ORIGINAL RESEARCH

Environmental, social and governance (ESG) rating prediction using machine learning approaches

Mohammad Ashraful Ferdous Chowdhury[1,2] [iD] · **Mohammad Abdullah[3]** ·
Md. Abul Kalam Azad[4] · **Zunaidah Sulong[3]** · **M. Nazmul Islam[5]**

Abstract

The study's objective is to predict the environmental, social, and governance (ESG) ratings of firms. Applying six machine learning algorithms, we collect a global data sample of 6166 firms in 73 countries from 2005 to 2019. We use firm-specific and macroeconomic predictors in the model and find that Random Forest Classifier provides the highest accuracy (78.50%) among the six machine learning algorithms by considering Kappa, area under the curve, receiver operating characteristic, and logLoss. The variable importance factor reveals that the lagged ESG score has the highest contribution to the model. Firm size has the second highest, and debt to equity ratio has the third-highest contribution, which indicates that a firm's total assets and a firm's financial leverage impact the ESG rating. In addition to a contribution to the growing body of ESG literature, the study's findings can help practitioners, firm regulators, and policymakers in social and environment-friendly decision-making and investors in investment decisions.

✉ Mohammad Ashraful Ferdous Chowdhury
ashraful_ferdous@yahoo.com; mohammad.chowdhury@kfupm.edu.sa

Mohammad Abdullah
si3861@putra.unisza.edu.my

Md. Abul Kalam Azad
kalam@iut-dhaka.edu

Zunaidah Sulong
zunaidah@unisza.edu.my

M. Nazmul Islam
nzmuljoy@gmail.com; islam.nazmul@bracu.ac.bd

[1] Interdisciplinary Research Center (IRC) for Finance and Digital Economy, KFUPM Business School, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

[2] Department of Business Administration, Shahjalal University of Science and Technology, Sylhet, Bangladesh

[3] Faculty of Business and Management, Universiti Sultan Zainal Abidin, Kuala Nerus, Terengganu, Malaysia

[4] Business and Technology Management Department, Islamic University of Technology, Gazipur, Bangladesh

[5] BRAC Business School, BRAC University, Dhaka, Bangladesh

🙋 Springer

1 Introduction

With the rapid growth of industrialization, firms are chasing profit and ignoring environmental pollution. Corporate activities have infiltrated the natural environment's equilibrium, triggering global warming, climate change, and harmful waste generation (Crace & Gehman, 2022; Dhasmana et al., 2023). The world has now begun to pay attention to sustainable development and the use of available resources without sacrificing the requirements of future generations. Since poor environmental quality can be a significant obstacle to economic development, all stakeholders, including institutions, customers, and investors, should think about how they can become more sustainable (Chung et al., 2023; Nekhili et al., 2021). As a result, environmental, social, and corporate governance (ESG) has emerged as a tool to assess the extent to which a corporation works on behalf of societal purposes that go beyond the corporation's function of maximizing profits on behalf of its shareholders (Billio et al., 2021; Kalaitzoglou et al., 2021; Liagkouras et al., 2020). ESG assesses a company's efforts in terms of energy efficiency, greenhouse gas emissions, waste, water, and resource management.

ESG rating goals include not only environmental issues, but also discrimination (such as gender and race), human rights issues (such as child labor), and incorporating information from smart sanctions lists issued by nations and global companies around the globe (D'Amato et al., 2021; D'Apice et al., 2021; Semenova & Hassel, 2019). In 2019, 300 mutual funds with ESG mandates collected $20 billion in net flows, which was four times the level in 2018, demonstrating investor interest in ESG (Sharma et al., 2020). Over 3000 institutional investors and service providers have joined the Principles of Responsible Investment (PRI), a commitment to include ESG factors in their analysis of investment and decision-making methods. However, financial institutions also use ESG rating as a lending decision tool. Their emphasis is not only to discourage companies with financial problems to keep their portfolios profitable, but there is also a rising interest in putting the burden on publicly traded companies to increase their ESG practices (Hisano et al., 2020).

The theoretical underpinnings of this study are grounded in stakeholder theory, agency theory, legitimacy theory, and signaling theory to illuminate the complexities of firm, industry and country level factors and their relationship to ESG (Baldini et al., 2018; Chung et al., 2023; Weber, 2014). Stakeholder theory underscores the positive connection between larger firm size and voluntary disclosure, driven by the aspiration to maintain harmonious stakeholder relationships and minimize environmental and social impacts (Ansoff, 1965). Signaling theory underscores factors such as firm size, auditor type, profitability, gearing, and industry dynamics as key determinants of disclosure, because firms strategically use information dissemination to address information asymmetry with external parties (Spence, 1978). Conversely, agency theory emphasizes ownership structure, specifically the impact of dispersed ownership on the need for robust information disclosure to align management and shareholder interests (Jensen & Meckling, 1976). The convergence of institutional and stakeholder theory sheds light on broader institutional pressures that influence firms to adopt sustainability reporting practices. This convergence fosters a comprehensive understanding of the motivations behind ESG-related disclosure decisions (DiMaggio & Powell, 1983).

Legitimacy theory explains that there is a relationship between firm size and a firm's intention to disclose voluntary information because of public pressure (Perrow, 1970).

Previous studies have shown the effect of ESG on different aspects of the firm, i.e., ownership characteristics, leadership characteristics, systematic risk, credit risk, firm profitability, and stock return (Billio et al., 2021; Broadstock et al., 2020; Đặng et al., 2022; Darnall et al., 2022; Drempetic et al., 2020; Nekhili et al., 2021; Sauerwald & Su, 2019; Sharma et al., 2020). Little attention has been given to developing ESG rating prediction methodology and theory (D'Amato et al., 2021, 2022). Prediction of an ESG score would offer different levels of advantage to investors, the environment and society. For example, an ESG prediction may advance the sustainable development of companies by identifying environmental issues, i.e., detecting emissions levels and energy consumption. Moreover, enhancing the arguments of corporate social performance (CSP) backgrounds, an ESG prediction will allow firm regulators to take the necessary action if firms' ratings tend to become lower. Investors and portfolio managers will be able to make investment decisions based on the rating. Financial institutions will be able to make the lending decisions and predict their borrower ESG rating. Therefore, we aim to predict ESG ratings using firm and macroeconomic variables. Traditional econometrics, i.e., logistic regression and the generalized linear model, allow only binary classification. However, machine learning models aid multiclass classification, which is quite important in the finance field, such as credit rating, bankruptcy prediction (Abdullah, 2021; Barbeito-Caamaño & Chalmeta, 2020; Jiang et al., 2022; Kumar et al., 2022; Wang et al., 2022). This study applies six machine learning algorithms to develop an ESG prediction model using 73 countries' 6166 firms' data and finds the random forest classifier outperforms other models with 78.50% accuracy.

This study makes several contributions. First, the study contributes to the growing body of literature by adding evidence of ESG rating forecasting. Earlier studies mostly focused on ESG rating forecasting with a small sample (D'Amato et al., 2021, 2022, 2023), and ESG text classification (Lee & Kim, 2023; Sokolov et al., 2021). We fill the gap of ESG rating classification using global data and wide range of multiclass classification models. Second, our study extends the studies by D'Amato et al., (2021, 2022), who attempt to forecast ESG scores using a machine learning random forest regression model for a time series model using 109 (D'Amato et al., 2021) and 401 (D'Amato et al., 2022) firm-level data from STOXX Europe 600 Index from 2014 to 2018 (D'Amato et al., 2021) and 2009–2019 (D'Amato et al., 2022). Our study is more comprehensive and considers 6166 firms from 73 countries from 2005 to 2019 and uses six machine learning classification models for a robust ESG rating prediction model.

Third, the insights gained from the variable importance analysis provide a more in-depth understanding of the fundamental drivers behind ESG ratings. This discovery is a valuable resource for policymakers, practitioners, and regulatory bodies. It allows them to develop well-informed policies and strategies to maintain and improve ESG ratings. Finally, our findings are substantiated by well-established theories including the stakeholder theory, agency theory, legitimacy theory, and signaling theory. This emphasizes the importance of factors specific to each firm and country in determining ESG performance. As a result, our findings not only add to existing knowledge, but also provide useful insights for advancing theoretical frameworks in ESG studies. Our findings pave the way to develop an ESG score prediction model, which may be utilized by policymakers, practitioners, investors, regulatory bodies, portfolio managers, and stakeholders to obtain up-to-date information for policy development and investment decision-making.

The paper continues as follows: Sect. 2 provides a literature review. Section 3 elaborates the study's data and methodology. Section 4 presents the empirical results. Section 5 has a

detailed discussion of the results. Finally, Sect. 6 concludes this study with implications and future study directions.

2 Literature review

Corporation stakeholders have been consistently demanding assurance of "sustainability" throughout corporate operations (Goodell et al., 2021). This concern has forced corporations to adopt environmental, social and governance (ESG) practices as one of their core business functions (Crespi & Migliavacca, 2020). ESG practice, first introduced by the United Nation's Principles of Responsible Investment—UNPRI (Sharma et al., 2020), has been widely recognized as a standard to measure sustainable performance. The 'E' stands for environmentally-related components such as water use, renewable/non-renewable resources, and emissions. The 'S' includes societal and community-related elements, including health and safety issues, workplace diversity, child labor, and labor strikes. Management and board components comprising board meetings, diversity, board attendance form the 'G' factor for Governance. The significance of ESG information disclosure can be observed through the lenses of corporations' financial value. Such disclosure is used as a tool to fulfill multiple stakeholders' expectations regarding business (Lokuwaduge & Heenetigala, 2017).

