Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4, 1962-1974
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i4.6431

Machine learning and statistical modeling in firm value prediction

Trang, Do Thi Van[1*], Huyen, Giang Thi Thu[2]
[1,2]Banking Academy of Vietnam; trangdtv@hvnh.edu.vn (T.D.T.V.) huyengtt@hvnh.edu.vn (H.G.T.T.)

Abstract: Determining the best approach model—machine learning or statistical model—to solve the analysis challenges of firm value prediction has garnered attention from many scholars. This article used both traditional statistical models and machine learning to predict the firm value of 435 non-financial companies that are listed on the Vietnam Security Exchange during 2014-2021. Based on the empirical results, the machine learning models provided evidence that they forecast firm value better than traditional models and identify the number of firm value determinants. The paper applied six machine learning models to find the best-performing one, including the multiple regression model (LM), Lasso, generalized additive model (GAM), random forests (RF), gradient boosting regression trees (GBM), and neural networks (NNET). The findings indicated that the RF is the best-performing model, selecting firm size, ROA, tangibility, GDP, quality, financial leverage, and inflation as reliable predictors of market firm value. The results suggest several recommendations for internal managers, investors, and creditors when choosing the appropriate model to forecast firm value for making financial decisions.

Keywords: *Firm Value, Forecasting, Machine learning.*

1. Introduction

In recent years, with the development of technology, machine learning has been used to solve many of the challenges of analytics in many fields. In the financial industry, these techniques are applied to a variety of research in order to obtain the data needed to garner valuable insights, solve the problems of non-linear relationships, and solve complicated issues efficiently without the requirement of data [1]. Particularly, applying machine learning approach seems to be more efficient in forecasting financial issues such as financial distress [2] financial risk [3] firm value [4] etc. Moreover, the increasing availability of data conferred the opportunity to utilize a number of analytical techniques and methodologies to explore patterns, often hidden, that might be useful for financial prediction issues. A variety of machine learning techniques have been applied in quantitative finance applications, such as support vector machines, neural networks [5, 6] hybrid, and ensemble classifiers [7]. Such techniques are based on inductive inference rather than on classical statistics [8]. Although there has been doubt regarding the efficacy of these methods, researchers have demonstrated that one of the primary advantages of machine learning techniques lies in their ability to effectively and accurately analyze vast amounts of data. Due to unclear dependencies within variables, identifying relationship between them is a very difficult task that can be tackled properly by machine learning models.

Among the issues of corporate finance, firm value is always taken into account, not only by internal managers but also by external users. Predictions related to firm value are an important part of information for making decisions for many purposes. For managers, this information guides the internal manager who plans to widen their business scale or intends to list their stock in the foreign security market. Understanding the firm value will help investors identify the direction of their investments and optimize their portfolios. For financial analysts, it is useful to know the determinants that affect firm value so that they can suggest and make commentary for their customers. Firm value prediction has

History: Received: 11 February 2025; Revised: 7 April 2025; Accepted: 11 April 2025; Published: 21 April 2025
* Correspondence: trangdtv@hvnh.edu.vn

attracted consideration from number of scholars of diverse backgrounds, such as Singh and Bansal [9]; Kuzey, et al. [4]; Aggarwal and Padhan [10]; Dang, et al. [11]; Huynh, et al. [12]; Dang and Do [13] and Juca and Fishlow [14] etc. Those studies address how firm characteristics influence firm value; others take into account how macroeconomic factors affect firm value. Many studies have identified and explained the impact of both microeconomic and macroeconomic determinants on firm value. The objective of this study is to address the influence of both microeconomic and macroeconomic factors on the firm value of 435 non-financial enterprises listed on the Vietnam Stock Exchange from 2014 to 2021 based on traditional statistical and machine learning approaches. The comparison between two approaches is conducted in order to find out which approach is more efficient than the others in the firm value prediction problems. In addition, the study concentrates mainly on the machine learning approach by applying six methods to choose the best-performing one for forecasting firm value.

The remainder of this research is organized as follows: The second section represents the literature review and makes the case for the novelty of this study. Section 3 describes the methodology that was used for the analysis. Section 4 details the sample selection and variables. Section 5 shows the empirical findings. Finally, Section 6 concludes the paper and makes recommendations for the findings.

2. Literature Review

2.1. Firm Value: the Traditional Prediction Approach

Firm value has been taken into account not only by investors but also by many scholars all over the world. This issue is given attention in Vietnam by applying different traditional econometric models. Vo and Ellis [15] examined the impact of capital structure on firm value based on the regression model to find the negative impact of capital structure on the firm value of listed companies on Vietnamese Stock Exchanges. The authors emphasized that only low-leveraged firms generate value for shareholders. Huynh, et al. [12] compared the firm value of enterprises and determinants that influence firm value between Vietnam, an emerging country, and England, a developed country, by applying the linear regression model. The findings indicated that the firm value was lower than that of UK firms. The average leverage, revenue, tangible assets, and information asymmetry variables in Vietnamese firms were much higher than those in UK firms. In particular, the information asymmetry variable had a significant negative influence on firm value in Vietnam. Dang and Do [13] conducted the GMM model to forecast the firm value under the impact of capital structure among different industries, such as the food and beverage industry, wholesale trade, construction, and real estate. The research was employed on 435 nonfinancial enterprises from 2012 to 2019 listed on the Vietnamese stock exchange. Vu and Le [16] have studied the impact of tax planning on the firm value of non-financial firms in Vietnam through regression analysis with the GLS model. Most of the studies on firm value in Vietnam have applied regression models with different forms, such as Ha [17]; Nguyen, et al. [18]; Nguyen and Doan [19]; Luu [20] and Nguyen, et al. [21].

