Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3, 219-233
2026
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v10i3.12339

Applying artificial intelligence to forecast the market capitalization of global companies based on ESG and financial indicators

Hoan Thi Duong[1*], Lam Tung Nguyen[2]
[1,2]School of Economics, Hanoi University of Industry, Hanoi, Vietnam; hoandt@haui.edu.vn (H.T.D.)
nguyentlam98765@gmail.com (L.T.N.).

Abstract: This study investigates the effectiveness of artificial intelligence in forecasting the market capitalization of publicly listed firms by integrating Environmental, Social, and Governance (ESG) indicators with traditional financial variables. Grounded in Stakeholder Theory and Signaling Theory, the research evaluates whether ESG information enhances predictive performance beyond conventional financial metrics. The analysis employs a global panel dataset comprising more than 11,000 firm-year observations from approximately 1,000 companies across multiple industries and regions during 2015–2025. Several machine learning models, including Random Forest, CatBoost, Extreme Gradient Boosting, Light Gradient Boosting Machine, Extra Trees, and Linear Regression, have been developed using log-transformed financial variables and dimension-reduced ESG components derived through principal component analysis. Time series validation is applied to ensure temporal robustness. The findings indicate that tree-based ensemble models significantly outperform linear regression, with Random Forest explaining 84.53% of the variation in future market capitalization. Financial indicators, particularly revenue and profit margin, remain dominant predictors, while ESG factors contribute limited short-term incremental value. The results highlight the complementary role of ESG reporting in supporting long-term transparency and sustainable competitiveness.

Keywords: *Artificial intelligence, Environmental, Financial performance, Governance, Machine learning, Market capitalization forecasting, Social.*

1. Introduction

Over the past decade, the focus of corporate decision-making has expanded beyond profit maximization to encompass the broader economic, social, and environmental consequences of business activities. Environmental, Social, and Governance (ESG) considerations have evolved from voluntary initiatives into essential components of corporate financial strategy and risk management [1]. This transition reflects contemporary governance perspectives that emphasize the central role of investors and stakeholders in shaping long-term corporate value [2]. As a result, firms increasingly allocate resources to ESG-related activities to enhance resilience, reputation, and strategic alignment in an uncertain global business environment.

Despite this transformation, approaches to forecasting firm value, commonly measured by market capitalization, have not fully adapted to these structural changes. Traditional valuation and forecasting models continue to rely heavily on historical financial information such as revenues, leverage, and accounting ratios. Prior studies indicate that such methods face notable limitations when applied to volatile and complex market conditions [3]. Although financial indicators remain fundamental, they often fail to capture intangible value drivers related to corporate reputation, workforce management, and environmental resilience [4]. Existing empirical research generally reports a positive association between ESG performance and financial efficiency [5]. However, much of this literature relies on linear regression techniques, which assume proportional and stable relationships between ESG indicators and

History: Received: 1 January 2026; Revised: 6 February 2026; Accepted: 9 February 2026; Published: 6 March 2026
* Correspondence: hoandt@haui.edu.vn

firm value [6]. In practice, ESG value relationships are frequently nonlinear and dynamic, as ESG investments may generate delayed effects or influence firm valuation only after reaching certain thresholds. Consequently, conventional linear models may overlook important interaction effects and hidden patterns.

Recent advances in Artificial Intelligence (AI) and Machine Learning (ML), including Random Forest and Gradient Boosting algorithms, offer powerful alternatives for modeling complex and nonlinear relationships in large datasets [7]. While these methods have been increasingly applied in financial forecasting, limited research has explicitly integrated ESG indicators into AI-based models to predict market capitalization [8]. Against this backdrop, this study aims to develop and compare advanced machine learning models for forecasting market capitalization by jointly incorporating traditional financial variables and ESG indicators.

This study addresses the following research questions:

RQ1: Does incorporating ESG indicators enhance the accuracy of market capitalization forecasting beyond traditional financial variables?

RQ2: Do machine learning models outperform linear regression in capturing non-linear relationships between firm attributes and market capitalization?

RQ3: How do financial indicators and ESG factors comparatively contribute to predicting future market capitalization?

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 outlines the data and methodology. Section 4 reports and discusses the empirical results. Section 5 concludes the paper and suggests directions for future research.

2. Literature Review

2.1. ESG Performance and Market Capitalization

Environmental, Social, and Governance (ESG) ratings aggregate multiple sustainability dimensions into a single metric and are widely interpreted as indicators of firm resilience and long term financial stability [9]. Investors and asset managers increasingly rely on ESG ratings to differentiate firms with stronger adaptive capacity to regulatory, environmental, and socio-political pressures, thereby influencing capital allocation decisions and liquidity conditions. As a result, ESG considerations are expected to play an expanding role in shaping global investment portfolios. From a theoretical perspective, Stakeholder Theory conceptualizes firms as systems shaped by diverse interest groups, with corporate value dependent on the ability to meet stakeholder expectations. Within this framework, Edmans [10] argues that high ESG performance reduces information asymmetry, enhances investor confidence, and contributes to more stable and predictable cash flows [10]. Complementary empirical evidence suggests that ESG performance also strengthens corporate reputation, which further supports firm valuation [11]. Moreover, firms with superior ESG ratings tend to exhibit more accurate earnings forecasts due to enhanced risk assessment and governance practices [12]. In emerging markets, however, the ESG financial performance relationship is moderated by institutional quality and the effectiveness of corporate governance mechanisms [2, 13].

Firm value is influenced by financial performance, tangible assets, and non-financial factors such as reputation and sustainability practices. As socially responsible investment behavior expands, ESG indicators have emerged as important tools guiding investors in evaluating firms that balance financial objectives with ethical considerations [14].

2.2. ESG Disclosure and Financial Market Outcomes

Legitimacy Theory and Signaling Theory provide complementary explanations for how ESG performance translates into market valuation outcomes. Firms disclose ESG information to align with societal expectations, maintain social acceptance, and reinforce organizational legitimacy [15]. From a signaling perspective, ESG disclosure conveys credible information regarding a firm's commitment to sustainable development, thereby reducing uncertainty and attracting long-term investors [16]. Prior

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

studies further indicate that companies with strong ESG performance are perceived as more trustworthy and credible by capital market participants, enhancing access to both equity and debt financing [17]. Through these mechanisms, ESG performance and disclosure exert an indirect yet meaningful influence on firm value and market capitalization.

