Asian Journal of Civil Engineering (2024) 25:4269–4279
https://doi.org/10.1007/s42107-024-01046-4

RESEARCH

Research on applying machine learning models to predict and assess return on assets (ROA)

Pham Vu Hong Son[1,2] · **Le Tung Duong**[1,2]

Received: 19 March 2024 / Accepted: 27 March 2024 / Published online: 27 April 2024

Abstract

Return on Assets (ROA), a profitability measure, is crucial in corporate finance for assessing how efficiently a company uses assets to generate profit. Currently, the prediction of the ROA index at present is a tedious, manual process. It usually involves making educated guesses or waiting for the accurate data, which becomes available only after financial reports have been compiled. This paper introduces a machine learning model for predicting the ROA index. The model draws data from 78 companies listed on the Vietnam Stock Exchanges (HOSE and HNX) over the span of 2012 to 2022.The random forest (RF) model was put to the test using datasets from selected Vietnamese businesses in 2023. The results demonstrated a high level of precision, with an error rate of less than 1%, an R^2 value of 0.9762, and a root mean square error (RMSE) of 0.5826. These findings indicate potential real-world uses in predicting and boosting business performance. In conclusion, the integration of machine learning in financial analysis and prediction represents substantial progress. It enhances both accuracy and efficiency and holds promise for future advancements in financial management practices. This study aims to encourage more research and development in this area, leading to more advanced and efficient financial management tools.

Keywords Profit · Working capital · Debt ratio · Growth rate · Vietnamese construction enterprises · Machine learning models · Optimization

Introduction

In recent years, the application of information technology in the construction industry has piqued the interest of scientists. Numerous studies, both domestic and international, have explored the use of artificial intelligence (AI) in various types of projects. These include civil projects (Nguyen Dang et al., 2024; Son et al., 2023, 2024a, 2024b; Son & Pham, 2024b), transport (Son & Pham, 2023a) and electricity (Son & Pham, 2023b).

In the evolving field of construction management, the application of machine learning (ML) techniques has marked a significant stride toward enhancing the predictability and understanding of profitability within the industry. Recent studies highlight the burgeoning role of ML in forecasting financial outcomes, underscoring a technological shift that promises to refine strategic planning and decision-making processes.

Adinyira et al. (2021) pioneered the use of a support vector regression algorithm (SVRA) to predict construction project profit margins in Ghana, showcasing a commendable predictive accuracy of 73.66%. This study not only demonstrates the applicability of ML in construction profitability forecasts but also sets a benchmark for future research in similar emerging markets. Following a similar trajectory, Zhang et al. (2015) employed principal component analysis (PCA) and a support vector machine (SVM) to navigate the complex financial landscapes of Chinese construction firms, achieving an impressive accuracy rate exceeding 80%. These findings illuminate the potential of ML algorithms to dissect and interpret multifaceted financial data effectively. Adding to the body of knowledge, Mahfouz (2012) introduced a decision support system designed to estimate productivity rates in construction projects through the integration

✉ Le Tung Duong
jarvisle11@gmail.com

Pham Vu Hong Son
pvhson@hcmut.edu.vn

[1] Faculty of Civil Engineering, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet, Ward 14, District 10, Ho Chi Minh City, Vietnam

[2] Vietnam National University Ho Chi Minh City (VNUHCM), Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam

of SVM and Naive Bayes models. This approach not only underscores the versatility of ML in handling diverse construction management challenges but also reinforces the importance of predictive accuracy in enhancing project outcomes.

In a study focusing on the Vietnamese context, Le et al. (2020) meticulously identified a set of critical factors influencing the profitability of local construction companies. Variables such as company age, debt ratio, growth rate, asset utilization performance, company size, and the proportion of fixed assets were highlighted as determinants of financial success. This research enriches the discourse on profitability drivers in construction, offering valuable insights into the Vietnamese market's unique characteristics.

Further more, Wassie (2020) investigates the impact of capital structure on the profitability of construction firms in Ethiopia, revealing that both debt-to-equity and long-term debt-to-total assets ratios exhibit a significant positive correlation with ROE and ROA. This finding aligns with the broader discourse on the pivotal role of capital structure decisions in influencing company worth and operational costs, thereby echoing the essentiality of strategic financial planning in the construction domain. Moreover, a conceptual review by Ngo and Ngoc (2023) on working capital management practices in the construction industry highlights the intricate relationship between working capital components and firm profitability, further substantiating the critical role of financial management in the sector's sustainability. This aligns with Soa La Nguyen et al. (2023) exploration into the associations between ROE, ROA, liquidity, and debt, emphasizing the nuanced impact of short-term loans on firm profitability and underscoring the potential of ML to refine these predictive analyses.

Collectively, these studies underscore the transformative potential of ML in the construction industry, presenting a promising avenue for future research and application. By leveraging advanced analytical techniques, the construction sector can achieve greater accuracy in profitability predictions, enabling more informed strategic decisions that drive sustainable growth and financial stability.

This study introduces a suite of machine learning (ML) models, chosen based on criteria such as data characteristics, computational efficiency, and the specific objectives of the analysis. The investigation encompasses a variety of algorithms, including supervised regression techniques like lasso, ridge regression (RR), k-nearest neighbors regressor (KNR), and support vector regression (SVR), as well as ensemble methods such as random forest (RF), extra trees regressor (ETR), gradient boosting regressor (GBR), and extreme gradient boosting (XGBoost). Additionally, the study evaluates an artificial neural network approach through the multilayer perceptron (MLP) model. The effectiveness of these models will be quantitatively measured using R square (R^2) and root mean square error (RMSE) metrics. The forthcoming section will detail the ML methodologies under review, the research approach employed, and a discussion on the findings derived from the analysis.

Research methodology

A range of studies have demonstrated the effectiveness of the random forest algorithm in various financial applications. Cai et al. (2020) and Wu and Chen (2012) both found that the algorithm outperformed other methods in predicting enterprise return on net assets and diagnosing assets impairment, respectively. Zhu et al. (2019) applied the algorithm to forecast fund return rate direction and select stocks, with both studies reporting positive results. Sevil and Güven (2020) and McGroarty et al. (2014) further extended the algorithm's application to predicting IPO initial returns and developing an automated trading system, respectively, with both studies finding the algorithm to be superior to other methods. Scornet and Erwan (2016) and Breiman (1999) provided comprehensive overviews of the algorithm, highlighting its versatility and robustness.