2.2. Firm Value: Machine Learning Approach

Morden research methods have been applied in the economic field since the early 1990s by applying the artificial neural network for forecasting purposes [22]. In recent years, machine learning applications have proven useful in predicting financial issues. Firm value prediction through machine learning models has been mentioned in corporate finance and represents more accuracy compared to traditional financial models. Kuzey, et al. [4] have investigated the firm's value through the impact of multinationals as measured by the foreign sales ratio. Machine learning techniques have shown evidence that they are efficient methods to identify the importance of determinants that impact the firm's valuation. The decision tree and neural network algorithms are indicated to help detect the ranked order among the independent variables. van Witteloostuijn and Kolkman [23] relied on machine learning models to predict the value of firm growth by using a dataset of 168,055 enterprises from the Netherlands and Belgium. The authors indicated that the random forest analysis that is developed by machine learning techniques had outperformed the multivariate OLS and other regression models.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

Zhang, et al. [24] predicted the value of firms for the energy industry in China using different machine learning models and made comparisons to consider whether the machine learning models are more accurate than the traditional financial evaluation ones; moreover, among the different machine learning methods, which methods outperformed firm value prediction better than others. Zhang, et al. [25] have applied the valuation model to venture capital investors in the early stages. The research has addressed the impact of either one feature or a few features on firm financing. The empirical results have shown that the machine learning model is more useful for forecasting firm valuations compared to traditional regression models from the perspective of venture capital.

3. Methodology

The basic regression problem in predicting enterprise value (EV) is to estimate the function $g(x_{i,t}) = E(EV_{i,t}|x_{i,t})$, where $EV_{i,t+1} = g(x_{i,t}) + \varepsilon_{i,t+1}$ is the ith enterprise value in year t+1 and $\varepsilon_{i,t+1}$ is a random error component. The regression function $E(EV_{i,t}|x_{i,t})$ is the conditional expectation of $EV_{i,t+1}$ conditioned on the vector of covariates $x_{i,t}$. The vector of covariates $x_{i,t}$ has 7 components, including: (1) $LEV_{i,t}$ (Capital structure), (2) $Quality_{i,t}$ (Quality), (3) $Size_{i,t}$ (Firm size), (4) $ROA_{i,t}$ (Firm profitability), (5) $Grow_{i,t}$ (Firm growth), (6) $Tang_{i,t}$ (Firm tangibility), and (7) $Liquidity_{i,t}$ (Firm liquidity). The objective of this study is to estimate $g(x_{i,t})$ with six methods in two classes, including (1) linear models and (2) machine learning models. The first class of models are variants of a standard linear model.

3.1. Linear Models

3.1.1. Multiple Regression Model (MRL)

Multiple regression models are used to explain the relationship between a dependent variable with more than one independent variable. Multiple regressions can be linear or nonlinear.

In this article, we employed the standard multiple regression model and estimated regression parameters by ordinary least squares (OLS). The multiple regression model has the form

$$g(x_{i,t}, \beta) = x_{i,t}.\beta$$

where $x_{i,t}$ vector of covariates mentioned before and β is the regression parameters. β is estimated by using OLS

$$\hat{\beta}^{OLS} = argmin_{\beta}||EV - x.\beta||_2^2$$

where $||a - b||_2$ is the distance between two vectors a and b (Euclidean distance).

3.1.2. LASSO

There are two drawbacks to using a standard multiple regression model, including (1) sensitivity to multicollinearity and (2) overfitting. To overcome these drawbacks, Lasso (the least absolute shrinkage and selection operator model) was introduced. LASSO [26] selects a reduced set of the known covariates for use in a model. A set of chosen covariates is considered more appropriate to predict the outcome. Thus, LASSO should be able to select the appropriate variables in our regression model, by reducing and selecting a subset of variables that have a significant role in predicting enterprise value.

LASSO does it by using the l1-penalty term ($\lambda||\beta||_1 = \lambda \sum_{i=1}^{n} |\beta_i|$) where λ is a positive value called a tuning parameter. The estimated parameters in the LASSO model are given by

$$\hat{\beta}^{LASSO} = argmin_{\beta}||EV - x.\beta||_2^2 + \lambda||\beta||_1$$

The l1-penalty term plays a decisive role in the model parameter value. If $\lambda = 0$, then $\hat{\beta}^{LASSO} = \hat{\beta}^{OLS}$, and as $\lambda \to \infty$, the penalty term forces the coefficients to zero. Therefore, choosing the tuning parameter λ is an important step for LASSO to be effective.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

3.1.3. Generalized Additive Model (GAM)

Without pre-specified manual intervention, OLS and LASSO models are not well suited to accurately model nonlinear relationships between the covariates and the dependent variable. A class of nonparametric models that can handle potential nonlinearities is known as a generalized additive model, or GAM. The regression function of the GAM model is given by

$$g(x_{i,t}) = \beta_0 + \sum_{j=0}^{n} f_i(x_{ij,t})$$

where f_i is *ith* non-linear function. One of the key advantages of using GAM is that the additive, functional effects of the covariates x_j on y are readily available and are often interpretable.