2.3. Artificial Intelligence, ESG Analytics, and Market Capitalization Forecasting

Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have significantly transformed ESG analysis and financial forecasting. AI-driven approaches enhance the reliability and strategic depth of ESG governance by automating data processing and translating heterogeneous sustainability information into standardized ESG key performance indicators. Advanced models such as Long Short-Term Memory networks exploit complex textual and news-based information to forecast ESG ratings, while other frameworks integrate ESG indicators to predict financial distress and risk exposure [18]. Digital integration further amplifies the positive impact of ESG on financial outcomes [19, 20].

Empirical studies consistently demonstrate that firms with strong ESG performance achieve superior financial efficiency, particularly in emerging markets, benefiting from lower capital costs and improved profitability [21, 22]. Beyond improving analytical precision, AI and ML also support strategic decision-making by capturing complex, nonlinear relationships between ESG indicators, financial variables, and firm value. Collectively, the literature highlights the growing importance of ESG performance and financial indicators in shaping firm value and market capitalization. However, prior empirical studies predominantly rely on linear or parametric methods that may be inadequate for capturing the complex, nonlinear interactions inherent in ESG financial relationships. Advances in Artificial Intelligence provide powerful tools to address these limitations by modeling high-dimensional data and uncovering latent patterns. Building on these insights, the following section presents the methodological framework and AI-based models employed to forecast market capitalization by jointly incorporating ESG and financial indicators. Table 1 summarizes prior studies examining the relationships among ESG performance, financial indicators, and firm value, as well as the application of AI and machine learning models in market capitalization forecasting.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

Table 1.
Summary of Research Studies on Credit Card Fraudulent Transactions and Machine Learning.

Authors (Year)	Key Findings	Dependent Variable (Output)	Independent Variables (Inputs)	AI Model	References
Sandberg et al. [23]	ESG ratings indicate firm resilience and long-term stability.	Firm value/market cap	ESG ratings	Conceptual	Sandberg et al. [23]
Edmans [10]	High ESG reduces information asymmetry and stabilizes cash flows.	Firm value	ESG performance	Conceptual/Empirical	Edmans [10]
Clementino and Perkins [11]	ESG enhances corporate reputation.	Firm value	ESG performance	Regression models	Clementino and Perkins [11]
Luo and Wu [12] and Xu et al. [24]	High ESG improves earnings forecast accuracy.	Earnings / firm value	ESG ratings	Forecasting models	Luo and Wu [12] and Xu et al. [24]
Ahmad et al. [2]	Institutional quality moderates the ESG performance relationship.	Firm performance	ESG, governance, institutions	Moderation / Regression	Ahmad et al. [2]
Praveen Kumar and Manoj Kumara [25]	Market capitalization reflects a firm's value under shocks.	Market capitalization	Macroeconomic & firm variables	Time-series models	Praveen Kumar and Manoj Kumara [25]
Ewing and Thompson [26]	Operational fundamentals affect market capitalization.	Market capitalization	Operational & financial metrics	Regression models	Ewing and Thompson [26]
Friede et al. [19] and Fu and Li [20]	Digital integration amplifies ESG's financial impact.	Firm performance	ESG + digitalization	Integrated ESG models	Friede et al. [19] and Fu and Li [20]
Gürsoy and Erbuğa [22]	High ESG improves profitability and lowers capital costs.	Financial performance	ESG indicators	Empirical performance models	Gürsoy and Erbuğa [22]

Although prior studies have examined ESG performance, financial indicators, and firm value, most research treats these factors separately and relies primarily on linear econometric models. Existing evidence on the joint impact of ESG and financial indicators on market capitalization remains limited, particularly within AI-driven analytical frameworks. Moreover, few studies provide global-scale evidence on how nonlinear interactions between ESG and financial performance influence market capitalization. Therefore, there is a clear need for an integrated AI-based approach that combines ESG and financial indicators to enhance the accuracy and explanatory power of market capitalization forecasting.

3. Methodology

Figure 1 illustrates the overall research workflow, outlining the sequential steps from dataset collection and data preprocessing to model evaluation and market capitalization forecasting.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

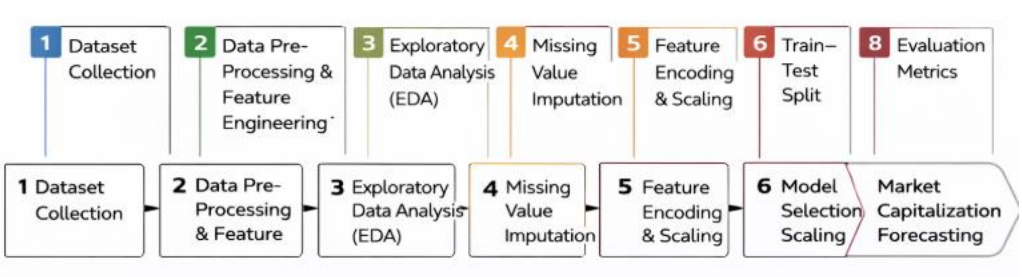

Figure 1.
Research workflow.

The workflow includes dataset collection, data preprocessing and feature engineering, exploratory data analysis, missing value imputation, feature encoding and scaling, model selection, and performance evaluation for market capitalization forecasting.

3.1. Dataset Collection

This study employs a comprehensive dataset of financial and ESG-related indicators collected from globally recognized and publicly accessible data sources. ESG scores and sustainability-related metrics were obtained from Refinitiv ESG Database and MSCI ESG ratings, which are widely used in academic research and financial analysis. Financial and market-related variables were retrieved from Yahoo Finance, World Bank Open Data, and FRED (Federal Reserve Economic Data), ensuring reliability and international comparability. The sample consists of approximately 1,000 publicly listed companies across multiple industries and geographic regions, covering the period from 2015 to 2025. This time span captures both stable economic conditions and major global disruptions, including the COVID-19 pandemic and geopolitical conflicts, thereby enhancing the robustness of the analysis. The final dataset comprises more than 11,000 firm-year observations. The independent variables utilized in this study are detailed in Table 2 below:

Table 2.
Description of variables in the dataset.