Return on assets (ROA) forecasting model

Predicting return on assets (ROA) enables businesses to develop risk management strategies and adjust their financial posture more effectively by forecasting future outcomes. This aids businesses in:

- Making critical strategic decisions for the company's future, guiding the business towards sustainable and financially secure growth.
- Planning and adjusting financial ratios (such as leverage ratio, working capital, etc.) to optimize profits.

The aim of this study is to contribute to financial management practices within the construction industry, while also providing a trained dataset for predicting and assessing the ROA for the construction sector in Vietnam.

Linear regression analysis model

The single linear regression analysis model, a statistical method, is employed to examine the correlation between one or more independent variables, also known as feature variables, and a dependent variable, or target variable. The goal is to anticipate the value of the dependent variable based on the independent one by identifying the optimal straight line to reduce the discrepancy between the projected and actual values of the dependent variable. Khoi and Pointer (2019) identified internal variables such as firm size, return

on equity, and earnings per share as significant predictors of ROA in the Vietnamese stock market by using a basic regression model. Moverover, the predictability of real estate asset returns using a vector regression model which incorporates financial spreads was found to be reduced over longer forecasting horizons (Tsolacos & Sotiris, 2010). The linear regression models used in our study encompass lasso regression, ridge regression, k neighbors regression, and support vector regression (SVR).

Ensemble model

The ensemble model, first introduced in the early 1990s Hansen and Salamon (1990), saw significant development in the early twenty-first century (Friedman, 2001). This model is a departure from the independent linear regression analysis model. Instead, it is a machine learning approach that combines the predictions of multiple individual models. The aim of the ensemble model is to enhance generalization and accuracy by combining the predictions of base estimators built with a certain learning algorithm (Fig. 1).

The two main methods, averaging and augmentation, are typically differentiated as follows:

Averaging involves independently constructing multiple estimators and then averaging their predictions. Generally, the combined estimator in averaging tends to outperform any single base estimator as it reduces variance. Models like the RF and GBR operate on this principle.

On the other hand, with augmentation, base estimators are built sequentially with an aim to decrease the bias of the combined estimator. The objective here is to amalgamate several weak models to produce a potent ensemble. Models such as GBR and XGBoost fall under this category.

Artificial neural network model

Artificial neural networks (ANNs) are computational models inspired by biological neural networks. Their origins date back to the 1940s and 1950s when Warren McCulloch and Walter Pitts first conceptualized the neuron model (McCulloch & Pitts, 1943).This work laid the ground work for initial artificial neural models. The field progressed significantly in the 1990s with the introduction of advanced models like the convolutional neural network (CNN) and the counter propagation neural net (CPN) (Kaveh & Iranmanesh, 1998). These models have been applied to various tasks, including speech recognition, natural language processing, and image and video analysis. Today, ANNs are prevalent in both industry and research, with the field evolving rapidly (Kaveh & Khavaninzadeh, 2023) and (Kayakus et al. 2023) found that artificial neural networks (ANNs) and support vector regression (SVR) were successful in predicting ROA in the iron and steel industry.

The multilayer perceptron (MLP), a type of ANN with multiple hidden layers, is particularly useful for regression problems. This model evolved from the perceptron model of the 1950s and 1960s. The 1980s marked the emergence of multilayer neural network architectures used for tackling complex problems. Today, MLP is still a commonly used machine learning model.

Model evaluation indicators

RMSE, or root mean square error, is used to measure the average magnitude of the error, also known as the residual, between the predicted value and the actual value. The smaller the RMSE value, the smaller the error, indicating a high level of estimation. This suggests that the model is reliable. Formula refer to: (Hodson, 2022)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_i^n \left(\hat{y}_i - y_i\right)^2}{n}} \tag{1}$$

R^2, or the coefficient of determination, measures the proportion of the target data's variance that our regression model can explain. An R^2 value close to 1 is typically considered to indicate good predictive ability. Formula refer to: Chicco et al. (2021)

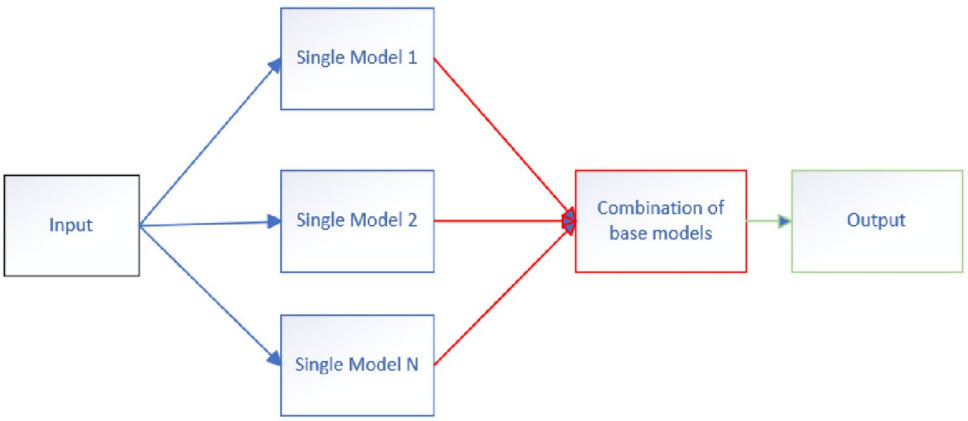

Fig. 1 Basic structure of combined model Source: Rincy and Gupta (2020)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (2)$$

In these formulas, y_i is the actual value of observation i, $\hat{y_i}$ is the predicted value of observation i, $\bar{y}$ is the average value of y, and n is the number of sample data.

Evaluation of machine learning models

Data set overview

Data collection

In accordance with the analysis requirements and the model under construction, we'll select the necessary data. Identifying the right data source, its format, and the collection method forms a crucial part of the process. The types of data we collect will be dictated by the purpose of our research and the results we're aiming for. Here are the specifics:

– We'll gather data on the company's business results from financial reports and balance sheets (revenue, cost of goods sold, financial costs, sales costs, profit, interest,…) available at https://finance.vietstock.vn/.
– We'll compute financial variables using specific formulas.

