Corporate Social Responsibility and Environmental Management

WILEY

Corporate Social Responsibility and Environmental Management

RESEARCH ARTICLE OPEN ACCESS

Predicting ESG Controversies in Banks Using Machine Learning Techniques

Anna Rita Dipierro[1] | Fernando Jimenéz Barrionuevo[2] | Pierluigi Toma[1]

[1]Department of Economics and Management, University of Salento, Lecce, Italy | [2]Department of Information and Communications Technology, University of Murcia, Murcia, Spain

Correspondence: Anna Rita Dipierro (annarita.dipierro@unisalento.it)

Received: 21 May 2024 | **Revised:** 31 October 2024 | **Accepted:** 20 January 2025

Keywords: banks | controversies | ESG | governance | machine learning | risk

ABSTRACT

Mistreating environmental, social, and governance (ESG) concerns has serious drawbacks in organizations of any type, and even more in banks. Deeply revolutionized in its taxonomy of risks, banking sector is herein evaluated in its integration of ESG parameters that, when lacking, leads to ESG-related controversies (ESGC). Thereby, this research approaches the almost uncharted territory of ESGC in banks, by means of machine learning. Aiming at selecting the set of features that are relevant in ESGC prediction, techniques belonging to feature selection are used over a real panel dataset of 140 banks evaluated for a wide set of features over 2011–2020 time-span. We find the power that governance-employees dynamics detains in making out-of-sample predictions and forecasting of ESGC banks' risk. Finally, we provide implications for researchers, practitioners and regulators, further confirming the need for the rapid inroads that machine learning tools are actually making in the banking toolkit and in the regulatory technology.

1 | Introduction

In adding *more light* toward the critical junctures of Machine Learning (ML)—a sub-branch of Artificial Intelligence (AI)—that economists and econometricians are, may and should be interested in, Chan and Mátyás (2022) seem providing us the lens to scrutinize ML in its ability to transform management decision. This work wants to answer the call for more light on the dynamic taxonomy of risks of banks, by means of ML. The ability to deal with large, and complex dataset and pursuing prediction tasks (Kinywamaghana and Steffen 2021) makes ML adequately mature do be deployed in the sector. The banking sector abounds of traditional econometric approaches, whereas there is a large room to apply ML tools, which are adequately mature do be deployed in. Actually, the European Money and Finance Forum (traced the rapid inroads that ML tools are making in the banks toolkit and in regulatory technology (Doerr, Gambacorta, and Serena 2021), becoming a research trend in the financial stability regulation (Chao et al. 2022).

Led by the question of Hirsch (2018), "what then of the potential use of AI in reputation risk management?," this work explores an almost uncharted category of risks in banks, through a novel approach. We focus on scandals. The questionable behaviors of certain banks revealed a shady image of the sector (Nirino et al. 2021). Scandals do abound coming form the financial accounting side (Pilkington 2022), as from the under-researched scandal named London Whale in 2012, from a group of traders that operated on account of JP Morgan Chase & Co. Also, disruptive scandals are emerging from the employees' side complaining unsustainable conditions, as in Goldman Sachs in 2013 and 2021, and Bank of America in 2013.

Anna Rita Dipierro carried out this research during her Ph.D. at LUM University (Casamassima, BA, Italy) and her fellowship at University of Salento (Lecce, LE, Italy). Anna Rita Dipierro is now affiliated to University of Calabria as a Post-Doctoral Researcher. She receives support within the GRINS project–Growing Resilient, INclusive and Sustainable from the European Union Next-Generation EU (GRINS PE00000018, CUP: H23C24000110006, Spoke 4 Sustainable Finance).

In detail, we investigate the mistreatment of resources inherent to environment, social, governance (ESG) factors. The attention toward ESG issues has increased in management science and practice over years (DasGupta and Roy 2023). Specifically, as a recurrent phenomena of high interest, we investigate the controversies related to ESG issues' mistreatment (hereinafter ESGC) in the banking industry, which is almost uncharted compared with business sector (e.g., Nirino et al. (2021)). ESGC represents a new and important risk component (Galletta and Mazzù 2023; Agnese et al. 2024) which deserves a particular attention firstly in light of the particular role of effective risk governance in the financial services industry (Hiebl et al. 2018), and due to the unprecedented exposure-to-media of banks. The voice of scandals is turning into noise for their survival. It follows the urgency of a more aligned with reality view than that applying ESG disclosed level.

We explore ESGC, specifically in what lies behind its prediction. We propose a methodological approach to pursue such aims in a panel dataset. We opt for a family of ML techniques, that is, Feature Selection (FS), which allows to remove from the dataset irrelevant and/or redundant attributes according to the specific task (Liu and Motoda 1998). Empirically, FS process is applied to a sample of 140 world-leading banks, measured over 28 variables (features), and the ESGC score (target to be predicted) in the 2011–2020 period. In so doing, we answer to the following Research Question (RQ):

RQ. *To what extent governance-employees dynamics matter in ESGC prediction and forecasting, in face of operational and financial concerns?*

Both ESGC prediction (prediction for a newly added bank in the same time-span), and ESGC time series forecasting (future prediction for the sampled banks through the multistep ahead prediction process) are performed.

Pursuing such goals requires a great methodological effort, since to the best of authors' knowledge there exists no study questioning the prevalence of certain indicator in predicting ESGC performance, and, most importantly, literature tackling panel data through ML-FS techniques is in its very beginning. We pave the way for new efforts in ESGC concern and in risk governance in banks that may test the proposed methodology in the spirit of knowledge. We also try facing lexical discrepancies between econometrics and ML in management that may threat the mutual understanding, exploring heterogeneous terminologies addressing same concepts. Beyond it, the study is of aid of practitioners and regulators, arguing on the transformation risk management and governance, by considering a new category of risks of primary importance for banks' managers and customers (Ecer and Pamucar 2022). We improve the awareness of internal resources' and their concerns' power in risk management, on the basis of which assess priority in actions (Table B1).

The remainder of this work is organized as follows: Section 2 presents the literature review. Section 3 propose the methodology. Section 4 provides the empirical application and its results. Section 5 finally draws the conclusions.

2 | Literature Review

Disrupted events in banks, from the oldest to the newest past, arise also from the mistreatment of ESG factors. The awareness of the unprecedented exposure to systemic risk from one entities' troubles is at the core of banks' regulators and managers. A large scale break-down leads to huge economic and social costs (Elsinger, Lehar, and Summer 2006). Further, the unprecedented social attention and media exposure surround the entire system.

However, the way in which story just goes on and repeats its mistakes raises a number of concerns. Scandals threaten the way in which banks are managed and regulated, and mirrors in turn in consumers' trust. Over the years, regulators are approaching ML tools to supervise and manage issue-specifically and time-urgently, given the crises and scandals of the past. Purposes of the research go through reviewing the main literature on ML in banks and in the existent approaches to ESG issues and ESGC threats.

2.1 | Machine Learning in Banks: Focus and Gaps

Recent years have seen a surge of the ML focus on the way, in which banks' risks are being detected, measured, reported and managed (Leo, Sharma, and Maddulety 2019). The opacity that banks' risk are experiencing through their increasing in size and organizational complexity (Agarwal et al. 2019) enhances this need. Before exploring ML literature, we should recall that ML consists of supervised techniques (i.e., techniques requiring the presence of a target variable to predict or estimate, given a set of predictors), unsupervised (that lack of the dependent variable), and semi-supervised that lie between the two. Algorithm, which is not used as synonymous of models, since not every kind of algorithm produces a model, can both perform regression tasks (the goal is to predict a numeric value), classification (the goal is to predict a class or a class probability) or both. Further, in ML terminology, target/outcome refers to the econometrics' dependent variable, and features are those that econometrics addresses as regressors, explanatory or independent variables.

What already exists in ML literature in banking industry concerns credit risk. This happens since ML tools are confirmed to yield better predictive performance than traditional quantitative models (Alonso-Robisco and Carbó 2022). Credit risk management is the most studied in ML studies applied into banking literature. In this area of research, Artificial Neural Network (ANN), Support Vector Machine (SVM) and Random Forest (RF) were applied to predict credit scoring (Addo, Guegan, and Hassani 2018; Ala'raj and Abbod 2016). Recently, stock price prediction (Hanauer, Kononova, and Rapp 2022), quality of service prediction (Castelli, Manzoni, and Popovič 2016), several areas of operational risk that enable its mitigation, such as fraud and suspicious detections (Sharma and Choudhury 2016; Peters et al. 2017; Khrestina et al. 2017) were delved into by means of ML. Further, light was shed upon the exploration of consumer trust (Adamyk et al. 2019) and satisfaction (Alkhatib and Abualigah 2020). Respectively, RF, linear regression (LR), logistic regression, decision trees, XGBoost Model were developed under a prediction model. Predicting the subscription of a term deposit in Portuguese Banks was the objective of Alexandra and Sinaga (2021), using Naive-Bayes (NB) and RF algorithms.

There also are studies assessing features importance, addressing the prediction of stock returns (Mohapatra et al. 2022), bankruptcy (Shrivastav 2019) in Indian banks, and insolvencies in United States (US)-based financial institutions (Petropoulos et al. 2020). In this latter case, six different algorithms were implemented, and RF outperformed, as also from Durand and Le Quang (2022) that exploring the impact of different regulatory variables on banks' performance demonstrated that RF outperformed Lasso, SVMs, and NNs.

Some other applications are enriching ML literature in the banking sector. See for instance Agarwal et al. (2019) that, by means of a mixed supervised (Ridge Regression, Lasso, RF) and unsupervised techniques (clustering), questioned the changing nature of risk culture from publicly available news for the financial institutions participating in the Federal Reserve stress tests. Interestingly, a number of studies are implementing evolutionary computing methods, which are capable of extracting meaning from data that are imprecise and contextually detect trends that humans or conventional technique are not capable to discover (Metawa, Hassan, and Elhoseny 2017). It happens since these techniques are inspired by organic mechanisms, such as natural selection or mutation, for example, Genetic Algorithm (GA), or swarm intelligence, for example, Particle Swarm Optimization (PSO). The former was already used by Lappas and Yannacopoulos (2021) to assess credit scoring in bank experimenting also FS task on an illustrative example-dataset in the financial sector. The experiments of Metawa, Hassan, and Elhoseny (2017) to organize bank lending decisions within a highly competitive environment further showed GA capability to well handle this task. PSO has even more application to real dataset on the financial sector (Doumpos et al. 2022; Reddy 2022; Tissaoui 2019; Guo et al. 2019) under a multiobjective variant configuration; Shie, Chen, and Liu (2012) and from Pai, Hsu, and Wang (2011) in a combined methodological framework with SVM as from Doumpos et al. (2022) trying to smooth the weaknesses of which SVM individually suffer when dealing with noise and outliers hosted by real-world data. Nowadays, multiobjective evolutionary FS algorithm should be also considered, given the already effectively usage in business applications (Jiménez et al. 2017) by providing successful result under the optimization of two objectives simultaneously. Even if slowly applied into the banking sector (Papouskova and Hajek 2019), it may deserve a particular attention in light of task of the present research, which will follow the directions depicted by Doumpos et al. (2022), for example, investigating misconduct in banks, and incorporating ESG when evaluating bank performance and operations testing the performance of algorithms in specific tasks.

2.2 | Constructive and Disruptive Approaches to ESG Factors

ESG factors show their uniqueness in the way they build reputation. Nevertheless, they are the same that may destroy it. A growing body of research fueled the debate about the first and constructive perspective of uniqueness. In this connection, several contributions investigated ESG performance of banks, both in terms of effects and determinants. Traditional econometric approaches abound and reflect upon the governance-employees concerns, as well as operational and financial domain. Board composition (Eagly, Johannesen-Schmidt, and Van Engen 2003; Barako and Brown 2008; Zhang, Zhu, and Ding 2013; Shakil 2021) and frequency (Birindelli et al. 2015), assets profitability (Soana 2011; Miralles-Quirós, Miralles-Quirós, and Redondo Hernández 2019) and quality (Wu and Shen 2013), context (Barako and Brown 2008; Miralles-Quirós, Miralles-Quirós, and Redondo Hernández 2019), size (Shakil 2021; Flanagan and O'Shaughnessy 2005) were studied in their relation with ESG.

