

What drives corporate greenwashing? A machine learning lens on pressure and opportunity in China's A-share listed firms

China Accounting
and Finance
Review

393

Yao Xiao

Department of Space Support, Space Engineering University, Beijing, China

Miao Yu

*International Business College, Dongbei University of Finance and Economics,
Dalian, China*

Mingji Liu

*International Business School, Beijing Foreign Studies University,
Beijing, China, and*

Kaijie Jiang

*Surrey International Institute, Dongbei University of Finance and Economics,
Dalian, China*

Received 9 May 2025

Revised 18 January 2026

Accepted 22 March 2026

Abstract

Purpose – This study aims to systematically identify the drivers of corporate greenwashing through the institution, market, organization and cognition (IMOC) framework. Grounded in institutional theory, resource-based view, agency theory and upper echelons theory, this research examines how factors across these four dimensions drive greenwashing through the dual mechanisms of institutional pressures and organizational opportunities, while establishing an effective prediction framework using machine learning techniques.

Design/methodology/approach – Employing a comparative ML approach, we comprehensively identify the factors affecting corporate greenwashing based on the pressure and opportunity dimensions and then explore how these factors influence corporate greenwashing by a sample of A-share listed firms in China from 2011 to 2022. The LightGBM, XGBoost and random forest algorithms are benchmarked against traditional logistic regression, with SHapley Additive exPlanations (SHAP) values used for feature interpretation.

Findings – The paper finds that (1) LightGBM outperforms other models in greenwashing prediction. (2) Institutional ownership concentration emerges as the strongest positive predictor, while moderate digital transformation and market competition exhibit inhibitory effects. (3) Digital economy development amplifies media monitoring efficacy and reduces principal-agent conflicts' impact on greenwashing.

Practical implications – Regulators should prioritize institutional investors' governance roles and establish digital transformation thresholds for environmental, social and governance (ESG) compliance. Investors can utilize our ML framework to assess greenwashing risks in portfolio companies.

© Yao Xiao, Miao Yu, Mingji Liu and Kaijie Jiang. Published in *China Accounting and Finance Review*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/>.

Funding: This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (No: 71802042) and the Fundamental Research Funds for the Central Universities (No: 2024JX032).

Declaration of interest statement: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Mingji Liu is a co-first author of this manuscript.



China Accounting and Finance Review

Vol. 28 No. 4, 2026

pp. 393-419

Emerald Publishing Limited

e-ISSN: 2307-3055

p-ISSN: 1029-807X

DOI 10.1108/CAFR-05-2025-0060

Originality/value – This research pioneers the integration of institutional theory with explainable artificial intelligence in greenwashing detection, revealing the non-linear impacts of digital transformation. The proposed SHAP-empowered framework enables dynamic monitoring of emerging ESG risks.

Keywords Corporate greenwashing, Opportunities, Pressures, Machine learning, SHAP interpretation

Paper type Research article

1. Introduction

Growing ecological challenges and frequent extreme weather events have made sustainable development a global priority. In September 2020, China formally announced its “dual carbon” goals (carbon peaking and neutrality), later reinforced by the 20th National Congress of the Communist Party of China, which emphasized green economic transformation and harmonious coexistence between humans and nature (Niu & Wang, 2024; Fang, Nolan, & Linggui, 2019). However, green transformation faces significant barriers, including high capital demands, long payback periods and investment risks. Under regulatory pressure, firms may engage in *greenwashing* – misleading environmental claims that appear more sustainable than they are (Du, 2015). Such practices harm consumer rights, market competition and policy effectiveness (Zeng, 2024), posing a significant barrier to sustainable development.

Academic research considers greenwashing as inherently fraudulent, with scholars predominantly applying fraud theory frameworks (triangular/four-factor models) to investigate its determinants. Empirical evidence identifies multi-level drivers encompassing organizational characteristics (firm lifecycle and leadership style) and external pressures (regulatory gaps, market competition and consumer expectations) (Zhang, Yang, & Wang, 2024; Zhang *et al.*, 2024). While prior research examines greenwashing drivers (e.g. regulatory pressures and firm traits; Zhang *et al.*, 2024), gaps remain in *predictive* modeling, especially amid evolving environmental, social and governance (ESG) complexities (e.g. supply chain or strategic greenwashing).

Contemporary corporate green governance confronts evolving contextual challenges, necessitating systematic identification of critical determinants of environmental conduct and their predictive mechanisms to advance market-oriented ecological regulation. In multifaceted decision-making landscapes, stakeholders increasingly prioritize developing predictive models for greenwashing risks – a strategic imperative enabling cost-efficient preventive measures and market-based internalization of sustainability risks.

This study addresses these gaps by employing machine learning (ML) to predict greenwashing behavior among Chinese A-share firms (2011–2022). Unlike traditional methods prioritizing unbiasedness, ML optimizes predictive accuracy and generalizability (Tiffin, 2019). To systematically disentangle these complexities, we propose the institution, market, organization and cognition (IMOC) framework, integrating institutional theory, the resource-based view, agency theory and upper echelons theory. This framework categorizes drivers into four hierarchical levels: institution, market, organization and cognition. Within this framework, we identify 27 key factors and employ ML to reveal how they drive greenwashing through the dual mechanisms of institutional pressures and organizational opportunities.

While emerging studies have begun to apply ML to predict greenwashing (e.g. Zeng, Wang, & Zeng, 2025; Li, He, & Li, 2025), our research differentiates itself in three pivotal aspects. First, we develop and empirically validate the Institution, Market, Organization, and Cognition (IMOC) framework, which synthesizes institutional, market, organizational and cognitive drivers into a coherent theoretical model. This move beyond identifying isolated predictors to offering a holistic understanding of the greenwashing ecosystem. Second, we incorporate a temporal dimension by examining how the predictive importance of key factors (e.g. media attention and agency costs) shifted after China’s “Digital China” initiative (2015), providing dynamic insights into the evolving governance landscape. Third, our analytical focus extends beyond predictive accuracy to interpretability. By leveraging Shapley Additive

exPlanations (SHAP), we uncover non-linear relationships and threshold effects (e.g. the conditional impact of digital transformation), thereby “opening the black box” of ML models to yield actionable, theory-informed insights.

Our contributions are threefold.

- (1) *Theoretical*: A unified framework of greenwashing drivers, expanding corporate social responsibility literature;
- (2) *Methodological*: Demonstrated superiority of ML in greenwashing prediction and
- (3) *Practical*: Insights for firms, investors and policymakers to mitigate risks and refine governance.

The paper proceeds as follows: [Section 2](#) reviews greenwashing and its drivers; [Sections 3–5](#) detail methods, results and analysis; [Section 6](#) concludes with implications.

2. Literature review

Greenwashing constitutes strategic information distortion across disciplinary contexts. (1) As symbolic impression management in business ethics, involving decoupled symbolic-substantive actions to accumulate moral capital ([Effron, O'Connor, Leroy, & Lucas, 2018](#)); (2) As selective disclosure in economic analysis, characterized by asymmetric environmental reporting ([Pizzetti, Gatti, & Seele, 2021](#)); (3) As hyperbolic marketing in management studies, manifesting through unsubstantiated eco-claims ([Parguel, Benoît-Moreau, & Larceneux, 2011](#)). This behavioral divergence stems from perceived green performance gaps, wherein firms employ environmental speculation tactics – from information manipulation to outright deception – when actual sustainability outcomes lag expectations ([Wang, Zeng, & Li, 2022](#)). Synthesizing fraud theory frameworks, we posit greenwashing as institutionally enabled fraud driven by pressure-opportunity dynamics, where organizational pressures catalyze deception and structural opportunities facilitate its execution.

2.1 Pressures for corporate greenwashing

Environmental performance is a key element for modern corporations to manage their public impression ([Yue & Li, 2023](#)). Greenwashing can improve a firm's image, meet regulatory requirements and reduce financial costs by attracting government subsidies. Such “advantages” may encourage more conduction of greenwashing and decrease very green and sustainable behaviors ([Zhang, 2022](#)). In addition, pressure has become a behavioral motivation for firms' implementation of corporate greenwashing and drives the corporates' direct interest.

Pressure can be categorized as exogenous and endogenous. First, from the outside, market pressure may lead to greenwashing ([Kim & Lyon, 2015](#); [Delmas, Nairn-Birch, & Lim, 2015](#)). As competition intensifies, companies increasingly strive to “demonstrate their commitment to the environment and comply with external environmental regulations, such as public scrutiny, by cultivating a positive environmental corporate image” ([Li, Li, Seppänen, & Koivumäki, 2023](#)). Nevertheless, [Lyon and Montgomery \(2013\)](#) find that the external pressure generated by companies' public participation and social media involvement can curb their greenwashing behavior. Second, financial precarity and agency conflicts constitute primary internal catalysts for greenwashing. Resource-constrained firms, particularly in environmentally sensitive industries, frequently adopt low-cost sustainability signaling to mitigate financing pressures and regulatory compliance costs ([Zhang, 2022, 2023](#); [Hu, Yu, & Han, 2023b](#)). This fiscal maneuvering manifests through two channels: (1) principal-agent divergence where shareholders prioritize short-term gains via symbolic environmental commitments ([Ferrón-Vílchez et al., 2021](#)) and (2) executive myopia incentivizing superficial compliance over substantive green investments ([Kim & Lyon, 2015](#); [Li, Wang, & He, 2023](#)).

2.2 Opportunities for corporate greenwashing

Opportunities are the internal and external conditions under which enterprises can transform their greenwashing motives into actual actions. First, from an external perspective, weak enforcement regimes create permissive environments for symbolic compliance (Bernini, Giuliani, & La Rosa, 2023). Zhang (2022) demonstrates that financial constraints arising from the financial environment drive companies to engage in greenwashing, making these constraints key determinants of such behavior. However, it remains uncertain whether fintech has a positive or negative effect on corporate ethical behavior (Li, Cao, & Wang, 2024a; Li, Miao, & Xu, 2024b; Xie, Chen, Liu, & Wang, 2023). Zhou, Jin, Li, & Tao (2024) show that if companies are politically associated, their tendency to greenwash is even more severe. Therefore, the higher the greenwashing in cohort firms, the higher the greenwashing in target firms (Chen & Dagestani, 2023).