Data processing

(a) Missing values:

To manage missing values while preserving the data structure, two methods are proposed:

(i) Replace the missing value with the column's mean or median. This technique is generally applied when the number of missing values is insignificant.
(ii) Exclude the row containing the missing value from the data series.

(b) Data analysis:

This study applies the cross-validation technique for assessing the machine learning model. We've used five folds for optimizing resources. When compared with the conventional technique (which splits data into training and testing parts), here's what we find:

– Data use efficiency

- Cross-validation technique: each dataset sample becomes a test set once and a training set (k – 1) times—maximizing data usage, which is crucial for small datasets.
- Traditional split: the data is divided in a fixed manner, typically 70% for training and 30% for testing. Some data is only used either for training or testing, which doesn't make the best use of all data.

– Bias and variance reduction

- Cross-validation technique: since each sample is used for both training and testing, this method helps curb bias and variance for a more accurate model performance estimate.
- Traditional split: there could be high bias or variance if the dataset split doesn't accurately represent the dataset's full structure or distribution, particularly if the dataset is small or lacks diversity.

– Flexibility and general applicability

- Cross-validation technique: offers a more flexible and general model evaluation method, good for various data types and different problems. It also allows fold number adjustment to balance computational efficiency and model performance estimate accuracy.
- Traditional split: suitable when there's plenty of data and a quick model evaluation is needed. However, this method may not accurately reflect the model's applicability to new data.

– Computation and time

- Cross-validation technique: needs more computational resources and time as the model is trained k times. But, it's considered a fair trade-off for a more accurate model performance estimate.
- Traditional split: requires fewer resources and less time, making it a good option when computational resources are limited or quick results are needed. But, this may compromise the accuracy of the model performance estimate.

Data description

Information about the financial database

The forecast model is built on data collected from the financial reports of 76 construction firms listed on the Vietnam Stock Exchange (HOSE, HNX) spanning from 2012 to 2022.

You'll find the data, which delves into comprehensive business outcomes and balance sheets.

Description of financial database

The financial results of companies play a significant role in determining financial indicators. Summarized financial report information can be accessed from the website: https://finance.vietstock.vn/. This website, established on 02/08/2002, aims to provide accurate data about corporate finance, stocks, bonds, and macro information. The data from this website is used in this article for calculations.

Variables in the predictive model

The variables considered to be included in the model include: 14 dependence variables (features) and 01 independence variable (target) (Table 1).

The characteristics of the dataset are illustrated by a graph (Fig. 2) which shows the frequency of values.

Results and discussion

In the process of preparing data for machine learning algorithms, an essential step involves exploring, visualizing, and preprocessing the available features. This step is crucial as it gives a clear overview of the data that will be fed into the machine learning model, helping to identify any areas that may need further refinement before proceeding.

The dataset utilised in this process consists of a substantial 740 rows and 15 columns. Such a volume of data provides a rich resource for the machine learning algorithms to draw from, thus increasing the potential accuracy of the forecasts produced.

In order to examine the details of the available columns in the dataset, the pandas data analysis library is employed. This powerful tool provides an efficient and effective means of handling and exploring the data in detail, allowing for a thorough examination of the dataset in its entirety.

This article goes a step further and delves into the relationships between the data by studying the correlation between different variables. This is done in order to select the most suitable data for the forecasting model. In this instance, the Pearson correlation is considered a useful tool for identifying quantities that have a high correlation with the ROA value. The results of this analysis can be seen in Fig. 3. This step is crucial as it ensures that only the most relevant and impactful data is included in the forecasting model, thus increasing the probability of accurate and meaningful results.

The correlation matrix reveals that:

Variables DSO, DIO, LEV, GRO, ROE, EBIT, GROS, and RE have a strong correlation with ROA.

Variables DPO, CCC, NWC, SIZ, CR, and QR have a weak correlation with ROA.

Even with these correlations, all variables are included in the machine learning model to guarantee accurate predictions, including non-linear ones.

The results of models predicting return on assets (ROA) are given in Table 2. Among these, the RF model yields the best results when cross-validated with five iterations (n_splits).

The results of the models applied to predict the return on assets are presented in (Table 2).

Table 1 Statistics of variables included in the model Source: Synthesized by the authors Ngo and Ngoc (2023)

No.	Name/variable name	Unit	Measurement
1	Return on asset/ROA	%	Net income to average total asset
2	Days sales outstanding/DSO	Days	Average account receivables balance to net sales) × 365
3	Days inventory outstanding/DIO	Days	Average account inventory balance costs of goods sold) × 365
4	Days payable outstanding/DPO	Days	Average account payables balance costs of goods sold) × 365
5	Cash conversion cycle/CCC	Days	DSO + DIO − DPO
6	Net working capital/NWC	Billions of VND	Current assets − current liabilities
7	Size/SIZ	Billions of VND	Natural logarithm of net sales
8	Financial leverage/LEV	%	Total debt/total asset
9	Current ratio/CR	%	Current assets/current debt
10	Growth rate/GRO	%	Percentage change in net sales
11	Return on equity/ROE	%	Net income/shareholder's equity
12	Earnings before interest taxes/EBIT	Billions of VND	Revenue − operating expenses (excluding interest and taxes)
13	Gross margin/GROS	%	(Total revenue − cost of goods sold)/total revenue
14	Quick ratio/QR	%	(Current assets − inventories)/current liabilities
15	Return/RE	Billions of VND	Net profit margin