Early studies reveal that better ESG disclosure attracts both financial and non-financial stakeholders, resulting in enhanced firm financial structures (Henrique et al., 2019). Both types of stakeholder expect that firms maintain transparency in their company ESG information disclosure. Asset managers and investors, a type of financial stakeholder, increasingly integrate ESG materials into their investment decision-making (Friede et al., 2015). Since the ESG data are relevant to investment performance, financial stakeholders are paying more attention to firms with high ESG disclosure (Chauhan & Amit, 2014; Drempetic et al., 2020; Hörisch et al., 2015; Yu & Choi, 2016).

The literature also suggests that the determinants of ESG disclosure are derived from three relatively related theories: the institutional, accountability, and legitimacy theories (Baldini et al., 2018; Weber, 2014). From the institutional theory, an organization resorts to ESG practice because of pressure from institutions such as non-government and independent organizations. Such institutional or broader social structure determines the survival of the organization. The accountability perspective concerns organizations report on ESG issues because of their obligations to stakeholders. Central to the legitimacy theory is the organization's fulfillment of society's needs and desires. Thus, an organization tries to secure or increase legitimacy by prioritizing society's requirements. One way to ensure legitimacy is for the organization to undertake sustainability management activities. Previous studies use that theoretical stance to identify a list of country- and firm-level ESG disclosure determinants. The country-level drivers include the country's political, labor, and cultural system, and the firm-level drivers comprise the firm's visibility to both investors and non-investors (Baldini et al., 2018).

Another stream of studies examines the relationship between ESG and sustainability. Buallay (2019) examine the association between ESG and European bank performance. They documented that ESG factors positively influence bank performance, suggesting that European banks are placing a greater focus on transparency, which in turn leads to improved financial performance. Using the Thomson Reuters ESG rating dataset, Drempetic et al. (2020) examine the impact of firm size on ESG rating and find a positive relationship between ESG and firm size. They concluded that larger corporations are more sustainable than smaller

corporations, and larger corporations have a greater advantage in terms of focusing on reporting transparency. Escrig-Olmedo et al. (2019) analyzed the criteria of ESG ratings and how an ESG rating contributes to sustainable development. Their findings indicate that rating agencies should incorporate a broader range of sustainability indicators to ensure a comprehensive assessment of sustainable development.

Alsayegh et al. (2020) find ESG disclosure enhance corporate sustainability performance. They found that transparent ESG disclosure, delivering accurate information about a company, creates better prospects for building trust among stakeholders, ultimately resulting in enhanced company performance. Johnson (2020) investigates ESG data, its challenges, and proposed solutions for establishing a shared ESG language to effectively measure sustainable outcomes in global investments. Using the resource-based view and the stakeholder capitalism theory, Bhandari et al. (2022) find a concave shaped relationship between ESG and competitive advantage. They contends that current attributes of a resource base for sustained competitive advantage, based on the resource-based view, overlook the crucial aspect of "ESG friendliness" in a resource. Garcia et al. (2019) examined the relationship between ESG rating and firm financials and find firm market capitalization is a significant determinant of ESG performance, with larger companies generally exhibiting better performance. Additionally, firms in sensitive industries demonstrate strong environmental performance even after accounting for size and location.

Several mathematical tools and methods have been in use for prediction. Machine learning is becoming a widely used methodological tool to explain and forecast market trends in the financial industry (Henrique et al., 2019). It widely covers multiple facets of an algorithm for recognizing patterns and making decisions (Henrique et al., 2019). This tool, integrated with an artificial intelligence system, tries to extract patterns learned from historical data in a way regarded as training and learning. From traditional hedge fund managers to contemporary fintech service providers, all are extensively applying machine learning expertise (Holzinger et al., 2018). Today's financial system, in other words, is supported by the generation of machine-readable data that act as a catalyst to disrupt and transform the financial industry.

An amount evidence exists regarding machine learning applications in the financial industry. Abdullah et al. (2023b) use a deep learning model with textual analysis for stock price forecasting. Barboza et al. (2017) combine traditional analytical tools (logistic and discriminant analysis) with the machine learning to analyze credit risk and bankruptcy issues. Ciampi et al. (2021) review SME-default prediction models. Abdullah et al. (2023a) apply machine learning models to forecast bank non-performing loans and find the superior performance of the random forest model. This list implies persistent growth of the application of machine learning tools and techniques in financial firms to explore multiple, diverse facets of the financial system.

In the implication of machine learning in ESG studies, there is growing body of literature that uses machine learning. Raman et al. (2020) use machine learning algorithms to examine earning call transcripts and document that ESG factors are vital for corporate policy. Unlike other studies, D'Amato et al. (2022) forecast ESG scores using STOXX Europe 600 index 401 firm data. They find the random forest model output better at forecasting the ESG score, indicating random forest model have better capacity to grasp nonlinear of the ESG model. Sokolov et al. (2021) develop an ESG rating model using deep learning that can predict the ESG rating using unstructured text. Lanza et al. (2020) propose a machine learning model to examine contradictions in ESG scores. They find information extracted from ESG factors significantly enhances the portfolio optimization model by providing better information about the companies' capacity to handle climate change risk, specifically transition risk.

De Lucia et al. (2020) develop a European firm financial performance prediction model using ESG indicators and macro-economic factors and document a positive relationship with firm financial performance and ESG practices. They argued that company-specific conditions establish a structural framework that connects practices with performance, and this correlation is influenced by a company's growth opportunities, which in turn are heightened by higher levels of transparency and reporting. Antoncic (2020) finds that big data analytics can avoid bias in ESG ratings and can eliminate greenwashing. Recently, Lee and Kim (2023) developed an ESG text classifier to extract ESG related information from Korean sustainability reports. Their proposed natural language processing model can achieve 86.66% accuracy.

Overall, the above literature review suggests that studies are mostly based on stakeholder, agency, legitimacy, and signaling theories. There is growing body of literature that uses machine learning and other big data analytics approaches in ESG related studies. However, studies are yet to be conducted to discover the implications of machine learning in ESG rating predictions using multiple countries' data.

3 Data and methodology

We aim to develop a machine learning-based ESG rating prediction model that policymakers and stakeholders can apply using a combination of firm-specific and macroeconomic predictors. The paper's framework is presented in Fig. 1. It elaborates the feature selection, data collection, data cleaning, feature elimination, model selection, and cross-validation processes.
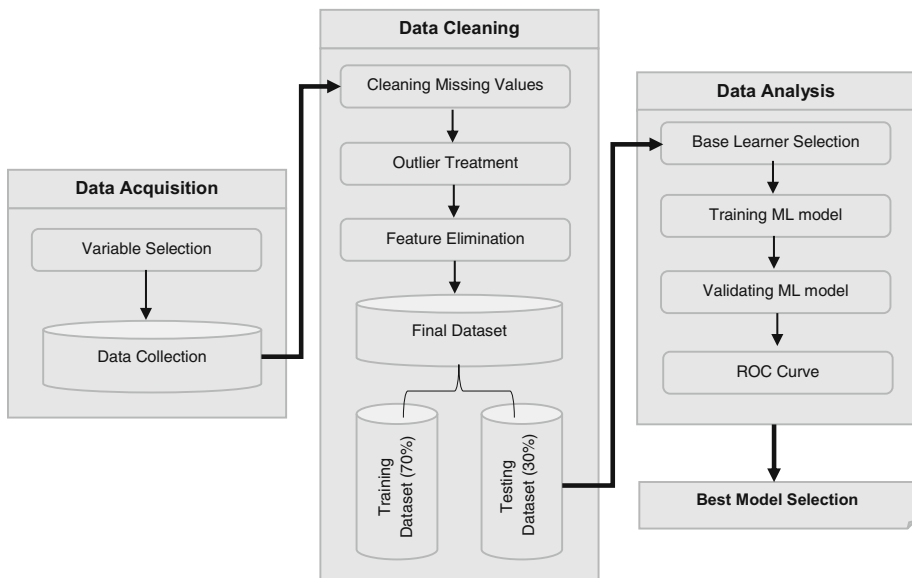

Fig. 1 Classification framework used in this study of machine learning models

Table 1 Specification of the variables used in the models

Variable	Description	Varimp*	Data Source
Outcome variable			
ESG	ESG rating (A, B, C, D)	N/A	Datastream
Predictors			
LESG	Lag ESG score	15,063.23	Datastream
SIZE	Size = Log total asset	1108.91	Datastream
DER	Total liability/total equity	455.25	Datastream
EPS	(Net Income- preferred Dividend)/Shares Outstanding	309.07	Datastream
TIE	Earnings before interest, taxes, depreciation, and amortization/interest expense	84.87	Datastream
ROE	Return on equity	0.00	Datastream
ROA	Return on assets	0.00	Datastream
NPM	Net income/ revenue	0.00	Datastream
GDP	Log gross domestic product	174.42	World Bank
GDPG	GDP growth rate	25.19	World Bank
UNEM	Unemployment rate	22.22	World Bank
INF	Inflation rate	0.00	World Bank

Varimp Variable importance. Variables with lower Varimp are removed from final model

3.1 Data and pre-processing

Before data collection, the authors conducted a rigorous review of the literature to determine the related ESG score predictors. The selection of variables is motivated from the theoretical foundation of the stakeholder, agency, legitimacy, and signaling theories. Predictors are then categorized into firm-specific and macroeconomic predictors of the ESG rating. Table 1 presents the selected predictors with ESG categorical rating as the main outcome variable. Refinitiv captures and calculates over 500 business-level ESG measures for ESG score calculation. It shortlists 186 of the most comparable metrics per industry driving a company's overall appraisal and scoring process. Refinitiv offers a company rating that ranges from A + to D− and is divided into 12 categories. However, we modify their standard rating system into four rating categories A, B, C, and D. The descriptions are presented in Table S.1. A total of 12 predictors are selected where eight predictors are chosen in the firm-specific predictor category, and four are selected in the macroeconomic predictor category.