Notwithstanding the great interest of researchers in ESG, the same cannot be stated whether approaching concerns related to the second and disruptive perspective of ESG factors. When not exploited as resources, ESG factors may let controversies arise. In spite of several scandals of oldest and recent past, ESGC seems an almost uncharted territory when compared to the brightness that surrounds ESG. To be fair, the debate is getting intense in the business sector, where, for instance, the relationship between ESG scandals and firm value has been investigated, finding no impact of the former on the latter (Aouadi and Marsat 2018). Rodríguez-Fernández et al. (2019) demonstrated the absence of a significant relationship between the financial performance and ESGC, but rather the existence of a positive impact of ESGC on Return on Assets (ROA). Schiemann and Tietmeyer (2022) studied the association of ESGC, ESG disclosure and forecast accuracy on a sample of firms over 2008–2017 period, showing that analyst forecast errors are generally greater for firms that are highly exposed to ESGC. Li et al. (2019) proved a striking difference between public and private firms in these concerns, finding the former as the more likely to engage in ESG related actions with no subsequent real impact. Over a sample of oil and gas firms from 2010 to 2018, Shakil (2021) found that ESG controversies moderates ESG-total risk nexus.

Despite this attention, very few works dealt with ESGC in banking. Flood (2019) and Li et al. (2019), respectively, found the ESGC impact on bank financial performance and reputation, and on the detriment of goodwill. Bătae, Dragomir, and Feleagă (2020) over a sample of 108 European banks in 2018, indicated that larger banks tend to have higher scores in environmental, social, and governance dimensions individually, yet they are also associated with a greater number of controversies. Shakil, Tasnia, and Mostafiz (2021) analyzed 37 US banks over 2013–2017 period, by finding a significant positive relationship between board gender diversity and the ESG performance of US banks and contextually a nonsignificant moderating effect of ESG controversies on the board gender diversity—ESG performance nexus. In the most recent years, the ESGC banking literature is enriching itself, reflecting the growing recognition of ESG as a new risk component. In their analysis of banks from the US and Europe within 2016–2021 time window, Agnese et al. (2023) demonstrated a positive and statistically significant relationship between ESG governance and ESGC scores. Focusing on a sample of 135 listed European banks from 2002 to 2020, Bruno, Iacoviello, and Giannetti (2023) highlighted that effective management of ESG practices significantly enhances stability. This result was further corroborated by Galletta and Mazzù (2023) that explored the performance of worldwide listed banks from 2011 to 2020. Nevertheless, a study by Agnese et al. (2024) on a sample of European banks from 2015 to 2022 reveals that banks with a higher incidence of ESGC tend to attain superior performance metrics. This underscores a potentially opportunistic approach among credit institutions, indicating that they may prioritize profitability over effectively addressing ESG controversies. Dipierro, Toma, and Frittelli (2024) evaluated the efficiency of a sample of worldwide leading banks, finding that being inefficient in the ESGC domain is not a necessary evil to achieve productive efficiency.

It is clear that more evidence is needed in the ESGC field. In a recent bibliometric review, Shakil (2024) advocated for the application of alternative methodologies to the topic, including those derived from supervised ML techniques. We aim to contribute to this promising line of inquiry, by proposing the methodological framework described in the following section, to seek answering to our RQ.

3 | Methodology

Perhaps no task is more prevalent, or more useful, in economics than the prediction of a numerical value through panel data (Chen 2021). However, we should review the few but still valuable advices coming from the very sporadically and partly literature in the field of FS for longitudinal data. In spite of recent popularity of ML methods for longitudinal data, FS for these methods remains understudied (Chikersal et al. 2021; Speiser 2021). However, FS procedures prove further to be not only beneficial but are often necessary for development of prediction models in real world applications. Literature in the field was criticized, since not even applied to real performance but rather only simulation study or theoretical suggestion with no practical recommendation. Useful insights both theoretically and computationally come from Tian and Wang (2019). In this line, average methods seems the most suitable choice to summarize a genes' expression value at the gene set level or a gene's expression values over time. According to the authors, FS process for longitudinal data degrades to a conventional cross-sectional FS process and successfully conquers the longitudinal feature selection problem.

Therefore, reduction should be considered as the suitable approach to handle panel data for FS scopes. Computing ML techniques directly on panel data should be left in favor of reduction processes succeeding to maintain the key information hidden over time and instances. In this light, our methodology proposes to tackle with panel data, by aiming at pursue, independently, ESGC prediction and ESGC time series forecasting tasks. The methodological flowchart is displayed in Figure 1 and consists of: exploratory analysis, dataset transformation, data split, feature selection, statistical tests, construction of the best ESGC prediction and time series forecasting models, attribute importance identification and validation of the models.

Exploratory analysis: to kick off the analysis, data preparation is needed, as an extremely important step of pre-processing. From the original dataset, missing value imputation will be computed,

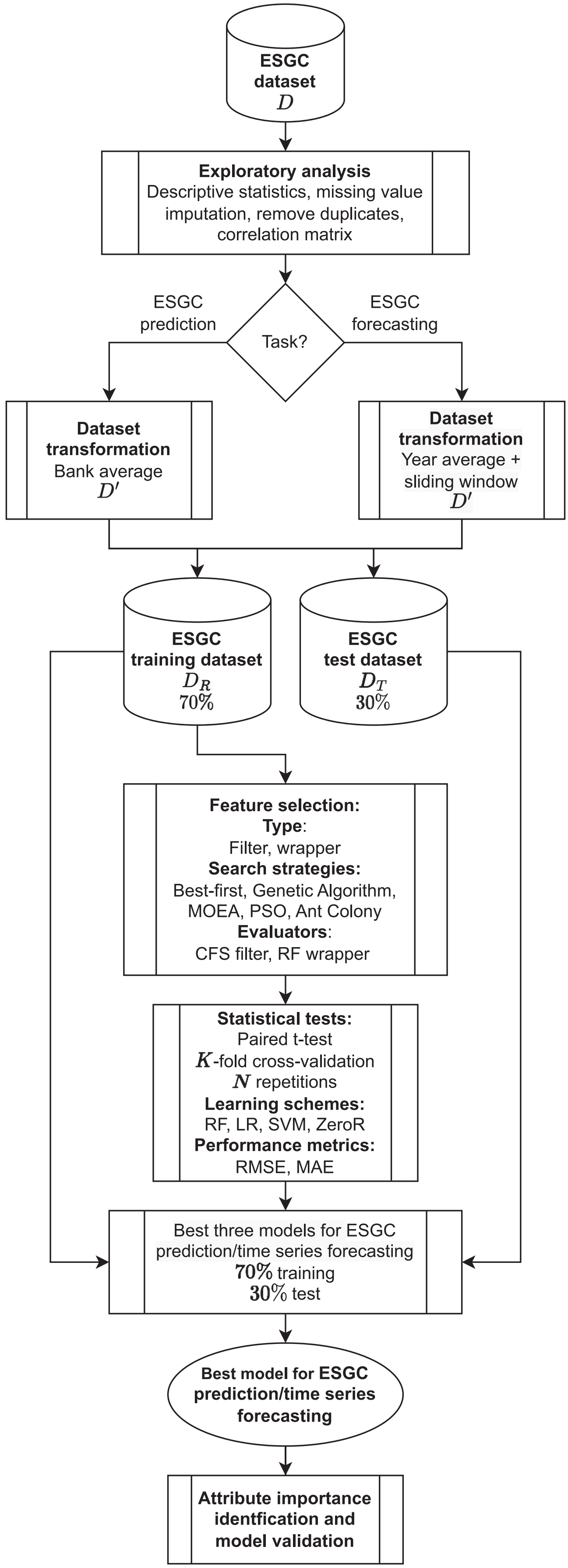

FIGURE 1 | Proposed methodology for ESGC prediction and time series forecasting.

duplicated instances will be removed. Then, descriptive statistics and correlation matrix are provided.

Dataset transformation: the panel dataset D is transformed, based on the task. Under ESGC prediction task, D is transformed into cross-sectional dataset by averaging the bank data, obtaining the transformed dataset D'. For the special purpose of ESGC time series forecasting task, the dataset transformation is computed by firstly averaging the bank data per year transforming the panel dataset D into a time series dataset, and then applying the *sliding window* method, finally obtaining the transformed dataset D'. Sliding window method is performed in order to remove time dependencies in the data, thus allowing the use of conventional machine learning algorithms for the construction of time series forecasting models as a supervised learning problem Brownlee (2020b). More in detail, let l be the *window size* (WS), that is, the number of previous time steps considered for data transformation. Given the dataset $D = \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n\}$, with $\boldsymbol{d}_t = \{x_t^1, \ldots, x_t^p, y_t\}$, $t = 1, \ldots, n$, where n is the number of different years in dataset D, p is the number of input attributes, and $y_t \in R$ is the average ESGC per year (target), the sliding window transformation process builds the following dataset D':

$$D' = \left\{ \left\{ x_{t-1}^1, \ldots, x_{t-l}^1 \right\}, \ldots, \left\{ x_{t-1}^p, \ldots, x_{t-l}^p \right\}, \left\{ y_{t-1}, \ldots, y_{t-l} \right\}, y_t \right\}, \quad t = 1, \ldots, n \tag{1}$$

Data split: each of the transformed datasets derived previously (D') are split into two datasets D_R and D_T, respectively, using 70% of the instances for training purposes and 30% for testing. D_R is used for the FS process and later for the training of the predictions models. The dataset D_T is used to test the prediction models. In this way, both the FS process and the training process do not involve the test dataset D_T, which will allow checking the overfitting ratio and the generalization error of the prediction models.

Feature selection: FS is the process of removing attributes from the dataset that are *irrelevant* and/or *redundant* with regarding the task to be performed Liu and Motoda (1998). Speeding up the training time of the prediction model while also improving response times, FS methods can improve the quality of the prediction in terms of accuracy. Highly desirable to facilitate the interpretation of results, FS methods can be classified as *attribute subset evaluation* and *attribute evaluation*. The former evaluates successive subsets of attributes that must be generated using a predefined *search strategy*. The latter evaluates each feature individually and assign a relevance level to each feature to return a feature ranking (hence, they are often called *feature ranking* methods). The FS methods that do not consider dependencies between attributes are *univariate*, while the methods that take into account dependencies between attributes are *multivariate*. Attribute evaluation methods can be univariate or multivariate, while attribute subset evaluation methods are always multivariate. Also, FS methods can be divided into *wrapper*, *filter*, and *embedded*. Wrappers use the predictive accuracy of a predetermined learning algorithm to determine the quality of selected attributes. Filters separate the attribute selection from the learning algorithm so that the influence of the learning algorithm does not interact with the FS algorithm. Embedded methods accomplish model fitting and attribute selection simultaneously. FS can be also expressed as a *boolean combinatorial optimization problem*, which the Appendix A provides further details and notation on. Optimization problems (from Equations A1 to A4) are *NP-hard* problems, where the search space contains 2^p possible solutions (attribute subsets). These types of problems are usually solved approximately by using *heuristics* or *meta-heuristics*. Over the two training datasets D_R, one for each task, we compute *best-first* (BF), *genetic algorithm* (GA), *particle swarm optimization* (PSO), and *ant colony* (AC) (Equations A1 and A3), and *multiobjective evolutionary algorithm* (MOEA) (Equations A2 and A4), both in filter and wrapper configuration. Hence, a set of 10 FS methods are used in this work, as shown in Table 1. BF is a heuristic algorithm that searches the space of attribute subsets by greedy hill-climbing augmented with a backtracking facility (Kohavi and John 1997). GA (Goldberg 1989) are meta-heuristics inspired by Darwin's natural selection and genetics. MOEA (Jiménez et al. 2017) is an evolutionary algorithm for multiobjective optimization, which once the Pareto front is identified, returns

TABLE 1 | Feature selection methods used in this study.