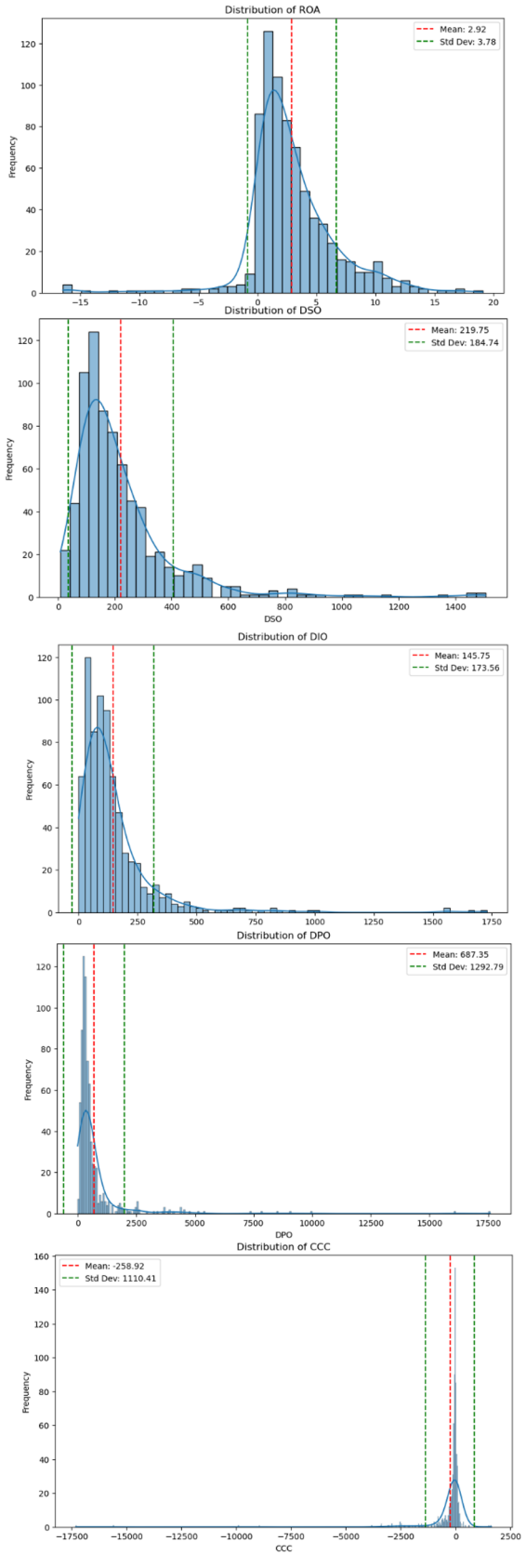

Fig. 2 Descriptive chart of data distribution

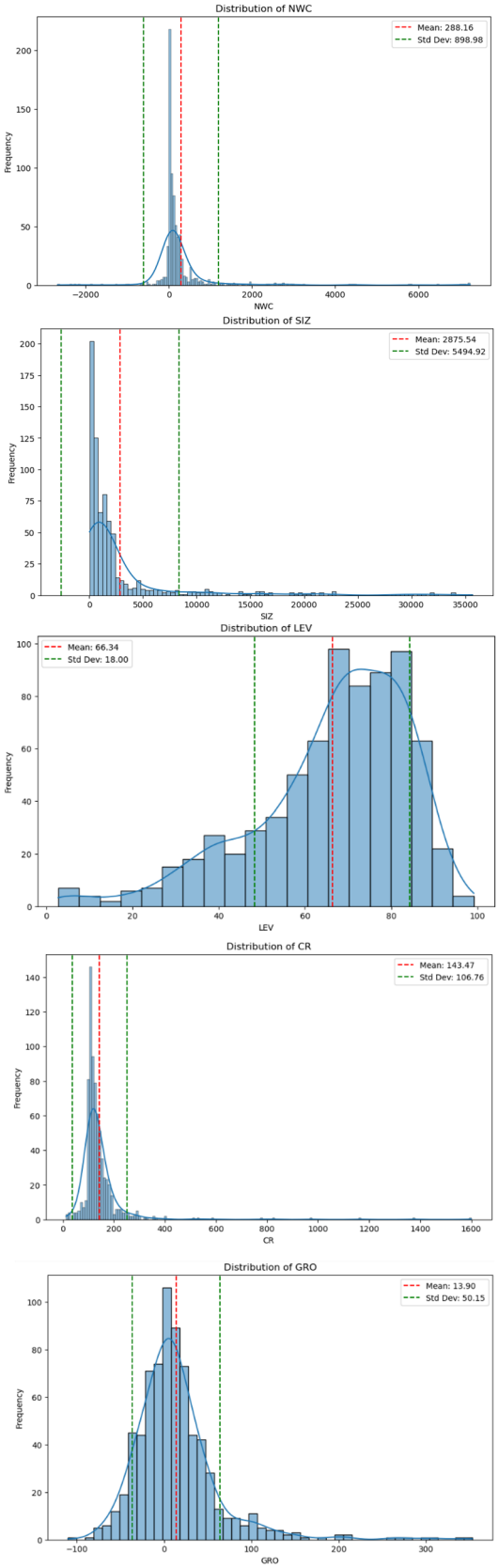

Fig. 2 (continued)