Data of ESG ratings and firm-specific predictors were collected from the Thomson Reuters DataStream database. Thomson Reuters, ESG rating report has 7096 firms' data in the most recent report. Fifteen fiscal year data from 2005 to 2019 were collected to consider before and after the financial crisis. At first, observations with missing values are excluded from the dataset, which produce an unbalanced panel data set of 45,175 observations from 6166 firms (see details in Table S.2). Outliers were Winsorized[1] to produce the final shape of data as in Fig. 2. The Recursive Partitioning and Regression Trees (RPART) algorithm was applied

---

[1] Winsorizing is the process of converting statistics by restricting extreme values in the data to lessen the impact of potentially erroneous outliers (Nyitrai & Virag, 2019). We have used the 1% and 99% winsorizing.
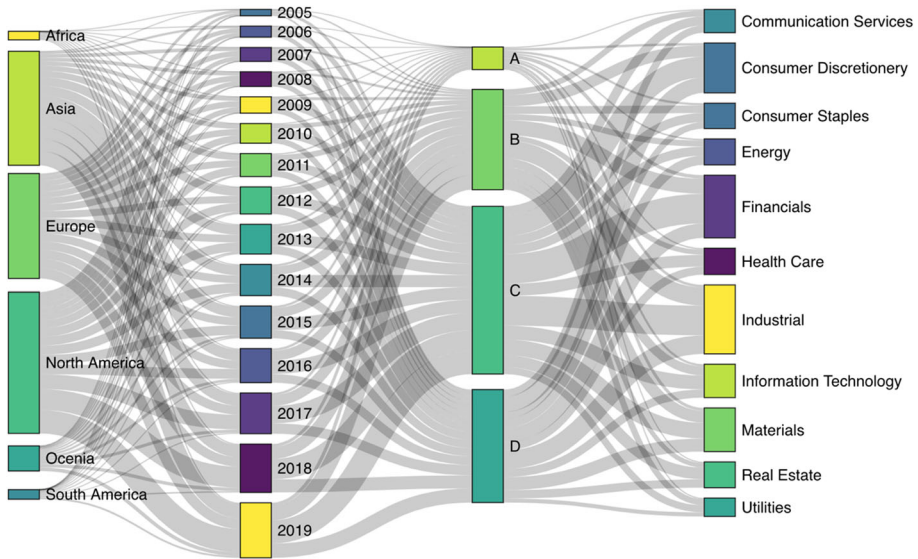
Fig. 2 Continent, year and industry of environment, society and governance scores' distribution

for feature elimination by calculating the variable importance factor (VarImp).[2] VarImp results are presented in Table 1 (VarImp low-value variables are removed from the dataset). Therefore, ROA, ROE, NPM, and INF are removed from the dataset. The dataset is then split into training and testing datasets by considering 70% of the data (2005–2014) for training and the rest, 30% (2015–2019), for the out-of-sample testing.

3.2 Methodology and hyperparameter optimization

Our methodology involves multiclass classification, which is different from binary classification and cannot be modeled using conventional methods (Abdullah, 2021; Barbeito-Caamaño & Chalmeta, 2020; Jiang et al., 2022; Kumar et al., 2022; Wang et al., 2022). Therefore, we use multiclass classification machine learning models. There is a wide range of machine learning models available for classification (Kotsiantis et al., 2006). Following the literature (Abdullah et al., 2023a; Barboza et al., 2017; D'Amato et al., 2022), we chose the Artificial Neural Networks Classifier, Bagging Classifier, k-Nearest Neighbors Classifier, Naive Bayes Classifier, Random Forest Classifier, and Support Vector Machines Classifier.

These selected models have proven to be superior in several financial classification problems, e.g., bankruptcy prediction, credit rating, and credit default prediction (Abdullah et al., 2023a; Barboza et al., 2017; D'Amato et al., 2022). The motivation behind choosing multiple models is to select the best model for ESG rating prediction from these widely used models. We applied the R statistical software caret packages for machine learning classification model training and testing.[3] Selected classifiers and their base learners are listed in Table

---

[2] Variable importance is a metric that quantifies the significance of a variable's contribution to a model. It is determined by assessing how much the introduction of a variable into a model reduces errors, measured by the decrease in error that occurs when the variable is added(Archer & Kimes, 2008).

[3] caret R package is available at: https://github.com/topepo/caret.

Table 2 Selected machine learning classifiers and base learners

Model	Method of caret	Tuning parameters	Base learner
Artificial Neural Networks Classifier (ANNC)	nnet	Size, decay	Neural network
Bagging Classifier (BGC)	treebag	None	Bagged CART
k-Nearest Neighbors Classifier (KNNC)	kknn	kmax, distance, kernel	k-nearest neighbors
Naive Bayes Classifier (NBC)	nb	fL, usekernel, adjust	Naive Bayes
Random Forest Classifier (RFC)	rf	mtry, ntree	Random Forest
Support Vector Machines Classifier (SVMC)	svmLinear	C	SVM with Linear Kernel

Size, Number of hidden units; decay, Weight decay; kmax, Maximum number of neighbors; distance, Distance; kernel, Kernel; fL, Laplace correction; usekernel, Distribution type; adjust, Bandwidth adjustment; Mtry, Number of randomly selected predictors; ntree, number of trees; C, Cost

2. Hyperparameters[4] are optimized by ten times cross-validation and five repetitions. Other tuning parameters and model details are discussed as follows:

3.2.1 Artificial neural networks classifier (ANNC)

Artificial neural networks are simple electronic networks of neurons based on the human brain's neural structure. They go through each observation one at a time, learning by comparing their categorization to the real categorization of the observation (May et al., 2010). The first pool of observation classification errors is passed back into the network and used to tweak the network's algorithm for subsequent iterations. Neurons are structured into three layers: input, hidden, and output layers. The algorithm starts with the input layer where there are only a few neurons that take input. They then pass them to the hidden layer. Finally, observations go to the output layer according to their classification. Equations (1) and (2) illustrate the ANNC, where X is the input matrix, and W is the weight matrix. Moreover, z_k defines the output function where nH denotes the number of perceptions in the hidden layer and w_0 denotes the errors. For this study, the neural network is taken as the base learner with size $= 10$ and decay $= 0.01$ as hyperparameter tuning after a grid search.

$$f(X) = \begin{cases} 1 & W^T X + w_0 > 0 \\ 0 & W^T X + w_0 \leq 0 \end{cases} \tag{1}$$

$$z_k = f(\sum_{j=1}^{nH} w_{kj} f\left(\sum_{i=1}^{d} x_i w_{ji} + w_{j0}\right) + w_{k0} \tag{2}$$

3.2.2 Bagging classifier (BGC)

A bagging classifier is an ensemble algorithm that assigns base classifiers to random subgroups of the main dataset and then combines their distinct predictions to generate a final

---

prediction. This type of algorithm reduces error by introducing random inputs for prediction. Each base classifier of the bagging classifier trains multiple models by parallel taking a different sample from the training dataset (Breiman, 1996). Equation (3) presents the construct of the bagging classifier here, where $\varphi(x, L)$ denotes predictors, and j denotes the class. For this study, Bagged CART is selected as the base learner of Bagging Classifier; no hyperparameter is available for this learner.

$$Q(j|x) = P(\varphi(x, L) = j) \tag{3}$$

3.2.3 k-Nearest neighbors classifier (KNNC)

K-nearest neighbors is one of machine learning's most basic but crucial classification algorithm. KNNC is a straightforward non-parametric technique used for classification and regression. In the k-nearest neighbors algorithm, different categorical classes are assigned corresponding to the prediction features of its k-nearest neighbor. Keller et al. (1985) proposed the first testing version of KNNC and named it the "fuzzy k-nearest neighbors classifier algorithm"Click or tap here to enter text. According to this methodology, all datasets are distributed to a different class. The formula for this method is:

$$u_i = \frac{\sum_{j=1}^{k} u_{ij} \left(1/\|x - x_j\|^{2/(m-1)}\right)}{\sum_{j=1}^{k} \left(1/\|x - x_j\|^{2/(m-1)}\right)} \tag{4}$$

Here $i = 1, 2, .., n$, and $j = 1, 2, .., k$ where n is the class number and k is the nearest neighbor. Here, m is denoted as a fuzzy strength parameter and is used to determine how strongly the relationship is weighted at the time of calculation of each neighbor's contribution to the group value. The term u_{ij} is the group degree of the model, x_i belongs to the training set of class i, between the k nearest neighbors of x. The following tuning parameters are used for final model training after grid search: distance = 2, kmax = 9, and kernel = optimal.

3.2.4 Naive Bayes classifier (NBC)

The Naive Bayes Classifier is classified as a probabilistic classifier that incorporates Bayes' hypothesis to achieve reliable predictions among input variables. The architecture of NBC consists of nodes and edges, wherein the nodes symbolize the inputs, and the edges establish connections between nodes leading to the output. This configuration forms a directed acyclic graph (Pearl, 1988). A different set of conditional probabilities works as a probabilistic input in the NBC model. For example, the credit rating of the company model will be trained based on previously rated companies, and NBC will output the rating by predicting the likelihood of the prediction class.

$$P(A|B) = P(B|A)P(A)/P(B) \tag{5}$$

Equation (5) demonstrates Bayes' main theory. Here, $P(A|B)$ is the function of two inputs that represent the proposition of probabilities. $P(B|A)$ represents the likelihood and $P(A)$ is a function of class probability and, finally, $P(B)$ represents the prediction probabilities. The Naive Bayes Classifier model assumes the following condition of independence:

$$P_i \perp \{P_1, P_2, \ldots, P_i - 1, P_i + 1, \ldots.P_n\}|V \tag{6}$$

where: $i = 1, 2, \ldots, n$. This theory declares that predictors, $P_1, P_2, \ldots, P_n$ are conditionally equally autonomous for prediction (Sun & Shenoy, 2007); fL = 0, usekernel = TRUE and adjust = 1 parameters are used for hyperparameter tuning after the grid search.