Short name	Type	Search strategy	Evaluator
BF-CFS	Filter	Best-first	Correlation-based FS
GA-CFS	Filter	Genetic algorithm	Correlation-based FS
MOEA-CFS	Filter	Multiobjective evolutionary algorithm	Correlation-based FS
PSO-CFS	Filter	Particle swarm optimization	Correlation-based FS
AC-CFS	Filter	Ant colony	Correlation-based FS
BF-RF	Wrapper	Best-first	RF(10 trees), RMSE, 5-fold CV
GA-RF	Wrapper	Genetic algorithm	RF(10 trees), RMSE, 5-fold CV
MOEA-RF	Wrapper	Multiobjective evolutionary algorithm	RF(10 trees), RMSE, 5-fold CV
PSO-RF	Wrapper	Particle swarm optimization	RF(10 trees), RMSE, 5-fold CV
AC-RF	Wrapper	Ant colony	RF(10 trees), RMSE, 5-fold CV

TABLE 2 | Features of ESGC dataset.

Name	Description	Source	Category
ESGC (target)	ESG Controversies Score	Refinitiv	
Employees	Number of full time equivalents	Bloomberg	Governance
Emp_pro	Employee protection's mechanisms	Bloomberg	Governance
Wom_board	Women in the board of directors	Bloomberg	Governance
Meet_board	Annual meetings of board of directors	Bloomberg	Governance
Npl	Non Performing Loans on total loans	Bloomberg	Capital and Assets
Roaa	Return on Average Assets	BankFocus	Capital and Assets
Roae	Return on Average Equity	BankFocus	Capital and Assets
Roa	Return on Assets	Bloomberg	Capital and Assets
Roe	Return on Equity	Bloomberg	Capital and Assets
Fin_lev	Average assets to average equity	Bloomberg	Capital and Assets
Capad	Equity on assets	Bloomberg	Capital and Assets
Equity	Total equity	Bloomberg	Capital and Assets
NPL/Assets	Non Performing Loans on assets	Bloomberg	Capital and Assets
Mktcap	Market capitalization	Bloomberg	Market
MTBV	Market to Book Value ratio	Bloomberg	Market
P/E	Price on Earnings ratio	BankFocus	Market
P/BV	Price on Book Value ratio	BankFocus	Market
Deposits	Total deposits	Refinitiv	Operations
Operat_exp	Operating expenses	Bloomberg	Operations
Int_exp	Interest expenses on total expenses	Bloomberg	Operations
Int&Div_inc	Income from interest and dividends	Refinitiv	Operations
Nim	Net interest Margin	Bloomberg	Operations
Loans/Assets	Total loans on total assets	Bloomberg	Liquidity
Ebitadj	Ebit adjusted for abnormal items	Bloomberg	Earnings
NetProfit	Net Income	Bloomberg	Earnings
Earnings	Net income on Total Assets	Bloomberg	Earnings
Geo	Headquarter	Refinitiv	Context

Note: 0–100; 0: low performance; 100: high performance. 1: bank has mechanisms of protection of employees; 0: otherwise. 1: Oceania; 2: Africa; 3: South America; 4: North America; 5: Asia; 6: Europe.

the solution with the best RMSE, in Problem (A2), and with the best trade-off between relevance and redundancy, in Problem (A4). PSO (Moraglio, Di Chio, and Poli 2007) are bio-inspired metaheuristics that optimize a problem by iteratively moving solutions (particles) in the search space according to a simple mathematical formula about the position and velocity of the particle, influenced by the best known local position of each particle and by the best known global position of the swarm. AC (Fong, Biuk-Aghai, and Millham 2018) are also bio-inspired metaheuristics that solve computational problems by using pheromone-based communication of biological ants. Correlation-based FS (CFS) and Random Forest (RF) are, respectively, used as evaluators for filter and wrappers methods, as in Appendix A.

Statistical tests: we perform a *paired t*-test to compare the reduced datasets obtained in the FS phase. Each reduced dataset (along with the original dataset D_R) is evaluated with N repetitions of K-fold cross-validation (such that $K \cdot N$ is, at least, 30 prediction models in total) using the learning schemes RF, LR (Freedman 2005), SVM (Cortes and Vapnik 1995), (Cortes and Vapnik 1995) and *ZeroR* (Brownlee 2020a), and the RMSE, mean squared error (MAE) and correlation coefficient (CC) performance metrics. Each dataset is compared with the rest, producing a ranking according to the difference *wins–losses*, where *wins* is the number of times that the dataset obtained statistically significant differences in favor, conversely for *losses*. The same process is performed with the learning schemes.

TABLE 3 | Descriptive statistics.

Name	Min	Q1	Median	Mean	Q3	Max	SD
2011–2020							
Employees	138	14199.25	37166.00	60014.65	61553.50	503082.00	80242.00
Emp_pro	0.00	1.00	1.00	0.76	1.00	1.00	0.43
Wom_board	0.00	8.33	18.75	18.78	27.40	55.56	12.83
Meet_board	0.00	8.00	10.00	12.07	14.25	75.00	9.08
Npl	0.00	0.82	1.58	2.64	2.88	40.47	3.27
Roaa	−12.28	0.68	1.00	1.00	1.26	20.22	1.05
Roae	−34.20	1.76	2.98	4.93	4.93	58.13	5.86
Roa	−12.28	0.61	0.95	0.97	1.26	20.37	1.06
Roe	−33.80	8.04	11.52	11.46	15.58	58.23	7.12
Fin_lev	1.83	9.86	13.26	14.12	17.28	56.53	5.93
Capad	0.00	0.06	0.08	0.09	0.11	0.99	0.05
Equity	328.43	26,471	57,278	2,947,833	224,311	$208{,}784 \times 10^{3}$	$162{,}778 \times 10^{2}$
NPLonAssets	0.00	0.00	0.01	0.01	0.02	0.14	0.02
Mktcap	513	27,483	72,255	$775{,}155 \times 10^{1}$	298,587	$834{,}572 \times 10^{3}$	$538{,}660 \times 10^{2}$
MTBV	0.00	0.87	1.29	1.44	1.76	8.36	0.87
P/E	0.00	8.78	12.05	16.19	16.04	994.84	43.77
P/BV	0.00	0.85	1.26	1.43	1.73	37.88	1.29
Tot_Deposits	$565{,}800 \times 10^{1}$	$179{,}500 \times 10^{3}$	$459{,}500 \times 10^{3}$	$362{,}800 \times 10^{5}$	$207{,}400 \times 10^{4}$	$261{,}900 \times 10^{7}$	$197{,}125 \times 10^{6}$
Operat_exp	137.94	4380.25	12,445	653536.56	60997.06	$521{,}638 \times 10^{2}$	$366{,}903 \times 10^{1}$
Int_Exp	3918	$229{,}000 \times 10^{1}$	$111{,}400 \times 10^{2}$	$796{,}000 \times 10^{3}$	$640{,}300 \times 10^{2}$	$476{,}800 \times 10^{5}$	$433{,}326 \times 10^{4}$
Int&Div_Income	2116	$737{,}300 \times 10^{1}$	$272{,}400 \times 10^{2}$	$154{,}700 \times 10^{4}$	$126{,}800 \times 10^{3}$	$121{,}800 \times 10^{6}$	$932{,}642 \times 10^{4}$
Nim	−0.03	1.62	2.57	2.67	3.12	33.55	2.19
LoansonAssets	0.00	0.45	0.56	0.54	0.65	0.85	0.16
Ebitadj	−3334	2205	6486	363,801	33,963	$175{,}787 \times 10^{2}$	$175{,}787 \times 10^{2}$
NetProfit	−189,567	1555	4718	436,960	19,806	$343{,}726 \times 10^{2}$	$283{,}646 \times 10^{1}$

(Continues)

TABLE 3 | (Continued)

Name	Min	Q1	Median	Mean	Q3	Max	SD
Earnings	−0.12	0.01	0.01	0.01	0.01	0.17	0.01
Geo	1.00	4.00	5.00	4.72	5.00	6.00	1.09
ESGC	0.49	68.52	94.66	76.84	100.00	100.00	31.58
2021							
Employees	215.00	15458.25	38338.00	59999.70	67271.00	459000.00	77992.33
Emp_pro	0.00	1.00	1.00	0.81	1.00	1.00	0.40
Wom_board	0.00	12.78	23.08	22.96	35.29	54.55	14.37
Meet_board	0.00	9.00	13.00	15.89	18.00	75.00	11.21
Npl	0.00	0.76	1.50	2.26	2.63	18.32	2.45
Roaa	−10.83	0.60	0.85	0.85	1.17	10.44	1.38
Roae	−19.81	6.85	9.78	9.65	13.05	27.16	6.03
Roa	−10.83	0.41	0.75	0.64	0.97	2.72	1.12
Roe	−19.81	5.64	8.64	8.15	11.49	21.08	5.75
Fin_lev	1.83	9.71	13.17	13.51	16.69	31.84	5.07
Capad	0.00	0.06	0.08	0.09	0.11	0.57	0.05
Equity	1687.13	39908.00	77470.22	$446,453\times10^{1}$	397723.01	$199,911\times10^{3}$	$2,409,997\times10^{1}$
NPLonAssets	0.00	0.00	0.01	0.01	0.01	0.10	0.01
Mktcap	2118.18	27321.03	75840.29	$130,878\times10^{2}$	275210.49	$834,572\times10^{3}$	$865,543\times10^{2}$
MTVB	0.00	0.63	1.07	1.32	1.55	8.36	1.10
P/E	0.00	8.48	12.83	15.80	16.42	131.50	16.16
P/BV	0.00	0.60	1.05	1.50	1.51	37.88	3.19
Tot_Deposits	$153,219\times10^{2}$	$257,820\times10^{3}$	$587,993\times10^{3}$	$500,665\times10^{5}$	$259,332\times10^{4}$	$261,895\times10^{7}$	$260,414\times10^{6}$
Operat_exp	296.04	4704.25	14165.00	924265.13	60997.06	$521,637\times10^{2}$	$528,138\times10^{1}$
Int_Exp	10647.00	$198,859\times10^{1}$	$918,350\times10^{1}$	$865,962\times10^{3}$	$687,985\times10^{2}$	$377,226\times10^{5}$	$459,908\times10^{4}$
Int&Div_Income	3884.00	$936,271\times10^{1}$	$288,758\times10^{2}$	$210,464\times10^{4}$	$149,825\times10^{3}$	$116,933\times10^{6}$	$125,114\times10^{4}$
Nim	0.57	1.52	2.26	2.47	2.97	17.24	1.73

(Continues)

TABLE 3 | (Continued)

Name	Min	Q1	Median	Mean	Q3	Max	SD
LoansonAssets	0.02	0.44	0.56	0.53	0.63	0.79	0.15
Ebitadj	−3334.00	2222.25	6486.50	363800.99	33963.30	$1{,}757{,}867\times10^{1}$	$191{,}818\cdot10^{1}$
NetProfit	−189567.00	1281.08	4535.50	514679.05	20112.52	$271{,}311\times10^{2}$	$312{,}977\times10^{1}$
Earnings	−0.09	0.00	0.01	0.01	0.01	0.03	0.01
Geo	1.00	4.00	5.00	4.72	5.00	6.00	1.09
ESGC	0.89	42.08	100.00	73.27	100.00	100.00	35.78

Best ESGC prediction/time series forecasting model: It follows the selection of the three best ESGC prediction and forecasting models, to be evaluated on testing data D_T to derive the best-in-class model for each task.

Attribute importance identification: Once outlined the best ESGC prediction and time series forecasting models, the predictive power of selected attributes is scored using *RF internal importance* and *Permutation based feature importance*. Suffering from being computed on statistics on the training dataset, the former produce importance that may be high even for features that later result as not predictive, as long as the model has the capacity to use them to overfit. The latter stands for a valid alternative that can mitigate these risks, being a model inspection technique reporting the decrease in a model score when a single feature value is randomly shuffled (Breiman 2001).

Model validation:finally, results are validated taking into account the main literature in the problems' domains already illustrated in Section 2.