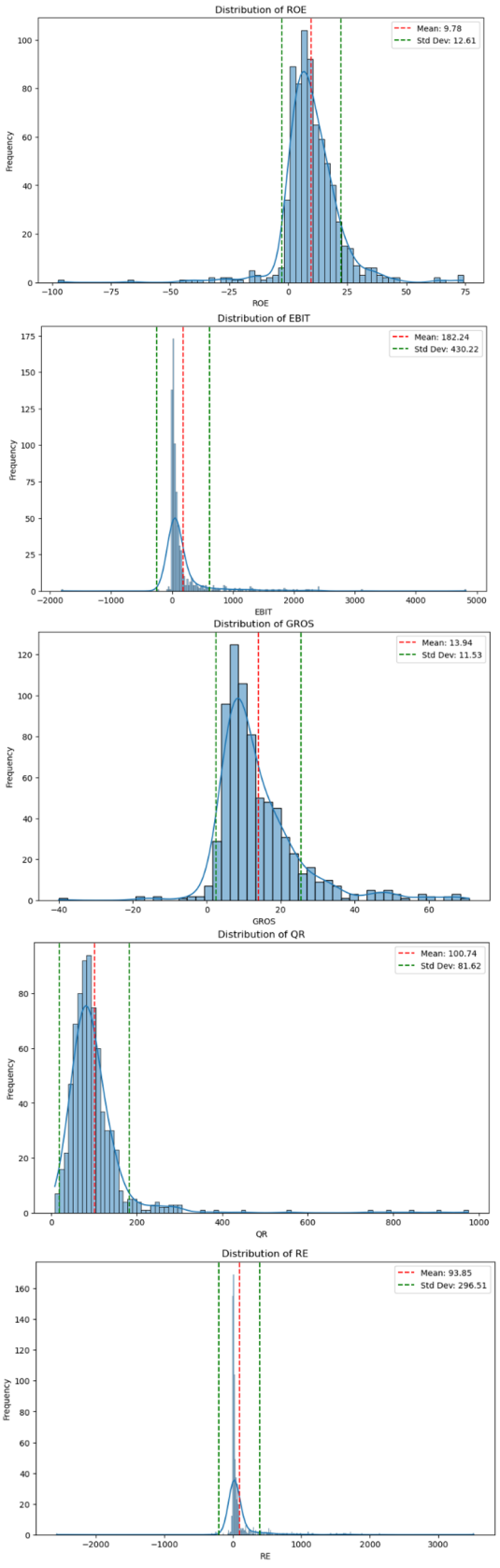

Fig. 2 (continued)

The results of ranking models R^2 and RMSE indexes are shown in Fig. 4.

The model will be evaluated and selected based on three criteria:

– R score: An optimal score is closer to 1.
– RMSE score: A lower score is better.
– Learning curves chart: This chart displays the number of training and test points.

The closer these two lines are, the better the model generalizes.

In deciding between the RF and XGBoost models, both of which have commendable R^2 and RMSE scores, we observe a larger gap between the training and testing lines in XGBoost (1 unit) compared to RF (approximately 0.7 units). As such, this document will use the RF model for computation.

Calibration of RF model

The GridSearchCV tool gives you the power to discover the optimal parameters for your machine learning model. It achieves this by testing a variety of parameters and determining the model's performance for each set, using cross-validation techniques to evaluate the model across diverse training and validation datasets.

In simple terms, GridSearchCV breaks down the parameters into separate values and creates various parameter sets by combining these values. The model is then trained using each parameter set, with its performance evaluated through methods like cross-validation. The best parameter set is then selected based on the model's performance on the validation dataset.

- max_depth = 10.
- min_samples_leaf = 2.
- min_samples_split = 5.
- n_estimators = 200.

Looking at the ranking of important features, it's evident that the ROE index has the most significant impact on the ROA. Additionally, the variables LEV, EBIT, and CR also have a substantial influence on the ROA index (Fig. 5).

From the learning curves (refer to Fig. 6), we can make the following observations:

The training score and cross-validation score remain relatively stable. As more samples are used for training, both scores tend to converge or nearly converge. This suggests that the model has strong generalization capabilities.

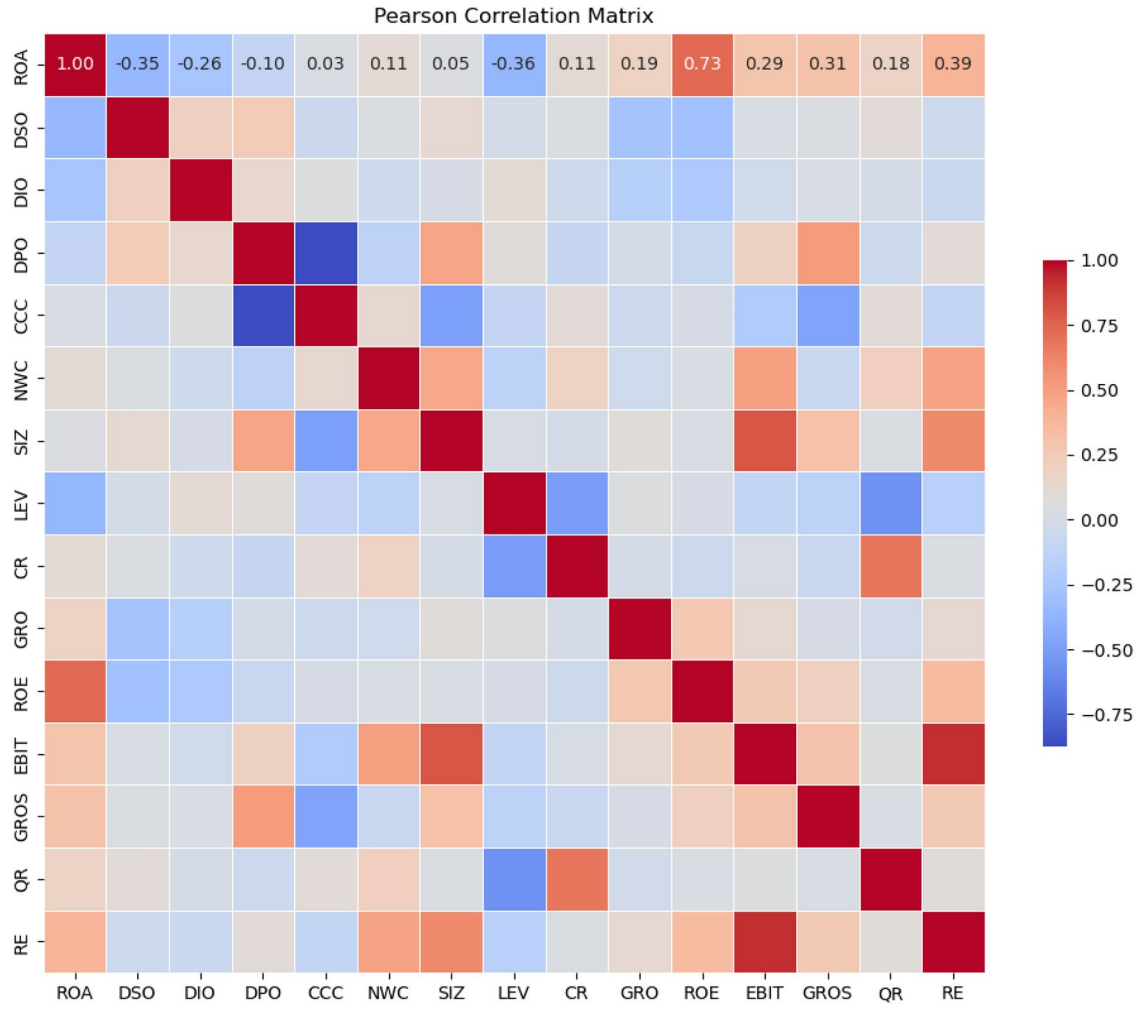

Fig. 3 Pearson correlation matrix

Table 2 Comparison results of models

No.	Model	R^2	RMSE
1	Lasso	0.5908	2.001
2	RR	0.7194	2.0012
3	KNR	0.452	2.79
4	SVR	0.3968	2.9339
5	RF	0.9762	0.5826
6	GBR	0.97	0.6528
7	XGBoost	0.9822	0.504
8	MLP	0.838	1.517

