e-ISSN: 3047-857X

*Proceedings of the International Symposium on Management (Volume 22, 2025)*

# Machine Learning in Financial Performance Prediction: The Role of ESG

Hien Thi Thu Hoang[1,*], Trieu Dang Viet[1], Trang La Thi Thu[1]

[1]*Banking Academy of Vietnam*
[*]*Corresponding author. Email: hienhtt@hvnh.edu.vn*

**ABSTRACT**
This study explores the application of machine learning techniques to predict corporate Environmental, Social, and Governance (ESG) scores, with a particular focus on identifying the most influential factors derived from company reports. Three predictive models - linear regression, random forests, and gradient boosting - were employed to estimate ESG risk scores. The experimental results demonstrate that the gradient boosting model outperforms the other approaches in predictive accuracy. Analysis using Shapley Additive Explanations (SHAP) reveals that industry classification is the most significant determinant of ESG scores, followed by key financial indicators such as Price/Sales ratio, Price/Book ratio, and Market Capitalization. The proposed predictive framework offers valuable insights for investors and corporations, facilitating informed investment decisions and strategic enhancements in ESG performance.

***Keyword:*** *ESG, Random Forest, Decision Tree, Logistic Regression, financial performance, machine learning*

## 1. INTRODUCTION

In recent years, Environmental, Social, and Governance (ESG) factors have gained substantial prominence in investment decisions and corporate financial performance assessment (Friede, Busch, & Bassen, 2015). Research suggests that firms with higher ESG scores often exhibit superior financial performance due to improved risk management and enhanced market reputation (Eccles, Ioannou, & Serafeim, 2014). Concurrently, the rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML) has transformed traditional financial analysis methodologies. ML models, with their ability to detect complex, nonlinear patterns in data, are progressively supplementing or replacing conventional forecasting techniques (Gu et al., 2020). Despite the growing body of research on ESG and its impact on financial performance, a notable gap remains in integrating ESG data with modern analytical techniques, such as Machine Learning (ML) and time-series models, to enhance financial forecasting accuracy. While traditional models have provided valuable insights, the increasing complexity and volume of ESG-related data necessitate the application of more sophisticated approaches. This study aims to bridge this gap by exploring the potential of ML in financial prediction, particularly in comparison with conventional forecasting methods.

## 2. LITERATURE REVIEW

### *2.1. ESG impact on financial performance*

In recent years, numerous studies have highlighted the significance of ESG (Environmental, Social, and Governance) disclosure, demonstrating its effectiveness in linking sustainability reporting with business performance (Adams, 2017). Several studies have examined the influence of ESG factors, either individually or collectively, on organizational performance, the results yet not convergent. For instance, Velte (2017) concluded that ESG positively impacts performance. Furthermore, ESG is recognized as a key driver of innovation and competitive advantage, which can improve future operational performance (Porter & Kramer, 2006). Research by Achim et al. (2016) on companies listed on the Bucharest Stock Exchange revealed a positive relationship between corporate governance quality and market value, highlighting that high governance scores can optimize corporate value. However, some studies highlight the negative or negligible impact of ESG on company performance. Garcia et al. (2017) found a negative relationship between environmental performance and profitability among companies in BRICS countries. Similarly, Jain et al. (2017) observed a negative relationship between ESG scores and business performance. Achim and Borlea (2016) also found that environmental investments increased internal financial burdens, leading to a decline in financial performance.

### *2.2. The application of machine learning in financial forecasting*

Machine Learning (ML) has increasingly become a crucial tool in financial analysis due to its ability to process nonlinear data and uncover complex patterns that traditional methods often fail to capture (Gu et al., 2020). In the context of ESG, ML can help identify hidden relationships between ESG data and financial performance, which traditional linear regression models may overlook (Berg et al., 2019). For instance, a study by Feng et al. (2022) utilized Deep Learning to forecast the impact of ESG factors on stock prices and found that ML-based models outperformed traditional forecasting approaches. Furthermore, Explainable AI (XAI) techniques such as SHAP (Shapley Additive Explanations) offer insights into the contribution of individual ESG factors in financial prediction models, making ML-based forecasts more interpretable and actionable (Molnar, 2022).

## 3. METHODOLOGY

### *3.1. Applying machine learning to assess ESG and financial performance*

This study uses 3 machine learning techniques to predict ESG impact on firm performance, namely Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF). To select the most suitale teachnique, $R^2$ Score is the index used to evaluate the model's accuracy. Impacts of ESG on firm performance then reported using the most suitable machine learning technique.