3.2.5 Random forest classifier (RFC)

The random forest classifier algorithm is constructed with many distinct decision trees that work together as an ensemble. Each tree in the random forest produces a distinct class prediction. The class with the maximum importance output is the prediction (Ho, 1995). The following equation constructs the random forest importance model as:

$$ni_j = w_j c_j - W_{L(j)} C_{L(j)} - W_{R(j)} C_{R(j)} \tag{7}$$

Here, j denotes each node of the tree, ni is the importance function, w_j denotes weight, C_j denotes nodes impurity, $L(j)$ and $R(j)$ denotes left and right node split. In this study mtry = 2 and ntree = 550 is used as final hyperparameter.

3.2.6 Support vector machines classifier (SVMC)

This classifier is a hybrid statistical learning model. Cortes and Vapnik (1995) introduced support vector machine models in 1995. They defined SVM as converting the original data into high-dimensional data to create prediction classesClick or tap here to enter text. Support vector machines are suitably straightforward to examine them statistically. SVMC aims to identify a hyperplane in a dimensional sample space that accurately classifies data points. Hence, the support vector machine is constructed as:

$$\frac{1}{2} w^T w + c \sum_{i=1}^{M} \in_i \tag{8}$$

$$y_i \left[w^T \varphi(x_i) + b \right] \geq 1 - \in_i \tag{9}$$

where $i = 1, 2, 3 \ldots, N, \xi i \geq 0$ are the errors associated with classification cost; c, y_i are the outcomes variables; and $\varphi(x)$ is transform space. We apply the SVMC algorithm as a base learner with C = 1 used for hyperparameter tuning after the grid search.

3.3 Performance metrics

Models are compared using different performance metrics after model training and testing. We use logLoss, AUC, prAUC, Accuracy, Kappa, F1, Sensitivity, and Specificity to measure model performance. Sensitivity measures a model's power to predict true positive rates through all classes and specificity measures a model's power to predict true negative rates across all classes. Equations (10) and (11) depict the sensitivity and specificity:

$$Sensitivity = NTP/(NTP + NFN) \tag{10}$$

$$Specificity = NTN/(NTN + NFN) \tag{11}$$

where the number of true positive classifications correctly classified is NTP, and the number of true negative classifications correctly classified is NTN. NFN is the number of false-positive classifications, and NFP is the number of false-positive classifications from Eq. (11). If the classification error is low, the sensitivity and specificity values will be close to one.

The efficiency of a machine learning classification model with a binary or multiclass prediction class is measured by logarithmic loss (logLoss). The primary objective of a machine learning model is to reduce the logLoss value as much as possible. When the logLoss is close to 0, the model is considered perfect. The area under the receiver operating characteristic (ROC) curve are represented by AUC. The ROC curve is a graphical representation of a classifier's discriminant threshold value. A relatively high AUC value indicates that the model has a high predictive power. The Precision recall AUC (prAUC) determines the predicted positive class percentage. Cohen's Kappa statistics, which indicate the degree of good model fitness, is an enhanced indicator for multiclass predictive modeling. According to Landis and Koch (1977), the value of Kappa can be classified into six categories: (i) 0 percent suggests that the model is not a good fit; (ii) 0 percent to 20 percent indicates low significance; (iii) 21 percent to 40 percent indicates fair significance; (iv) 41 percent to 60 percent indicates moderate significance; (v) 61 percent to 80 percent suggests considerable significance; and (vi) 81 percent indicates significant significance.

4 Empirical results

4.1 Exploratory data analysis results

First, the descriptive statistics are applied to confirm the validity of the dataset. The results are presented in Table 3. The different sectors and continents plot the yearly ESG score in Fig. 2. Figure 2 indicates that most of Thompson's companies rated in the ESG rating are from North America. There are fewer observations from the dataset start period (2005–2007). Also, the amount of A category companies is very low compared with other rating scales. Nevertheless, the descriptive statistics results in Table 3 indicate low variation among the predictors throughout the full training and testing dataset. However, there is high variation in LESG, and TIE among all datasets. Except for SIZE and GDP, the predictors are positively skewed.

Pearson correlation analysis is used to validate the correlations among the predictors; the results are presented in Table 3. The main study outcome variable is ESG rating, a categorical variable that is removed from the correlation analysis. However, we consider the lagged effect of the ESG score (LESG) that describes possible correlations among the variables. The results indicate that there is a statistically significant correlation among all variables with LESG. However, except for GDP, TIE, and GDPG that are negatively correlated, all variables are positively correlated with LESG. These statistical indicators define the validity of the training and testing dataset. Thus, we proceed with further analysis by developing six machine learning models.

4.2 Model training and testing results

Six machine learning models were trained and validated on confirming the dataset rationality to develop the machine learning models. Figure 3 illustrates all models' confusion matrix. The results of the ANNC confusion matrix show that there is less misclassification. For (i) ESG rating A, there are 743 actual observations in the testing dataset and the model predicted 816; (ii) for B, there are 3409 actual observations, whereas the model predicted 3340; (iii) for C, there are 5676 observations and the ANNC model predicted 5638; and iv) for rating D the model predicted 3759 where there are 3725 observations. Thus, these confusion matrix

Table 3 Descriptive statistics and correlation analysis of the firm sample and sub–samples

Variable	N	Mean	Median	SD	Min	Max	Skw	Kur	LESG	SIZE	DER	EPS	GDP	TIE	GDPG	UNEM
Panel A: full dataset																
LESG	45,175	40.96	38.65	20.50	0.19	95.07	0.32	− 0.78	1							
SIZE	45,175	22.29	22.32	1.76	18.86	25.55	− 0.07	− 0.58	0.37**	1						
DER	45,175	2.59	1.46	2.96	0.25	11.65	1.98	3.02	0.13**	0.45**	1					
EPS	45,175	1.66	0.88	2.01	− 0.28	7.25	1.48	1.35	0.15**	0.25**	0.06**	1				
GDP	45,175	28.72	28.70	1.52	20.56	30.50	− 0.75	0.62	− 0.05**	− 0.03**	0.01	0.17**	1			
TIE	45,175	21.65	5.71	42.68	− 3.60	174.88	2.77	6.70	− 0.03**	− 0.16**	− 0.17**	0.05**	0.03**	1		
GDPG	45,175	0.02	0.02	0.02	− 0.22	0.25	0.50	8.36	− 0.07**	0.09**	− 0.02**	− 0.13**	0.01	0.03**	1	
UNEM	45,175	0.06	0.05	0.04	0.04	0.29	3.31	13.64	0.12**	0.08**	0.06**	− 0.05**	− 0.17**	− 0.04**	− 0.18**	1
Panel B: training dataset																
LESG	31,622	40.98	38.62	20.56	0.19	94.77	0.33	− 0.78	1							
SIZE	31,622	22.29	22.32	1.76	18.86	25.55	− 0.07	− 0.58	0.37**	1						
DER	31,622	2.60	1.46	2.98	0.25	11.65	1.96	2.92	0.14**	0.45**	1					
EPS	31,622	1.66	0.88	2.02	− 0.28	7.25	1.48	1.35	0.15**	0.25**	0.06**	1				
GDP	31,622	28.73	28.70	1.52	20.56	30.50	− 0.75	0.62	− 0.05**	- 0.04**	0.01	0.16**	1			
TIE	31,622	21.80	5.73	42.85	− 3.60	174.88	2.75	6.60	− 0.03**	− 0.16**	− 0.17**	0.05**	0.03**	1		
GDPG	31,622	0.02	0.02	0.02	− 0.22	0.25	0.51	8.19	− 0.08**	0.09**	− 0.01*	− 0.13**	0.01	0.03**	1	
UNEM	31,622	0.06	0.05	0.04	0.03	0.29	3.33	13.91	0.12**	0.08**	0.05**	− 0.05**	− 0.17**	− 0.04**	− 0.18**	1
Panel C: testing dataset																
LESG	13,553	40.93	38.70	20.35	0.36	95.07	0.31	− 0.76	1							
SIZE	13,553	22.28	22.32	1.75	18.86	25.55	− 0.08	− 0.57	0.37**	1						
DER	13,553	2.56	1.46	2.92	0.25	11.65	2.02	3.24	0.13**	0.45**	1					
EPS	13,553	1.65	0.89	2.00	− 0.28	7.25	1.47	1.36	0.15**	0.25**	0.06**	1				
GDP	13,553	28.72	28.70	1.53	20.56	30.50	− 0.74	0.63	− 0.05**	− 0.03**	0.01	0.17**	1			
TIE	13,553	21.29	5.68	42.28	− 3.60	174.88	2.81	6.95	− 0.02*	− 0.15**	− 0.16**	0.05**	0.04**	1		
GDPG	13,553	0.02	0.02	0.02	− 0.22	0.25	0.49	8.77	− 0.07**	0.09**	− 0.03**	− 0.12**	0.02	0.03**	1	
UNEM	13,553	0.06	0.05	0.04	0.03	0.28	3.28	13.06	0.12**	0.08**	0.07**	− 0.06**	− 0.18**	− 0.04**	− 0.17**	1