4 | Experiments and Results

4.1 | ESGC Dataset

The sample is drawn from the Bloomberg World Banks Index (BWBANK), a capitalization-weighted index of the leading banks stocks that involves 165 banks. Each sampled bank was measured over many dimensions, consisting of a target and 27 features. Additionally, due to data availability, the final sample consists of 140 banks over 2011–2020, with 1440 instances.

4.1.1 | Target

Most recently, scores on controversies joined the entire sphere of ESG measures in the disparate datasets available, either for free and not. In this scenario, Refitiniv provides ESG databases, which consists of ESG, ESG Controversies, and ESG-C scores. Specifically, ESG-C is obtained by doing the weighted average of the ESG scores and ESG Controversies score per fiscal period. For the main purpose of this research, the focus will be on ESG Controversies, chosen as our target, under the label ESGC. Collected in October 2022 via Refinitiv, ESGC is based on 23 ESG controversy topics, related to community, human rights, management, product responsibility, resource use, shareholders, and workforce, ESGC fits for the analysis since it catches the penalization consequential to a scandal. Scores are provided both on a percentile basis and a letter grade. Scores go from 0 to 100, and so from D− to A+, by respectively signaling poor/excellent ESGC level and insufficient/high degree of transparency in reporting ESGC data publicly.

4.1.2 | Features

Notwithstanding the big debate regarding the superiority of certain indicators in predicting bank failures (Petropoulos et al. 2020), no indication in this sense emerges considering prediction of controversies performance. Several empirical studies

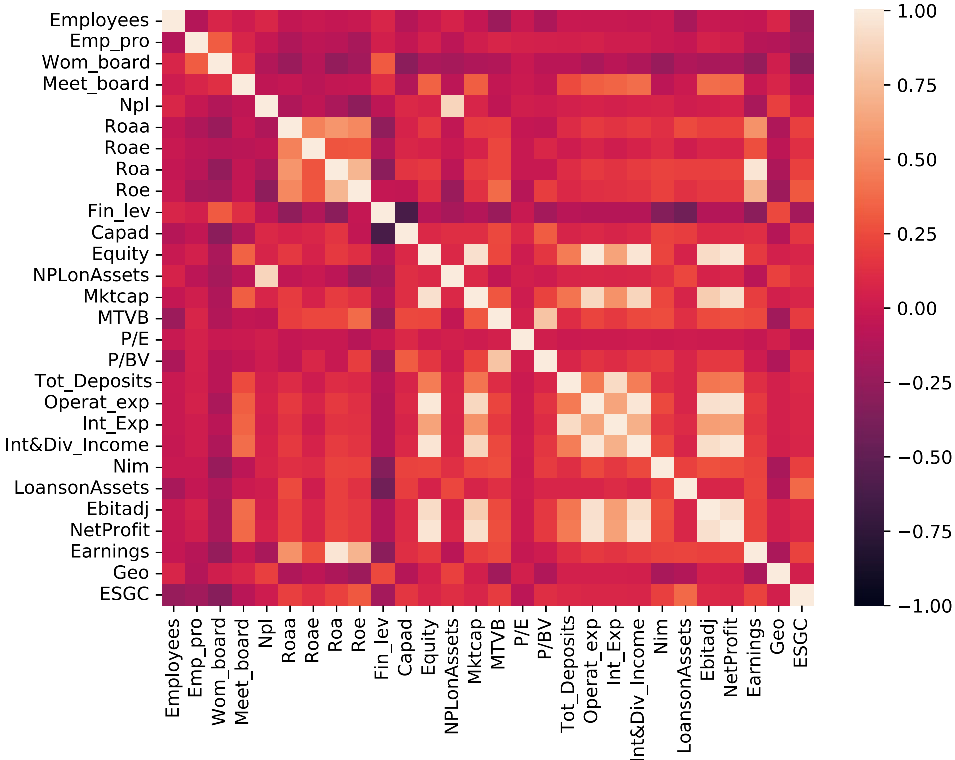

FIGURE 2 | Correlation matrix of all variables used in the study.

TABLE 4 | Hyperparameters of the search strategies used for FS in ESGC prediction problem.

Search strategy	Hyperparameters
BF	Direction = bi-directional; search termination = 5
GA	Population size = 20; generations = 1000; crossover prob. = 0.6; mutation prob. = 0.033
MOEA	Population size = 20; generations = 1000
PSO	Swarm size = 20; iterations = 1000; mutation type = bit-flip; mutation prob. = 0.01; cognitive weight = 0.34; social weight = 0.33; inertia weight = 0.33
AC	Colony size = 20; iterations = 1000; mutation type = bit-flip; mutation prob. = 0.01; chaotic coef. = 4.0; evaporation rate = 0.9; heuristic rate == 0.7; pheromone rate = 2.0

on bank failures refer to variables inherent to CAMELS—acronym of Capital, Asset Quality, Management, Earnings, Liquidity, and Sensitivity to market risk indicators—to assess the soundness of financial institutions (Durand and Le Quang 2022). Literature on bank reputation (Fang 2005; Jizi et al. 2014; Pérez and del Bosque 2015), its efficiency (Abate, Basile, and Ferrari 2021; Alam, Banna, and Hassan 2021; Belasri, Gomes, and Pijourlet 2020; Balcerzak et al. 2017; Wheelock and Wilson 2000), and failures (Petropoulos et al. 2020; Viswanathan, Srinivasan, and Hariharan 2020) was accounted for. It results an extended set of explanatory features, as in Table 2, capturing many banks' dimensions, grouped as follows: Governance, Capital and Assets, Market, Operations, Liquidity, Earnings, and Context.

4.2 | Results

4.2.1 | Explanatory Analysis

Within our final sample of 1440 instances, which consists of 140 banks measured over 10 years through 27 features and the target ESGC, few imputation of missing values were computed. Hence, we derived—up in the Table 3 the descriptive statistics for the whole set of features, and target, for the whole time-span 2011–2020. However, given the long-term horizon covered by our panel dataset, and the upheaval that Covid-19 pandemic arose throughout the economy, descriptive statistics for 2020 are also provided—in the lower part of the 3. Variables are explored in terms of minimum (Min), first quartile (Q1), second quartile (Median), mean (Mean), third quartile (Q3), maximum (Max), and standard deviation (SD).

Across years, high SD values are registered for some variables. Even if for some ML techniques normalization of features is required, the chosen set of search strategies enable to avoid this step. In so doing, avoiding normalization more meaningful insights can be drawn from descriptive statistics and we can directly link values to single instances. Thus, exploring examples ought to be noted, Q3 of *Wom_board* stops at 27.40

over the 2011–2020 period. Explicitly, only the 25% of the instances exceeds the 27.40% of women on the board of directors, proving the few observations hosted by the tails of the distribution. The bank that counts for more than one observation above 50% is Societe Generale. The situation in terms of board gender diversity is encouraging in 2020, where Q3 reaches 35.29%, even if only Skandinaviska Enskilda Banken has more than 50% of woman on board. *Npl* also demonstrates interesting issues, in terms of the huge gap between Q3 and Max, respectively, about 3% and 40% of unproductive loans. The worrying value of 40.47% of *Npl* is detained by Morgan Stanley (MS) early in the considered time-span. Even if it is true that in 2011–2012, some analysts were worried about MS positions in Europe and PIIGS (Portugal, Ireland, Italy, Greek, Spain), they soon become tolerable for MS. However, across 2011–2020, MS was stained by countless scandals and angered in accusations. From further comparisons between the upper and lower portion of the Table 3, *Operat_exp* suggests that the average operating expenses for the sampled banks in 2020 is 140% of the operating expenses over the 10 years considered. It mainly be due by Covid-19 pandemic, which challenged operativity, by also decreasing liquidity as from *LoansonAssets*, earnings and market capitalization whether comparing the Q3 of *Earnings* and *Mktcap* across tables.

Attention should be also provided to Figure 2, which presents the heatmap of correlation matrix for the overall set of features and target, by means of meaningful color differences. Its support helps derive insights related to the direction followed by variables when compared twice at time. Moreover, computing the correlation indexes between variables enables to exclude multicollinearity. In our methodological approach, the highly positive correlation values (light color) for some variables and the highly negative (dark color) for *Fin_lev* and *Capad* and for *Fin_lev* and *LoansonAssets* do not represent a problem for the implemented search strategies and developed models.

4.2.2 | ESGC Prediction Task

As above stated, reduction is the suitable approach to handle longitudinal data. In this case, the problem was faced by firstly transforming the dataset into a cross-sectional dataset, by therefore including the bank average for all values over the considered long time-span. Given the ratio explained in Section 3, after the data split, the FS process is implemented in the training dataset. After having performed the 10 attribute subset evaluation methods for FS with the hyperparameters in Table 4, statistical tests succeed to provide the ranking of the datasets depending on the difference wins-losses. Table 5 shows the results that emerge from the comparison of each reduced dataset obtained in FS phase using RF, LR, SVM, and ZeroR learning schemes and RMSE, MAE, and CC performance metrics.

TABLE 5 | Ranking of datasets based on difference wins-losses in ESGC prediction problem.

Metric	Wins	Losses	Difference	Result set
RMSE	0	2	−2	BF-RF
	0	2	−2	AC-RF
	0	1	−1	GA-RF
	0	1	−1	PSO-CFS
	0	1	−1	MOEA-CFS
	0	1	−1	GA-CFS
	0	1	−1	BF-CFS
	0	1	−1	AC-CFS
	0	0	0	Test-base
	2	0	2	PSO-RF
	8	0	8	MOEA-RF
MAE	1	0	1	AC-RF
	1	0	1	PSO-CFS
	1	0	1	MOEA-CFS
	1	0	1	GA-CFS
	1	0	1	AC-CFS
	1	0	1	Test-base
	0	0	0	PSO-RF
	0	0	0	GA-RF
	0	0	0	BF-RF
	0	0	0	BF-CFS
	6	0	6	MOEA-RF
CC	2	0	2	MOEA-RF
	1	0	1	GA-RF
	1	0	1	PSO-CFS
	1	0	1	MOEA-CFS
	1	0	1	GA-CFS
	1	0	1	BF-CFS
	1	0	1	AC-CFS
	0	0	0	PSO-RF
	0	0	0	Test-base
	0	1	−1	AC-RF
	0	7	−7	BF-RF

Let test dataset join the training dataset for the best three ESG prediction models. MOEA-RF resulted as the best ESGC prediction model, from Table 5, using D_R. Nevertheless, the best dataset in D_T is produced by PSO-RF. It results that PSO-RF performed slightly better than MOEA-RF, concerning data in D_T not evaluated in the optimization algorithm. Derived by the application of the PSO search strategy and a RF wrapper evaluator, PSO-RF selected a set of 15 variables that includes the target and the following features randomly ordered: *Employees*, *Emp_pro*, *Wom_board*, *Meet_board*, *Roe*, *Equity*, *Mktcap*, *P/E*, e *Tot_Deposits*, *Operat_exp*, *Int&Div_Income*, *LoansonAssets*, *NetProfit*, *Geo*.