Second, *greenwashing* behavior is closely relevant to a firm's size. Compared with smaller ones, large firms are more likely to overcome regulatory controls (Bernini *et al.*, 2023). They are more inclined to greenwash to meet stakeholder demands (Testa, Boiral, & Iraldo, 2018). In pursuit of profits, companies may select the positive information regarding their green development to report instead of negative information (Hu, Wang, & Du, 2023c). Management characteristics can also lead to greenwashing. Blome, Foerstl and Schleper (2017) found that authoritative leadership styles influence greenwashing, which is inhibited by moral incentives. Moreover, executives' ages, risk preferences and experience can affect greenwashing (Xia, Chen, Yang, Li, & Zhang, 2023; Zhang, Qin, & Zhang, 2023). Additionally, customers become more conscious about climate change and environmental protection (Lin & Niu, 2018; Kumar, Manrai, & Manrai, 2017). Consequently, corporations may resort to wash their environmental performance to attract this kind of customers and enhance their social good image (Szabo & Webster, 2021; Ioannou, Kassinis, & Papagiannakis, 2023).

Recent methodological advances have seen the application of ML to greenwashing prediction. For instance, Zeng *et al.* (2025) employed optimized ML algorithms to identify key predictors of ESG greenwashing, while Li *et al.* (2025) developed a predictive model focusing on heavy-pollution industries (HPIs). These studies underscore the value of ML in handling complex, non-linear relationships. However, they primarily focus on predictive performance and feature ranking, leaving room for a more theoretically integrated framework that explains why and under what conditions these factors matter across different levels of analysis. Furthermore, the temporal evolution of these drivers in the context of rapid digitalization remains underexplored.

2.3 Integrated theoretical framework: a multi-level perspective on the driving mechanisms of greenwashing

Through a systematic review of existing literature, this study finds that corporate greenwashing is influenced by a complex array of factors across multiple levels. In order to establish a clear analytical structure for this research, we have developed an integrated theoretical framework that categorizes the driving factors into four interrelated theoretical dimensions: the institutional level, market level, internal organizational level and individual cognitive level. This framework moves beyond the traditional "pressure-opportunity" dichotomy by offering a more nuanced classification system to clarify the theoretical rationale for including each variable.

- (1) Institutional level. Grounded in institutional theory, this dimension emphasizes the role of the external institutional environment in shaping corporate behavior. It includes regulatory pressure, normative pressure and cognitive mimicry. Specifically,

regulatory pressure stems from the intensity of government environmental regulations and policies, leading firms to engage in symbolic compliance to gain legitimacy. Normative pressure arises from societal expectations, such as those from the media and the public, prompting companies to manage environmental impressions to protect their reputation. Cognitive mimicry refers to the tendency of firms, under conditions of uncertainty, to imitate the behaviors (including greenwashing) of other firms in the same industry, resulting in isomorphic practices within the sector.

- (2) Market level. Derived from industrial organization theory and resource dependence theory, this dimension focuses on the competitive and resource environments in which firms operate. It encompasses competitive pressure and financing pressure. Competitive pressure typically arises from intense market competition and performance expectation gaps, driving firms to use green claims for differentiation or to alleviate performance pressures. Financing pressure refers to the motivation for firms to cultivate a positive green image to attract green investments or improve financing conditions.
- (3) Internal organizational level. Primarily based on agency theory, this dimension examines internal governance, resources and managerial characteristics. The framework further refines this level by highlighting governance deficiencies: weak corporate governance mechanisms create opportunities for managerial opportunism. In terms of resources and capabilities, if a firm lacks the resources or technical capacity to implement substantive green innovations, it is more likely to opt for low-cost symbolic strategies. Regarding managerial cognition and motivation, executives' personal backgrounds, characteristics and incentive structures significantly influence their preferences and judgments in environmental decision-making.
- (4) Individual cognitive level. Drawing on behavioral ethics, this dimension delves into the psychological motivations of individual decision-makers. Although variables at this level are often difficult to observe directly, their underlying mechanisms are critical. For instance, myopic management, characterized by an excessive focus on short-term goals, may lead to the sacrifice of long-term environmental value. Similarly, moral disengagement allows decision-makers to rationalize unethical behaviors, such as greenwashing, through cognitive restructuring mechanisms like shifting responsibility or minimizing consequences.

In summary, this framework not only provides a clear theoretical basis and justification for the predictor variables selected in this study, ensuring systematic and comprehensive variable selection, but also establishes a solid theoretical foundation for interpreting the predictive results of ML in subsequent sections. We posit that greenwashing is not driven by a single factor but rather results from the interplay and accumulation of factors across these four levels.

3. Sample and variables

3.1 Sampling and data collection

Based on the available time period of the ESG rating data, we take all the listed companies from 2011 to 2022 in the A-share market as the initial sample. Referring to conventional practice in the existing literature, we exclude some special samples as follows: (1) companies with abnormal status, including ST, *ST and PT (particular transfer firms), (2) companies in the financial field and (3) observations with missing research variables. After that, we get a final sample containing 5,326 observations. We winsorize all continuous variables at 1% and 99%. We collect the ESG performance rating data from the “*Morgan Stanley Capital International ESG ratings*” in the Wind database, the ESG disclosure ratings data from the Bloomberg database and all other data from the China Stock Market & Accounting Research (CSMAR) database. [Table 1](#) presents the annual and industrial distributions of the sample.

Table 1. Distribution of the sample by a year and industry

Year	Sample size	Sample proportion (%)	Industry	Sample size	Sample proportion (%)
2012	378	7.097	A	59	1.108
2013	404	7.585	B	293	5.501
2014	399	7.492	C	3,670	68.907
2015	479	8.994	D	207	3.887
2016	479	8.994	E	162	3.042
2017	512	9.613	F	143	2.685
2018	542	10.176	G	143	2.685
2019	637	11.960	H	12	0.225
2020	634	11.904	I	309	5.802
2021	743	13.950	K	100	1.878
2022	119	2.234	L	50	0.939
			M	28	0.526
			N	50	0.939
			Q	33	0.620
			R	52	0.976
			S	15	0.282
Total	5,326	100		5,326	100

The annual distribution shows an increasing trend year by year, especially after 2017, and the annual sample that can be used to calculate greenwashing is more than 500. This reflects that, on the one hand, China has attached great importance to ecological issues in recent years, and the capital market further echoes the relevant policy orientation. On the other hand, in January 2017, Hong Kong Exchanges and Clearing Limited completed the revision of its ESG Guidelines, which made it mandatory for both A-share and H-share listed companies to disclose their ESG-related reports (Liao, Sun, & Xu, 2023). This act of disclosure also serves as a demonstration for A-share listed companies only.

Regarding industry distribution, C (manufacturing) has the largest sample size, accounting for nearly 70% of the total, which rightly reflects the fact that China has many listed companies in the manufacturing industry. Four industries – that is, H (accommodation and food service), M (scientific research and technical services), Q (health and social work industry) and S (general) – have a smaller sample size, accounting for a combined share of about 1.65%. This is because the number of listed companies in these four industries is already relatively low (Chen, Du, Wu, & Zhou, 2017), and the main businesses of these industries have a relatively small negative impact on the environment and society (Baudry, Bukowski, & Lament, 2024); thus, they may pay less attention to ESG practices.

3.2 Variables and measurement

3.2.1 Dependent variable. We use a calculation formula to measure the extent of corporate greenwashing as follows:

$$GW_{i,t} = \left(\frac{ER_{dis i,t} - \overline{ER}_{dis i,t}}{\sigma_{dis}} \right) - \left(\frac{ER_{per i,t} - \overline{ER}_{per i,t}}{\sigma_{per}} \right),$$

where $ER_{dis i,t}$ denotes the environmental disclosure score of firm i in year t and is measured using the ESG rating score from the Bloomberg database; $\overline{ER}_{dis i,t}$ denotes the mean of the environmental disclosure scores of the same industry in the same year; σ_{dis} is the standard deviation of the firms' environmental disclosure scores in the current year; $ER_{per i,t}$ denotes the

actual score of the environmental performance; $\overline{ER}_{per,i,t}$ is the mean of the actual environmental performance scores for the same industry in the same year and σ_{per} is the standard deviation of the actual scores of the firms. When the GW value is positive, firms tend to conceal their poor environmental performance by releasing misleading environmental information.

3.2.2 Independent variables. In line with ML practices, where all input variables are treated as features for prediction, we select a comprehensive set of potential predictors for corporate greenwashing based on established literature. These features can be broadly categorized based on the pressure-opportunity framework and other key firm characteristics identified in prior studies. However, it is important to note that all features are treated collectively and their influence is assessed simultaneously by the models. The selected features are defined as follows.

- (1) *Industry competition:* Herfindahl–Hirschman Index (*HHI*) is used to measure industry competition according to the usual practice. The lower the value of this index, the higher the extent of competition in the industry where the firm is involved.
- (2) *External attention:* We use two indicators to measure the intensity of external attention towards the environment: governmental attention (*GOV*) and media attention (*MEDIA*). Specifically, *GOV* is measured by the magnitude of the low-carbon policy intensity in prefecture-level cities (Dong, Wang, Zhang, Zhang, & Xia, 2024), while *MEDIA* is measured by the total number of news relative to the company in the year (Hu & Liu, 2023; Yue & Li, 2023).
- (3) *Environmental uncertainty:* We use the environmental uncertainty (*EU*) index (Kim, Fairclough, & Dibrell, 2017) to measure *EU*. The larger the value, the higher the *EU* extent.
- (4) *Corporate finance:* A company's financial status and capabilities usually affect its strategic decision-making. We use *leverage* and return on assets *return on assets* as the representatives to measure the company's financial need and profitability (Xia et al., 2023).
- (5) *Company type:* Heavily polluting firms have more incentives to greenwash (Zhang, Qin, & Zhang, 2023). We define the *HPIs* and code them based on the "Guidelines on Industry Classification of Listed Companies revised by the China Securities Regulatory Commission in 2012" (Guo, 2022).
- (6) *Agency problem.* The two types of agency costs are introduced to measure the firm's agency problem. The first type (*AC1*) is calculated by the ratio of expenses on sales and administration to operational revenues, while the second type (*AC2*) is calculated by the ratio of other receivables to total assets (Zhao & Wang, 2024; Zhang, Wang, & Jiang, 2023). The higher the agency cost, the more serious the agency problem.
- (7) *Green investor:* Green investor (*GI*) plays an active role in the ESG-driving market. We use "the Napierian Logarithm of the number of *GIs* plus one" to measure this variable (Feng and Yuan, 2024).
- (8) *Bankruptcy risk:* Bankruptcy risk is a crucial force to challenge a firm's ethic. We use the Z-score index (*ZSCORE*) [1] to measure the corporate bankruptcy risk (Zhang, 2023).

Second, the following variables are used for opportunities.

- (1) *Financial technology:* Referring to Wen, Wang and He (2024), we use the provincial aggregate index of "the Peking University Digital Financial Inclusion Index" [2] to measure the level of digital financial technology development (*DFIIC*).