SD Standard Deviation, *Skw* Skewness, *Kur* Kurtosis, *Min* Minimum, *Max* Maximum, ** denotes 0.01 significance level
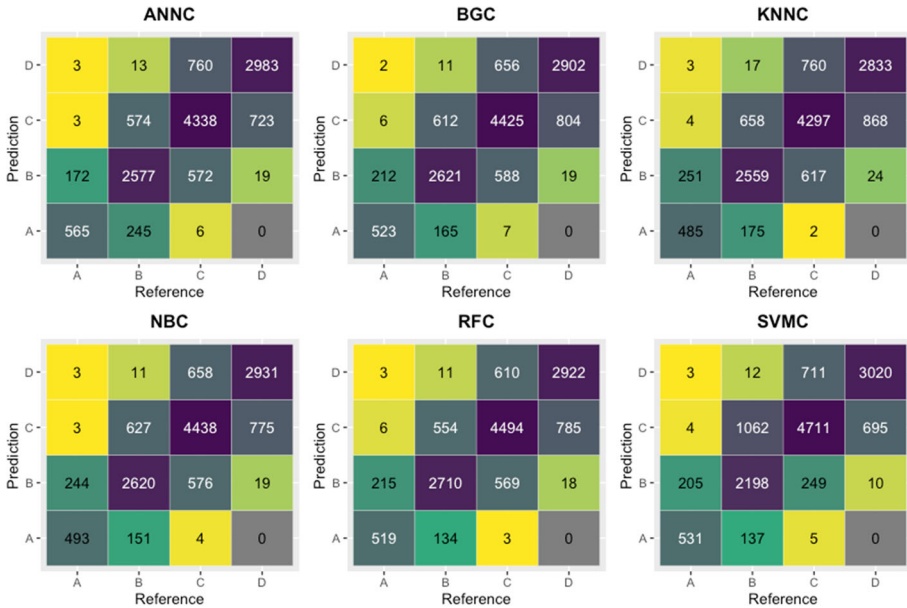*Denotes 0.05 significance level

Fig. 3 Confusion matrix of all machine learning models

results indicate there is low prediction error. Other models output similar results from the confusion matrix. However, it is hard to determine which model is the best for ESG rating prediction with a confusion matrix. Therefore, we analyze the performance metrics of each model.

All machine learning models' training and testing performance metrics are presented in Table 4. The models' training phase shows that the random forest classifier has the highest accuracy (78.3%). The testing phase accuracy output is a little bit higher at 78.5% for the random forest classifier. The rest of the models produce similar results, but other performance metrics should be considered before selecting RFC as the best model. The logLoss value of the bagging and k-nearest neighbor classifier models is over 1, indicating a higher error rate. However, other models have lower logLoss values; RFC has the lowest logLoss value for the training and testing phase (Abdullah, 2021; Bishop, 1995; Ho, 1995).

The area under the curve (AUC) value also supports that RFC is the best model because the AUC value is 94.1%; this is the highest value among all models. AUC values of all models are plotted in the ROC curve in Fig. 4. The plots show that all models have similar areas under the curve since all lines are near each other. The testing phase outputs provide results comparable with the training phase. The Kappa value of the RFC model is highest with a value of 68.2%, which indicates considerable significance for the training and testing phase (Abdullah, 2021; Ciampi et al., 2021). Sensitivity and specificity values are near zero, indicating a low misclassification rate. The F1 scores of all models are near one, indicating low false-positive and low false negative errors in the predictions. Interestingly, the artificial neural network model took the lowest training time of 119.162 min, and the random forest took the longest time of 221.669 min. This outcome is expected because this study used multiclass classification, leading to higher splits of trees for the random forest algorithm (D'Amato et al., 2022; Krauss et al., 2017; Priya et al., 2018).

Table 4 Training and testing results of seven different machine learning models

Model	logLoss	AUC	prAUC	Accuracy	Kappa	F1	Sensitivity	Specificity
Panel A: training								
ANNC	0.550	0.937	0.815	0.776	0.672	0.768	0.748	0.916
BGC	1.005	0.929	0.712	0.771	0.664	0.762	0.756	0.914
KNNC	1.348	0.915	0.660	0.742	0.621	0.729	0.718	0.903
NBC	0.558	0.934	0.816	0.768	0.660	0.757	0.747	0.913
RFC	**0.534**	**0.941**	**0.837**	**0.783**	**0.682**	**0.775**	**0.765**	**0.919**
SVMC	0.634	0.934	0.820	0.771	0.660	0.762	0.746	0.912
Panel B: testing								
ANNC	0.581	0.931	0.786	0.772	0.667	0.763	0.770	0.915
BGC	0.992	0.930	0.707	0.773	0.666	0.764	0.758	0.914
KNNC	1.327	0.916	0.662	0.751	0.633	0.739	0.730	0.906
NBC	0.544	0.936	0.822	0.773	0.667	0.761	0.750	0.915
RFC	**0.538**	**0.941**	**0.842**	**0.785**	**0.684**	**0.777**	**0.767**	**0.919**
SVMC	0.636	0.934	0.823	0.772	0.660	0.765	0.750	0.911

logLoss, Log-loss; AUC, Area under the curve; prAUC, Area under the precision-recall curve; F1 = F1 accuracy score. Best model in bold font
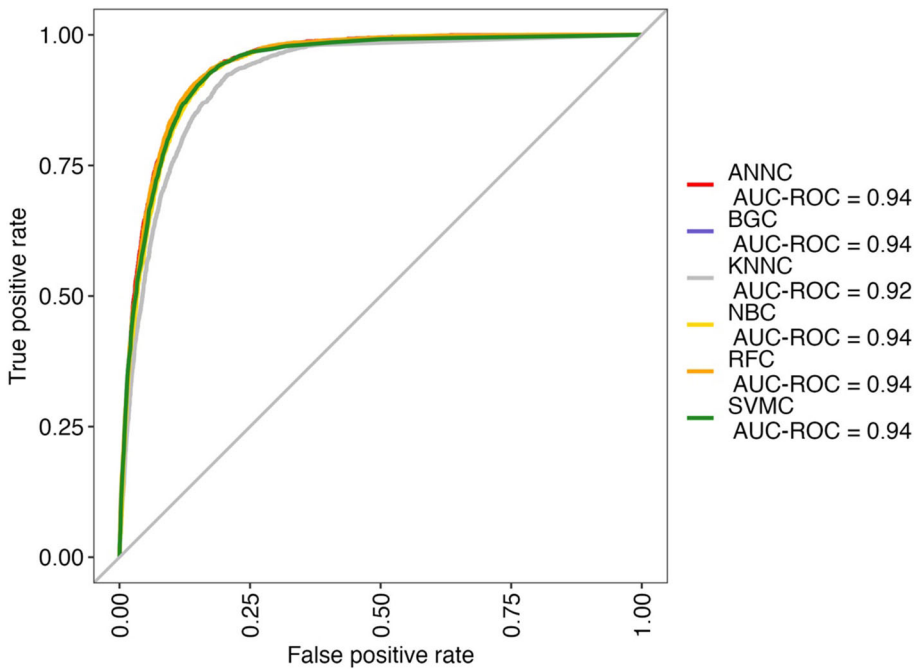

Fig. 4 The receiver operating characteristic curve of all machine learning models

Variable	Description	Variable's importance
LESG	Lag ESG score	5059.9945
SIZE	Size	814.6731
DER	Financial leverage	675.8992
EPS	Earnings per share	649.3823
TIE	Times interest earned	647.6524
GDP	Log gross domestic product	387.3302
GDPG	GDP growth rate	505.6987
UNEM	Unemployment rate	480.5083

Table 5 The importance of the variable in the random forest classifier model

Overall, the results indicate that the random forest classifier (RFC) outperforms all other machine learning models in ESG rating prediction. We have robust results across the training and testing phases of machine learning model development. Considering all the performance metrics, such as logLoss, AUC, prAUC, Accuracy, Kappa, F1, Sensitivity, Specificity, and ROC curve values, random forest classifier is the superior machine learning approach for ESG score prediction. Many previous studies find greater accuracy with the random forest machine learning algorithms in other different financial issues (D'Amato et al., 2022; Krauss et al., 2017; Priya et al., 2018).

4.3 Variable importance results

Based on the fact that the random forest classifier outperforms all models, Table 5 presents the result for variable importance factor of the random forest classifier to examine how much each variable contributes to the model (Archer & Kimes, 2008). The results indicate the lagged ESG score is the top contributing variable, firm size is the second-highest and financial leverage is the third-highest. Yu and Choi (2016) also find firm size and financial leverage are important factors in ESG ratingsClick or tap here to enter text. Macroeconomic variables are the lowest contributors to the model.

4.4 Robustness test results

To examine the robustness of our model, we used different sets of training–testing splits. Our initial analysis was based on a 70% (train) and 30% (test) sample division. Following the approach of Abdullah et al. (2023a), we also used 50:50 and 90:10 test splits to assess both the stability and performance of our models. Table 6 presents the results of the robustness test using different training–testing splits. Our findings indicate that the RFC consistently achieves the greatest accuracy across all models. Other performance metrics, such as AUC and Kappa, also confirm the superior performance of the RFC.