3.2. Machine Learning Models

Next, we evaluate the utility of three machine learning models: a random forest [27] a gradient boosting machine [28] and a feed-forward neural network with a single hidden layer [29]. Machine learning algorithms are especially helpful for capturing hidden interactions in these covariates and modeling nonlinear relationships between dependent and independent variables.

3.2.1. Random Forests

Random Forest belongs to the ensemble learning method. A random forest algorithm combines multiple decision trees to make more accurate predictions. Decision trees are a fundamental component of random forests. They are hierarchical structures that recursively split the data based on feature values, creating a tree-like flowchart of decision rules. Each node represents a feature, and the branches represent the possible feature values.

The main strength of Random Forest lies in its ability to reduce overfitting and improve generalization performance by combining multiple decision trees. It is robust, less sensitive to noise and outliers, and can handle large datasets with high-dimensional features. Random Forest is widely used in various fields such as finance, healthcare, and natural language processing due to its versatility and effectiveness in handling complex and diverse datasets.

3.2.2. Gradient Boosting Regression Trees

Gradient Boosting Regression Trees (GBRT) is a specific component of Gradient Boosting Machine (GBM). GRBT is a powerful machine learning algorithm belonging to the ensemble learning family. Unlike Random Forest, which builds trees independently, GRBT builds trees sequentially, with each tree correcting the errors made by its predecessor. GBRT plays a vital role within the GBM framework, contributing to the overall performance and accuracy of the model.

The main advantage of GBRT lies in its ability to create a strong predictive model by gradually improving the weaknesses of its base learners. It handles complex relationships between features and target variables, and by adjusting the learning rate, it can control the contribution of each tree to the final prediction, preventing overfitting.

3.2.3. Neural Networks

Our final choice of machine learning algorithms is a single-layer, feed-forward neural network (1-FFNL). A single-layer, feed-forward neural network, also known as a perceptron, consists of a single layer of neurons, where each neuron is connected to the input features directly without any hidden layers.

This simple architecture is suitable for linearly separable problems, such as binary classification tasks. However, its limitations arise when dealing with complex and non-linear data, where deeper networks with multiple layers are needed for better representation and performance.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

3.3. Model Fitting

We use default parameters for each model. We begin by splitting the training set and test set in a 70:30 ratio. Then we fit each candidate model to the entire training set from 2014 to 2021. We use these models to make predictions each year. Last, we evaluate the prediction results by two closely related metrics: mean squared forecast error ($MSFE$) and the out-of-sample R-squared (R^2_{OS}), defined as:

$$MSFE = \frac{1}{N}\sum_{i=1}^{N} \quad (y_i - \hat{y}_i)^2$$

$$R^2_{OS} = 1 - \frac{\sum_{i=1}^{N} \quad (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} \quad (y_i - \underline{y_i})^2}$$

where N is the number of observations, y_i is the *ith* actual response, $\hat{y}_i$ is the *ith* fitted value based on the selected model, and $\underline{y_i}$ is the average of responses.

4. Sample Selection and Variables

4.1. Data

The dataset was obtained from FinnPro, the largest financial database in Vietnam, which provides a variety of financial data, software, and analysis tools to users. Initially, the dataset for this study consisted of more than six hundred listed Vietnamese nonfinancial companies for the period from 2014 to 2021. The sample has just concentrated on the non-financial firms and the firms that have audited their financial statements during the period. For all firms that were missing data, we eliminated them from the sample. Thus, after processing the data, the dataset consists of 435 nonfinancial companies. The number of unique cases/records retrieved from the database was 4932. After that, the data was analyzed for missing values, and 1452 cases had a large number of missing values for critical financial indicators; therefore, they were eliminated. There were also 38 cases with unexplainably large values (identified as outliers), which were also eliminated from the dataset. After the data pre-processing, the final dataset consisted of 3480 cases. The final dataset included proper values for all financial indicators from 2014 to 2021.

4.2. Variables

This research has not used endogenous and exogenous variables in terms of the classical econometric modeling approach but rather employed them in terms of the machine learning approach as endogenous (dependent, target, or predicted) variables and exogenous (independent, explained, or predictor) variables. Nevertheless, the firm value variable was conducted as an endogenous variable, and leverage, quality, size, return on assets, sale growth, tangibility, liquidity, GDP, and inflation were employed as exogenous variables. The determinant that affects the firm value has been taken into account by hundreds of researchers; thus, this article has chosen some of them in order to illustrate the comparison between two approaches: statistical and machine learning. Table 1 summarizes the literature on endogenous exogenous variables that are examined in this model. Table 2 illustrates independent variables and briefly defines all variables that are exerted in this paper. These variables are associated with the firm value that has been examined widely in previous firm value research.

The dependent variable (firm value) is measured by enterprise value, which is equal to market capitalization plus book value of debt minus cash and cash equivalents [9-11, 13].

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

Table 1.
The literature on financial variables included in this study.