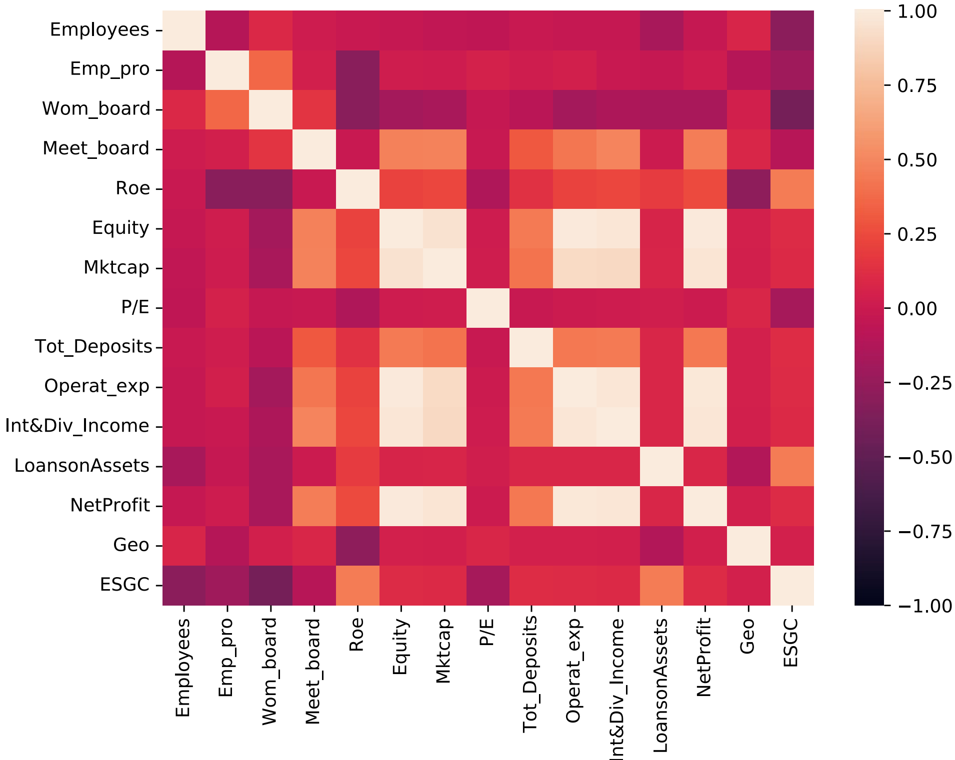

FIGURE 3 | Correlation matrix of variables selected by PSO-RF in ESGC prediction problem.

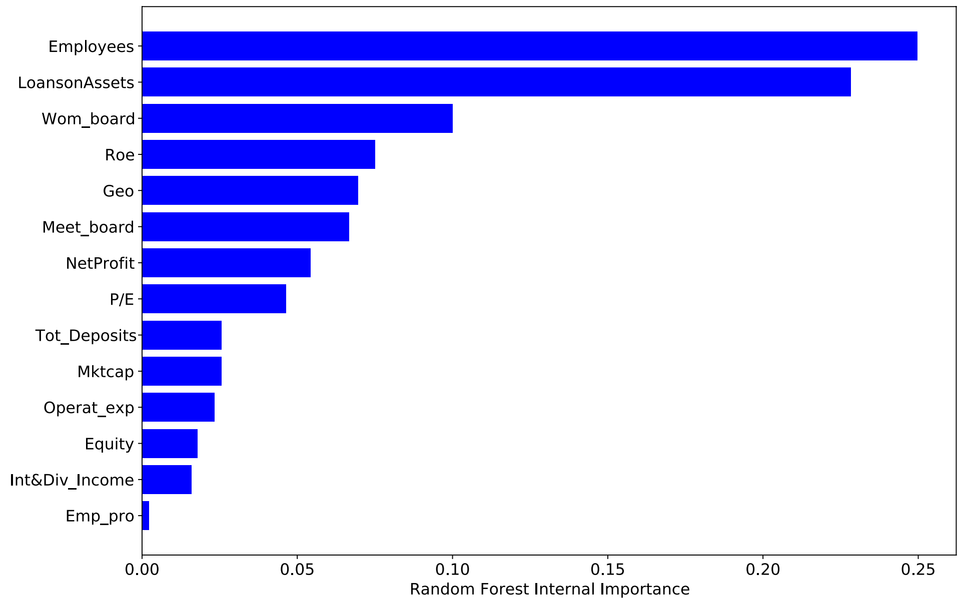

FIGURE 4 | Random Forest internal importance in ESGC prediction problem.

Afterwards, it is useful computing the correlation matrix of the resultant attributes that were selected and the target ESGC, as in Figure 3. Among the selected features, *Employees*, *Wom_board Emp_pro*, and *P/E* are the most highly negative correlated with the target, whereas positive and medium-high values of correlation indexes are registered between the target and *Roe*, and the target and *LoansonAssets* are highly negative correlated. The set of selected attributes can be also evaluated in terms of the importance of each attribute in predicting the target. Each attribute's explanatory power is therefore evaluated producing rankings. According to the relevant literature in the field (Durand and Le Quang 2022), attributes' importance was evaluated depending on the following macrocriteria: internal importance that RF provides to each attribute without applying any method, and the permutation-based importance that is instead calculated after a model is fitted. In this latter case, three scoring metrics were used: R^2, negative RMSE and negative MSE. Accordingly, Figure 4 suggests insights looking at the RF-internal importance, and Figures 5–7 show the Permutation-based importance using R^2, negative RMSE and negative MSE.

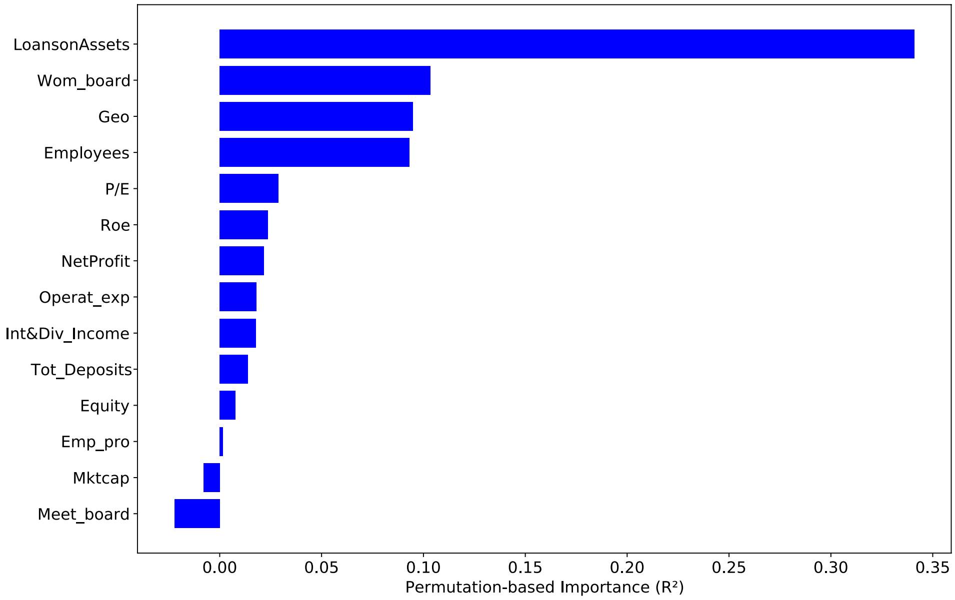

FIGURE 5 | Permutation-based importance in ESGC prediction problem (R^2).

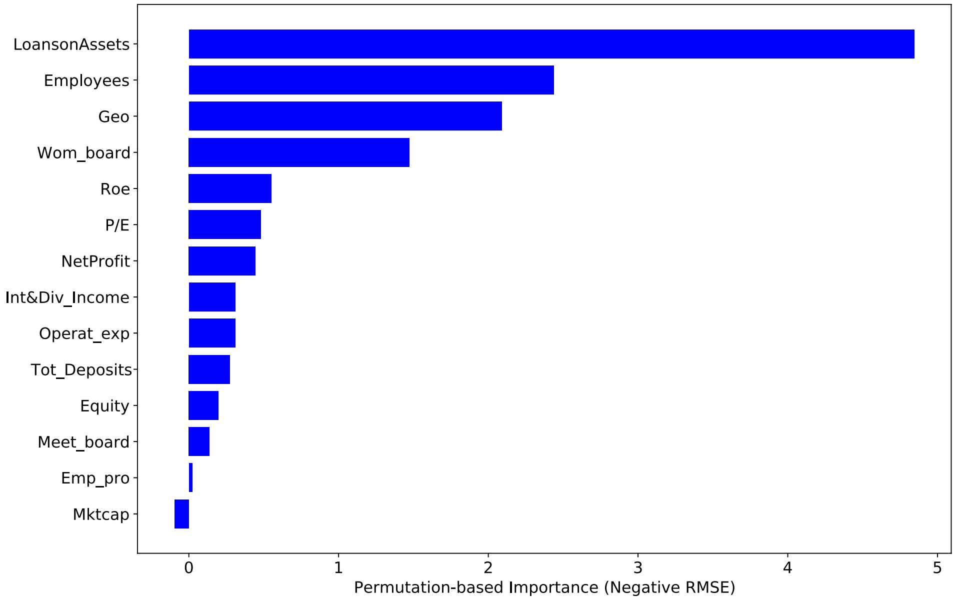

FIGURE 6 | Permutation-based importance in ESGC prediction problem (negative RMSE).

4.2.3 | ESGC Time Series Forecasting Task

For the lagged transformation, we use the *sliding window* method. The same search strategies' set implemented in the previous task are computed, as in Table 6.

From statistical tests, no significant statistical differences result amongst the produced datasets. The choice is that of the combination that outperforms in terms of the involved metrics (RMSE, MAE, CC), that is, AC-RF, with the 36 attributes in Table 7.

To be more confident, AC-RF dataset is compared with the base-line dataset that lagged the whole variables. Tables 8 and 9 shows the future predictions from training and test data at 1, 2, and 3 years ahead. AC-RF outperforms the All-attributes model, as proved by the overfitting ratio (errors from training divided by errors from testing data) in Table 10. Figure 8 allows for a graphical visualization of ESGC actual and predicted values from AC-RF model from the end of the training data. Average ESGC predicted values in the three-time periods are 78.5572, 77.6033, and 77.8213.

5 | Conclusions

In its construction, our work addresses some methodological and empirical gaps. Exploiting the potential of FS in panel datasets, we address ESGC topic, which is still uncharted in banks. In detail, we perform prediction and time series

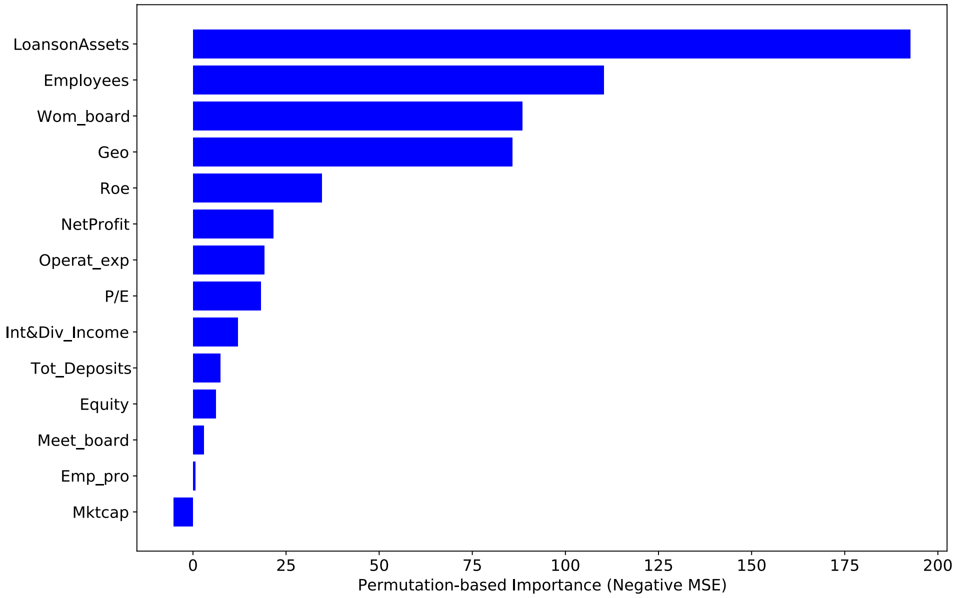

FIGURE 7 | Permutation-based importance in ESGC prediction problem (negative MSE).

TABLE 6 | Hyperparameters of the search strategies in ESGC time series forecasting problem.

Search strategy	Hyperparameters
BF	Direction = bi-directional; search termination = 5
MOEA	Population size = 20; generations = 1000
GA	Population size = 20; generations = 1000 Crossover prob. = 0.6; mutation prob. = 0.033
PSO	Swarm size = 20; iterations = 1000 Mutation type = bit-flip; mutation prob. = 0.01 Cognitive weight = 0.34; social weight = 0.33 Inertia weight = 0.33
AC	Colony size = 20; iterations = 1000 Mutation type = bit-flip; mutation prob. = 0.01 Chaotic coef. = 4.0; evaporation rate = 0.9 Heuristic rate == 0.7; pheromone rate = 2.0

TABLE 7 | Features selected by AC-RF method in ESGC time series forecasting problem.