These results further reinforce our initial findings based on the 70:30 train-test split, highlighting the RFC as the superior machine learning approach for predicting ESG scores. These outcomes are consistent with earlier studies that demonstrate RFC model outperforms others in the context of various financial modelling (Abdullah et al., 2023a; D'Amato et al., 2022; Krauss et al., 2017; Priya et al., 2018). A comparison with related studies highlights the best-performing model (see Table 7). The table shows the prevailing accuracy of the

Table 6 Robustness across different training and testing splits of the machine models

Model	50:50 train-test split			90:10 train-test split		
	AUC	Accuracy	Kappa	AUC	Accuracy	Kappa
Panel A: training						
ANNC	0.938	0.777	0.674	0.938	0.776	0.671
BGC	0.929	0.771	0.664	0.930	0.771	0.664
KNNC	0.914	0.743	0.623	0.916	0.747	0.629
NBC	0.936	0.770	0.663	0.935	0.770	0.663
RFC	**0.940**	**0.783**	**0.681**	**0.941**	**0.784**	**0.683**
SVMC	0.937	0.780	0.674	0.936	0.774	0.664
Panel B: testing						
ANNC	0.939	0.780	0.677	0.938	0.777	0.671
BGC	0.929	0.771	0.663	0.931	0.775	0.670
KNNC	0.913	0.740	0.617	0.920	0.752	0.636
NBC	0.933	0.768	0.659	0.936	0.771	0.664
RFC	**0.939**	**0.781**	**0.678**	**0.942**	**0.783**	**0.682**
SVMC	0.935	0.771	0.660	0.936	0.773	0.664

AUC Area under the curve. Best model in bold font

Table 7 Machine learning model performance in related studies

Citation	Objective	Dataset	Best model	Performance
Our result	ESG rating prediction	Global 6166 firms	Random Forest	Accuracy = 78.5%
D'Amato et al. (2021)	ESG score forecasting	109 European firms	Random Forest	RMSE = 8.02
D'Amato et al. (2022)	ESG score forecasting	401 European firms	Random Forest	RMSE = 11.15
Sokolov et al. (2021)	ESG text labeling	6000 Tweets	BERT language model	ROC AUC = 90%*
Lee and Kim (2023)	ESG text classifier	1641 Korean sustainability reports	KLUE-RoBERTa-large	Accuracy = 86.66%
Abdullah et al. (2023a)	Nonperforming loan forecasting	Emerging countries 322 banks	Random Forest	R^2 = 76.10%
Barboza et al. (2017)	Bankruptcy prediction	10,000 firm year observation of North American firms	Random Forest	Accuracy = 87.5%

*The result of the ESG risk category and full list can be found in Exhibit 2 in Sokolov et al. (2021)
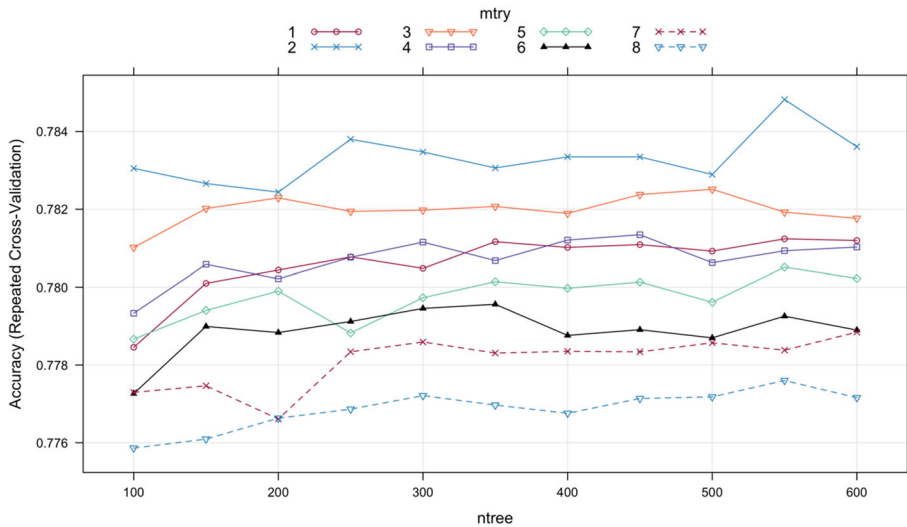
Fig. 5 Random Forest Classifier model: An accuracy across grid search

RFC model, both in ESG analysis and across broader fields of finance-related research. The efficacy of the RFC's performance can be attributed to the nature of its algorithm. By creating and pruning trees, this algorithm adeptly addresses multi-class classification challenges, such as our task of predicting ESG ratings.

We then performed a parallelized evaluation using our initial grid of hyperparameters to diagnose the performance of the RFC model. Our grid search included a range of "mtry" (from 1 to 8) and "ntree" (from 100 to 600) values, all of which were applied to our baseline sample (70:30 split). Figure 5 illustrates the RFC's accuracy results. The results show a decreasing trend in accuracy as the number of mtrys increases. This implies that there may be a trade-off between accuracy and complexity introduced by higher mtry values. The behavior of ntree is variable, with instances of both increase and decrease as mtry values increase. The peak accuracy, 78.5%, was achieved when mtry was set to 2 and ntree was set to 550. This configuration represents a compromise in which the model's accuracy is optimized while taking into account the number of features and decision trees used.

Figure 6 presents the Kappa results from the RFC. The findings again highlight the highest Kappa value is when mtry is set to 2 and ntree is set to 550. The Kappa statistic is a concordance measure that considers the possibility of concordance occurring by chance. In our context, a high Kappa value indicates a high level of agreement between predicted and actual ESG ratings, especially when the mtry and ntree values are considered.

Figure 7 illustrates the AUC results from the RFC. Again, the results highlight that the highest AUC value is when mtry is set to 2 and ntree is set to 550. In our context, the AUC metric indicates the model's ability to discriminate between different classes, which translates into its effectiveness in distinguishing various ESG ratings. An increased AUC value emphasizes RFC's robust discriminatory power, especially when the mtry and ntree settings are set to 2 and 550, respectively. Overall, the above results reinforce the RFC's efficacy in accurately predicting ESG ratings under these specific parameter conditions.
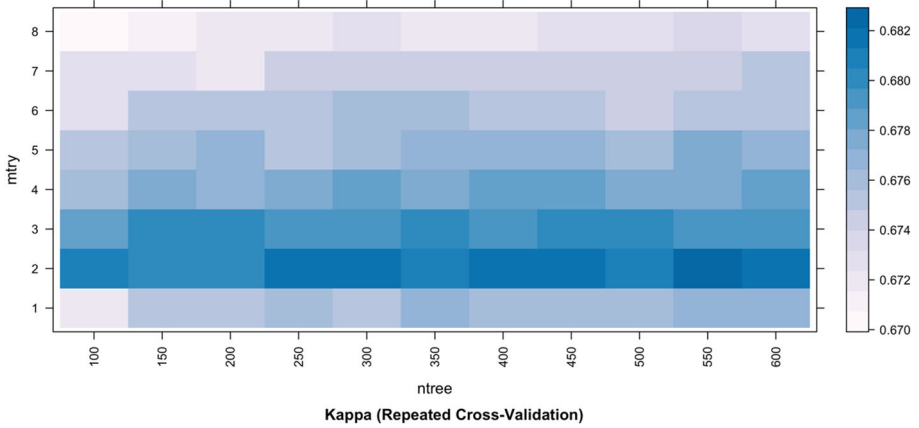
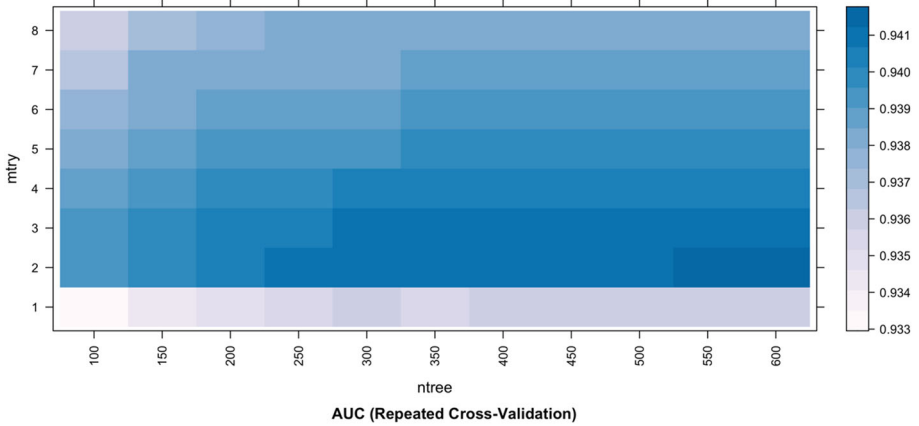**Fig. 6** Random Forest Classifier model: Kappa across grid search

Fig. 7 Random Forest Classifier model: Area under the curve across grid search

5 Discussion

Over the last decade, environmental, social, and governance (ESG) rating has been becoming a key indicator for investment decisions by investors, i.e., institutional investors, portfolio managers, rational investors, and firm mangers to assess sustainable development goals. We try to predict ESG ratings using firm and macroeconomic variables by applying six machine learning algorithms. We find that the RFC is the superior algorithm for predicting ESG ratings among the other algorithms considered (the logLoss, AUC, prAUC, Kappa, F1, Sensitivity, Specificity, and ROC curve). The accuracy results indicate that the RFC achieves an ESG rating accuracy of 78.5%.

The primary discovery of this study is that the ESG score from the preceding year exerts the most influential effect on forecasting increasing ESG ratings. This indicates that ESG ratings are greatly influenced by how well a company performed in the previous year. This

finding aligns with ideas from institutional, accountability, and legitimacy theories (Baldini et al., 2018; Weber, 2014). It also suggests that the increasing attention paid to ESG ratings is motivating company leaders to either maintain their previous year's rating or work on improving it. In this way, they can uphold or even enhance the positive perception of their commitment to sustainability among investors and stakeholders.