Variable Type	Variable	Description
Dependent Variable	MarketCap_Future	Future market capitalization (year t+1), used as the prediction target
Financial Variable	Revenue_Log	Log-transformed annual revenue representing firm scale
Financial Variable	ProfitMargin	Net profit margin indicates firm profitability
Financial Variable	GrowthRate	Year over year revenue growth rate reflecting the firm's growth
Financial Variable	MarketCap	Current market capitalization (year t) representing firm size
ESG Variable	ESG_Environmental	Environmental performance score of the firm
ESG Variable	ESG_Social	Social responsibility score of the firm
ESG Variable	ESG_Governance	Corporate governance quality score
ESG Variable	WasteEmissions	Composite environmental impact index derived from PCA
Control Variable	Industry	Industry sector classification
Control Variable	Region	Geographic region of the firm
Control Variable	Year	Reporting year
Identification Variable	CompanyID	Unique identifier for each firm (excluded from modeling)
Identification Variable	CompanyName	Firm name (excluded from modeling)

Table 2 summarizes the variables employed in the predictive model. The dependent variable is future market capitalization (MarketCap_Future). The independent variables consist of financial indicators and ESG-related measures capturing environmental, social, and governance dimensions. Industry, region, and year are included as control variables to account for structural and temporal effects. Identification variables are excluded from the modeling process to prevent information leakage and ensure model robustness.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

3.2. Data Pre-Processing and Feature Engineering

The data preprocessing procedure focused on constructing the prediction target and preparing the feature set for machine learning models. The dependent variable, MarketCap_Future, was generated by shifting firm-level market capitalization forward by one year, enabling the prediction of future firm value based on contemporaneous financial and ESG information. This lag-based target construction is widely adopted in financial forecasting to avoid look-ahead bias and ensure temporal consistency [27, 28]. Observations without future values were excluded to ensure valid target labels. Financial indicators and ESG pillar scores were selected as explanatory variables, while industry, region, and year were incorporated as control variables to account for structural and temporal heterogeneity. Identification variables were excluded from model estimation to prevent information leakage, consistent with recent machine learning practices in financial modeling.

To enhance model robustness, skewed financial variables were log-transformed, categorical variables were encoded, and multicollinearity among ESG indicators was mitigated using principal component analysis (PCA). These preprocessing and feature engineering steps are aligned with recent studies emphasizing the integration of ESG information and advanced data transformation techniques in predictive financial analytics [29].

3.3. Exploratory Data Analysis (EDA)

An exploratory analysis was conducted to understand distributions and variable relationships. Initially, financial variables (Revenue, MarketCap) and the target variable were found to be severely right-skewed. The distribution of the dependent variable is illustrated in Figure 1, revealing a highly right-skewed pattern that motivates the use of a logarithmic transformation in subsequent modeling.

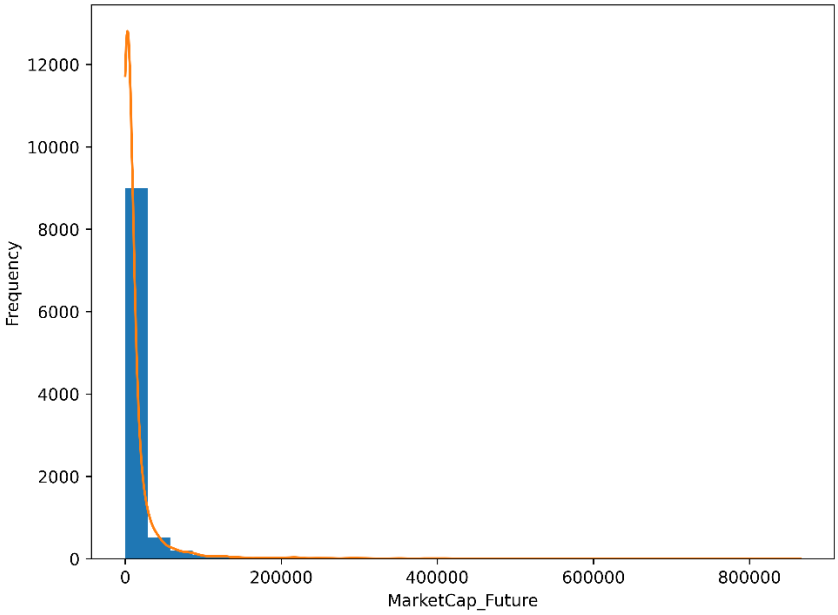

Figure 2.
Distribution of MarketCap_Future.

Figure 2 reveals a highly right-skewed distribution of MarketCap_Future, where the mean value (13,809.57) is more than four times the median (3,190.30), accompanied by extreme skewness (8.77) and kurtosis (107.94), clearly indicating the presence of heavy tails and significant outliers in the dataset.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

The correlation matrix also reveals the non-linearity of the data, and the traditional Linear Regression model might not be the most suitable to solve the problem. This is illustrated in Table 3 and Figure 3.

Table 3.
VIF Index for Independent Variables.

Data Field	VIF
Revenue	3.34
ProfitMargin	2.63
GrowthRate	1.3
ESG_Overall	2 909 102.21
ESG_Environmental	388 861.22
ESG_Social	365 051.43
ESG_Governance	332 757.41
CarbonEmissions	302.15
WaterUsage	31.4
EnergyConsumption	388.84