- (2) *Digital transformation*: We use “the Digital Transformation Index” from the CSMAR Database to measure a firm’s digital transformation level (*DT*). The comprehensive index integrates the strategy-driven, technology-enabled, organization-enabled, environment-enabled, digital achievement and digital application scores.
- (3) *Analyst following*: According to [Zhang \(2022\)](#), we use “the Napierian logarithm of the number of analysts following plus one” to measure analyst following (*ANA*). The higher the *ANA*, the lower is the level of information asymmetry.
- (4) *Economic policy uncertainty*: According to [Baker, Bloom and Davis \(2016\)](#), “the Economic Uncertainty index” [3] is chosen to measure economic policy uncertainty.
- (5) *Corporate governance*: We use indicators such as institutional ownership (*INSTITUTION*), internal control (*IC*) and the proportion of independent directors (*INDBOARD*) ([Nadeesha, 2022](#)) to measure corporate governance characteristics.
- (6) *Management characteristics*: Studies show that management characteristics can significantly influence firms’ greenwashing behavior ([Yue & Li, 2023](#); [Zhang, Qin, & Zhang, 2023](#)). We consider the chief executive officer’s (CEO’s) age (*CEO_AGE*), gender (*GENDER*), academic background (*ACADEMIC_BACK*), financial background (*FINANCE_BACK*), green background (*ENV_BACK*) and diversity of functional experience (*COMP_BACK*) as variables predicting greenwashing behavior.
- (7) *Political affiliation*: We mainly use the CEOs’ government background to measure their corporate political affiliation (*GOV_BACK*). Specifically, the value is taken as 1 if the CEO has served in a government department and 0 otherwise.
- (8) *Family-owned company*: We take family-owned company as a dummy variable to measure whether the company is controlled by family members. It takes the value of 1 if the company is a family business and 0 otherwise ([Habbash, 2013](#)). A family business is one in which at least one family member with kinship ties – besides the actual controller – holds shares in, manages or controls the listed or controlling shareholder company ([Kim et al., 2017](#)).
- (9) *Technology innovation*: We adopt R&D intensity (*RD*) to measure a firm’s technological innovation capability. Specifically, “the ratio of expenditure on research and development to operational revenue” is used to calculate the *RD* ([Liu, Liu, Hu, & Xie, 2020](#)). The higher the *RD*, the higher the firm’s technological innovation ability.
- (10) *Firm-characteristic features*: Based on the literature on ESG and corporate governance, we also include several fundamental firm-level characteristics as features. These include firm size (*SIZE*), measured as the natural logarithm of total assets; firm age (*AGE*), calculated as the natural logarithm of the current year minus the year of incorporation plus one; total asset turnover (*TAT*), which is operating revenue divided by total assets, and management shareholding (*MHOLD*), the proportion of management’s shareholding to total share capital. We also incorporate Tobin’s Q (*TobinQ*), the ratio of market value to total assets; whether the firm is a state-owned enterprise (*SOE*), a dummy variable equal to 1 for SOEs and 0 otherwise and CEO-chairman duality (*DUAL*), which is 1 if the chairman and CEO are the same person and 0 otherwise. The detailed definitions and descriptive statistics of all variables are presented in [Appendix Table A1](#).

4. Research design

4.1 Model selection

We selected three machine learning models – Random Forest, XGBoost and LightGBM – which are currently widely used in the field. Moreover, we use a multiple linear regression model as a benchmark for cross-reference to compare the predictive performance of the ML and traditional econometric models.

The selection of these established ML algorithms is motivated not by algorithmic novelty but by their proven efficacy in capturing complex, non-linear patterns in high-dimensional data – a task for which traditional linear models are often ill-suited. While the absolute improvement in predictive metrics (e.g. R^2) over linear regression may appear modest (see Table 2), the primary value of our ML approach lies in its explanatory power. By combining these algorithms with SHAP interpretation, we shift the focus from mere prediction to understanding the nuanced, conditional and non-linear influences of various drivers on greenwashing – insights that are typically obscured in conventional linear models.

4.2 Model evaluation indicators

In the ML domain, regression tasks are extensively utilized to establish a connection between the continuous target variable and one or more independent variables. To assess the performance of ML algorithms in regression tasks, four commonly used evaluation metrics are applied (Jabeur, Mefteh-Wali, & Viviani, 2024): root mean squared error, mean squared error, mean absolute error and R-squared (R^2). These metrics provide quantitative insights into the accuracy and generalizability of the trained models.

4.3 Machine-learning interpretability

SHAP is a method for interpreting the predictions of ML models. It can provide insights into how individual features contribute to specific predictions. This interpretability is crucial for building trust in ML models and understanding their decision-making processes.

4.3.1 SHAP value. SHAP values are based on Shapley values, which were originally used in cooperative game theory to fairly distribute the “payout” among players based on their contributions. The SHAP value for the feature is given by

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)],$$

where ϕ_i represents the SHAP value for feature i , N is the set of all features, S is a subset of features excluding feature i and $f(S)$ is the prediction function when applied to the subset S . This formula computes the average marginal contribution of a feature over all possible combinations of features, ensuring a fair allocation of the prediction across all features.

4.3.2 SHAP feature importance. SHAP feature importance is determined by averaging the absolute SHAP values for each feature across all samples. This provides a global view of the most influential features of the model:

Table 2. Comparison of the prediction performance

Model	RMSE	MSE	MAE	R^2
LinearRegression	1.277	1.630	1.000	7.179%
RandomForest	1.103	1.217	0.860	30.698%
XGboost	1.082	1.171	0.828	33.352%
LightGBM	1.074	1.155	0.827	34.274%

Note(s): RMSE = root mean squared error, MSE = mean squared error and MAE = mean absolute error

$$Feature\ Importance\ (i) = \frac{1}{M} \sum_{j=1}^M |\phi_{ij}|,$$

where M denotes the number of samples, and ϕ_{ij} is the SHAP value for feature i in sample j . This formula aggregates the SHAP values to measure the overall importance of each feature in the model.

5. Results and analysis

5.1 Performance of machine learning models

The dataset is randomly split into a training set and a test set at an 80:20 ratio. To mitigate the impact of randomness and obtain a more robust assessment of model performance, a five-fold cross-validation is implemented on the training set. This involves iteratively dividing the training set into five folds: four folds for training and one for evaluation. Each algorithm’s performance is averaged across five folds to provide a more reliable estimate of its generalization ability. To further enhance the ML algorithms’ performance, hyperparameter optimization was conducted using Optuna [4], as shown in Figure 1.

Table 2 summarizes the prediction performance achieved by each ML algorithm on the testing set. These scores reflect the ability of each algorithm to accurately predict the target variable for unseen data points. LightGBM outperforms the other algorithms in all metrics.

5.2 Robustness checks

To ensure the robustness of our findings, we conducted multiple supplementary analyses. First, we transformed the continuous greenwashing measure into a binary indicator (GW_Dummy) and reframed the task as a classification problem. As detailed in Appendix 2, the LightGBM model maintained superior performance (AUC = 0.763), and the relative importance of key predictors remained consistent. Second, we performed subsample analyses by (1) industry type (manufacturing vs. non-manufacturing) and (2) ownership (state-owned vs. non-state-owned). The core findings regarding the top predictors (INSTITUTION, DT and HHI) were qualitatively unchanged across subsamples. Third, we varied the train-test split ratios (70:30, 85:15) and obtained similar model performance rankings. These tests collectively affirm the reliability and generalizability of our primary results.

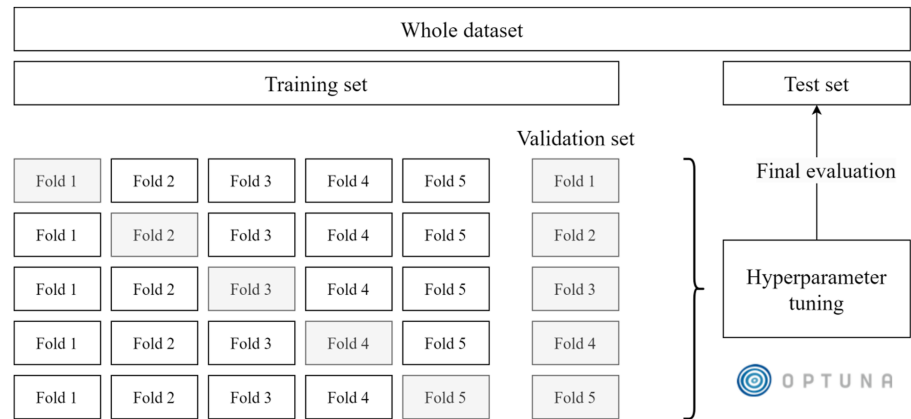


Figure 1. Workflow of K-fold cross-validation with hyperparameter tuning

5.3 Opportunities vs. pressures: importance comparison between groups

To investigate the difference in predictive power between opportunities and pressures, we calculate the feature importance of the opportunities, pressures and baseline variables sets using SHAP feature importance analysis. The sets' results are 0.529, 0.571 and 0.523, respectively. This indicates that there is no significant difference between the two groups. We then observe the status of each group.

As shown in Figure 2, among the opportunities, institutional ownership is the most important in predicting corporate greenwashing, followed by the digital transformation and regional fintech. The results indicate that external shareholders, digital strategy and regional financial facilities have a more predictive power than other factors such as managerial characteristics.

As shown in Figure 3, among the pressures, market competition is the most important feature in predicting corporate greenwashing, followed by type-II agency problems and the debt-to-assets ratio. The results indicate that market competitive pressures, agency cost pressures and debt risk pressures are critical motivators for greenwashing because they have more predictive power than other factors such as external concerns, pollution status and firm profitability.

5.4 Finding the key forces: importance comparison among all the features

We further employ SHAP to compare the importance of the variables individually. As shown in Figure 4, among all the opportunities and pressures, institutional ownership, a firm's digital transformation and market competition have the best predictive power for corporate greenwashing.