Moreover, firm size is the second-highest contributing variable to ESG rating prediction. Grounded in the findings of Drempetic et al. (2020), Chauhan et al. (2014), and Hörisch et al. (2015), big firms with higher liquidity tend to invest more in ESG. Larger firms may be more aware of sustainability than smaller firms because they cause more damage to the environment. This viewpoint is consistent with the ideas advanced by the stakeholder and signaling theories. According to these theories, larger corporations make an effort to maintain positive relationships with various stakeholders and to reduce their environmental and social footprint (Ansoff, 1965). Larger firms frequently use their growth to signal their potential for future success. This means they may see investing in ESG practices as a way to demonstrate their commitment to sustainable practices, which can help their reputation and prospects (Spence, 1978).

Furthermore, we document that leverage emerges as the third-highest contributing variable in ESG rating predictions. Our result is in line with the findings of Yu and Choi (2016), which suggest that leverage is a significant determinant of ESG scores. This finding can be interpreted through the lens of the agency theory, which sheds light on this relationship by emphasizing the importance of ownership structure (Jensen & Meckling, 1976). It emphasizes that when ownership is distributed among shareholders, there is a greater need for transparent information disclosure to align the interests of management and shareholders. In essence, highly leveraged firms appear to be motivated by a desire to demonstrate their commitment to sustainability and a desire to maintain clear communications with various stakeholders, including investors. This intertwining of financial structure and ESG practices emphasizes the complex relationships that exist between corporate decisions and external pressures.

6 Concluding remarks

6.1 Conclusion and policy implications

With rising stakeholder awareness of firms' roles in society and an interest in social, environmental, and ethical issues, firm managers are under pressure from the public and private sectors to disclose their practices. ESG is also evolving as an investment decision tool for institutional investors and portfolio managers. This study develops ESG rating predictions using a machine learning approach. We use six machine learning classification models for ESG prediction with data on 6171 firms from 2005 to 2019. Of all models considered (RFC, logLoss, AUC, prAUC, Kappa, F1, Sensitivity, Specificity, and ROC curve), the RFC model demonstrates superior performance across various metrics. The accuracy achieved by this model is 78.50%. Our results consistently hold under robustness testing with different train-test splits and hyperparameter tuning. This suggests that the RFC model is the most suitable choice to predict ESG ratings.

The study's results have several significant implications. First, the significance of ESG ratings in investment decisions highlights the importance of consistent and reliable assessment methodologies. Given that our research identifies the RFC as the most effective algorithm

for predicting ESG ratings, regulators and rating agencies may consider endorsing or adopting this algorithm to ensure consistent and accurate evaluations across industries. Moreover, embracing machine learning tools into enterprise resource planning systems, may enable companies to systematically track their ESG ratings over time. The transparency and comparability in ESG ratings could be improved by establishing a preferred algorithm, assisting investors in making informed decisions aligned with their sustainable development goals. Second, the significant impact of the previous year's ESG score on future ratings emphasizes the importance of sustained commitment to ESG practices. The policymakers and industry groups could encourage companies to view ESG ratings as continuous improvement targets rather than static assessments. Recognizing the positive correlation between prior-year performance and future ratings, companies may be incentivized to focus on maintaining or improving their ESG scores, fostering a culture of sustainable practices, and enhancing their reputation among stakeholders.

Third, our findings highlight the relationship between firm size, liquidity, and ESG investing. Policymakers and industry regulators could encourage larger corporations to adopt ESG practices through targeted incentives and reporting requirements. Encouraging large corporations to invest proactively in ESG practices can have a positive impact on sustainability, given their significant influence on the environment and society. These corporations can demonstrate leadership by aligning growth strategies with sustainability goals, thereby contributing to a more sustainable business landscape. Fourth, the relationship between leverage and ESG scores reveals a way for businesses to demonstrate their commitment to sustainability. Regulatory bodies can promote greater transparency in financial disclosures related to leverage, encouraging companies to be accountable for their financial structures and ESG practices. Aligning financial and sustainability goals could encourage responsible behavior and strategic decision-making that benefits both investors and larger stakeholder groups. Finally, regulatory bodies and policymakers may consider the study's conclusions when formulating policies related to ESG performance and disclosure, thereby cultivating a more sustainable and responsible corporate environment. By leveraging our ESG score prediction model, policymakers, practitioners, investors, regulatory bodies, portfolio managers, and stakeholders may develop their prediction models, facilitating the acquisition of current information for policy formulation and informed investment choices.

6.2 Limitations and future research directions

This study's focus is clearly on ESG rating prediction based on Thomson Reuters ESG rating data, which lack consideration of most emerging nations' firms. Future studies can contribute by developing a new model that considers more predictors to develop an ESG rating model for all public-listed firms not listed in Thomson Reuters. This will allow decision-makers to assess all public-listed firms' ESG ratings. Future studies can extend our findings by developing an ESG prediction model for small and medium enterprises.

Author contributions Conceptualization: [MAFC, MA]; Methodology: [MA]; Formal analysis and investigation: [MA]; Writing—original draft preparation: [MA, MAFC, MAK, ZS]; Writing—review and editing: [MAFC, MNI]; Resources: [MAFC, MA]; Supervision: [MAFC, ZS].

Data availability The datasets analyzed in this study are privately available in the Refinitiv Datastream repository, https://www.refinitiv.com/en/products and macroeconomic variables data are available at the World Bank database, https://data.worldbank.org/.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states no conflict of interest.

References

Abdullah, M., Sulong, Z., & Chowdhury, M. A. F. (2023). Explainable Deep Learning Model for Stock Price Forecasting Using Textual Analysis. *Available at SSRN 4355596*. https://doi.org/10.2139/ssrn.4355596

Abdullah, M. (2021). The implication of machine learning for financial solvency prediction: An empirical analysis on public listed companies of Bangladesh. *Journal of Asian Business and Economic Studies, 28*(4), 303–320. https://doi.org/10.1108/JABES-11-2020-0128

Abdullah, M., Chowdhury, M. A. F., Uddin, A., & Moudud-Ul-Huq, S. (2023). Forecasting nonperforming loans using machine learning. *Journal of Forecasting*. https://doi.org/10.1002/for.2977

Alsayegh, M. F., Abdul Rahman, R., & Homayoun, S. (2020). Corporate economic, environmental, and social sustainability performance transformation through ESG disclosure. *Sustainability*. https://doi.org/10.3390/su12093910

Ansoff, H. I. (1965). *Corporate Strategy*. McGraw Hill.

Antoncic, M. (2020). Uncovering hidden signals for sustainable investing using Big Data: Artificial intelligence, machine learning and natural language processing. *Journal of Risk Management in Financial Institutions, 13*(2), 106–113.

Archer, K. J., & Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis, 52*(4), 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015

Baldini, M., Maso, L. D., Liberatore, G., Mazzi, F., & Terzani, S. (2018). Role of country- and firm-level determinants in environmental, social, and governance disclosure. *Journal of Business Ethics, 150*(1), 79–98. https://doi.org/10.1007/s10551-016-3139-1

Barbeito-Caamaño, A., & Chalmeta, R. (2020). Using big data to evaluate corporate social responsibility and sustainable development practices. *Corporate Social Responsibility and Environmental Management, 27*(6), 2831–2848. https://doi.org/10.1002/csr.2006

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications, 83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Bhandari, K. R., Ranta, M., & Salo, J. (2022). The resource-based view, stakeholder capitalism, ESG, and sustainable competitive advantage: The firm's embeddedness into ecology, society, and governance. *Business Strategy and the Environment, 31*(4), 1525–1537. https://doi.org/10.1002/bse.2967

Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management, 28*(5), 1426–1445. https://doi.org/10.1002/csr.2177

Bishop, C. M. (1995). Training with noise is equivalent to Tikhonov regularization. *Neural Computation, 7*(1), 108–116. https://doi.org/10.1162/neco.1995.7.1.108

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140. https://doi.org/10.1007/bf00058655

Broadstock, D. C., Matousek, R., Meyer, M., & Tzeremes, N. G. (2020). Does corporate social responsibility impact firms' innovation capacity? The indirect link between environmental & social governance implementation and innovation performance. *Journal of Business Research, 119*, 99–110. https://doi.org/10.1016/j.jbusres.2019.07.014

Buallay, A. (2019). Is sustainability reporting (ESG) associated with performance? Evidence from the European banking sector. *Management of Environmental Quality: An International Journal, 30*(1), 98–115. https://doi.org/10.1108/MEQ-12-2017-0149

Chauhan, S. A. (2014). A relational study of firm's characteristics and CSR expenditure. *Procedia Economics and Finance, 11*, 23–32. https://doi.org/10.1016/S2212-5671(14)00172-5

Chung, R., Bayne, L., & Birt, J. L. (2023). Determinants of ESG disclosure among listed firms under voluntary and mandatory ESG disclosure regimes in Hong Kong. *Journal of Applied Accounting Research*. https://doi.org/10.1108/JAAR-07-2022-0179

Ciampi, F., Giannozzi, A., Marzi, G., & Altman, E. I. (2021). Rethinking SME default prediction: A systematic literature review and future perspectives. *Scientometrics, 126*(3), 2141–2188. https://doi.org/10.1007/s11192-020-03856-0

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Crace, L., & Gehman, J. (2023). What really explains ESG performance? Disentangling the asymmetrical drivers of the triple bottom line. *Organization & Environment, 36*(1), 150–178. https://doi.org/10.1177/10860266221079408

Crespi, F., & Migliavacca, M. (2020). The Determinants of ESG rating in the financial industry: The same old story or a different tale? *Sustainability*. https://doi.org/10.3390/su12166398

D'Amato, V., D'Ecclesia, R., & Levantesi, S. (2021). Fundamental ratios as predictors of ESG scores: A machine learning approach. *Decisions in Economics and Finance, 44*(2), 1087–1110. https://doi.org/10.1007/s10203-021-00364-5

D'Amato, V., D'Ecclesia, R., & Levantesi, S. (2022). ESG score prediction through random forest algorithm. *Computational Management Science, 19*(2), 347–373. https://doi.org/10.1007/s10287-021-00419-3

D'Amato, V., D'Ecclesia, R., & Levantesi, S. (2023). Firms' profitability and ESG score: A machine learning approach. *Applied Stochastic Models in Business and Industry*. https://doi.org/10.1002/asmb.2758

D'Apice, V., Ferri, G., & Intonti, M. (2021). Sustainable disclosure versus ESG intensity: Is there a cross effect between holding and SRI funds? *Corporate Social Responsibility and Environmental Management, 28*(5), 1496–1510. https://doi.org/10.1002/csr.2178

Đặng, R., Hikkerova, L., Simioni, M., & Sahut, J.-M. (2022). How do women on corporate boards shape corporate social performance? Evidence drawn from semiparametric regression. *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-04550-5

Darnall, N., Ji, H., Iwata, K., & Arimura, T. H. (2022). Do ESG reporting guidelines and verifications enhance firms' information disclosure? *Corporate Social Responsibility and Environmental Management, 29*(5), 1214–1230.