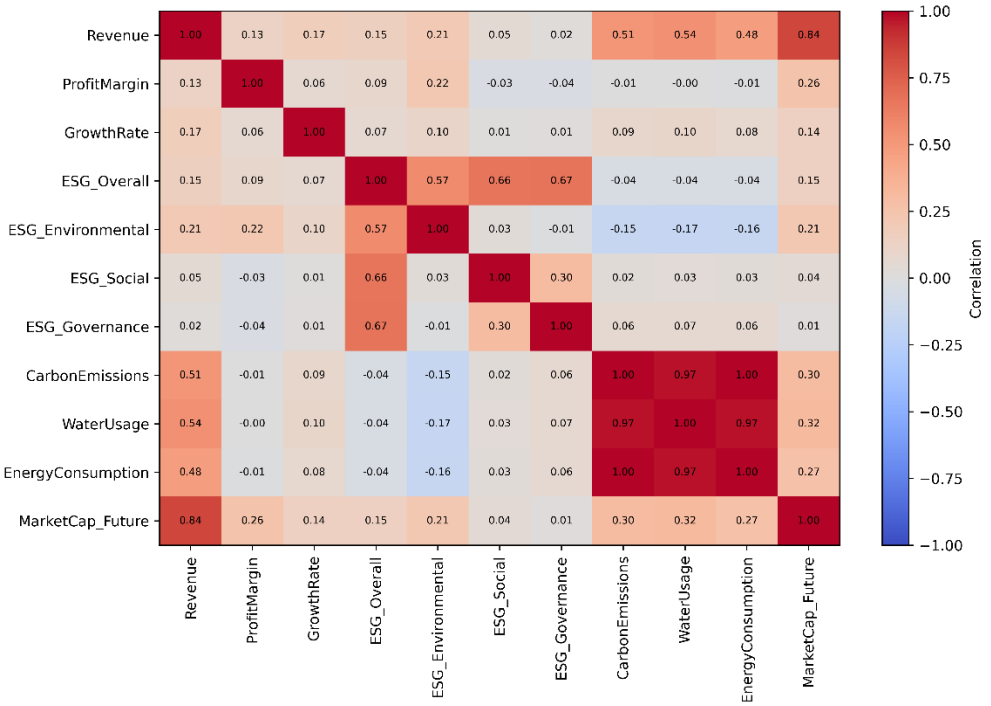

Figure 3.
Correlation matrix of numerical variables.

Table 3 and Figure 3 indicate severe multicollinearity among ESG and environmental variables. Figure 2 shows very high correlations between CarbonEmissions, WaterUsage, and EnergyConsumption (r = 0.97-1.00) and strong links between ESG_Overall and its components (0.57-0.67). This is confirmed by extreme VIF values for ESG_Environmental (388,861.22), ESG_Social (365,051.43), and ESG_Governance (332,757.41), while financial variables such as Revenue (3.34) and ProfitMargin (2.63) remain within acceptable limits.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

3.4. Missing Value Imputation

The data processing workflow addressed missing values efficiently. Specifically, 1,000 missing entries in the GrowthRate column were identified as belonging to 2015 (the first year of data). The research group decided that removing them would be the best way since it only accommodated a small portion of the entire data.

3.5. Feature Encoding

To resolve the multicollinearity, the variable ESG_Overall was removed because it is a linear combination of the three sub-component scores (Environmental, Social, Governance), creating perfect multicollinearity. Principal Component Analysis (PCA) was employed to synthesize the three environmental variables into a single variable named WasteEmissions. This new variable retained 99.98% of the original variance, effectively solving the multicollinearity issue without significant information loss. Afterward, the research group decided to use OneHotEncoding to encode category variables into dummy variables for each value of each category variable.

3.6. Feature Scaling

To address the severe right skewness observed in the financial variables (Revenue, MarketCap) and the target variable during EDA, a logarithmic transformation (np.log1p) was applied. This step was crucial to normalize the distribution and reduce the influence of outliers on the predictive models. Figure 4 presents the distribution of MarketCap_Future after logarithmic transformation to examine the normalization effect prior to model estimation.

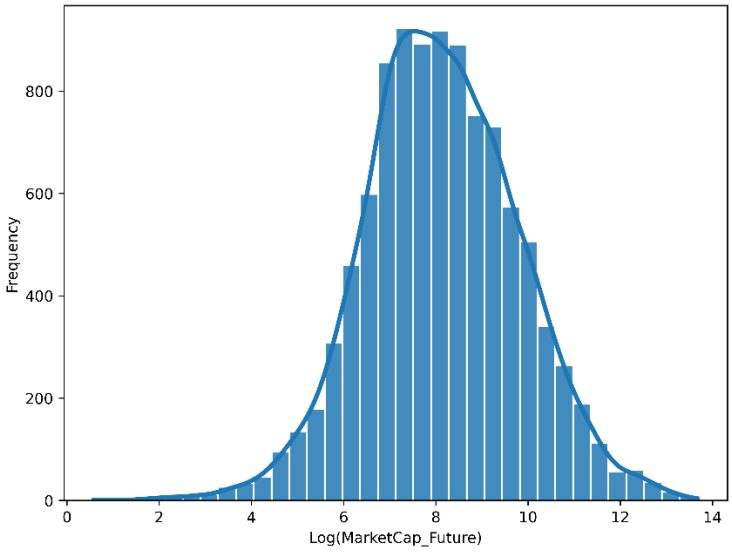

Figure 4.
Distribution of MarketCap_Future after log transformation.

Figure 4 shows that after log transformation, Log(MarketCap_Future) is concentrated around the range 7–9, forming an approximately symmetric, bell-shaped distribution. Compared with the original scale (skewness = 8.77), the log-transformed variable substantially reduces skewness and extreme dispersion, indicating improved normality and suitability for subsequent regression and machine learning models.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

3.7. Train-Test Split

The training and evaluation process followed a strict scientific protocol. Data was split chronologically to prevent data leakage and ensure the availability of ground truth values for validation. The training set comprised data from 2015 through 2023. The testing set utilized independent variables from the fiscal year 2024 to predict the Market Capitalization for 2025. This structure ensures that the model is evaluated on the most recent complete financial year, where actual market outcomes are known. Hyperparameter tuning was conducted via GridSearchCV combined with TimeSeriesSplit.

3.8. Model Selection and Architecture

To address the nonlinear complexities inherent in financial forecasting and ESG integration, this study employs a comprehensive suite of ensemble learning algorithms. The selection of these models is predicated on recent benchmarking, which demonstrates that for structured tabular data, tree-based ensembles consistently outperform deep learning architectures by preserving the orientation of data and effectively modeling irregular decision boundaries [30]. The algorithms utilized fall into two distinct methodological families: Bagging (Variance Reduction) and Boosting (Bias Reduction).