Selected features	Lag
Employees, *Roaa*, *NPLonAssets*, *PE*, *Nim*, *LoansonAssets*	−3
Roe, *MTBV*, *Int&Div_Income*, *Earnings*	−1
Wom_board, *Equity*, *PBV*, *NetProfit*	−1, −3
Capad, *Int_exp*	−2, −3
Npl, *Roae*, *Mktcap*, *Operat_exp*	−1, −2, −3

TABLE 8 | MAE and RMSE at 1, 2, 3-steps ahead predictions from AC-RF model.

Data	Metric	1-step-ahead	2-steps-ahead	3-steps-ahead
Training				
	MAE	1.2305	1.3716	0.4111
	RMSE	1.7416	1.9563	0.5197
Test				
	MAE	1.7163	2.4961	4.5543
	RMSE	2.6431	3.2352	4.5543

forecasting of the banks' ESGC level for a real and large-dimensional dataset.

Once defined the suitable ways to deal with the critical junctures that the duality of the task produces, that is, prediction and time series forecasting of the target ESGC, accurate results have been produced through PSO and AC wrapper models, respectively, for each task. Proving the relevance of governance-employees dynamics in this domain, we demonstrated the influence of the external environment on the organizational behavior. Interdependences with the outside world seem easier to be managed when the internal environment is correctly dealt with.

Relevant implications for researchers, practitioners and regulators follow.

The work contributes firstly to the main literature in addressing controversies related to ESG concerns. The study of ESGC is currently in the developmental phase. In this promising field, ESGC research within the banking sector is quite new. With such models, it has been made possible to both predict ESGC for a newly added bank out-of-sample in the same time-span, or producing the time series forecasting for the sampled banks. Authors aims at stimulating further tests of the proposed methodology. Performing FS in a panel dataset stands for a quite uncharted field. The intention is to open up for further applications in the spirit of knowledge and sharing, even in different sectors. Data from the recent and oldest past are treated through ad hoc techniques able to handle prediction tasks, in a reduced computational time and under complex configurations that real-world data host. Methodologically, the accuracy and reliability through which literature is increasingly entrusting bio-inspired metaheuristic algorithms as PSO and AC are further powered by this study. Also MOEA well-performs both in the prediction and time series forecasting tasks. Moreover, the research further stresses the fitness of the strategy to include ML methods in banks' toolkit (Doerr, Gambacorta, and Serena 2021), in the current on-going changes for risk management and governance. Indeed, this work proves that ML performs well when it comes to dealing with banks, also due to the large availability of balance sheet data, in line with Durand and Le Quang (2022), Petropoulos et al. (2020), Tanaka, Kinkyo, and Hamori (2016), and Kumar and Ravi (2007).

TABLE 9 | MAE and RMSE at 1, 2, 3-steps ahead predictions from All-attributes model.

Data	Metric	1-step-ahead	2-steps-ahead	3-steps-ahead
Training				
	MAE	1.2960	1.3292	0.2380
	RMSE	1.8246	2.0402	0.2782
Test				
	MAE	2.1886	3.1178	4.7808
	RMSE	2.8914	3.5336	4.7808

TABLE 10 | Overfitting ratios at 1, 2, 3-steps ahead predictions from AC-RF and all-attributes model.

Model	Metric	1-step-ahead	2-steps-ahead	3-steps-ahead
AC-RF				
	MAE	0.7169	0.5495	0.0902
	RMSE	0.6589	0.6047	0.1141
All-attributes				
	MAE	0.5922	0.4263	0.0498
	RMSE	0.6310	0.5774	0.0582

Variables inherent to governance-employees sphere resulted as particularly relevant in predicting ESGC and are worthy to be of further interests of researchers, supporting (Shakil 2021), as well as of practitioners. The huge importance of governance-employees dynamics in face of other operational and financial domains, when delving into their ESGC explanatory power, is a result that should push further practitioners. Banks themselves should increase their attention to avoid underestimation of governance-employees linkage and capital and assets adequacy, enhancing risk management. Promoting operational autonomy, job satisfaction and status may be of aid of employees and, certainly of the entire organizational climate. In so doing, emerging issues may be dealt with in a timely manner and appropriately with an effective oversight and by raising the attention toward areas concerning governance-employees domain. Investors, as well, can benefit from the provided findings in terms of the knowledge boost about the nature of features to take care of when approaching to banks, with a view toward controversies' effects prevention.

The research can be of support of authorities' toolkit to assess the financial system health, and proactively supervise entities in risk management. Improving monitoring toward the important dimensions for ESGC prediction is herein suggested, in light of the unique exposure to systemic risk summarized by the "knock-on effect" expression Le and Viviani (2018). Audit committees dynamics deserve attention in this regard (Shakil 2024).

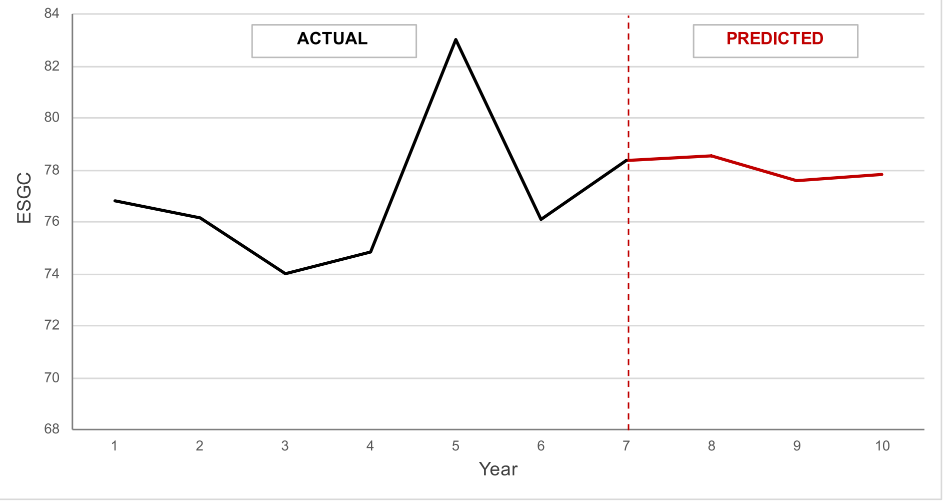

FIGURE 8 | Time-series of ESGC actual and predicted values.

Further, more support to banks' managers is needed to face everyday challenges, given the huge pervasiveness of media that amplifies scandals' noise. Promoting gender diversity, adopting mechanisms of employees' protection may lead to higher ESGC performance, when restoring reputation, or may prevent scandals' incurring. Moreover, even if not as largely evident as for governance-employees concerns, banks should be encouraged in cleaning their balance sheets and pursue adequate liquidity ratios and capital adequacy.

Despite its contributions, the study has some limitations. As first, the dataset transformation may threat the study. Nevertheless, despite transforming dataset allowed for loss of information, the generated models are hopeful. Empirically, data provision was quite challenging and allowed to obtain a sample of 1440 year-bank instances, referred to 140 banks in a 10 years time-span. Despite the main strengths of the ESGC used score, which made it the most used indicator for controversies in the literature (Agnese et al. 2023, 2024; Bruno, Iacoviello, and Giannetti 2023; Dorfleitner, Kreuzer, and Sparrer 2020; Dipierro, Toma, and Frittelli 2024), the standardized ESG reporting resulting from the EU's 2022 Corporate Sustainability Reporting Directive (CSRD) will ensure in the future utilizing expanded datasets with even more accurate information on such dynamics. As an additional limitation, this work avoided to involve macroeconomic variables of potential interest given the variety of contexts of the sampled banks. Thus, future researches may work on these directions. It will be useful applying the proposed methodology to enriched datasets, which ML has proved to perform better for. Also, starting from our limitation referred to the chosen sample of banks inquired, future works may let banks out of the giants join the sample, due to the high likelihood to incur in triggering events and break the news for large banks. Further developments may also focus on one geographical area, in light of the relevance that the context exerts on ESGC domain. Indeed, we in future studies we should consider the heterogeneity of ESG-related issues across countries, fostering the main findings of DasGupta and Roy (2023). Supporting Hiebl et al. (2018) and Sassen et al. (2018), solutions referred to risk management in the financial service industry need to be adapted to the institutional and legal setting. Future studies may also explore the sentiment of the general public, as well as of the investors, during the scandals. The models may be also enriched by data related to board member compensation or executive compensation. In line with Shakil (2024), future research can further leverage supervised ML-based methods. We suggest to implement FS, by looking at the local importance of features, to assess how correct and wrong predictions arise in specific cases. Further, the topic fits for features reduction and clustering tasks. Finally, future researches may opt for further increasing interpretation, by exploiting the potential of Explainable Artificial Intelligence. Decision-trees and rule-based evolutionary systems are worthy to be explored in evolutionary settings according to what Ribeiro, Singh, and Guestrin (2016) defined as an even more important part of data science: interpreting a model prediction.

Acknowledgments

Open access publishing facilitated by Universita della Calabria, as part of the Wiley - CRUI-CARE agreement.

References

Abate, G., I. Basile, and P. Ferrari. 2021. "The Level of Sustainability and Mutual Fund Performance in Europe: An Empirical Analysis Using Esg Ratings." *Corporate Social Responsibility and Environmental Management* 28: 1446–1455. https://doi.org/10.1002/csr.2175.

Adamyk, B., A. Skirka, K. Snihur, and O. Adamyk. 2019. "Analysis of Trust in Ukrainian Banks Based on Machine Learning Algorithms." In *2019 9th International Conference on Advanced Computer Information Technologies (ACIT)*, 234–239. Ceske Budejovice, Czech Republic: IEEE. https://doi.org/10.1109/ACITT.2019.8779974.

Addo, P. M., D. Guegan, and B. Hassani. 2018. "Credit Risk Analysis Using Machine and Deep Learning Models." *Risks* 6, no. 2: 38. https://doi.org/10.3390/risks6020038.

Agarwal, A., A. Gupta, A. Kumar, and S. G. Tamilselvam. 2019. "Learning Risk Culture of Banks Using News Analytics." *European Journal of Operational Research* 277: 770–783. https://doi.org/10.1016/j.ejor.2019.02.045.

Agnese, P., F. Battaglia, F. Busato, and S. Taddeo. 2023. "Esg Controversies and Governance: Evidence From the Banking Industry." *Finance Research Letters* 53: 103397. https://doi.org/10.1016/j.frl.2022.103397.

Agnese, P., M. Cerciello, R. Oriani, and S. Taddeo. 2024. "Esg Controversies and Profitability in the European Banking Sector." *Finance Research Letters* 61: 105042. https://doi.org/10.1016/j.frl.2024.105042.

Alam, A. W., H. Banna, and M. K. Hassan. 2021. "Esg Activities and Bank Efficiency: Are Islamic Banks Better?" *Journal of Islamic Monetary Economics and Finance* 8, no. 1: 65–68. https://doi.org/10.21098/jimf.v8i1.1428.

Ala'raj, M., and M. F. Abbod. 2016. "A New Hybrid Ensemble Credit Scoring Model Based on Classifiers Consensus System Approach." *Expert Systems With Applications* 64: 36–55. https://doi.org/10.1016/j.eswa.2016.07.017.

Alexandra, J., and K. P. Sinaga. 2021. "Machine Learning Approaches for Marketing Campaign in Portuguese Banks." In *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, 1–6. Makasar, Indonesia: IEEE. https://doi.org/10.1109/ICORIS52787.2021.9649623.

Alkhatib, K., and S. Abualigah. 2020. "Predictive Model for Cutting Customers Migration From Banks: Based on Machine Learning Classification Algorithms." In *2020 11th International Conference on Information and Communication Systems (ICICS)*, 303–307. Irbid, Jordan: IEEE. https://doi.org/10.1109/ICICS49469.2020.239544.