Dependent variable: Firm value	**Singh and Bansal [9]; Aggarwal and Padhan [10]; Dang, et al. [11] and Dang and Do [13]**
Independent variable: Leverage	Gill and Obradovich [30]; Vo and Ellis [15]; Juca and Fishlow [14]; Huynh, et al. [12]; Dang and Do [13] and Senan, et al. [31]
Quality	Singh and Bansal [9]; Kuzey, et al. [4]; Aggarwal and Padhan [10]; Juca and Fishlow [14]; Dang, et al. [11] and Dang & Do (2021)
Size	Cheng and Tzeng [32]; Cheng and Tzeng [33]; Singh and Bansal [9]; Kuzey, et al. [4]; Dang and Do [13]; Senan, et al. [31] and Juca and Fishlow [14]
ROA	Kuzey, et al. [4]; Singh and Bansal [9]; Juca and Fishlow [14]; Aggarwal and Padhan [10]; Dang, et al. [11] and Dang and Do [13]
Growth	Kuzey, et al. [4]; Aggarwal and Padhan [10]; Dang, et al. [11]; Huynh, et al. [12]; Dang and Do [13] and Senan, et al. [31]
Tangibility	Kuzey, et al. [4]; Huynh, et al. [12] and Dang and Do [13]
Liquidity	Kuzey, et al. [4]; Aggarwal and Padhan [10]; Dang and Do [13] and Senan, et al. [31]
GDP	Aggarwal and Padhan [10]; Usman, et al. [34]; Dang and Do [13]; Senan, et al. [31] and Faradila and Effendi [35]
Inflation	Cheng and Tzeng [32]; Cheng and Tzeng [33]; Aggarwal and Padhan [10]; Dang and Do [13] and Senan, et al. [31]

Table 2.
The list of variables included in this study

Independent variable: Leverage	**Total debt/total assets**
Quality	EBIT/total assets
Size	LnAssets: Natural logarithm of total assets
ROA	Net income/total assets
Growth	$(Sale_t - Sale_{t-1})/ Sale_{t-1}$
Tangibility	Net long-term assets/total assets
Liquidity	Current assets/current liabilities
GDP	General Statistics Office of Vietnam
Inflation	General Statistics Office of Vietnam

5. Results

5.1. The traditional approach

As can be seen from Table 3, all variables have been described based on the mean, maximum, minimum value, standard deviation, and some tests for the whole sample. The mean value of firm value is 2565.75 which is lower than the firm value of enterprises in England [12] Netherlands, and Belgium [23]. Since the enterprises in Vietnam are mainly medium and small-size ones and the number of listed companies is quite low compare to the developed financial market. The probabilities of Jarque-Bera tests are significant at the 1% level; thus, the distribution of all variables is non-normality. The firm value is an absolute number without modification and has much higher standard deviations compared to other variables. It means that the database also consists of negative figures as the minimum values of some variables are lower than 0, including firm value, quality, firm profitability, and firm growth.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

Table 3.
The statistical description.

	EV	Lev	Quality	Size	ROA	Growth	Tang	Liq	GDP	Inflation
Mean	2565.7500	5.6841	0.0495	6.6914	4.8184	0.1171	3.1942	1.9784	5.7718e-01	4.1187
Max	291754.8500	9.7061	0.5979	11.4855	99.3800	19.3362	9.7485	47.7700	1.0000	9.1000
Min	-1141.2006	1.1055	-0.6485	3.0165	-57.6700	-0.8416	3.2139	0.2600	0.0000	6.3000
Stdev	15391.1015	2.0917	0.0781	1.4541	8.8291	0.6340	2.0614	2.3483	3.6253	2.4421
Skewness	11.8298	-5.2167	-0.8664	0.0793	0.9535	22.9306	9.5170	9.5583	-3.8757	7.8004
Kurtosis	165.4505	-4.2518	16.1231	0.0099	20.3213	679.7822	3.7502	142.1294	-1.3365	-1.4671
Jarque-Bera Prob.	0.0000	0.0000	0.0000	0.0051	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

Furthermore, the F-test has a Sig value of 0.0000 (lower than 0.05), which represents that all the independent variables could be employed for the changes in the dependent variable. All independent variables are accepted in the model.

Table 4.
Regression results

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.	Collinearity statistics	
	B	Std. Error	Beta			Tolerance	VIF
Constant	-20361.986	7536.797		-2.702	0.007		
Leverage	-7243.011	2590.905	-0.098	-2.196	0.005	0.534	1.874
Quality	35646.52	9265.082	0.181	3.847	0.000	0.299	3.350
Size	3546.073	285.021	0.337	12.441	0.000	0.900	1.111
ROA	41.887	87.331	0.024	0.480	0.632	0.264	3.793
Growth	-855.217	635.739	-0.035	-1.345	0.179	0.965	1.037
Tangibility	3089.388	2118.108	0.041	1.459	0.145	0.822	1.216
Liquidity	-194.692	198.560	-0.030	-0.981	0.327	0.721	1.387
GDP	13905.859	99834.829	0.006	0.139	0.899	0.346	2.888
Inflation	-732.341	27559.257	-0.001	-0.027	0.979	0.345	2.902
Observation	1248						
R^2	0.2105						

According to Table 4, the independent variables (sig. > 0.05) that have influenced the firm value of listed enterprises are ROA, growth, tangibility, liquidity, GDP, and inflation. The quality and firm size variables have a positive impact, whereas the capital structure variable has a negative influence. The factor VIF of all variables is less than 10, thus, the model does not have multicollinearity.