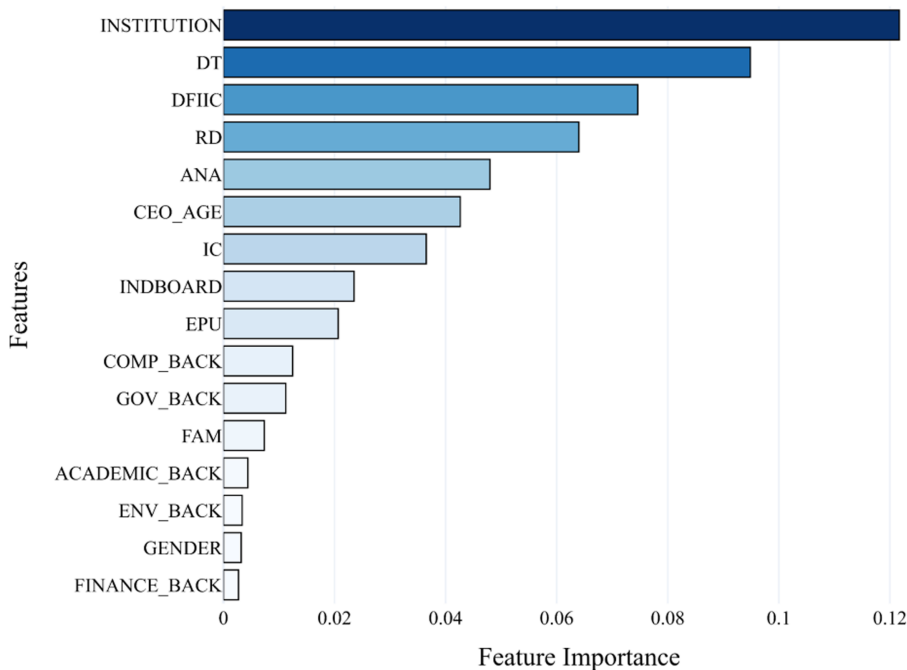


Figure 2. Feature importance of opportunities

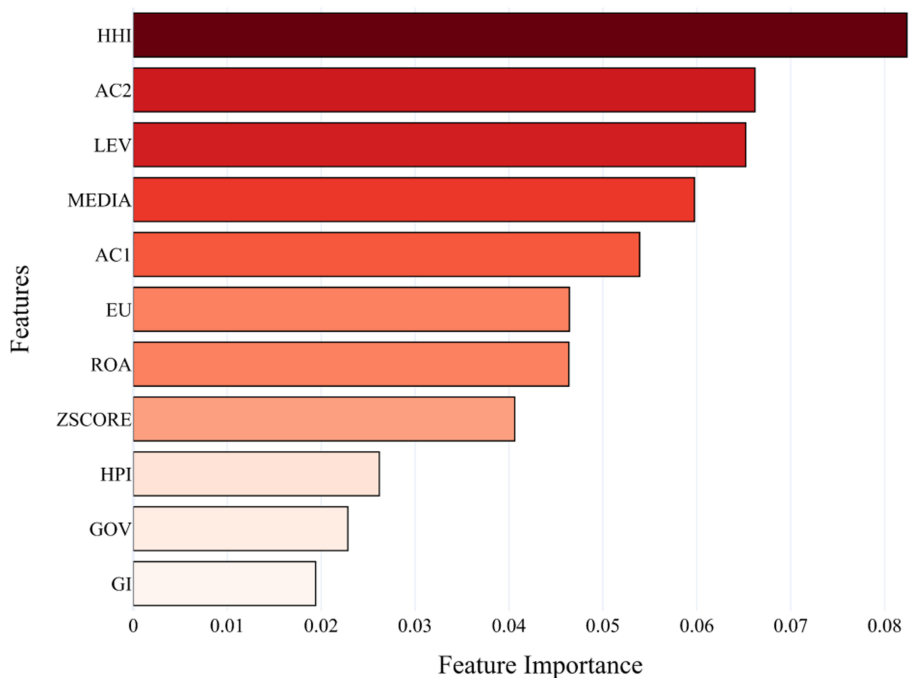


Figure 3. Feature importance of pressures

5.5 How do the key forces matter: predictive patterns by SHAP interpretation

The above analysis indicates that institutional ownership, digital transformation and market competition are key forces with the strongest capacity to predict corporate greenwashing. However, we still do not know exactly how these factors matter. Therefore, we combine the SHAP value interpretation to further explain the predictive pattern.

5.5.1 Institutional ownership. As shown in Figure 5, the SHAP value of institutional ownership increases as the shareholding proportion increases, indicating that institutional ownership may increase the risk of greenwashing. Specifically, from 0% to 40%, we can see a negative influence on greenwashing, which verifies the existing conclusion that institutional investors are sufficiently motivated and capable of playing the role of a monitor or governor against greenwashing (Crane, Koch, & Sebastien, 2019). However, when the proportion is over 40%, the negative influence decreases gradually, and the positive influence is mixed from 40% to 80%. This finding calls back the argument that whether institutional investors are watchdogs or profit grabbers in corporate social responsibility decision-making depends on the stability or dynamics of the market. Jiang and Kim (2015) suggest that in emerging countries, the capital market is not well-matured and institutional investment is less stable, which may easily lead to myopia.

In addition, we observe further dramatic results. As shown in Figure 5, above 80%, the influence of institutional shareholdings is positive and increases in step with the proportion. This means that the more ownership is concentrated among institutional investors, the more likely they are to engage in greenwashing. According to the agency theory, when ownership is over-concentrated, large shareholders are likely to use their absolute right to control manager behavior. Additionally, most Chinese firms are in a period of transition as well as face financial constraints. Thus, instead of developing real green innovations, firms may greenwash their

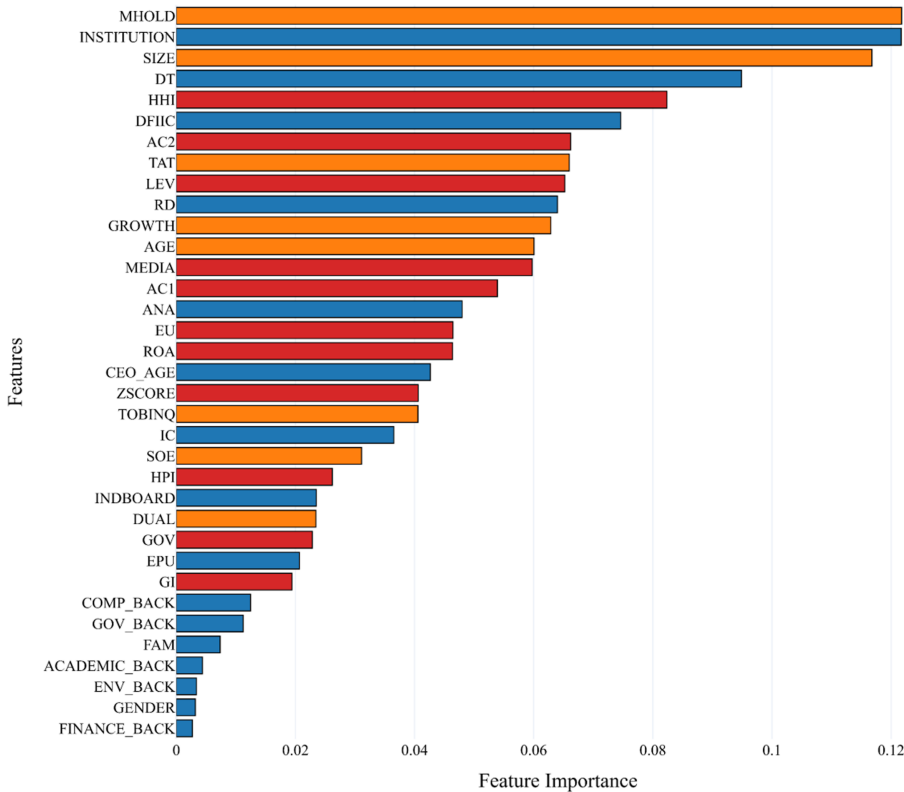


Figure 4. Feature importance of variables

environmental performance as a low-cost strategy to gain the benefit of a good reputation (Wang, Lin, & Xie, 2022).

5.5.2 Firm's digital transformation. Figure 6 shows the trend wherein the SHAP value decreases as the firm's digital transformation intensifies, which indicates that high digitalization may inhibit its greenwashing behavior. Note that the negative influence becomes significant from the score of 50, before which the influence is steadily positive. This result suggests that the digital technology empowerment effect inhibits greenwashing only when digital transformation reaches a certain level. Otherwise, when the firm is in the primary stage of digital exploration, transformational costs and challenges aggravate the burden of capital and other resources, which will likely crowd out the investment in green activities. Under pressure from both institutions and markets, a firm may greenwash to attract more investment and facilitate transformation.

An advanced level of digitalization, where digital technologies are deeply integrated with the operations, increases resource efficiency, reduces the cost of fulfilling environmental responsibilities and provides incentives to undertake substantive green activities. Artificial intelligence (AI)-enabled resource optimization reduces environmental compliance costs while improving profitability, thereby alleviating financial pressures that traditionally incentivize symbolic sustainability claims (Cao, Li, Hu, Wan, & Wang, 2023). The resultant standardized data streams foster unprecedented environmental disclosure transparency, enabling real-time public verification through advanced analytical tools that heighten greenwashing detection risks. Concurrently, digitally transformed organizations become

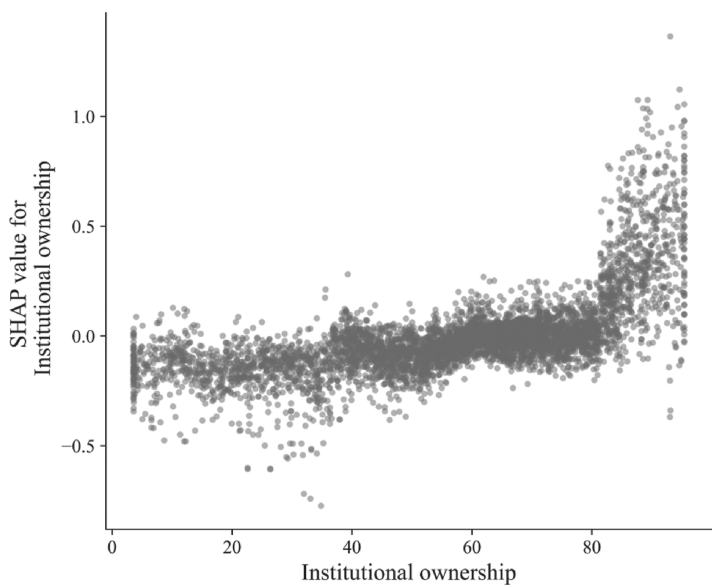


Figure 5. SHAP dependence plot of institutional ownership

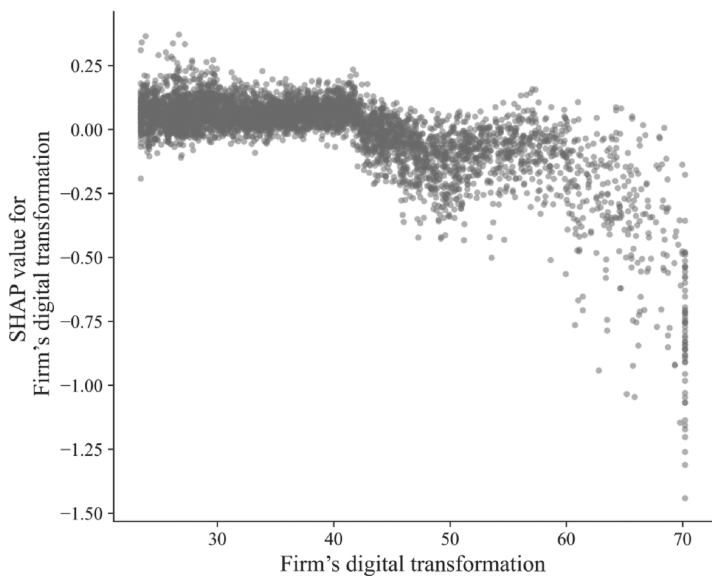


Figure 6. SHAP dependence plot for the digital transformation (DT) score

hyper-visible entities subjected to intensified multilateral scrutiny from media, investors and industry watchdogs – a phenomenon amplified by machine-readable sustainability reporting formats that enable automated compliance audits (Hu, Han, & Zhong, 2023a; Martínez-Peláez *et al.*, 2023). This technological triad systematically reconfigures the cost-benefit calculus,

making substantive green investments strategically preferable to deceptive environmental posturing.