### *3.2. Data, variables and sampling*

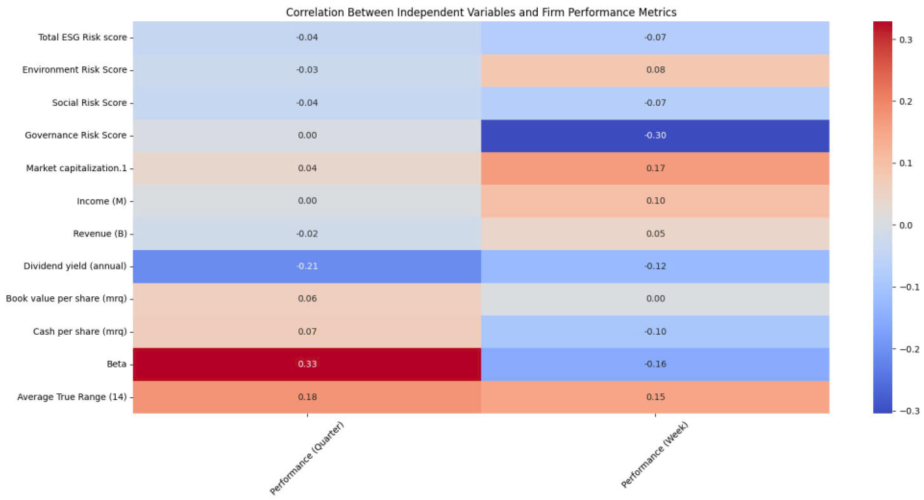
This study integrates two distinct datasets to explore the relationship between ESG risk and firm performance among S&P 500 companies. The first dataset, sourced from Kaggle, comprises ESG risk ratings, which quantify a company's exposure to environmental, social, and governance risks. The second dataset consists of comprehensive financial indicators for the same set of S&P 500 companies, also obtained from Kaggle. This includes metrics such as revenue, net income, return on equity (ROE), earnings per share (EPS), total assets, and market capitalization. Performance indivators were weekly, and quarterly based. Upon merging these two datasets based on a common identifier- typically the company name - a consolidated dataset was formed, containing 90 variables (attributes) and 461 observations (rows).

## 4. RESULT AND DISCUSSION

### *4.1. Correlations*

The correlation analysis assesses the strength and direction of the linear relationships between the ESG score and firm performance. Results can be seen in Table 1. The correlation analysis shows that Beta has the strongest positive relationship with quarterly performance (0.33), indicating that firms with higher market risk tend to deliver stronger short-term returns. Similarly, Average True Range (ATR) is positively correlated with both quarterly (0.18) and weekly performance (0.15), reinforcing the link between price volatility and return potential.

**Table 1.** Correlations




Source: Table calculated by authors

In contrast, Governance Risk Score shows a notable negative correlation with weekly performance (−0.30), suggesting that weak governance is associated with poorer short-term outcomes. Dividend Yield also has a negative relationship with both performance measures (−0.21 for quarter, −0.12 for week), indicating that high-dividend firms may underperform in terms of price appreciation. ESG variables such as Environmental, Social, and Total ESG Risk Scores show only weak or slightly negative correlations, suggesting limited predictive power for short-term returns.

### *4.2. Model comparison based on the linear evaluation metrics*

To assess the effectiveness of machine learning models in predicting firm performance based on environmental, social, governance (ESG) scores, R-squared ($R^2$) performance metric is used, the result is presented in Table 2.

**Table 2.** R-squared ($R^2$) performance metric

| Model | DT | LR | RF |
|---|---|---|---|
| Performance (Quarter) | -0.1015 | 0.0123 | 0.1961 |

The R-squared ($R^2$) metric, which measures the proportion of variance explained by the model, showed an even more pronounced difference. Random Forest was the only model that consistently achieved positive $R^2$ values across all target variables. It performed particularly well on short-term targets like weekly performance ($R^2 = 0.3191$) and quarterly performance ($R^2 = 0.1961$). Therefore, firm performance (quarter) is predicted by the Random Forest model, the results are presented in the next sub-section.