5.2. The Machine Learning Approach

The dataset has been trained by six prediction models, which included the multiple regression model (LM), Lasso, generalized additive model (GAM), random forests (RF), gradient boosting regression trees (GBM), and neural networks (NNET). Table 5 highlights the comparison of the independent variables for firm value prediction by using six machine learning methods. The criteria that can be utilized to test the accuracy of the model are R^2_{OS} and RMSE. For the R^2_{OS}, the higher the R^2_{OS} {0, 1}, the greater the accuracy of the model. The RMSE value indicates the average difference (error) value on the dataset after the model is trained; hence, the lower values indicate more accurate predictions of the model. The smaller the value of RMSE, the larger the value of R^2_{OS} will be, and vice versa. Table 5 shows the results of six prediction models using two indicators. The RF model with R^2_{OS} = 0.8599 (the highest one) and RMSE = 0.000509 (the lowest value) will be the most accurate prediction model. Compared to the R^2 value of 0.2105 or the traditional model, the machine learning models demonstrate a significantly higher level of accuracy than traditional statistical models.

Table 5.
The results of machine learning models.

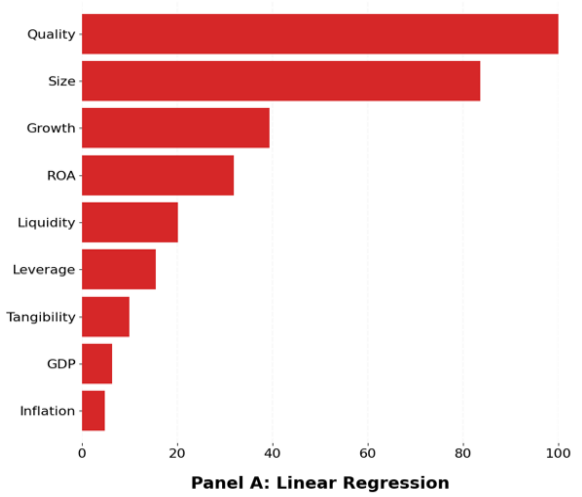
Model	R^2_{OS}	RMSE
LM	0.2204	0.002831
LASSO	0.2069	0.002879
GAM	0.839	0.000585
RF	0.8599	0.000509
GBM	0.7621	0.000864
NN	0.7531	0.000896

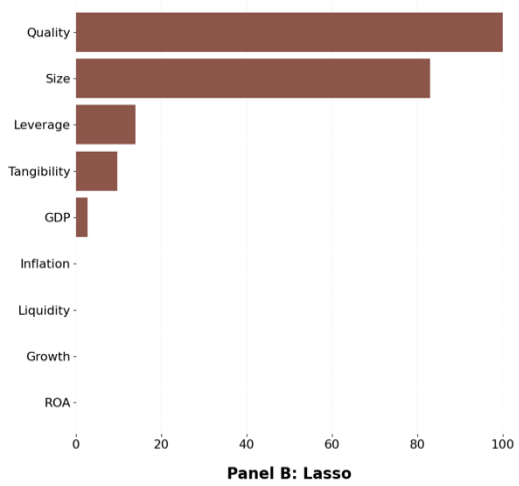
Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

For each model, there will be a set of weights according to the sample properties. Based on the properties of the model, each attribute will be assigned one or more weights (for example, in the neural network model, each attribute is assigned many weights because of the connection to intermediate nodes). This coefficient will reflect the influence (importance) of the attribute on the output of the model. In this paper, the importance is determined by summing the weights that are assigned to the attributes. Then, the results are standardized by the scale [0, 100], with 100 being the largest value among the attributes. Figure 1 has illustrated the importance of variables in the different machine learning models. The importance of variables has been represented as decreasing levels from the top to the bottom. According to two criteria, the RF model is the most accurate one compare to others. The results indicate that the size variable has the most influence on the firm value, and the other variables are ROA, tangibility, GDP, quality, leverage, and inflation.

For the model with the highest accuracy, the Random Forest (Random Forest), the variables with the greatest impact on firm value in descending order of importance are SIZE, ROA, TANGIBILITY, GDP, QUALITY, GROWTH, LIQUIDITY, LEVERAVAGE, and INFLATION. However, the importance of variable SIZE is distinctly different from the other variables. In the Random Forest model, variable importance is measured using the mean and standard deviation of the accumulation of impurity decrease within each decision tree. This shows that SIZE plays crucial roles in determining firm value, and the RF model uses these measures to optimize model performance in predicting firm value. However, overemphasizing a few variables may lead to a lack of generalization in the model, resulting in poorer predictive performance when applied to new or different cases.

As can be seen from Figure 1, the firm size variable is one of the most important factors across all models. Besides, quality variable has a high weight in LR, Lasso, GAM, GBM, and NN models. The remaining variables have different importance depending on the results of each model. The key difference between these models is that RF and GBM tend to drop off more rapidly. These two models just have one variable, whose weight is higher by 20%. The equitability of weights is present most clearly in NN. The NN importance of variables plot shows that the firm size variable is the most important, although ROA, quality, inflation, tangibility, financial leverage, liquidity, growth, and GDP variables have at least 50% weights, as well. Hence, NN more equally assigns variable importance than the remaining models.





Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

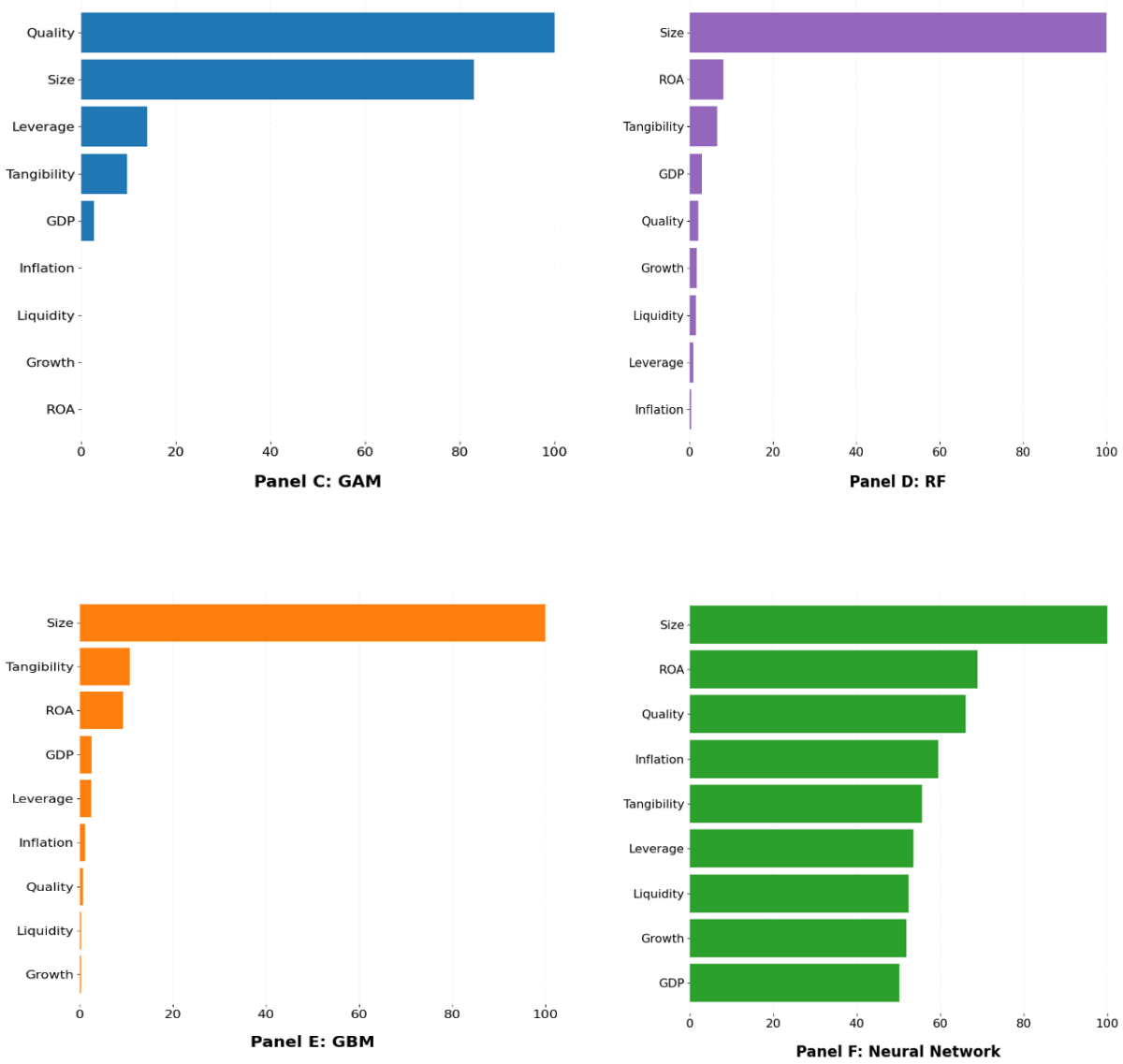


Figure 1.
The importance of variables.

Figure 2 illustrates the information fusion based sensitivity analysis result for firm value from 2014 to 2021. The result has shown the impact of all dependent variables of firm value, such as financial leverage ratio, quality, firm size, ROA, tangibility, liquidity, GDP, and inflation. This result is consistent with the result from Kuzey, et al. [4].

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

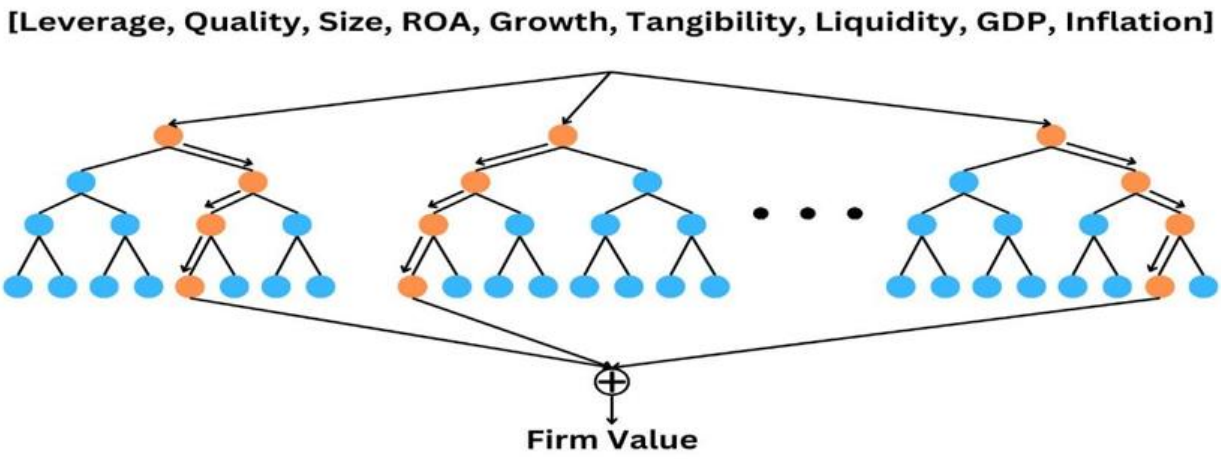


Figure 2.
Illustration of random forests model used in this study.