5.5.3 Market competition. As Figure 7 shows, changes in the SHAP value of market competition have a U-shaped impact on corporate greenwashing. This indicates that both extreme competition and monopoly stimulate corporate greenwashing, while market competition within an appropriate range ($HHI = 0.3\text{--}0.9$) effectively inhibits greenwashing.

First, a more competitive market allows a firm to have less pricing power. In this case, it is difficult for a firm to transfer the costs of its environmental activities to consumers. High prices trigger product market risks such as consumer shifting or substitution, which harm a firm's competitiveness and market share (Hu, Yu, & Han, 2023). To appeal to the consumer market and investor demand for green products, firms may use fake advertisements, exaggerated ESG performance and other greenwashing approaches to enhance their image and earnings (Nardi, 2022). In contrast, a few large corporations in a monopolized industry have the advantage of controlling the price and supply, which makes it difficult for consumers to bargain for green claims and thus drive innovation.

5.6 Further analysis

The global ascendancy of digital economies, now constituting core national competitiveness, finds particular manifestation in China's strategic "Digital China" initiative. Since its 2015 elevation to national policy, this technological paradigm shift has propelled China's digital economy to RMB 56.1 trillion (44% of gross domestic product) by 2023, fundamentally reconfiguring industrial operations and governance frameworks. Our temporal analysis of techno-economic paradigm shifts (pre-/post-2015 bifurcation) reveals evolving greenwashing predictors: institutional ownership and corporate digitalization demonstrate growing explanatory power, whereas media influence (*MEDIA*) and principal-agent conflicts (*AC2*) exhibit non-linear feature importance trajectories. This empirical evidence underscores digitalization's dual role as both an economic accelerator and a corporate accountability modulator within transitional economies.

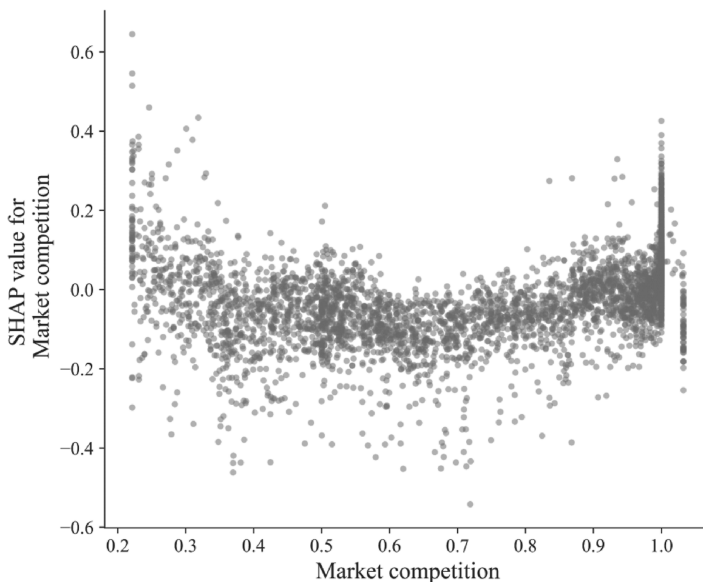


Figure 7. SHAP dependence plot for market competition

5.6.1 Media attention. Empirical analysis reveals media attention’s predictive importance for greenwashing ascended from 16th to 4th position post-2015 (Figures 8 and 9), surpassing market competition and trailing only institutional ownership and digital transformation. This growing influence reflects media’s reconfigured role in digital economies, where monitoring mechanisms have evolved from traditional reportage to AI-enhanced investigative analytics (Sun, Wang, Sun, & Zhang, 2023; Lungu, Georgescu, & Juravle, 2024). Technological convergence enables three transformative capabilities: (1) blockchain-enabled supply chain forensics ensuring commitment veracity (Jedynak, 2024); (2) cross-domain data integration generating multidimensional sustainability assessments and (3) algorithm-driven public engagement platforms amplifying grassroots environmental oversight. These advancements establish the media as both technological auditors and social accountability amplifiers, compelling enhanced disclosure transparency while democratizing green governance through crowdsourced vigilance.

5.6.2 Type-II agency problem. Empirical evidence demonstrates type-II agency problems’ diminishing predictive power for greenwashing post-2015 (Figures 8 and 9), reflecting digital economies’ capacity to counterbalance controlling shareholders’ expropriation tendencies. In concentrated ownership structures typical of emerging markets, majority shareholders traditionally exploit informational asymmetries to prioritize short-term tunneling over sustainable investments (La Porta, Lopez-de-Silanes, & Shleifer, 1999; Zhang, Meng, & Zhang, 2023). Digital transformation introduces three disciplinary mechanisms: (1) algorithmic auditing of historical decision patterns enables real-time tunneling detection;

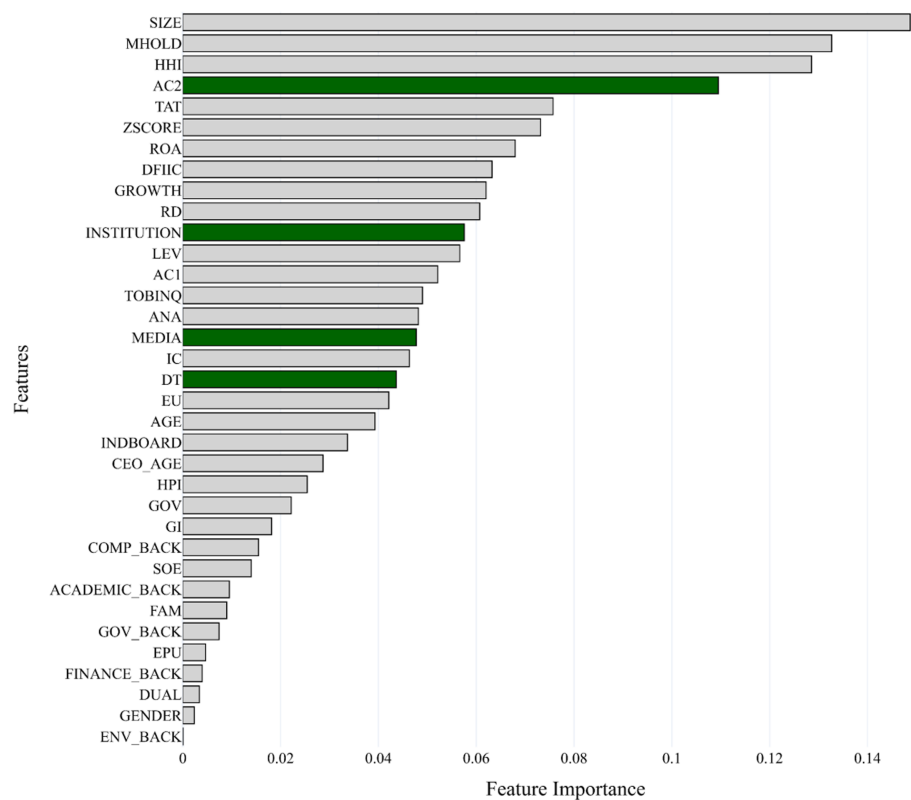


Figure 8. Feature importance (2012–2014)

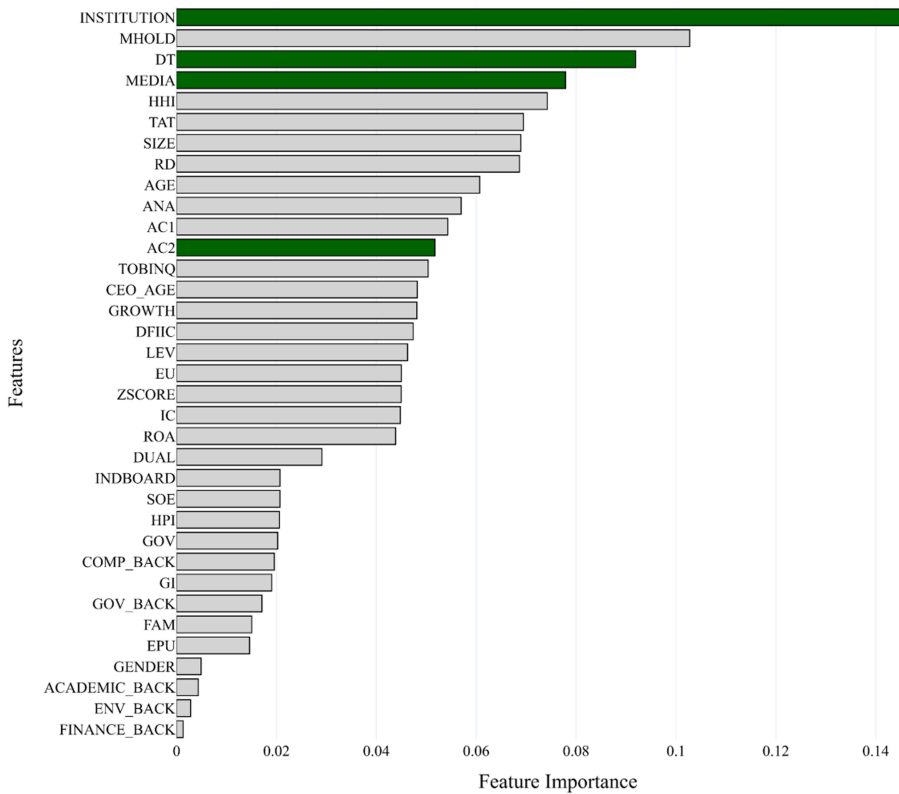


Figure 9. Feature importance (2015–2022)

(2) ML-enhanced text analysis deciphers hidden expropriation signals (Lafarre and Van der Elst, 2018) and (3) investor portals like “Hudongyi” of the Shenzhen Stock Exchange and “Yihudong” of the Shanghai Stock Exchange create multilateral oversight through crowdsourced governance (Wang & Tang, 2024). These technological safeguards increase tunneling risks while empowering minority shareholders through enhanced information parity, thereby reorienting dominant shareholders’ calculus toward long-term environmental commitments over immediate rent extraction.