The study utilizes Random Forest and Extra Trees to establish a baseline of stability. Random Forest operates by constructing a multitude of decorrelated decision trees, where each tree is trained on a bootstrap sample of the data, and node splitting is restricted to a random subset of features. This process reduces variance by averaging out individual tree errors [31]. Complementing this, Extra Trees (Extremely Randomized Trees) pushes stochasticity further by selecting split points completely at random rather than optimizing for information gain. This mechanism not only reduces computational cost but also creates smoother decision boundaries, making it particularly robust against noisy input features [32].

The study also implements AdaBoost and the big three Gradient Boosting Machines (GBM): XGBoost, LightGBM, and CatBoost [33]. Unlike bagging, these algorithms grow trees sequentially, where each new learner attempts to correct the residual errors of its predecessors [34]:

- AdaBoost (Adaptive Boosting): As the genesis of the boosting paradigm, AdaBoost serves as a crucial benchmark for iterative learning. Unlike bagging methods that treat all data points equally, AdaBoost systematically alters the distribution of the training data based on the performance of previous models. It functions by sequentially training a series of weak learners (typically shallow decision stumps). After each iteration, the algorithm increases the weights of misclassified instances and decreases the weights of correctly classified ones. This adaptive re-weighting mechanism forces subsequent learners to focus exclusively on the hard-to-predict examples that define the complex, nonlinear boundaries of the feature space. The final prediction is constructed as a weighted sum of these weak learners, effectively minimizing an exponential loss function to transform a collection of weak predictors into a single strong classifier [34].
- XGBoost: This algorithm minimizes a regularized objective function using a second-order Taylor expansion of the loss. By incorporating $L1$ and $L2$ regularization terms directly into the objective function, XGBoost effectively prevents overfitting. Furthermore, its sparsity-aware split finding algorithm allows it to handle missing data and sparse features without imputation, preserving the information encoded in the data's structure.
- LightGBM: Designed for scalability, LightGBM employs two novel techniques: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). GOSS accelerates training by focusing on instances with large gradients (larger errors) while randomly sampling those with small gradients, maintaining accuracy while significantly reducing data volume. Additionally, its leaf-wise tree growth strategy allows the model to converge faster on complex nonlinear patterns compared to the level-wise growth of traditional algorithms [35].
- CatBoost: Specifically selected for its proficiency with categorical variables, CatBoost addresses the issue of target leakage found in standard gradient boosting. It employs ordered Boosting, a

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

permutation-driven approach where the residual for the i-th sample is calculated using a model trained only on samples preceding it in the permutation. This unbiased gradient estimation is critical for ensuring the model's generalization capability on the categorical ESG and industry data used in this study [36].

3.9. Evaluation Metrics

To assess the predictive performance of the deployed models, this study utilizes a multi-metric evaluation framework consisting of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). These metrics are selected to provide a holistic view of model accuracy, encompassing both absolute error magnitude and relative explanatory power.

3.9.1. Mean Squared Error (MSE) and Root Mean Squared Error (RMSE)

MSE measures the average of the squares of the errors, that is, the average squared difference between the estimated values and the actual values. By squaring the errors, MSE heavily penalizes larger deviations, making it useful for identifying models that may produce significant outliers. RMSE is derived as the square root of MSE:

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

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

RMSE is preferred for interpretation as it expresses the error in the same units as the target variable (*MarketCap_Future*), facilitating a direct comparison of the magnitude of error against the financial values [37].

3.9.2. Mean Absolute Percentage Error (MAPE)

To provide a relative measure of accuracy independent of the data scale, MAPE is employed. It expresses the average absolute error as a percentage of the actual values:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (3)$$

This metric is particularly valuable for stakeholders to understand the average percentage deviation of the forecast from the actual market capitalization.

3.9.3. Coefficient of Determination (R^2)

Finally, the R^2 score is utilized to quantify the proportion of the variance in the dependent variable that is predictable from the independent variables:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

An R^2 value closer to 1 indicates that the model explains a high proportion of the variance in market valuation, confirming the model's goodness of fit [38].

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

4. Results and Discussions

Table 4 and Figure 5 summarizes the comparative performance of the machine learning models in predicting MarketCap_Future using multiple error metrics and goodness of fit measures.

Table 4.
Comparison of Machine Learning Model Performance.

Model	MSE	RMSE	MAPE (%)	R2 (%)
RandomForest	0.5	0.7	6.67	84.53
CatBoost	0.5	0.71	6.66	84.48
XGBoost	0.51	0.71	6.72	84.14
LightGBM	0.52	0.72	6.88	83.68
ExtraTrees	0.56	0.75	7.29	82.6
LinearRegression	0.74	0.86	8.35	76.95
AdaBoost	0.82	0.91	9.14	74.33