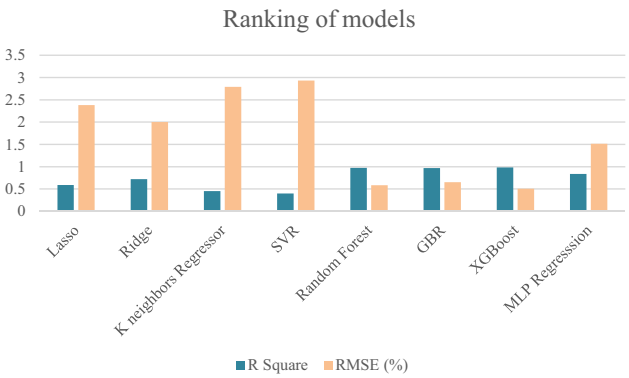

Fig. 4 Ranking of machine learning models

Prediction

The predictive model outlined in this article can assist business owners in the following ways:

Enhancing business decisions: The model provides a more accurate prediction of ROA when specific profit indicators have not been calculated, enabling businesses to predict future scenarios and analyze "What-If" situations. This supports planning and risk management.

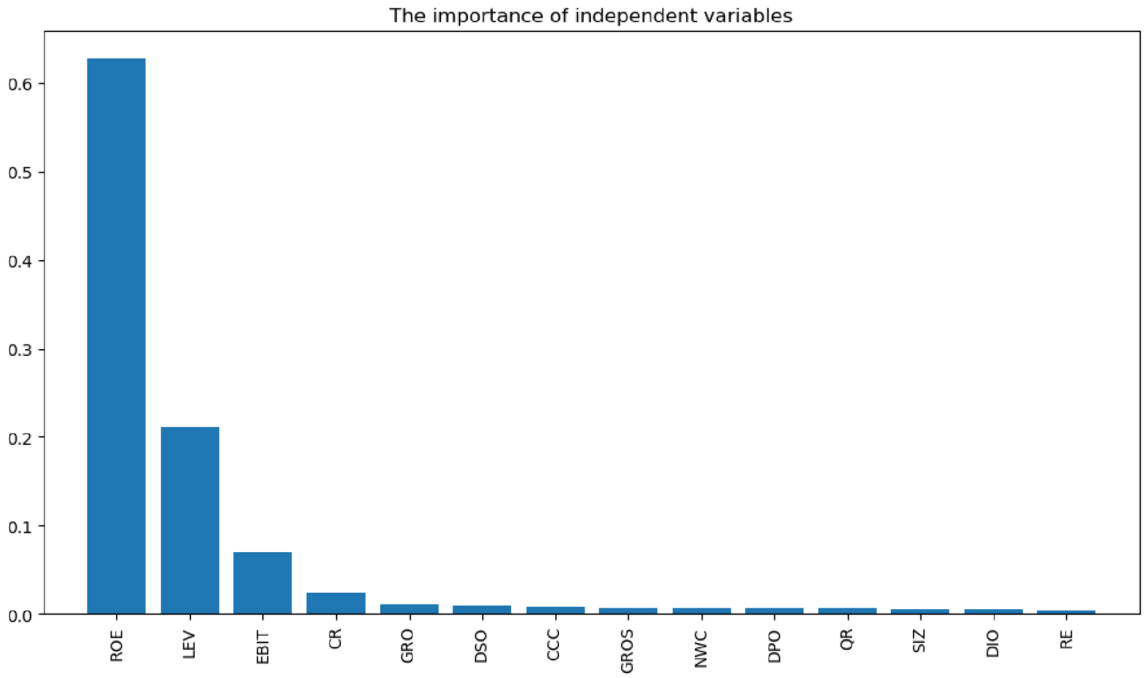

Fig. 5 Ranking of the importance of independent variables

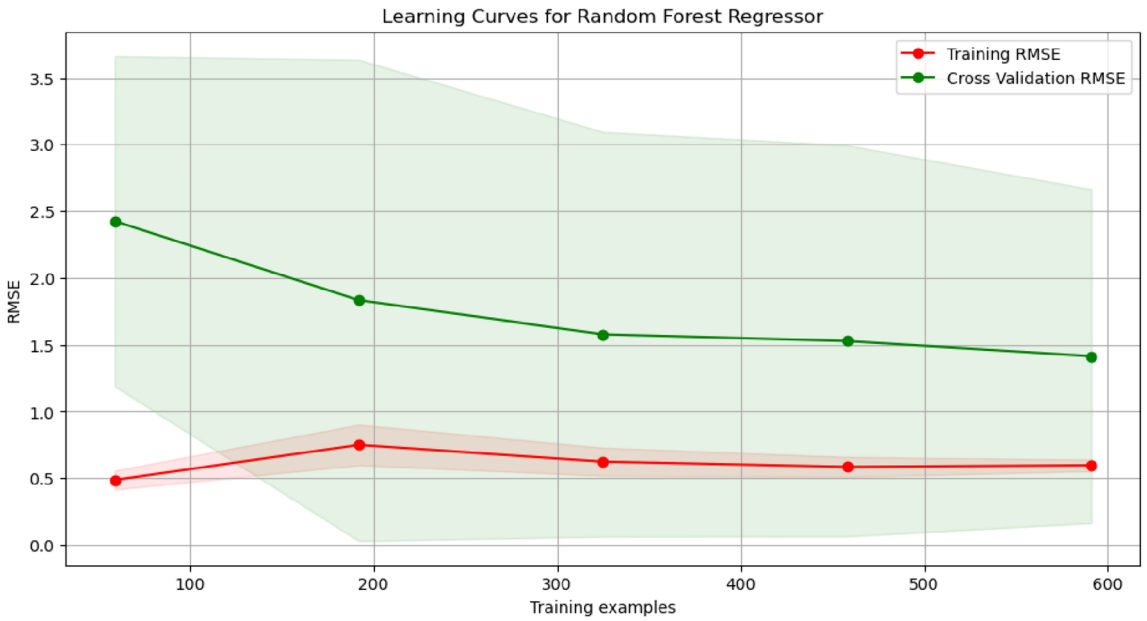

Fig. 6 Learning curve of GBR model

Streamlining and automating the prediction process: The application of machine learning models standardizes the prediction process, making it automated, quick, and less reliant on manual intervention.