6. Conclusions

Forecasting the firm's value has drawn interest from internal managers, investors, and other stakeholders. This paper addresses another methodology that might be useful for firm value prediction besides the traditional quantitative methodology. Although statistical models have been widely applied in much previous research, their limitations are related to strict data conditions. The development of computational science has induced machine learning models to develop, and they have been proven to be more efficient with higher accuracy and lower errors than traditional methods [1, 36]. The model is learned from the data, does not require much data structure, and can be applied flexibly.

This paper makes a comparison between traditional and machine learning methodologies to forecast the firm value of 435 non-financial listed enterprises on the Vietnamese Securities Exchange from 2014 to 2021. In addition, among the machine learning methods, this paper has compared six methods that can be used for predicting firm value. The prediction results are in accordance with previous studies showing that machine learning is superior to traditional methods, and the RF model is the best-performing compared to others. Overall, the usage of machine learning models has demonstrated better predictive capabilities. Thus, this result is meaningful for internal managers, shareholders, creditors, and potential investors in forecasting the firm value and helping all stakeholders make appropriate financial decisions. However, machine learning models still have some notable drawbacks, including: (1) difficulty in interpretation and (2) the potential for overfitting, which can lead to reduced model performance when applied to new or unseen data.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgement:

The author gratefully acknowledges the financial support from the Banking Academy of Vietnam.

References

[1] M. Bennett, K. Hayes, E. J. Kleczyk, and R. Mehta, "Similarities and differences between machine learning and traditional advanced statistical modeling in healthcare analytics," *arXiv preprint arXiv:2201.02469*, 2022. https://doi.org/10.48550/arXiv.2201.02469

[2] Y.-P. Huang and M.-F. Yen, "A new perspective of performance comparison among machine learning algorithms for financial distress prediction," *Applied Soft Computing*, vol. 83, p. 105663, 2019. https://doi.org/10.1016/j.asoc.2019.105663

[3] A. Mashrur, W. Luo, N. A. Zaidi, and A. Robles-Kelly, "Machine learning for financial risk management: A survey," *Ieee Access*, vol. 8, pp. 203203-203223, 2020. https://doi.org/10.1109/ACCESS.2020.3036571

[4] C. Kuzey, A. Uyar, and D. Delen, "The impact of multinationality on firm value: A comparative analysis of machine learning techniques," *Decision Support Systems*, vol. 59, pp. 127-142, 2014. https://doi.org/10.1016/j.dss.2013.11.009

[5] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Literature review: Machine learning techniques applied to financial market prediction," *Expert Systems with Applications*, vol. 124, pp. 226-251, 2019. https://doi.org/10.1016/j.eswa.2019.01.033

[6] F. Rundo, F. Trenta, A. L. Di Stallo, and S. Battiato, "Machine learning for quantitative finance applications: A survey," *Applied Sciences*, vol. 9, no. 24, p. 5574, 2019. https://doi.org/10.3390/app9245574

[7] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, "Machine learning in financial crisis prediction: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 421-436, 2011. https://doi.org/10.1109/TSMCC.2011.2124876

[8] V. Vapnik, *The nature of statistical learning theory*. New York: Springer Science & Business Media, 2013.

[9] A. K. Singh and P. Bansal, "Impact of financial leverage on firm's performance and valuation: A panel data analysis," *Indian Journal of Accounting*, vol. 48, no. 2, pp. 73-80, 2016.

[10] D. Aggarwal and P. C. Padhan, "Impact of capital structure on firm value: Evidence from Indian hospitality industry," *Theoretical Economics Letters*, vol. 7, no. 4, pp. 982-1000, 2017. https://doi.org/10.4236/tel.2017.75070

[11] H. N. Dang, V. T. T. Vu, X. T. Ngo, and H. T. V. Hoang, "Study the impact of growth, firm size, capital structure, and profitability on enterprise value: Evidence of enterprises in Vietnam," *Journal of Corporate Accounting & Finance*, vol. 30, no. 1, pp. 144-160, 2019. https://doi.org/10.1002/jcaf.22389

[12] T. L. D. Huynh, J. Wu, and A. T. Duong, "Information Asymmetry and firm value: Is Vietnam different?," *The Journal of Economic Asymmetries*, vol. 21, p. e00147, 2020. https://doi.org/10.1016/j.jeca.2020.e00147

[13] T. D. Dang and T. V. T. Do, "Does capital structure affect firm value in Vietnam?," *Investment Management & Financial Innovations*, vol. 18, no. 1, p. 33, 2021. https://doi.org/10.21511/imfi.18(1).2021.04

[14] M. Juca and A. Fishlow, "The impact of social capital on firm value," *Contemporary Economics*, vol. 16, no. 2, pp. 182-194, 2022. https://doi.org/10.5709/ce.1897-9254.541