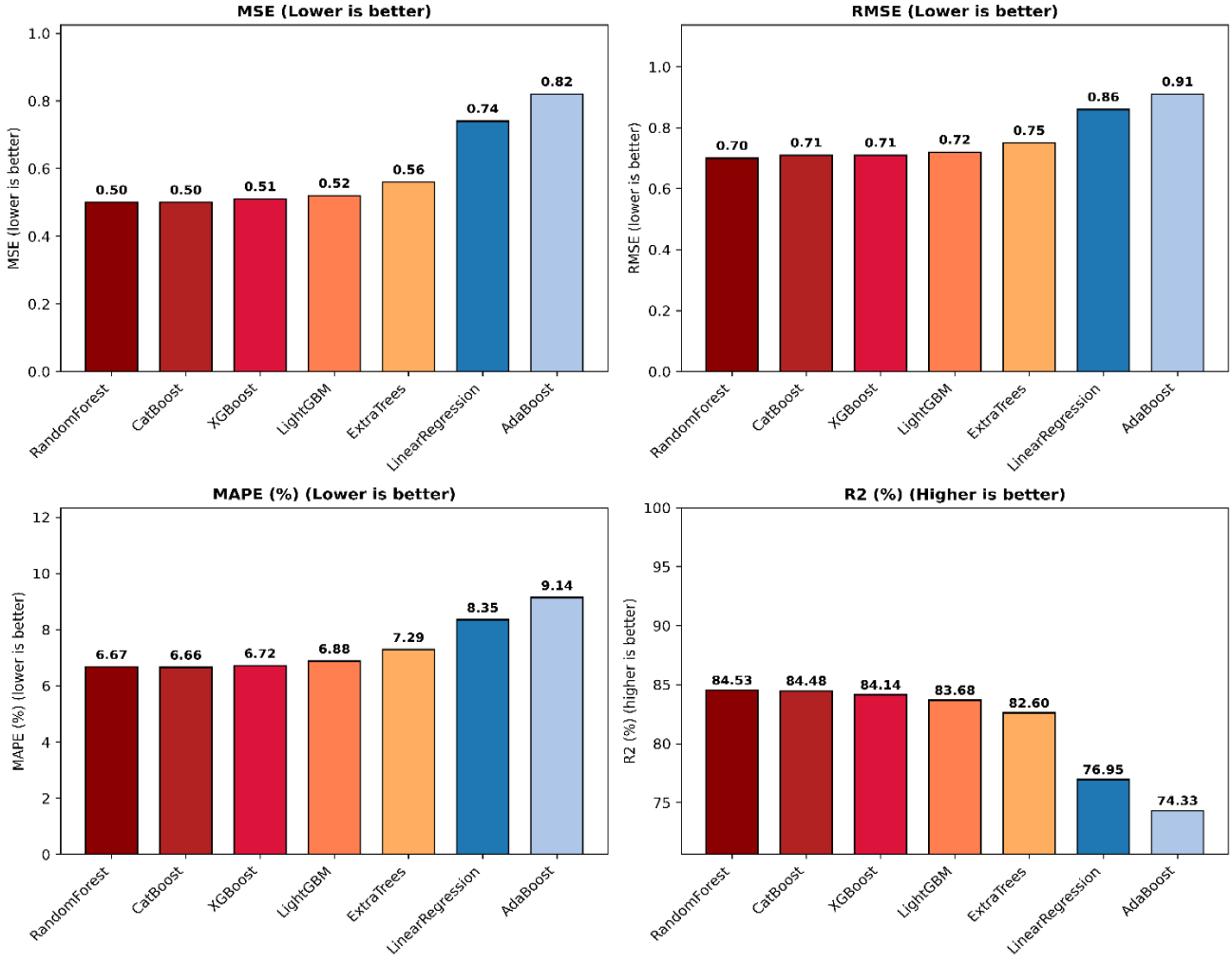

Figure 5.
Comparison of machine learning models.

The analysis reveals that Random Forest emerged as the most robust model, achieving the highest explanatory power with an R^2 of 84.53% and the lowest error rate with a MAPE of 6.67%. It was closely

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

followed by the Gradient Boosting models CatBoost (84.48%) and XGBoost (84.14%), indicating that tree-based ensemble methods are consistently effective for this dataset. To validate the stability of the Random Forest model, we analyzed the divergence between training and testing performance. The difference in R^2 between the training set (90.99%) and the test set (84.53%) was limited to 6.46%. This relatively low difference confirms that the model has successfully learned generalized patterns rather than merely memorizing historical data (overfitting), making it reliable for forecasting unseen future scenarios.

In contrast, the traditional Linear Regression model yielded an R^2 of only 76.95% and a higher MAPE of 8.35%. This highlights a critical limitation in conventional econometric modeling. While Linear Regression explained roughly three-quarters of the variance, it failed to capture the final, most complex quartile of market behavior that the AI models successfully decoded. As established by Shwartz-Ziv and Armon [30], tree-based models like Random Forest generally outperform both linear models and deep learning architectures on tabular data [30]. The logic for this superiority lies in the model's structure: while Linear Regression imposes a rigid, straight line relationship between variables (assuming, for instance, that every 1 point increase in ESG results in a fixed dollar increase in value), Random Forest builds decision boundaries based on logical rules and interactions.

The underperformance of Linear Regression, specifically that the relationship between corporate attributes (ESG) and market capitalization is inherently nonlinear. Linear models fail to recognize threshold effects, situations where, for example, ESG scores might only positively influence value after a company achieves a certain level of profitability or exceeds a specific sustainability score. By successfully identifying these complex, conditional interactions without the need for extensive manual preprocessing, the Random Forest model demonstrates why AI is becoming indispensable for modern valuation tasks where traditional linear statistics fall short.

To identify the relative contribution of each explanatory variable to the model, Figure 6 presents the feature importance results derived from the trained machine learning model.

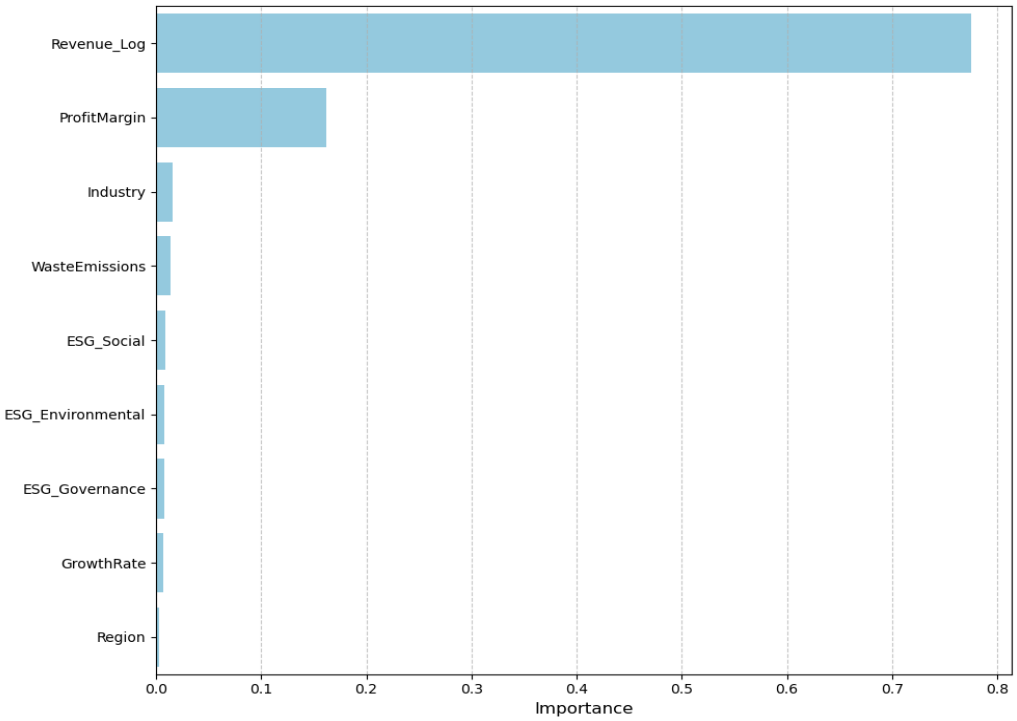

Figure 6.
Feature Importance - Random Forest.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