6. Discussion and conclusions

6.1 Key findings

Although ESG investment is booming in China’s A-share market, a report released by Standard & Poor’s [5] shows that investors are concerned about ESG investment, with more than 44% of investors believing that greenwashing is the biggest concern. This study systematically analyzes the drivers of corporate greenwashing by considering two dimensions: opportunity and pressure. Based on a sample of A-share listed companies in China from 2012 to 2022, we use ML approaches to comprehensively examine and compare the predictive effects of 27 specific factors on corporate greenwashing.

First, the results show no significant difference in the overall performance of the two categories of factors – opportunity and pressure – in predicting firms’ greenwashing. Among the opportunity factors, the feature importance of the institutional investors’ shareholdings,

digital transformation and regional fintech where the firms are located are ranked first, second and third, respectively, and are important conditions for the occurrence of greenwashing behavior. Among the pressure factors, the feature importance of market competition, type-II agency problem and asset-liability ratio ranked first, second and third, respectively, and constituted the key motives for greenwashing behavior.

Second, among the many influencing factors, three variables – institutional investor shareholdings, digital transformation and market competition – are the best predictors of corporate greenwashing. Specifically, institutional investor ownership positively predicts corporate greenwashing, which means that greenwashing behavior is more likely to be conducted as the institutional ownership is concentrated, and a certain level of digital transformation has a negative predictive effect on enterprises' greenwashing behavior. However, a U-shaped relationship exists between market competition intensity and greenwashing behavior. It demonstrates that over-dispersed and over-concentrated industries are prone to greenwashing, whereas moderate market competition inhibits it.

Third, the fast development of the digital economy profoundly changes the factors and modes of production and significantly impacts the strategic decision-making of corporate sustainable development. In particular, social media's governance effect was significantly enhanced, and internal agency problems' stimulus effect on greenwashing behavior was mitigated after 2015 as the digital transformation became a national strategy in China. Finally, by comparing the models, this study confirms that ML algorithms significantly outperform traditional regression models in terms of event prediction. In addition, we find that LightGBM is the most effective model for predicting corporate greenwashing issues.

6.2 Theoretical implications

First, unlike the existing research on the motivational effect of greenwashing, this study makes no harsh assumptions. Instead, it adopts the advanced ML integration algorithm to “learn” from the disordered information, develops the optimal function model, predicts enterprises' greenwashing behavior and obtains more systematic empirical conclusions.

Second, we advance the literature by constructing the IMOC framework, which integrates institutional theory, resource-based view, agency theory and upper echelons theory. Unlike prior studies that isolate specific drivers, our model systematically categorizes 27 factors into institution, market, organization and cognition levels, offering a holistic view of the greenwashing ecosystem. The model expands the understanding and analysis of the dynamics of greenwashing and enriches the literature on ESG and ESG greenwashing behavior.

Third, we move the nascent literature on ML applications in greenwashing (e.g. [Zeng et al., 2025](#); [Li et al., 2025](#)) from a primarily predictive focus to an explanatory and diagnostic one. By integrating SHAP interpretability with our IMOC framework, we decode the “black box” of ML predictions. We demonstrate, for instance, that organizational factors like ownership structure primarily function as opportunity modulators, while market factors like competition intensity act as structural pressure generators. This nuanced understanding refines the application of fraud theory in environmental contexts and provides a template for future explainable AI research in sustainability governance.

Fourth, this study obtains several interesting findings through an in-depth investigation of the predictive mechanisms of these key factors. While most existing studies emphasize the monitoring and governance roles of institutional investors, our study finds the opposite: The risk of greenwashing gradually increases as institutional ownership increases. The literature emphasizes market competition as an important pressure factor for firms to manipulate disclosures for impression management ([Bagnoli & Watts, 2010](#); [Markarian & Santalo, 2014](#)). However, our study finds that the relationship between market competition intensity and corporate greenwashing is not a simple linear relationship but rather a U-shaped relationship. Our study further supports the condition that digital transformation can effectively restrain

greenwashing behavior. We further suggest that this effect becomes apparent only after achieving a certain level of digital transformation.

In addition, our analysis of the time dynamics reveals that the rapid development of the digital economy has important implications for firms' internal and external governance environments. The governance effect of media supervision is significantly enhanced, and the stimulus of greenwashing behavior by intra-firm agency problems is mitigated.

6.3 Practical implications

The findings of this study provide the following practical insights. First, we should pay attention to institutional investors' governance and jointly discourage greenwashing behavior. The results suggest that the proportion of institutional investor shareholdings is the most important factor affecting greenwashing behavior. In particular, greenwashing risk increases as shareholding increases. Therefore, companies should take the initiative to disclose detailed environmental performance data to institutional investors, including information on carbon emissions, energy consumption and pollution control, all of which are important for improving corporate performance. This will increase the transparency of information and help firms realize that green development also has economic benefits. In addition, companies should formulate clear long-term green development strategies and integrate them into their overall development plans. By working with institutional investors to develop and implement these strategies, they can effectively advance green governance objectives.

Second, moderate industry competition guarantees green development. Our research shows that dynamic changes in market competition have a complex, non-linear impact on green behavior. Accordingly, the government should perform macro-control functions to guide market competition and prudently optimize resource allocation. Specifically, the government should adopt anti-monopoly policies, laws and regulations for industries with excessive market concentration. This will help prevent enterprises from forming monopolies and enhance market competition. Through policy incentives, large enterprises can be guided to assume more environmental responsibility and enhance their green governance. Moreover, the government can promote the optimization and upgrading of industrial structures and support the development of green and low-carbon industries through industrial policies and economic planning. Moderately regulating and transforming high-pollution and high-energy-consumption industries can reduce adverse impacts on the environment and promote industry-wide green transformation.

Third, enterprises' digital transformation is an important governance mechanism that can inhibit greenwashing behavior. The descriptive statistics indicate that the mean value of digital transformation is 38.369, which is still low. This study's results show that the deep integration of digital technology and enterprise operation and management can effectively exert their enabling effects only when the digital transformation score is higher than 40. Digital transformation can improve the quality of IC, reduce information asymmetry, improve enterprise productivity and profitability, alleviate financing constraints, inhibit enterprise greenwashing and promote substantive green performance.

Fourth, media has become an important subject of external governance with the development of the digital economy. The results show that the media's importance in predicting corporate greenwashing increased from the 16th to the fourth position after 2015, confirming that digital economy development further enhances the media's role in monitoring corporate greenwashing. This dynamic is vividly illustrated by real-world events; for example, the exposure of several carbon emission certificate frauds by Chinese authorities in 2022 [6]. Such incidents, once disclosed, are rapidly amplified by the digital media, leading to heightened public scrutiny and significant reputational damage for the firms involved. This reinforces the media's critical function as a powerful external governance mechanism in the digital era. Therefore, media should fully use digital technologies such as big data analytics, AI and blockchain to enhance the depth and breadth of corporate environmental reporting.

Specifically, through digital technology, media can track enterprises' environmental performance more accurately, expose greenwashing behavior and provide authoritative analyses and evaluations. This can effectively curb corporate greenwashing, encourage enterprises to enhance green practices and promote the society's sustainable development.

6.4 Limitations and prospects

This study has some limitations. First, our sample only contains data on listed companies in China. Because institutional and cultural differences affect corporate ethical behavior, our findings may lack global generalizability. In the future, we can expand the sample to conduct empirical research and incorporate institutional and cultural differences into the research framework for examination and comparison. Second, ML methods emphasize out-of-sample generalization, do not require prior assumptions and are characterized by a “black box” that lacks the explanatory advantages of the influence mechanism. However, we refer to the existing scholarly research and explain the prediction mechanism of key variables by combining ML methods with human theoretical knowledge. Moreover, the principle of SHAP scatter dependence may be challenging to understand owing to its difference from traditional measurement logic. In future research, a hybrid method combining traditional measurement and ML methods can be attempted to elucidate and interpret the influence mechanism.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author upon reasonable request.

Statements and declarations

All errors are our own.

Appendix 1

Variable definitions

Table A1. Variable definitions and descriptive statistics

Variable	Definition	Mean	Max	SD	Median	Min
GW	Corporate <i>greenwashing</i> behavior, model (1) for details of the calculation	−0.440	5.691	1.310	−0.492	−4.708
HHI	Herfindahl-Hirschman Index	0.782	1.032	0.251	0.915	0.221
GOV	Low-carbon policy intensity in prefecture-level cities	2.857	21.500	4.384	0.000	0.000
MEDIA	Ln (Company-related news +1)	3.793	5.743	0.825	3.951	1.609
EU	Standard deviation of sales revenues (industry-adjusted) over the five years	1.240	6.643	1.096	0.948	0.123
ROA	Net Income/Total Assets	0.055	0.229	0.055	0.046	−0.106
HPI	1 if firm is in heavy polluting industries and 0 otherwise	0.295	1.000	0.456	0.000	0.000
AC1	(Administration expenses + selling expenses)/ Operating revenue	0.141	0.576	0.116	0.106	0.013

(continued)