[15] X. V. Vo and C. Ellis, "An empirical investigation of capital structure and firm value in Vietnam," *Finance Research Letters*, vol. 22, pp. 90-94, 2017. https://doi.org/10.1016/j.frl.2017.05.009

[16] T. A. T. Vu and V. H. Le, "The effect of tax planning on firm value: A case study in Vietnam," *The Journal of Asian Finance, Economics and Business*, vol. 8, no. 2, pp. 973-979, 2021. https://doi.org/10.13106/jafeb.2021.vol8.no2.973

[17] P. Ha, "Cash holding, state ownership and firm value: The case of Vietnam," *International Journal of Economics and Financial Issues*, vol. 6, no. 6, pp. 110-114, 2016. https://doi.org/10.2139/ssrn.2842871

[18] D. P. Nguyen, X. V. Vo, T. T. A. Tran, and T. K. T. Tu, "Government cost and firm value: Evidence from Vietnam," *Research in International Business and Finance*, vol. 46, pp. 55-64, 2018. https://doi.org/10.1016/j.ribaf.2018.01.005

[19] A. H. Nguyen and D. T. Doan, "The impact of intellectual capital on firm value: Empirical evidence from Vietnam," *International Journal of Financial Research*, vol. 11, no. 4, pp. 74-85, 2020. https://doi.org/10.5430/ijfr.v11n4p74

[20] D. H. Luu, "The impact of capital structure on firm value: A case study in Vietnam," *The Journal of Asian Finance, Economics and Business*, vol. 8, no. 5, pp. 287-292, 2021. https://doi.org/10.13106/jafeb.2021.vol8.no5.287

[21] L. Nguyen, T. K. P. Tan, and T. H. Nguyen, "Determinants of firm value: An empirical study of listed trading companies in Vietnam," *The Journal of Asian Finance, Economics and Business*, vol. 8, no. 6, pp. 809-817, 2021. https://doi.org/10.13106/jafeb.2021.vol8.no6.809

[22] C. Serrano-Cinca, "Self organizing neural networks for financial diagnosis," *Decision Support Systems*, vol. 17, no. 3, pp. 227-238, 1996. https://doi.org/10.1016/0167-9236(96)00017-2

[23] A. van Witteloostuijn and D. Kolkman, "Is firm growth random? A machine learning perspective," *Journal of Business Venturing Insights*, vol. 11, p. e00107, 2019. https://doi.org/10.1016/j.jbvi.2019.e00107

[24] C. Zhang, H. Zhang, and D. Liu, "A contrastive study of machine learning on energy firm value prediction," *IEEE Access*, vol. 8, pp. 11635-11643, 2019. https://doi.org/10.1109/ACCESS.2020.2973743

[25] R. Zhang, Z. Tian, K. J. McCarthy, X. Wang, and K. Zhang, "Application of machine learning techniques to predict entrepreneurial firm valuation," *Journal of Forecasting*, vol. 42, no. 2, pp. 402-417, 2023. https://doi.org/10.1002/for.2774

[26] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267-288, 1996. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431

[27] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001. https://doi.org/10.1023/A:1010933404324
[28] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, pp. 1189-1232, 2001. https://doi.org/10.1214/aos/1013203451
[29] W. N. Venables and B. D. Ripley, *Modern applied statistics with S-PLUS*. New York: Springer Science & Business Media, 2013.
[30] A. Gill and J. Obradovich, "The impact of corporate governance and financial leverage on the value of American firms," *International Research Journal of Finance and Economics*, vol. 91, no. 2, pp. 46-56, 2012.
[31] N. A. M. Senan, M. A. S. Al-Faryan, S. Anagreh, E. A. Al-Homaidi, and M. I. Tabash, "Impact of working capital management on firm value: an empirical examination of firms listed on the Bombay Stock Exchange in India," *International Journal of Managerial and Financial Accounting*, vol. 14, no. 2, pp. 138-156, 2022. https://doi.org/10.1504/IJMFA.2022.119345
[32] M.-C. Cheng and Z.-C. Tzeng, "The effect of leverage on firm value and how the firm financial quality influence on this effect," *World Journal of Management*, vol. 3, no. 2, pp. 30-53, 2011.
[33] M.-C. Cheng and Z.-C. Tzeng, "Effect of leverage on firm market value and how contextual variables influence this relationship," *Review of Pacific Basin Financial Markets and Policies*, vol. 17, no. 01, p. 1450004, 2014. https://doi.org/10.1142/S0219091514400010
[34] M. Usman, M. Shaique, S. Khan, R. Shaikh, and N. Baig, "Impact of R&D investment on firm performance and firm value: Evidence from developed nations (G-7)," *Revista de Gestão, Finanças e Contabilidade*, vol. 7, no. 2, pp. 302-321, 2017. https://doi.org/10.18028/2237-5497/rgfc.v7i2.191
[35] S. Faradila and K. A. Effendi, "Analysis of financial performance and macroeconomic on firm value," *Jurnal Manajemen*, vol. 27, no. 2, pp. 276-296, 2023. https://doi.org/10.24912/jm.v27i2.13776
[36] B. Ratner, *Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data*. Boca Raton, FL: Chapman and Hall/CRC, 2017.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1962-1974, 2025
DOI: 10.55214/25768484.v9i4.6431