The results (as shown in Figure 5) show that financial variables are still the main drivers. Revenue_Log (Revenue after log - transformation) accounts for about 77% of the model's decisions, followed by ProfitMargin at 16%. In contrast, ESG indicators, including WasteEmissions, ESG_Social, and ESG_Governance, had very little effect, contributing only about 1-2% each. Since the model prioritizes the variable that best explains variance in absolute magnitude, the subtle premium offered by ESG efficiency is overshadowed by the sheer impact of commercial scale. This finding differs from the optimistic view often found in sustainable finance literature. Stakeholder Theory suggests that meeting social goals creates long-term value [10, 16] and argues that sharing ESG data attracts investors. However, our results show that for short-term valuation, the market is still mostly focused on financial success. Our results clarify that while ESG may influence valuation multiples, it does not act as a substitute for financial scale in determining total market value.

This difference can be explained by Signaling Theory. While Bamahros et al. [17] suggest ESG signals trust, our results imply it is a secondary signal [17]. Financial numbers (Revenue and Profit) act as the main signal that a company is healthy. While Del Gesso and Lodhi [15] argue that ESG data helps clarify a company's situation, our study suggests that in a model controlling for revenue scale, ESG acts as a marginal quality signal rather than a primary driver of size [15].

Furthermore, these results agree with Domanović [5], who noted that despite the rise of ESG, financial numbers remain the core language of valuation [5]. The difference between our findings and those of Luo and Wu [12], who found strong links between ESG and forecast accuracy, may be due to the methods used [12]. Previous studies often used linear regression, which might make the connection between ESG and value look stronger than it really is. By using AI to isolate the specific impact of each factor, this study reveals that once Revenue and Profit are known, ESG adds very little extra information for predicting immediate market capitalization.

5. Conclusion

This study examined the effectiveness of artificial intelligence–based models in forecasting market capitalization by jointly incorporating traditional financial indicators and Environmental, Social, and Governance factors. Addressing the first research question, the findings indicate that integrating Environmental, Social, and Governance information with financial variables improves the overall explanatory framework of market capitalization forecasting. However, the incremental predictive contribution of Environmental, Social, and Governance indicators remains limited in the short term when dominant financial factors are taken into account.

With respect to the second research question, the empirical results clearly demonstrate that machine learning models, particularly tree-based ensemble methods, outperform conventional linear regression in predicting future market capitalization. These models are more effective in capturing nonlinear relationships and interaction effects that characterize complex firm-level financial and sustainability data.

Regarding the third research question, the analysis shows that financial indicators, especially firm revenue and profit margin, are the primary determinants of future market capitalization. Environmental, Social, and Governance factors contribute only marginally to predictive accuracy, suggesting that their influence is secondary to financial scale and efficiency in explaining absolute market value.

Overall, the findings suggest that while market capitalization remains largely driven by financial fundamentals, Environmental, Social, and Governance considerations play an important complementary role in enhancing transparency and supporting long-term sustainable competitiveness. This study contributes to the literature by providing global-scale evidence on the relative roles of financial and Environmental, Social, and Governance factors within artificial intelligence-based forecasting frameworks. Future research may extend this approach by examining industry-specific effects or distinguishing between developed and emerging markets.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 219-233, 2026
DOI: 10.55214/2576-8484.v10i3.12339

Funding:
This study is the result of a student research project funded by Hanoi University of Industry.

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.

References
[1] M. S. Dincă, C.-D. Vezeteu, and D. Dincă, "The relationship between ESG and firm value. Case study of the automotive industry," *Frontiers in Environmental Science*, vol. 10, p. 1059906, 2022. https://doi.org/10.3389/fenvs.2022.1059906

[2] S. Ahmad, W. Mohti, M. Khan, M. Irfan, and O. K. Bhatti, "Creating a bridge between ESG and firm's financial performance in Asian emerging markets: Catalytic role of managerial ability and institutional quality," *Journal of Economic and Administrative Sciences*, 2024. https://doi.org/10.1108/JEAS-01-2024-0004

[3] M. J. Flannery and R. R. Bliss, *Market discipline in regulation: Pre and post crisis. In A. N. Berger, P. Molyneux, & J. O. S. Wilson (Eds.), The Oxford handbook of banking*, 2nd ed. Oxford, United Kingdom: Oxford University Press, 2019.

[4] V. Agarwal and R. Taffler, "Comparing the performance of market-based and accounting-based bankruptcy prediction models," *Journal of Banking & Finance*, vol. 32, no. 8, pp. 1541-1551, 2008. https://doi.org/10.1016/j.jbankfin.2007.07.014

[5] V. Domanović, "The relationship between ESG and financial performance indicators in the public sector: Empirical evidence from the Republic of Serbia," *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*, vol. 27, no. 1, pp. 69-80, 2022. https://doi.org/10.7595/management.fon.2021.0032

[6] A. F. Christine, D. F. Hakam, Y. A. Nainggolan, S. K. Wiryono, and L. I. Hakam, "Environmental, social, and governance (ESG) impact on corporate financial strategy of energy and utilities companies worldwide," *Energy Strategy Reviews*, vol. 62, p. 101916, 2025. https://doi.org/10.1016/j.esr.2025.101916

[7] X. Xu and H. Zhao, "An empirical study on ESG evaluation of Chinese energy enterprises based on high-quality development goals—A case study of listed company data," *Sustainability*, vol. 16, no. 15, p. 6602, 2024. https://doi.org/10.3390/su16156602

[8] C. Zhao *et al.*, "ESG and corporate financial performance: Empirical evidence from China's listed power generation companies," *Sustainability*, vol. 10, no. 8, p. 2607, 2018. https://doi.org/10.3390/su10082607

[9] D. K. Kamugisha and H. Sun, "Environmental, social, and governance ratings and corporate financial performance: Evidence from emerging economies," *Total Quality Management & Business Excellence*, vol. 36, no. 13-14, pp. 1480-1498, 2025. https://doi.org/10.1080/14783363.2025.2563668