Integrating data, knowledge, and context: While this model only considers 15 variables, integrating additional variables and contexts into the analysis process can better reflect the actual situation and the influences from the external business environment.

When applying these machine learning models to forecast the ROA for construction businesses in 2023, we obtained the following results: (Fig. 7).

Despite the impressive results (error < 1%), this study has not yet been validated with a larger dataset. This would

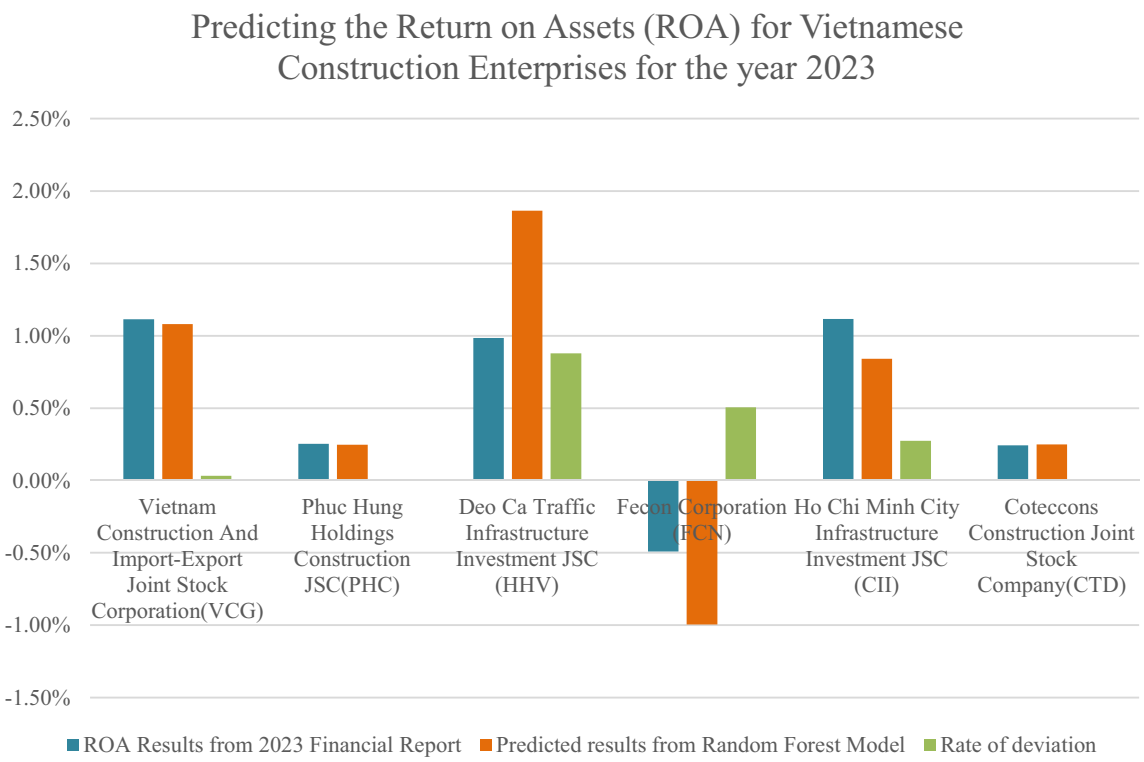

Fig. 7 Predicting the return on assets (ROA) for Vietnamese construction enterprises for the year 2023

involve a larger number of feature variables and require further enhancement of the computational model to improve the accuracy of the prediction tool.

Conclusions, application and proposal

This article presents a financial perspective in the field of construction, applying machine learning models such as simple linear regression, ensemble models, and neural network models. The results indicate that the RF model delivered more optimal results than the second-ranked GBR model (0.62%), as well as the XGBoost model, when considering the distance between the two training and testing lines with the Learning Curves chart. However, the SVR and KNN models were found to be unsuitable for this dataset, given their poor *R* square results (less than 0.5).

This study sets the stage for various research avenues that could be pursued in the near future, including: comparative analysis across industries, temporal stability of ML predictions, integration with other financial health indicators, impact of external variables, advanced ML techniques and algorithms, cross-country comparisons, ML interpretability and decision-making, sustainability and environmental considerations.

These proposed research avenues can expand the understanding built by my study. Moreover, they could offer valuable insights for professionals in the industry, policymakers, and scholars intrigued by the blend of finance, construction, and machine learning.

Acknowledgements For this work, we gratefully recognize the time and facilities provided by Ho Chi Minh City University of Technology (HCMUT), VNUHCM.