Table A1. Continued

Variable	Definition	Mean	Max	SD	Median	Min
AC2	Other accounts receivable/ Operating revenue	0.013	0.115	0.018	0.007	0.000
GI	Ln(The sum of green investor+1)	1.158	3.367	0.942	1.099	0.000
ZSCORE	Altman Z index	4.823	44.578	6.272	2.863	0.273
DFIIC	Provincial level financial inclusion index	242.763	351.532	65.717	250.624	91.000
DT	Digital transformation index	38.369	70.204	11.376	35.997	23.411
ANA	Ln(Analyst followings+1)	2.385	4.007	0.933	2.485	0.693
EPU	Economic Policy Uncertainty index from Baker et al. (2016)	436.948	791.874	232.109	460.470	113.897
INSTITUTION	The institutional investors' shareholdings/The total number of shares	57.060	95.383	22.535	60.633	3.563
IC	Internal control index of Shenzhen Dibo database	667.059	880.920	121.721	683.525	0.000
INDBOARD	Independent directors/Total directors	0.375	0.571	0.055	0.364	0.333
CEO_AGE	CEO Age	51.222	65.000	5.791	51.000	36.000
GENDER	1 if CEO is males and 0 for females	0.936	1.000	0.244	1.000	0.000
ACADEMIC_ BACK	1 if the CEO has worked in a university research organization and 0 otherwise	0.184	1.000	0.387	0.000	0.000
FINANCE_ BACK	1 if CEO has worked in banks, securities, or fund companies and 0 otherwise	0.044	1.000	0.206	0.000	0.000
ENV_BACK	1 if CEO's resume includes environment related experience and 0 otherwise	0.048	1.000	0.213	0.000	0.000
COMP_BACK	Number of CEOs with professional backgrounds	2.038	4.000	0.901	2.000	1.000
GOV_BACK	1 if the CEO has served in the government and 0 otherwise	0.103	1.000	0.303	0.000	0.000
FAM	1 if the listed company is a family business and 0 otherwise	0.323	1.000	0.468	0.000	0.000
RD	R&D expenditures/Operating revenue	0.040	0.281	0.046	0.030	0.000
SIZE	Ln (Total assets)	23.512	27.294	1.337	23.395	20.806
LEV	Total debts/Total assets	0.468	0.849	0.187	0.482	0.075
AGE	Ln(Current year – year of incorporation + 1)	18.959	33.000	5.719	19.000	6.000
TAT	Operating revenue/Total assets	0.669	2.350	0.399	0.584	0.115
MHOLD	Management shareholding/Total share capital	0.075	0.642	0.150	0.001	0.000
TOBINQ	Market value/Total assets	2.079	9.167	1.508	1.548	0.818
SOE	1 if firm is state-owned enterprise and 0 if it is a private enterprise	0.486	1.000	0.500	0.000	0.000
DUAL	1 if the chairman and CEO are the same, and 0 otherwise	0.233	1.000	0.423	0.000	0.000
GROWTH	(Sales revenue for the year t/sales revenue for the year t-1)-1	0.178	2.089	0.339	0.121	-0.392

Note(s): SD = standard deviation, CEO = chief executive officer and R&D = research and development

Appendix 2
Robustness check with binary classification

To address the inherent challenges in quantifying greenwashing and to ensure the robustness of our findings, we conducted an additional analysis using an alternative specification for our dependent variable. While our main analysis uses a continuous variable to capture the magnitude of the disclosure-performance gap, for this check, we transformed it into a binary variable to test the model’s ability to predict the occurrence of greenwashing.

We defined a dummy variable, *GW_Dummy*, which takes a value of 1 if the original greenwashing score *GW* is greater than zero and 0 otherwise. A value of 1 indicates that a firm’s standardized environmental disclosure outranks its substantive environmental performance, signifying a state of greenwashing. This transformed our task from a regression problem into a binary classification problem.

We then retrained the LightGBM, XGBoost and random forest models on this new target variable. We evaluated their performance using standard classification metrics: accuracy, precision, recall, F1-score and the area under the receiver operating characteristic curve (AUC). The results are presented in [Appendix Table A2](#). The results show that all machine learning models demonstrate strong predictive capability, significantly outperforming a baseline logistic regression. Consistent with our primary findings, LightGBM achieved the best performance. The consistency of these results confirms that our model and the identified key features are robust predictors of greenwashing behavior.

Table A2. Performance of models for binary classification task

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic regression	0.641	0.640	0.041	0.077	0.599
Random forest	0.708	0.832	0.254	0.389	0.748
XGBoost	0.727	0.744	0.387	0.509	0.759
LightGBM	0.733	0.731	0.426	0.538	0.763

Notes

1. The data is from CSMAR database.
2. The data information is from <https://idf.pku.edu.cn/yjcg/zsbg/513800.htm>
3. According to [Baker et al. \(2016\)](#), the index is conducted by newspaper-based indices of policy uncertainty in China.
4. Optuna is a Bayesian optimization framework that efficiently explores the hyperparameter space to identify the combination that yields the best performance. The optimized hyperparameters are then applied to the trained models on the testing set.
5. The specific report is that <https://www.spglobal.com/ratings/en/research/articles/210823-the-fear-of-greenwashing-may-be-greater-than-the-reality-across-the-global-financial-markets-12074863>
6. “China uncovers fraud scheme in carbon emission reporting,” Upstream Online, March 25, 2022. <https://www.upstreamonline.com/energy-transition/china-uncovers-fraud-scheme-in-carbon-emission-reporting/2-1-1184461>

Reference

- Bagnoli, M., & Watts, G. S. (2010). Oligopoly, disclosure and earnings management. *The Accounting Review*, 85(4), 1191–1214. doi: [10.2308/accr.2010.85.4.1191](https://doi.org/10.2308/accr.2010.85.4.1191).
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4), 1593–1636. doi: [10.1093/qje/qjw024](https://doi.org/10.1093/qje/qjw024).

- Baudry, M., Bukowski, S., & Lament, M. B. (2024). *Financial stability, economic growth and sustainable development*. London: Routledge.
- Bernini, F., Giuliani, M., & La Rosa, F. (2023). Measuring greenwashing: A systematic methodological literature review. *Business Ethics, the Environment & Responsibility*, 1(4), 1–19. doi: [10.1111/beer.12631](https://doi.org/10.1111/beer.12631).
- Blome, C., Foerstl, K., & Schleper, M. C. (2017). Antecedents of green supplier championing and greenwashing: An empirical study on leadership and ethical incentives. *Journal of Cleaner Production*, 152, 339–350. doi: [10.1016/j.jclepro.2017.03.052](https://doi.org/10.1016/j.jclepro.2017.03.052).
- Cao, Y., Li, X., Hu, H. L., Wan, G. Y., & Wang, S. Y. (2023). How does digitalization drive the green transformation in manufacturing companies? An exploratory case study from the perspective of resource orchestration theory. *Journal of Management World*, 39(3), 96–112+126+113. doi: [10.19744/j.cnki.11-1235/f.2023.0045](https://doi.org/10.19744/j.cnki.11-1235/f.2023.0045).
- Chen, P., & Dagestani, A. A. (2023). Greenwashing behavior and firm value—From the perspective of board characteristics. *Corporate Social Responsibility and Environmental Management*, 30(5), 2330–2343. doi: [10.1002/csr.2095](https://doi.org/10.1002/csr.2095).
- Chen, Y., Du, Y., Wu, Y., & Zhou, H. (2017). The study on zombie firms in China using modified FN-CHK method. In *2017 4th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS)* (pp. 1–7). IEEE. doi: [10.1109/IEIS.2017.8078672](https://doi.org/10.1109/IEIS.2017.8078672).
- Crane, A. D., Koch, A., & Sebastien, M. (2019). Institutional investor cliques and governance. *Journal of Financial Economics*, 133(11), 175–197. doi: [10.1016/j.jfineco.2018.11.012](https://doi.org/10.1016/j.jfineco.2018.11.012).
- Delmas, M. A., Nairn-Birch, N., & Lim, J. (2015). Dynamics of environmental and financial performance: The case of greenhouse gas emissions. *Organization & Environment*, 28(4), 374–393. doi: [10.1177/1086026615620238](https://doi.org/10.1177/1086026615620238).
- Dong, X., Wang, C., Zhang, F., Zhang, H., & Xia, C. (2024). China's low-carbon policy intensity dataset from national-to prefecture-level over 2007–2022. *Scientific Data*, 11(1), 213. doi: [10.1038/s41597-024-03033-5](https://doi.org/10.1038/s41597-024-03033-5).
- Du, X. (2015). How the market values greenwashing? Evidence from China. *Journal of Business Ethics*, 128(3), 547–574. doi: [10.1007/s10551-014-2122-y](https://doi.org/10.1007/s10551-014-2122-y).
- Effron, D. A., O'Connor, K., Leroy, H., & Lucas, B. J. (2018). From inconsistency to hypocrisy: When does 'saying one thing but doing another' invite condemnation. *Research in Organizational Behavior*, 38, 61–75. doi: [10.1016/j.riob.2018.10.003](https://doi.org/10.1016/j.riob.2018.10.003).
- Fang, C., Nolan, P., & Linggui, W. (2019). *Routledge handbook of the belt and road* (pp. 213–217). London: Routledge.
- Feng, J., & Yuan, Y. (2024). Green investors and corporate ESG performance: Evidence from China. *Finance Research Letters*, 60, 104892. doi: [10.1016/j.frl.2023.104892](https://doi.org/10.1016/j.frl.2023.104892).
- Ferrón-Vílchez, V., Valero-Gil, J., & Suárez-Perales, I. (2021). How does greenwashing influence managers' decision-making? An experimental approach under stakeholder view. *Corporate Social Responsibility and Environmental Management*, 28(2), 860–880. doi: [10.1002/csr.2095](https://doi.org/10.1002/csr.2095).
- Guo, X. (2022). Environmental regulation and green innovation of enterprises. *BCP Business and Management*, 25, 593–604. doi: [10.54691/bcpbm.v25i.1885](https://doi.org/10.54691/bcpbm.v25i.1885).
- Habbash, M. (2013). Earnings management, audit committee effectiveness and the role of blockholders ownership: Evidence from UK large firms. *International Journal of Business Governance and Ethics*, 8(2), 155–180. doi: [10.1504/IJBGE.2013.054418](https://doi.org/10.1504/IJBGE.2013.054418).
- Hu, Y., & Liu, Q. (2023). Local digital economy and corporate social responsibility. *Sustainability*, 15(11), 8487. doi: [10.3390/su15118487](https://doi.org/10.3390/su15118487).
- Hu, J., Han, Y. M., & Zhong, Y. (2023a). How corporate digital transformation affects corporate ESG performance. *Review of Industrial Economics*, 1, 105–123. doi: [10.19313/j.cnki.cn10-1223/f.20221104.001](https://doi.org/10.19313/j.cnki.cn10-1223/f.20221104.001).
- Hu, J., Yu, X. R., & Han, Y. M. (2023b). Can ESG rating promote green transformation of enterprises?. *Journal of Quantitative and Technological Economics*, 40(7), 90–111. doi: [10.20092/j.cnki.ddjryj.2023.06.001](https://doi.org/10.20092/j.cnki.ddjryj.2023.06.001).