Alonso-Robisco, A., and J. M. Carbó. 2022. "Can Machine Learning Models Save Capital for Banks? Evidence From a Spanish Credit Portfolio." *International Review of Financial Analysis* 84: 102372.

Aouadi, A., and S. Marsat. 2018. "Do Esg Controversies Matter for Firm Value? Evidence From International Data." *Journal of Business Ethics* 151: 1027–1047. https://doi.org/10.1007/s10551-016-3213-8.

Balcerzak, A. P., T. Kliestik, D. Streimikiene, and L. Smrcka. 2017. "Non-Parametric Approach to Measuring the Efficiency of Banking Sectors in European Union Countries." *Acta Polytechnica Hungarica* 14: 51–70. https://doi.org/10.12700/APH.14.7.2017.7.4.

Barako, D. G., and A. M. Brown. 2008. "Corporate Social Reporting and Board Representation: Evidence From the Kenyan Banking Sector." *Journal of Management and Governance* 12: 309–324. https://doi.org/10.1007/s10997-008-9053-x.

Bătae, O. M., V. D. Dragomir, and L. Feleagă. 2020. "Environmental, Social, Governance (ESG), and Financial Performance of European Banks." *Accounting and Management Information Systems* 19: 480–501. https://doi.org/10.24818/jamis.2020.03003.

Belasri, S., M. Gomes, and G. Pijourlet. 2020. "Corporate Social Responsibility and Bank Efficiency." *Journal of Multinational Financial Management* 54: 100612. https://doi.org/10.1016/j.mulfin.2020.100612.

Birindelli, G., P. Ferretti, M. Intonti, and A. P. Iannuzzi. 2015. “On the Drivers of Corporate Social Responsibility in Banks: Evidence From an Ethical Rating Model.” *Journal of Management and Governance* 19: 303–340. https://doi.org/10.1007/s10997-013-9262-9.

Breiman, L. 2001. “Random Forests.” *Machine Learning* 45: 5–32.

Brownlee, J. 2020a. “How to Estimate a Baseline Performance for Your Machine Learning Models in Weka.” https://machinelearningmastery.com/estimate-baseline-performance-machine-learning-models-weka/.

Brownlee, J. 2020b. “Time Series Forecasting as Supervised Learning.” https://machinelearningmastery.com/time-series-forecasting-supervised-learning/.

Bruno, E., G. Iacoviello, and C. Giannetti. 2023. “Bank Credit Loss and Esg Performance.” *Finance Research Letters* 59: 104719. https://doi.org/10.1016/j.frl.2023.104719.

Castelli, M., L. Manzoni, and A. Popovič. 2016. “An Artificial Intelligence System to Predict Quality of Service in Banking Organizations.” *Computational Intelligence and Neuroscience* 2016: 1–7. https://doi.org/10.1155/2016/9139380.

Chan, F., and L. Mátyás. 2022. *Econometrics With Machine Learning*. Cham: Springer International Publishing.

Chao, X., Q. Ran, J. Chen, T. Li, Q. Qian, and D. Ergu. 2022. “Regulatory Technology (Reg-Tech) in Financial Stability Supervision: Taxonomy, Key Methods, Applications and Future Directions.” *International Review of Financial Analysis* 80: 102023. https://doi.org/10.1016/j.irfa.2022.102023.

Chen, J. M. 2021. “An Introduction to Machine Learning for Panel Data.” *International Advances in Economic Research* 27: 1–16. https://doi.org/10.1007/s11294-021-09815-6.

Chikersal, P., A. Doryab, M. J. Tumminia, et al. 2021. “Detecting Depression and Predicting Its Onset Using Longitudinal Symptoms Captured by Passive Sensing: A Machine Learning Approach With Robust Feature Selection.” *ACM Transactions on Computer-Human Interaction* 28: 1–41. https://doi.org/10.1145/3422821.

Collette, Y., and P. Siarry. 2004. “Chapter Introduction: Multiobjective Optimization and Domination.” In *Multiobjective Optimization: Principles and Case Studies*, 15–43. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-08883-8_1.

Cortes, C., and V. Vapnik. 1995. “Support-Vector Networks.” *Machine Learning* 20: 273–297.

DasGupta, R., and A. Roy. 2023. “Firm Environmental, Social, Governance and Financial Performance Relationship Contradictions: Insights From Institutional Environment Mediation.” *Technological Forecasting and Social Change* 189: 122341. https://doi.org/10.1016/j.techfore.2023.122341.

Dipierro, A. R., P. Toma, and M. Frittelli. 2024. “Introducing Esg Controversies as the Polluting Factor of Banks' Activity: A Nonparametric Efficiency Approach.” *Journal of Economic Studies*. https://doi.org/10.1108/JES-02-2024-0080.

Doerr, B. S., L. Gambacorta, and J. M. Serena. 2021. “How Do Central Banks Use Big Data and Machine Learning.” *Suerf Policy Briefs* 67: 1–6.

Dorfleitner, G., C. Kreuzer, and C. Sparrer. 2020. “Esg Controversies and Controversial Esg: About Silent Saints and Small Sinners.” *Journal of Asset Management* 21: 393–412. https://doi.org/10.1057/s41260-020-00178-x.

Doumpos, M., C. Zopounidis, D. Gounopoulos, E. Platanakis, and W. Zhang. 2022. “Operational Research and Artificial Intelligence Methods in Banking.” *European Journal of Operational Research* 306: 1–16. https://doi.org/10.1016/j.ejor.2022.04.027.

Durand, P., and G. Le Quang. 2022. “Banks to Basics! Why Banking Regulation Should Focus on Equity.” *European Journal of Operational Research* 301: 349–372. https://doi.org/10.1016/j.ejor.2021.10.057.

Eagly, A. H., M. C. Johannesen-Schmidt, and M. L. Van Engen. 2003. “Transformational, Transactional, and Laissez-Faire Leadership Styles: A Meta-Analysis Comparing Women and Men.” *Psychological Bulletin* 129: 569–591. https://doi.org/10.1037/0033-2909.129.4.569.

Ecer, F., and D. Pamucar. 2022. “A Novel Lopcow-Dobi Multi-Criteria Sustainability Performance Assessment Methodology: An Application in Developing Country Banking Sector.” *Omega* 112: 102690. https://doi.org/10.1016/j.omega.2022.102690.

Elsinger, H., A. Lehar, and M. Summer. 2006. “Risk Assessment for Banking Systems.” *Management Science* 52: 1301–1314. https://doi.org/10.1287/mnsc.1060.0531.

Fang, L. H. 2005. “Investment Bank Reputation and the Price and Quality of Underwriting Services.” *Journal of Finance* 60: 2729–2761. https://doi.org/10.1111/j.1540-6261.2005.00815.x.

Flanagan, D. J., and K. C. O'Shaughnessy. 2005. “The Effect of Layoffs on Firm Reputation.” *Journal of Management* 31: 445–463. https://doi.org/10.1177/0149206304272186.

Flood, C. 2019. “Esg Controversies Wipe $500 Bn off Value of Us Companies.” *Financial Times*. https://https://www.ft.com/content/3f1d44d9-094f-4700-989f-616e27c89599.

Fong, S., R. P. Biuk-Aghai, and R. C. Millham. 2018. “Swarm Search Methods in Weka for Data Mining.” In *Proceedings of the 10th International Conference on Machine Learning and Computing, ICMLC 2018, Macau, China*, ACM, 122–127. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3195106.31951.

Freedman, D. 2005. “Statistical Models: Theory and Practice.” *Statistical Models: Theory and Practice*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139165495.

Galletta, S., and S. Mazzù. 2023. “Esg Controversies and Bank Risk Taking.” *Business Strategy and the Environment* 32: 274–288. https://doi.org/10.1002/bse.3129.

Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st ed. USA: Addison-Wesley Longman Publishing Co., Inc.

Guo, Y., J. He, L. Xu, and W. Liu. 2019. “A Novel Multi-Objective Particle Swarm Optimization for Comprehensible Credit Scoring.” *Soft Computing* 23: 9009–9023. https://doi.org/10.1007/s00500-018-3509-y.

Hall, M. A. 2000. “Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning.” In *Proceedings of the Seventeenth International Conference on Machine Learning*, 359–366. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Hanauer, M. X., M. Kononova, and M. S. Rapp. 2022. “Boosting Agnostic Fundamental Analysis: Using Machine Learning to Identify Mispricing in European Stock Markets.” *Finance Research Letters* 48: 102856. https://doi.org/10.1016/j.frl.2022.102856.

Hiebl, M., R. Baule, A. Dutzi, V. Stein, and A. Wiedemann. 2018. “Roles and Actors in Risk Governance.” *Journal of Risk Finance*. https://doi.org/10.1108/JRF-10-2018-0150.

Hirsch, P. B. 2018. “Tie Me to the Mast: Artificial Intelligence & Reputation Risk Management.” *Journal of Business Strategy* 39: 61–64. https://doi.org/10.1108/JBS-11-2017-0160.

Jiménez, F., G. Sánchez, J. M. García, G. Sciavicco, and L. Miralles. 2017. “Multi-Objective Evolutionary Feature Selection for Online Sales Forecasting.” *Neurocomputing* 234: 75–92. https://doi.org/10.1016/j.neucom.2016.12.045.

Jizi, M. I., A. Salama, R. Dixon, and R. Stratling. 2014. “Corporate Governance and Corporate Social Responsibility Disclosure: Evidence From the US Banking Sector.” *Journal of Business Ethics* 125: 601–615. https://doi.org/10.1007/s10551-013-1929-2.

Khrestina, M. P., D. I. Dorofeev, P. A. Kachurina, T. R. Usubaliev, and A. S. Dobrotvorskiy. 2017. “Development of Algorithms for Searching, Analyzing and Detecting Fraudulent Activities in the Financial Sphere." *European Research Studies Journal* 20, no. 4B: 484–498.

Kinywamaghana, A., and S. Steffen. 2021. "A Note on the Use of Machine Learning in Central Banking." FIRE Research Paper, July 13.

Kohavi, R., and G. H. John. 1997. "Wrappers for Feature Subset Selection." *Artificial Intelligence* 97: 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X.

Kumar, P. R., and V. Ravi. 2007. "Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques—A Review." *European Journal of Operational Research* 180: 1–28. https://doi.org/10.1016/j.ejor.2006.08.043.

Lappas, P. Z., and A. N. Yannacopoulos. 2021. "A Machine Learning Approach Combining Expert Knowledge With Genetic Algorithms in Feature Selection for Credit Risk Assessment." *Applied Soft Computing* 107: 107391. https://doi.org/10.1016/j.asoc.2021.107391.

Le, H. H., and J. L. Viviani. 2018. "Predicting Bank Failure: An Improvement by Implementing a Machine-Learning Approach to Classical Financial Ratios." *Research in International Business and Finance* 44: 16–25. https://doi.org/10.1016/j.ribaf.2017.07.104.

Leo, M., S. Sharma, and K. Maddulety. 2019. "Machine Learning in Banking Risk Management: A Literature Review." *Risks* 7: 29. https://doi.org/10.3390/risks7010029.

Li, J., Z. A. Haider, X. Jin, and W. Yuan. 2019. "Corporate Controversy, Social Responsibility and Market Performance: International Evidence." *Journal of International Financial Markets Institutions and Money* 60: 1–18. https://doi.org/10.1016/j.intfin.2018.11.013.

Liu, H., and H. Motoda. 1998. *Feature Selection for Knowledge Discovery and Data Mining*. Vol. 454. New York, NY: Springer Science & Business Media.

Metawa, N., M. K. Hassan, and M. Elhoseny. 2017. "Genetic Algorithm Based Model for Optimizing Bank Lending Decisions." *Expert Systems With Applications* 80: 75–82. https://doi.org/10.1016/j.eswa.2017.03.021.

Miralles-Quirós, M. M., J. L. Miralles-Quirós, and J. Redondo Hernández. 2019. "Esg Performance and Shareholder Value Creation in the Banking Industry: International Differences." *Sustainability* 11: 1404. https://doi.org/10.3390/su11051404.