Author contributions Both the authors wrote, prepared and reviewed the manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The corresponding author (LTD) and author (PVHS) are available to provide the data, model, or code underlying the findings of this study upon request, in accordance with reasonable conditions.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Adinyira, E., Adjei, E., Agyekum, K., & Fugar, F. (2021). Application of machine learning in predicting construction project profit in

Ghana using support vector regression algorithm (SVRA). *Engineering Construction and Architectural Management*. https://doi.org/10.1108/ECAM-08-2020-0618

Breiman, L. (1999). Random forests—Random features. Computer Science, Mathematics. https://www.stat.berkeley.edu/~breiman/random-forests.pdf

Cai, Y., Yin, Qi., Qian, Su., Huang, X., Zhang, Y., & Liu, T. (2020). Prediction method of enterprise return on net assets based on improved random forest algorithm. *Conference Series*. https://doi.org/10.1088/1742-6596/1682/1/012083

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*. https://doi.org/10.7717/peerj-cs.623

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*. https://doi.org/10.1214/aos/1013203451

Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/34.58871

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development. https://doi.org/10.5194/gmd-15-5481-2022

Kaveh, A., & Iranmanesh, A. (1998). Comparative study of backpropagation and improved counterpropagation neural nets in structural analysis and optimization. *International Journal of Space Structures, 13*(4), 177–185. https://doi.org/10.1177/026635119801300

Kaveh, A., & Khavaninzadeh, N. (2023). Efficient training of two ANNs using four meta-heuristic algorithms for predicting the FRP strength. *Structures, 52*, 256–272. https://doi.org/10.1016/j.istruc.2023.03.178

Kayakus, M., Tutcu, B., Terzioglu, M., Tala, H., & Ünal Uyar, G. F. (2023). ROA and ROE forecasting in iron and steel industry using machine learning techniques for sustainable profitability. *Sustainability*. https://doi.org/10.3390/su15097389

Khoi, L. V., & Pointer, P. (2019). Predictors of return on assets and return on equity for banking and insurance companies on Vietnam stock exchange. *Entrepreneurial Business and Economics Review*. https://doi.org/10.15678/EBER.2019.070411

Le, T. N. H., Mai, V. A., & Van Nguyen, C. (2020). Determinants of profitability: Evidence from construction companies listed on Vietnam securities market. *Management Science Letters*. https://doi.org/10.5267/j.msl.2019.9.028

Mahfouz, T. (2012). A productivity decision support system for construction projects through machine learning (ML). *Proceedings of the CIB W78 2012: 29th international conference*, Beirut, Lebanon, 17–19 October. https://itc.scix.net/pdfs/w78-2012-Paper-54.pdf

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*. https://doi.org/10.1007/BF02478259

McGroarty, A. B., & GerdingFrank, E. (2014). Automated trading with performance weighted random forests and seasonality. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2013.12.009

Ngo, T. Q. V., & Ngoc, C. (2023). Does working capital management matter? A comparative case between consumer goods firms and construction firms in Vietnam. *Cogent Business & Management*. https://doi.org/10.1080/23311975.2023.2271543

Nguyen Dang, N. T., Nguyen, V. N., & Pham, V. H. S. (2024). Achieving improved performance in construction projects: Advanced time and cost optimization framework. *Evolutionary Intelligence*. https://doi.org/10.1007/s12065-024-00918-7

Pointer, L. V., & Khoi, P. D. (2019). Predictors of return on assets and return on equity for banking and insurance companies on Vietnam stock exchange. *Entrepreneurial Business and Economics Review*. https://doi.org/10.15678/eber.2019.070411

Rincy, T. N., & Gupta, R. (2020). Ensemble learning techniques and its efficiency in machine learning: a survey. 2nd International Conference on Data, Engineering and Applications, https://doi.org/10.1109/IDEA49133.2020.9170675

Scornet, G. B., & Erwan. (2016). A random forest guided tour. *TEST*. https://doi.org/10.1007/S11749-016-0481-7

Sevil, B. B., & Güven. (2020). "Predicting IPO initial returns using random forest." *Borsa Istanbul Review*. https://doi.org/10.1016/J.BIR.2019.08.001

Soa, C., La, Nguyen, Van, P. T., Truong, T. V., Le Phi, L., & Vu, T. (2023). Relationship between capital structure and firm profitability: Evidence from Vietnamese listed companies. *International Journal of Financial Studies*. https://doi.org/10.3390/ijfs11010045

Son, L. N., Khoi, Q., & Hong, P. V. (2023). Optimization in construction management using adaptive opposition slime mould algorithm. *Advances in Civil Engineering*. https://doi.org/10.1155/2023/7228896

Son, L. N., Khoi, Q., & Hong, P. V. (2024a). Artificial intelligent support model for multiple criteria decision in construction management. *Opsearch*. https://doi.org/10.1007/s12597-024-00749-1

Son, N. D., Trinh, N., Van Nam, N., & Hong, P. V. (2024b). Advanced vehicle routing in cement distribution: A discrete salp swarm algorithm approach. *International Journal of Management Science and Engineering Management*. https://doi.org/10.1080/17509653.2024.2324172

Son, N. T. V., & Pham, V. H. (2024). Applying ant colony optimization algorithm to optimize construction time and costs for mass concrete projects. *Asian Journal of Civil Engineering*. https://doi.org/10.1007/s42107-024-00990-5

Son, N. V. N., & Pham, V. H. (2023a). Cement transport vehicle routing with a hybrid sine cosine optimization algorithm. *Advances in Civil Engineering*. https://doi.org/10.1155/2023/2728039

Son, T. H. D., & Pham, V. H. (2023b). Research on applying machine learning models to predict the electricity generation capacity of rooftop solar energy systems on buildings. *Asian Journal of Civil Engineering*. https://doi.org/10.1007/s42107-023-00722-1

Tsolacos, S., & Brooks, C. (2010). Forecasting real estate returns using financial spreads. *Journal of Property Research*. https://doi.org/10.1080/09599910110060037

Wassie, F. A. (2020). Impacts of capital structure: Profitability of construction companies in Ethiopia. *Journal of Financial Management of Property and Construction*. https://doi.org/10.1108/JFMPC-08-2019-0072

Wu, C. W., & Chen, C. L. (2012). Diagnosing assets impairment by using random forests model. *International Journal of Information Technology & Decision Making*. https://doi.org/10.1142/S0219622012500046

Zhang, H., Yang, F., Li, Y., & Li, H. (2015). Predicting profitability of listed construction companies based on principal component analysis and support vector machine—Evidence from China. *Automation in Construction, 53*, 22–28. https://doi.org/10.1016/J.AUTCON.2015.03.001

Zhu, Z. T., Yan, Z., & Guangwei, Z. (2019). Stock selection with random forest: An exploitation of excess return in the Chinese stock market. *Heliyon*. https://doi.org/10.1016/j.heliyon.2019.e02310