### *4.3. Analyse the effect of each factor to firm performance (Quarter)*

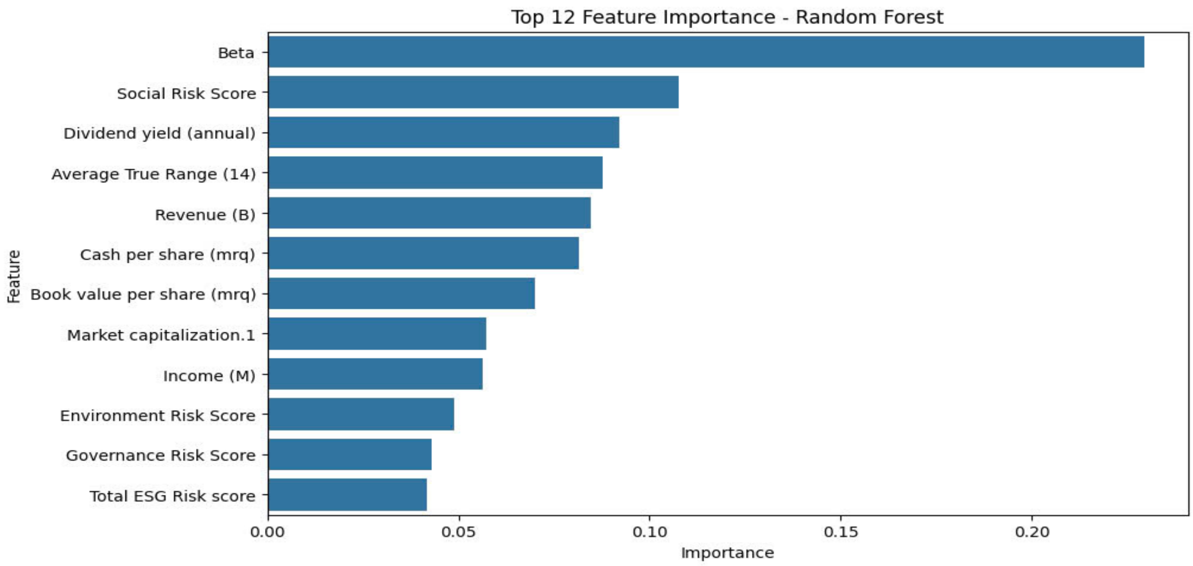



**Figure 1** Firm performance prediction.

As can be seen in Figure 1, it identifies Beta (≈ 0.23) as the strongest predictor of firm performance, highlighting the central role of market risk in short- and medium-term returns (Sharpe, 1964). The Social Risk Score (≈ 0.11) ranks second in predicting performance, reflecting the growing financial relevance of social practices such as employee treatment and stakeholder engagement (Eccles et al., 2014). Dividend Yield and ATR (both ≈ 0.09) follow, indicating the importance of income stability and price volatility. Financial indicators like Revenue, Cash per Share, and Book Value hold moderate weight (0.06–0.07), while broader ESG components including Environmental, Governance, and Total ESG Risk Scores—are least influential (~0.04), suggesting limited short-term explanatory power (Khan et al., 2016).

## 5. KEY FINDINGS AND CONCLUSION

This study explores the application of machine learning techniques in predicting ESG risk scores using three models: Linear Regression, Random forests, and Decision Tree. The results indicate that the Random Forest model outperforms the others in predictive accuracy. SHAP analysis highlights that industry classification is the most significant determinant of firm performance, followed by key financial indicators such as the Price/Sales (P/S) ratio, Price/Book (P/B) ratio, and Market Capitalization. The proposed ESG prediction framework offers valuable insights for investors and corporations, aiding in more informed investment decisions and strategic ESG enhancements. By identifying the key factors influencing firm performance, businesses can focus on crucial aspects to improve their ESG performance and attract investors. These findings underscore the critical role of industry classification and financial indicators in determining ESG risk levels and pave the way for further research on machine learning applications in ESG analysis. Furthermore, leveraging advanced predictive models enhances ESG reporting transparency and provides investors with a more accurate assessment of corporate sustainability risks and opportunities.

## REFERENCES

Achim, M. V., Borlea, S. N. (2016). Economic and Financial Crime: Corruption, Shadow Economy, and Money Laundering. Springer.

Achim, M. V., et al. (2016). Corporate Governance and Market Value: Empirical Evidence from the Bucharest Stock Exchange. Economic Research-Ekonomska Istraživanja, 29(1), 68-77.

Adams, C. A. (2017). The sustainable development goals, integrated thinking and the integrated report. Journal of Sustainability Accounting and Management, 5(2).

Berg, F., Koelbel, J. F., & Rigobon, R. (2019). Aggregate confusion: The divergence of ESG ratings. MIT Sloan School of Management Working Paper.

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). "The impact of corporate sustainability on organizational processes and performance." Management Science, 60(11), 2835-2857.

Feng, G., Giglio, S., & Xiu, D. (2022). Taming the factor zoo: A test of new factors. Journal of Finance, 77(4), 2417-2476.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. Journal of Sustainable Finance & Investment, 5(4), 210-233.

Garcia, A. S., Mendes-Da-Silva, W., & Orsato, R. J. (2017). Sensitive industries produce better ESG performance: Evidence from emerging markets. Journal of Cleaner Production, 150, 135-147.

Gu, S., Kelly, B., & Xiu, D. (2020). "Empirical asset pricing via machine learning." The Review of Financial Studies, 33(5), 2223-2273.

Jain, T., Aguilera, R. V., & Jamali, D. (2017). Corporate Governance and Environmental, Social, and Governance (ESG) Performance: The Moderating Role of Emerging Markets. Journal of Business Ethics, 145(2), 471-490.

Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Leanpub.

Porter, M. E., & Kramer, M. R. (2006). Strategy and Society: The Link Between Competitive Advantage and Corporate Social Responsibility. Harvard Business Review, 84(12), 78-92.

Velte, P. (2017). Does ESG performance have an impact on financial performance? Evidence from Germany. Sustainable Development, 25(4), 242-254.