Mohapatra, S., R. Mukherjee, A. Roy, A. Sengupta, and A. Puniyani. 2022. "Can Ensemble Machine Learning Methods Predict Stock Returns for Indian Banks Using Technical Indicators?" *Journal of Risk and Financial Management* 15: 350. https://doi.org/10.3390/jrfm15080350.

Moraglio, A., C. Di Chio, and R. Poli. 2007. "Geometric Particle Swarm Optimisation." In *Genetic Programming*, edited by M. Ebner, M. O'Neill, A. Ekárt, L. Vanneschi, and A. I. Esparcia-Alcázar, 125–136. Berlin, Heidelberg: Springer Berlin Heidelberg.

Nirino, N., G. Santoro, N. Miglietta, and R. Quaglia. 2021. "Corporate Controversies and Company's Financial Performance: Exploring the Moderating Role of Esg Practices." *Technological Forecasting and Social Change* 162: 120341. https://doi.org/10.1016/j.techfore.2020.120341.

Pai, P. F., M. F. Hsu, and M. C. Wang. 2011. "A Support Vector Machine-Based Model for Detecting Top Management Fraud." *Knowledge-Based Systems* 24: 314–321. https://doi.org/10.1016/j.knosys.2010.10.003.

Papouskova, M., and P. Hajek. 2019. "Two-Stage Consumer Credit Risk Modelling Using Heterogeneous Ensemble Learning." *Decision Support Systems* 118: 33–45. https://doi.org/10.1016/j.dss.2019.01.002.

Pérez, A., and I. R. del Bosque. 2015. "How Customer Support for Corporate Social Responsibility Influences the Image of Companies: Evidence From the Banking Industry." *Corporate Social Responsibility and Environmental Management* 22: 155–168. https://doi.org/10.1002/csr.1331.

Peters, G., P. V. Shevchenko, R. Cohen, and D. Maurice. 2017. "Statistical Machine Learning Analysis of Cyber Risk Data: Event Case Studies." *SSRN Electronic Journal*: 3073704. https://doi.org/10.2139/ssrn.3073704.

Petropoulos, A., V. Siakoulis, E. Stavroulakis, and N. E. Vlachogiannakis. 2020. "Predicting Bank Insolvencies Using Machine Learning Techniques." *International Journal of Forecasting* 36: 1092–1113. https://doi.org/10.1016/j.ijforecast.2019.11.005.

Pilkington, M. 2022. "The London Whale Scandal Under New Scrutiny: A Critical Appraisal of the Concept of Retro-Causality." *International Review of Financial Analysis* 80: 102050. https://doi.org/10.1016/j.irfa.2022.102050.

Reddy, N. S. 2022. "Optimal Feature Selection and Hybrid Deep Learning for Direct Marketing Campaigns in Banking Applications." *Evolutionary Intelligence* 15: 1969–1990. https://doi.org/10.1007/s12065-021-00604-y.

Ribeiro, M. T., S. Singh, and C. Guestrin. 2016. "Model-Agnostic Interpretability of Machine Learning." *arXiv*: 1606.05386. https://doi.org/10.48550/arXiv.1606.05386.

Rodríguez-Fernández, M., E. M. Sánchez-Teba, A. A. López-Toro, and S. Borrego-Domínguez. 2019. "Influence of Esgc Indicators on Financial Performance of Listed Travel and Leisure Companies." *Sustainability* 11: 5529. https://doi.org/10.3390/su11195529.

Sassen, R., M. Stoffel, M. Behrmann, W. Ceschinski, and H. Doan. 2018. "Effects of Committee Overlap on the Monitoring Effectiveness of Boards of Directors: A Meta-Analysis." *Journal of Risk Finance* 19: 379–395. https://doi.org/10.1108/JRF-11-2017-0187.

Schiemann, F., and R. Tietmeyer. 2022. "Esg Controversies, Esg Disclosure and Analyst Forecast Accuracy." *International Review of Financial Analysis* 84: 102373. https://doi.org/10.1016/j.irfa.2022.102373.

Shakil, M. H. 2021. "Environmental, Social and Governance Performance and Financial Risk: Moderating Role of Esg Controversies and Board Gender Diversity." *Resources Policy* 72: 102144. https://doi.org/10.1016/j.resourpol.2021.102144.

Shakil, M. H. 2024. "Environmental Social and Governance Controversies: A Bibliometric Review and Research Agenda." *Finance Research Letters* 70: 106325. https://doi.org/10.1016/j.frl.2024.106325.

Shakil, M. H., M. Tasnia, and M. I. Mostafiz. 2021. "Board Gender Diversity and Environmental, Social and Governance Performance of Us Banks: Moderating Role of Environmental, Social and Corporate Governance Controversies." *International Journal of Bank Marketing* 39: 661–677. https://doi.org/10.1108/IJBM-04-2020-0210.

Sharma, S., and A. R. Choudhury. 2016. "Fraud Analytics: A Survey on Bank Fraud and Fraud Prediction Using Unsupervised Learning Based Approach." *International Journal of Innovations in Engineering Research and Technology* 3: 1–9.

Shie, F. S., M. Y. Chen, and Y. S. Liu. 2012. "Prediction of Corporate Financial Distress: An Application of the America Banking Industry." *Neural Computing and Applications* 21: 1687–1696. https://doi.org/10.1007/s00521-011-0765-5.

Shrivastav, S. K. 2019. "Measuring the Determinants for the Survival of Indian Banks Using Machine Learning Approach." *FIIB Business Review* 8: 32–38. https://doi.org/10.1177/2319714519825939.

Soana, M. G. 2011. "The Relationship Between Corporate Social Performance and Corporate Financial Performance in the Banking Sector." *Journal of Business Ethics* 104: 133–148. https://doi.org/10.1007/s10551-011-0894-x.

Speiser, J. L. 2021. "A Random Forest Method With Feature Selection for Developing Medical Prediction Models With Clustered and Longitudinal Data." *Journal of Biomedical Informatics* 117: 103763. https://doi.org/10.1016/j.jbi.2021.103763.

Tanaka, K., T. Kinkyo, and S. Hamori. 2016. "Random Forests-Based Early Warning System for Bank Failures." *Economics Letters* 148: 118–121. https://doi.org/10.1016/j.econlet.2016.09.024.

Tian, S., and C. Wang. 2019. "Feature Selection for Longitudinal Data by Using Sign Averages to Summarize Gene Expression Values Over

Time." *BioMed Research International* 2019: 1–12. https://doi.org/10.1155/2019/1724898.

Tissaoui, K. 2019. "Forecasting Implied Volatility Risk Indexes: International Evidence Using Hammerstein-Arx Approach." *International Review of Financial Analysis* 64: 232–249. https://doi.org/10.1016/j.irfa.2019.06.001.

Viswanathan, P., S. Srinivasan, and N. Hariharan. 2020. "Predicting Financial Health of Banks for Investor Guidance Using Machine Learning Algorithms." *Journal of Emerging Market Finance* 19: 226–261. https://doi.org/10.1177/0972652720913478.

Wheelock, D. C., and P. W. Wilson. 2000. "Why Do Banks Disappear? The Determinants of Us Bank Failures and Acquisitions." *Review of Economics and Statistics* 82: 127–138. https://doi.org/10.1162/003465300558560.

Wu, M. W., and C. H. Shen. 2013. "Corporate Social Responsibility in the Banking Industry: Motives and Financial Performance." *Journal of Banking & Finance* 37: 3529–3547. https://doi.org/10.1016/j.jbankfin.2013.04.023.

Zhang, J. Q., H. Zhu, and H. B. Ding. 2013. "Board Composition and Corporate Social Responsibility: An Empirical Investigation in the Post Sarbanes-Oxley Era." *Journal of Business Ethics* 114: 381–392. https://doi.org/10.1007/s10551-012-1352-0.

Appendix A

FS Methods

In Section 3, we consider four Boolean combinatorial optimization problems for FS, of which two are of type wrapper and two filter. For each type of wrapper and filter problem, one of them is a *single-objective* and another is a *multiobjective* optimization problem, expressed mathematically as follows:

- Single-objective Boolean problem for wrapper FS.

$$\text{Minimize} \quad \mathcal{F}_1(\mathbf{x}) = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{n_R^k}\sum_{t=1}^{n_R^k}\left(y_t^k - \hat{y}_t^k(\mathbf{x})\right)^2} \tag{A1}$$

where $\mathbf{x} = (x_1, \ldots, x_p)$ is a vector of decision variables with $x_i \in \{\text{false}, \ \text{true}\}, i = 1, \ \ldots, p$. A decision variable $x_i = \text{true}$ represents that the attribute i has been selected, while $x_i = \text{false}$ represents that the attribute i has not been selected. In Problem (A1), the function $\mathcal{F}_1(\mathbf{x})$ is the root mean squared error (RMSE) of a prediction model trained with a learning algorithm and the dataset D_R on K-fold cross-validation with the selected attributes in $\mathbf{x}$, y_t^k is the observed output value for the sample t of the fold k, $\hat{y}_t^k(\mathbf{x})$ is the model-predicted value, trained on the set of folds different from fold k, for the sample t of k with the selected attributes in $\mathbf{x}$, and n_R^k is the number of samples of the fold k in D_R. We use $K = 5$ in the optimization phase.

- Multiobjective Boolean problem for wrapper FS.

$$\begin{aligned} &\text{Minimize} \quad \mathcal{F}_1(\mathbf{x}) \\ &\text{Minimize} \quad \mathcal{F}_2(\mathbf{x}) = \sum_{i=1}^{p}\mathcal{N}(x_i) \end{aligned} \tag{A2}$$

where $\mathcal{F}_2(\mathbf{x})$ is the cardinality of the attribute subset selected in $\mathbf{x}$. $\mathcal{N}(x_i)$ is a function that transforms the Boolean value x_i into numeric ($true = 1$ and $false = 0$). The solution to multiobjective optimization problems is drawn from a set of *non-dominated* solutions that constitute the *Pareto front* (Collette and Siarry 2004), in "a posteriori" decision-making process.

- Single-objective Boolean problem for filter FS.

$$\text{Maximize} \quad \mathcal{F}_3(\mathbf{x}) = \frac{s\cdot\sigma_c(\mathbf{x})}{\sqrt{s + s\cdot(s-1)\cdot\sigma_f(\mathbf{x})}} \tag{A3}$$

where s is the number of attributes selected in $\mathbf{x}$ (i.e., $s = \mathcal{F}_2(\mathbf{x})$), $\sigma_c(\mathbf{x})$ is the mean of the correlations between each attribute selected in ($\mathbf{x}$) and the class attribute, and $\sigma_f(\mathbf{x})$ is the average correlation between each of the $\binom{s}{2}$ possible attribute pairs selected in $\mathbf{x}$. In other words, the numerator indicates the predictive degree of a set of attributes while the denominator indicates the redundancy between the attributes. This FS method is known in the literature as *correlation-based feature selection* (CFS) (Hall 2000).

- Multiobjective Boolean problem for wrapper FS.

The CFS method can also be solved as a multiobjective problem as follows:

$$\begin{aligned} &\text{Maximize} \quad \mathcal{F}_3(\mathbf{x}) \\ &\text{Minimize} \quad \mathcal{F}_2(\mathbf{x}) \end{aligned} \tag{A4}$$

Appendix B

Abbreviations

TABLE B1 | Used abbreviations.

Abbreviation	Meaning
ESG	Environmental, social, governance
ESGC	Controversies related to environmental, social, governance issues
ML	Machine learning
AI	Artificial intelligence
FS	Feature selection
ANN	Artificial neural network
SVM	Support vector machine
RF	Random forest
LR	Linear regression
NB	Naive-Bayes
BF	Best-first
GA	Genetic algorithm
PSO	Particle swarm optimization
AC	Ant colony
MOEA	Multiobjective evolutionary algorithm
WS	Window size
CFS	Correlation-based feature selection
RMSE	Root mean squared error
MAE	Mean absolute error
CC	Correlation coefficient
CAMELS	Capital, asset quality, management, earnings, liquidity, sensitivity to market risk