De Lucia, C., Pazienza, P., & Bartlett, M. (2020). *Does good esg lead to better financial performances by firms?* Machine Learning and Logistic Regression Models of Public Enterprises in Europe. https://doi.org/10.3390/su12135317

Dhasmana, S., Ghosh, S., & Kanjilal, K. (2023). Does investor sentiment influence ESG stock performance? Evidence from India. *Journal of Behavioral and Experimental Finance, 37*, 100789. https://doi.org/10.1016/j.jbef.2023.100789

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147–160. https://doi.org/10.2307/2095101

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: corporate sustainability ratings under review. *Journal of Business Ethics, 167*(2), 333–360. https://doi.org/10.1007/s10551-019-04164-1

Escrig-Olmedo, E., Fernández-Izquierdo, M. Á., Ferrero-Ferrero, I., Rivera-Lirio, J. M., & Muñoz-Torres, M. J. (2019). Rating the raters: Evaluating how ESG rating agencies integrate sustainability principles. *Sustainability*. https://doi.org/10.3390/su11030915

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. https://doi.org/10.1080/20430795.2015.1118917

Garcia, A. S., Mendes-Da-Silva, W., & Orsato, R. J. (2019). Corporate sustainability, capital markets, and ESG performance BT - individual behaviors and technologies for financial innovations. In W. Mendes-Da-Silva (Ed.), (pp. 287–309). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-91911-9_13

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance, 32*, 100577. https://doi.org/10.1016/j.jbef.2021.100577

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications, 124*, 226–251. https://doi.org/10.1016/j.eswa.2019.01.012

Hisano, R., Sornette, D., & Mizuno, T. (2020). Prediction of ESG compliance using a heterogeneous information network. *Journal of Big Data, 7*(1), 22. https://doi.org/10.1186/s40537-020-00295-9

Ho, T. K. (1995). Random decision forests Tin Kam Ho perceptron training. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*.

Holzinger, A., Kieseberg, P., Weippl, E., & Tjoa, A. M. (2018). Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: From Machine Learning to Explainable AI BT - Machine

Learning and Knowledge Extraction. In A. Holzinger, P. Kieseberg, A. M. Tjoa, & E. Weippl (Eds.), (pp. 1–8). Cham: Springer International Publishing.

Hörisch, J., Johnson, M. P., & Schaltegger, S. (2015). Implementation of sustainability management and company size: A knowledge-based view. *Business Strategy and the Environment, 24*(8), 765–779. https://doi.org/10.1002/bse.1844

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Jiang, M., Jia, L., Chen, Z., & Chen, W. (2022). The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm. *Annals of Operations Research, 309*(2), 553–585. https://doi.org/10.1007/s10479-020-03690-w

Johnson, C. (2020). The measurement of environmental, social and governance (ESG) and sustainable investment: Developing a sustainable new world for financial services. *Journal of Securities Operations & Custody, 12*(4), 336–356.

Kalaitzoglou, I., Pan, H., & Niklewski, J. (2021). Corporate social responsibility: How much is enough? A higher dimension perspective of the relationship between financial and social performance. *Annals of Operations Research, 306*(1), 209–245. https://doi.org/10.1007/s10479-020-03834-y

Keller, J. M., Gray, M. R., & Givens, J. A. (1985). A fuzzy K-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, SMC, 15*(4), 580–585. https://doi.org/10.1109/TSMC.1985.6313426

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review, 26*(3), 159–190. https://doi.org/10.1007/s10462-007-9052-3

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research, 259*(2), 689–702. https://doi.org/10.1016/j.ejor.2016.10.031

Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2022). Past, present, and future of sustainable finance: Insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04410-8

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363–374.

Lanza, A., Bernardini, E., & Faiella, I. (2020). Mind the gap! Machine learning, ESG metrics and sustainable investment. *Machine learning, ESG metrics and sustainable investment (June 26, 2020). Bank of Italy Occasional Paper*, (561).

Lee, J., & Kim, M. (2023). ESG information extraction with cross-sectoral and multi-source adaptation based on domain-tuned language models. *Expert Systems with Applications, 221*, 119726. https://doi.org/10.1016/j.eswa.2023.119726

Liagkouras, K., Metaxiotis, K., & Tsihrintzis, G. (2020). Incorporating environmental and social considerations into the portfolio optimization process. *Annals of Operations Research*. https://doi.org/10.1007/s10479-020-03554-3

Lokuwaduge, C. S. D. S., & Heenetigala, K. (2017). Integrating environmental, social and governance (ESG) disclosure for a sustainable development: An Australian study. *Business Strategy and the Environment, 26*(4), 438–450. https://doi.org/10.1002/bse.1927

Mai, L., Koliousis, A., Li, G., Brabete, A. O., & Pietzuch, P. (2019). Taming hyper-parameters in deep learning systems. *Operating Systems Review (ACM), 53*(1), 52–58. https://doi.org/10.1145/3352020.3352029

May, R. J., Maier, H. R., & Dandy, G. C. (2010). Data splitting for artificial neural networks using SOM-based stratified sampling. *Neural Networks, 23*(2), 283–294. https://doi.org/10.1016/j.neunet.2009.11.009

Nekhili, M., Boukadhaba, A., & Nagati, H. (2021). The ESG–financial performance relationship: Does the type of employee board representation matter? *Corporate Governance: An International Review, 29*(2), 134–161. https://doi.org/10.1111/corg.12345

Nyitrai, T., & Virág, M. (2019). The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Economic Planning Sciences, 67*, 34–42. https://doi.org/10.1016/j.seps.2018.08.004

Pearl, J. (1988). Chapter 3 - Markov and Bayesian networks: Two graphical representations of probabilistic knowledge. In J. B. T.-P. R. I. S. Pearl (Ed.), (pp. 77–141). San Francisco (CA): Morgan Kaufmann. https://doi.org/10.1016/B978-0-08-051489-5.50009-6

Perrow, C. (1970). Departmental power and perspectives in industrial firms. *Power in Organizations, 7*, 59–89.

Priya, K. U., Pushpa, S., Kalaivani, K., & Sartiha, A. (2018). Exploratory analysis on prediction of loan privilege for customers using random forest. *International Journal of Engineering Technology, 7*(2.21), 339–341.

Raman, N., Bang, G., & Nourbakhsh, A. (2020). Mapping ESG trends by distant supervision of neural language models. *Machine Learning and Knowledge Extraction*. https://doi.org/10.3390/make2040025

Sauerwald, S., & Su, W. (2019). CEO overconfidence and CSR decoupling. *Corporate Governance: An International Review, 27*(4), 283–300. https://doi.org/10.1111/corg.12279

Semenova, N., & Hassel, L. G. (2019). Private engagement by Nordic institutional investors on environmental, social, and governance risks in global companies. *Corporate Governance: An International Review, 27*(2), 144–161. https://doi.org/10.1111/corg.12267

Sharma, P., Panday, P., & Dangwal, R. C. (2020). Determinants of environmental, social and corporate governance (ESG) disclosure: A study of Indian companies. *International Journal of Disclosure and Governance, 17*(4), 208–217. https://doi.org/10.1057/s41310-020-00085-y

Sokolov, A., Mostovoy, J., Ding, J., & Seco, L. (2021). Building machine learning systems for automated ESG scoring. *The Journal of Impact and ESG Investing, 1*(3), 39–50.

Spence, M. (1978). Job market signaling. In P. Diamond & M. B. T.-U., E. Rothschild (Eds.), (pp. 281–306). Academic Press. https://doi.org/10.1016/B978-0-12-214850-7.50025-5

Sun, L., & Shenoy, P. P. (2007). Using Bayesian networks for bankruptcy prediction: Some methodological issues. *European Journal of Operational Research, 180*(2), 738–753. https://doi.org/10.1016/j.ejor.2006.04.019

Wang, L., Jia, F., Chen, L., & Xu, Q. (2022). Forecasting SMEs' credit risk in supply chain finance with a sampling strategy based on machine learning techniques. *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-04518-5

Weber, O. (2014). The financial sector's impact on sustainable development. *Journal of Sustainable Finance & Investment, 4*(1), 1–8. https://doi.org/10.1080/20430795.2014.887345

Yu, Y., & Choi, Y. (2016). Stakeholder pressure and CSR adoption: The mediating role of organizational culture for Chinese companies. *The Social Science Journal, 53*(2), 226–234. https://doi.org/10.1016/j.soscij.2014.07.006