/10.1145/3197517.3201388

Wang, Q., Ma, Y., Zhao, K., and Tian, Y. (2020). A Comprehensive Survey of Loss Functions in Machine Learning. *Annals of Data Science,* 1-26.

Zhou, X., Liu, X., Zhai, D., Jiang, J., and Ji, X. (2023). Asymmetric Loss Functions for Noise-Tolerant Learning: Theory and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 45(7), 8094-8109. https://doi.org/10.1109/TPAMI.2023.3236459

Appendix 1. Asymmetric MSE with Varying Penalty Values



Model/Penalty	× 2	× 4	× 6	× 8	× 10	× 12	× 14	× 16	× 18	× 20
OLS	0.011	0.018	0.025	0.032	0.040	0.046	0.053	0.060	0.067	0.074
GBM	0.006	0.010	0.013	0.017	0.021	0.024	0.028	0.032	0.036	0.040
Customized GBM	0.005	0.007	0.009	0.010	0.011	0.012	0.013	0.014	0.015	0.016

The vertical dotted line denotes a ten-times higher penalty for overvaluation. The figure shows that increasing the penalty value leads to performance degradation across all three models, which is anticipated as increasing the penalty value intentionally introduces bias. However, the customized GBM shows relatively less degradation, thus indicating its robustness against higher penalty values.