[10] A. Edmans, "The end of ESG," *Financial Management*, vol. 52, no. 1, pp. 3-17, 2023. https://doi.org/10.1111/fima.12413

[11] E. Clementino and R. Perkins, "How do companies respond to environmental, social and governance (ESG) ratings? Evidence from Italy," *Journal of Business Ethics*, vol. 171, no. 2, pp. 379-397, 2021. https://doi.org/10.1007/s10551-020-04441-4

[12] K. Luo and S. Wu, "Corporate sustainability and analysts' earnings forecast accuracy: Evidence from environmental, social and governance ratings," *Corporate Social Responsibility and Environmental Management*, vol. 29, no. 5, pp. 1465-1481, 2022. https://doi.org/10.1002/csr.2284

[13] B. Qian, S. Poshakwale, and Y. Tan, "'E' of ESG and firm performance: Evidence from China," *International Review of Financial Analysis*, vol. 96, p. 103751, 2024. https://doi.org/10.1016/j.irfa.2024.103751

[14] B. A. Alareeni and A. Hamdan, "ESG impact on performance of US S&P 500-listed firms," *Corporate Governance*, vol. 20, no. 7, pp. 1409-1428, 2020. https://doi.org/10.1108/CG-06-2020-0258

[15] C. Del Gesso and R. N. Lodhi, "Theories underlying environmental, social and governance (ESG) disclosure: A systematic review of accounting studies," *Journal of Accounting Literature*, vol. 47, no. 2, pp. 433-461, 2024. https://doi.org/10.1108/JAL-08-2023-0143

[16] D. Z. X. Huang, "Environmental, social and governance factors and assessing firm value: Valuation, signalling and stakeholder perspectives," *Accounting & Finance*, vol. 62, pp. 1983-2010, 2022. https://doi.org/10.1111/acfi.12849

[17] H. M. Bamahros *et al.*, "Corporate governance mechanisms and ESG reporting: Evidence from the Saudi stock market," *Sustainability*, vol. 14, no. 10, p. 6202, 2022. https://doi.org/10.3390/su14106202

[18] R. Y. C. Seow, "Transforming ESG analytics with machine learning: A systematic literature review using TCCM framework," *Corporate Social Responsibility and Environmental Management*, vol. 32, no. 6, pp. 7358-7389, 2025. https://doi.org/10.1002/csr.70089

[19] G. Friede, T. Busch, and A. Bassen, "ESG and financial performance: Aggregated evidence from more than 2000 empirical studies," *Journal of Sustainable Finance & Investment*, vol. 5, no. 4, pp. 210-233, 2015. https://doi.org/10.1080/20430795.2015.1118917

[20] T. Fu and J. Li, "An empirical analysis of the impact of ESG on financial performance: The moderating role of digital transformation," *Frontiers in Environmental Science*, vol. 11, p. 1256052, 2023. https://doi.org/10.3389/fenvs.2023.1256052

[21] M. Aydoğmuş, G. Gülay, and K. Ergun, "Impact of ESG performance on firm value and profitability," *Borsa Istanbul Review*, vol. 22, pp. S119-S127, 2022. https://doi.org/10.1016/j.bir.2022.11.006

[22] A. Gürsoy and G. S. Erbuğa, *A literature review on ESG score and its impact on firm performance. In Sustainability Development through Green Economics (Contemporary Studies in Economic and Financial Analysis)*. Leeds, United Kingdom: Emerald Group Publishing Ltd, 2024.

[23] H. Sandberg, A. Alnoor, and V. Tiberius, "Environmental, social, and governance ratings and financial performance: Evidence from the European food industry," *Business Strategy and the Environment*, vol. 32, no. 4, pp. 2471-2489, 2023. https://doi.org/10.1002/bse.3259

[24] J. Xu, W. Wu, and X. Feng, "The impact of ESG performances on analyst report readability: Evidence from China," *International Review of Financial Analysis*, vol. 102, p. 104056, 2025. https://doi.org/10.1016/j.irfa.2025.104056

[25] M. Praveen Kumar and N. V. Manoj Kumara, "Market capitalization: Pre and post COVID-19 analysis," *Materials Today: Proceedings*, vol. 37, pp. 2553-2557, 2021. https://doi.org/10.1016/j.matpr.2020.08.493

[26] B. T. Ewing and M. A. Thompson, "The role of reserves and production in the market capitalization of oil and gas companies," *Energy Policy*, vol. 98, pp. 576-581, 2016. https://doi.org/10.1016/j.enpol.2016.09.036

[27] F. Berg, J. F. Kölbel, and R. Rigobon, "Aggregate confusion: The divergence of ESG ratings," *Review of Finance*, vol. 26, no. 6, pp. 1315-1344, 2022.

[28] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," *The Review of Financial Studies*, vol. 33, no. 5, pp. 2223-2273, 2020.

[29] Z. Sautner, L. Van Lent, G. Vilkov, and R. Zhang, "Firm-level climate change exposure," *The Journal of Finance*, vol. 78, no. 3, pp. 1449-1498, 2023. https://doi.org/10.1111/jofi.13219

[30] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84-90, 2022. https://doi.org/10.1016/j.inffus.2021.11.011

[31] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001. https://doi.org/10.1023/A:1010933404324

[32] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3-42, 2006. https://doi.org/10.1007/s10994-006-6226-1

[33] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937-1967, 2021. https://doi.org/10.1007/s10462-020-09896-5

[34] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 1997. https://doi.org/10.1006/jcss.1997.1504

[35] G. Ke *et al.*, *LightGBM: A highly efficient gradient boosting decision tree. In I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 30*. Long Beach, CA, USA: Curran Associates, Inc, 2017.

[36] L. Ostroumova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, *CatBoost: Unbiased boosting with categorical features. In Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Red Hook, NY: Curran Associates, Inc, 2017.

[37] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247-1250, 2014. https://doi.org/10.5194/gmd-7-1247-2014

[38] T. O. Kvålseth, "Cautionary note about R 2," *The American Statistician*, vol. 39, no. 4, pp. 279-285, 1985. https://doi.org/10.1080/00031305.1985.10479448