- Hu, S., Wang, A., & Du, K. (2023c). Environmental tax reform and greenwashing: Evidence from Chinese listed companies. *Energy Economics*, 124, 106873. doi: [10.1016/j.eneco.2023.106873](https://doi.org/10.1016/j.eneco.2023.106873).
- Ioannou, I., Kassinis, G., & Papagiannakis, G. (2023). The impact of perceived greenwashing on customer satisfaction and the contingent role of capability reputation. *Journal of Business Ethics*, 185(2), 333–347. doi: [10.1007/s10551-022-05151-9](https://doi.org/10.1007/s10551-022-05151-9).
- Jabeur, S. B., Mefteh-Wali, S., & Viviani, J. L. (2024). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 334(1), 679–699. doi: [10.1007/s10479-021-04187-w](https://doi.org/10.1007/s10479-021-04187-w).
- Jedynak, M. (2024). *Corporate social responsibility and the supply chain: CSR collaboration with Suppliers*. New York: Routledge.
- Jiang, F., & Kim, K. A. (2015). Corporate governance in China: A modern perspective. *Journal of Corporate Finance*, 32, 190–216.
- Kim, E. H., & Lyon, T. P. (2015). Greenwash vs. brownwash: Exaggeration and undue modesty in corporate sustainability disclosure. *Organization Science*, 26(3), 705–723. doi: [10.1287/orsc.2014.0949](https://doi.org/10.1287/orsc.2014.0949).
- Kim, J., Fairclough, S., & Dibrell, C. (2017). Attention, action, and greenwash in family-influenced firms? Evidence from polluting industries. *Organization & Environment*, 30(4), 304–323. doi: [10.1177/1086026616673410](https://doi.org/10.1177/1086026616673410).
- Kumar, B., Manrai, A. K., & Manrai, L. A. (2017). Purchasing behaviour for environmentally sustainable products: A conceptual framework and empirical study. *Journal of Retailing and Consumer Services*, 34, 1–9. doi: [10.1016/j.jretconser.2016.09.004](https://doi.org/10.1016/j.jretconser.2016.09.004).
- Lafarre, A., & Van der Elst C. (2018). Blockchain technology for corporate governance and shareholder activism. European Corporate Governance Institute-Law Working Paper, (390), 1-26. doi: [10.2139/ssrn.3135209](https://doi.org/10.2139/ssrn.3135209).
- La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (1999). Corporate ownership around the world. *The Journal of Finance*, 54(2), 471–517. doi: [10.1111/0022-1082.00115](https://doi.org/10.1111/0022-1082.00115).
- Li, Q., Wang, R., & He, Z. C. (2023a). Can the exit threat of institutional investors play green governance? Based on the greenwashing behavior of enterprises. *Economy and Management*, 37 (04), 72–82. Available from: <https://link.cnki.net/urlid/13.1032.F.20230613.1349.004>
- Li, W., Li, W., Seppänen, V., & Koivumäki, T. (2023b). Effects of greenwashing on financial performance: Moderation through local environmental regulation and media coverage. *Business Strategy and the Environment*, 32(1), 820–841. doi: [10.1002/bse.3177](https://doi.org/10.1002/bse.3177).
- Li, Y., Cao, X., & Wang, M. (2024a). Can the cumulative effect of technological resources promote green technology collaborative innovation in resource-based regions?. *Journal of Cleaner Production*, 461, 142589. doi: [10.1016/j.jclepro.2024.142589](https://doi.org/10.1016/j.jclepro.2024.142589).
- Li, Z., Miao, S., & Xu, L. (2024b). Digital transformation and environmental, social, and governance greenwashing: Evidence from China. *Journal of Environmental Management*, 365, 121460. doi: [10.1016/j.jenvman.2024.121460](https://doi.org/10.1016/j.jenvman.2024.121460).
- Li, Q., He, Z., & Li, H. (2025). Uncovering corporate greenwashing: A predictive model based on Chinese heavy-pollution industries. *Sustainability Accounting, Management and Policy Journal*, 16(1), 137–167. doi: [10.1108/SAMPJ-11-2023-0813](https://doi.org/10.1108/SAMPJ-11-2023-0813).
- Liao, F., Sun, Y., & Xu, S. (2023). Financial report comment letters and greenwashing in environmental, social and governance disclosures: Evidence from China. *Energy Economics*, 127, 107122. doi: [10.1016/j.eneco.2023.107122](https://doi.org/10.1016/j.eneco.2023.107122).
- Lin, S. T., & Niu, H. J. (2018). Green consumption: E nvironmental knowledge, environmental consciousness, social norms, and purchasing behavior. *Business Strategy and the Environment*, 27(8), 1679–1688. doi: [10.1002/bse.2233](https://doi.org/10.1002/bse.2233).
- Liu, T., Liu, M., Hu, X., & Xie, B. (2020). The effect of environmental regulations on innovation in heavy-polluting and resource-based enterprises: Quasi-natural experimental evidence from China. *PLoS One*, 15(12), e0239549. doi: [10.1371/journal.pone.0272098](https://doi.org/10.1371/journal.pone.0272098).

- Lungu, A. E., Georgescu, M. R., & Juravle, D. (2024). A bibliometric analysis of digital entrepreneurship. *Journal of the Knowledge Economy*, 15(4), 1–29. doi: [10.1007/s13132-024-01885-1](https://doi.org/10.1007/s13132-024-01885-1).
- Lyon, T. P., & Montgomery, A. W. (2013). Tweetjacked: The impact of social media on corporate greenwash. *Journal of Business Ethics*, 118(4), 747–757. doi: [10.1007/s10551-013-1958-x](https://doi.org/10.1007/s10551-013-1958-x).
- Markarian, G., & Santalo, J. (2014). Product market competition, information and earnings management. *Journal of Business Finance & Accounting*, 4(5-6), 572–599. doi: [10.1111/jbfa.12064](https://doi.org/10.1111/jbfa.12064).
- Martínez-Peláez, R., Ochoa-Brust, A., Rivera, S., Félix, V. G., Ostos, R., Brito, H., & Mena, L. J. (2023). Role of digital transformation for achieving sustainability: Mediated role of stakeholders, key capabilities, and technology. *Sustainability*, 15(14), 11221. doi: [10.3390/su151411221](https://doi.org/10.3390/su151411221).
- Nadeesha, D. H. D. (2022). *The effectiveness of board diversity reforms and corporate practices: The moderating role of a familial culture*. Prevalence of Family Businesses and Family Ownership. (Doctoral dissertation, The University of Liverpool, United Kingdom).
- Nardi, L. (2022). The corporate social responsibility price premium as an enabler of substantive CSR. *Academy of Management Review*, 2, 282–308. doi: [10.5465/amr.2019.0425](https://doi.org/10.5465/amr.2019.0425).
- Niu, D., & Wang, Z. (2024). Can ESG ratings promote green total factor productivity? Empirical evidence from Chinese listed companies. *Heliyon*, 10(7), e29307. doi: [10.1016/j.heliyon.2024.e29307](https://doi.org/10.1016/j.heliyon.2024.e29307).
- Parguel, B., Benoît-Moreau, F., & Larceneux, F. (2011). How sustainability ratings might deter ‘greenwashing’: A closer look at ethical corporate communication. *Journal of Business Ethics*, 102(1), 15–28. doi: [10.1007/s10551-011-0901-2](https://doi.org/10.1007/s10551-011-0901-2).
- Pizzetti, M., Gatti, L., & Seele, P. (2021). Firms talk, suppliers walk: Analyzing the locus of greenwashing in the blame game and introducing ‘vicarious greenwashing’. *Journal of Business Ethics*, 170(1), 21–38. doi: [10.1007/s10551-019-04406-2](https://doi.org/10.1007/s10551-019-04406-2).
- Sun, Z. Y., Wang, W., Sun, M. X., & Zhang, Y. (2023). Impact of media coverage on corporate greenwashing—mediating role of executive characteristics and internal supervision. *Journal of Beijing Institute of Technology (Social Sciences Edition)*, 25(1), 67–79. doi: [10.15918/j.jbitss1009-3370.2023.0636](https://doi.org/10.15918/j.jbitss1009-3370.2023.0636).
- Szabo, S., & Webster, J. (2021). Perceived greenwashing: The effects of green marketing on environmental and product perceptions. *Journal of Business Ethics*, 171(4), 719–739. doi: [10.1007/s10551-020-04461-0](https://doi.org/10.1007/s10551-020-04461-0).
- Testa, F., Boiral, O., & Iraldo, F. (2018). Internalization of environmental practices and institutional complexity: Can stakeholders pressures encourage greenwashing?. *Journal of Business Ethics*, 147(2), 287–307. doi: [10.1007/s10551-015-2960-2](https://doi.org/10.1007/s10551-015-2960-2).
- Tiffin, A. J. (2019). *Machine learning and causality: the impact of financial crises on growth*. Washington, DC: International Monetary Fund.
- Wang, Z., & Tang, P. (2024). Substantive digital innovation or symbolic digital innovation: Which type of digital innovation is more conducive to corporate ESG performance?. *International Review of Economics and Finance*, 93, 1212–1228. doi: [10.1016/j.iref.2024.05.023](https://doi.org/10.1016/j.iref.2024.05.023).
- Wang, L., Zeng, T., & Li, C. (2022a). Behavior decision of top management team and enterprise green technology innovation. *Journal of Cleaner Production*, 367, 133120. doi: [10.1016/j.jclepro.2022.133120](https://doi.org/10.1016/j.jclepro.2022.133120).
- Wang, H., Lin, W. F., & Xie, R. (2022b). Environmental backgrounds of executives and green investor entry. *Journal of Quantitative and Technological Economics*, 39(12), 173–194. doi: [10.13653/j.cnki.jqte.2022.12.009](https://doi.org/10.13653/j.cnki.jqte.2022.12.009).
- Wen, Y. D., Wang, R., & He, Z. C. (2024). Digital finance and enterprise ESG performance: Effects, mechanisms and ‘greenwash’ test. *Statistics and Decision*, 40(1), 142–147. doi: [10.13546/j.cnki.tjyc.2024.01.025](https://doi.org/10.13546/j.cnki.tjyc.2024.01.025).

- Xia, F., Chen, J., Yang, X., Li, X., & Zhang, B. (2023). Financial constraints and corporate greenwashing Rong strategies in China. *Corporate Social Responsibility and Environmental Management*, 30(4), 1770–1781. doi: [10.1002/csr.2453](https://doi.org/10.1002/csr.2453).
- Xie, J., Chen, L., Liu, Y., & Wang, S. (2023). Does fintech inhibit corporate greenwashing behavior? Evidence from China. *Finance Research Letters*, 55, 104002. doi: [10.1016/j.frl.2023.104002](https://doi.org/10.1016/j.frl.2023.104002).
- Yue, J., & Li, Y. (2023). Media attention and corporate greenwashing behavior: Evidence from China. *Finance Research Letters*, 55, 104016. doi: [10.1016/j.frl.2023.104016](https://doi.org/10.1016/j.frl.2023.104016).
- Zeng, W. Q. (2024). A review of corporate ‘greenwashing’ research: Behavioral identification, motivational analysis and economic effects. *Finance and Accounting Monthly*, 45(10), 64–70. doi: [10.19641/j.cnki.42-1290/f.2024.10.010](https://doi.org/10.19641/j.cnki.42-1290/f.2024.10.010).
- Zeng, F. L., Wang, J. T., & Zeng, C. Y. (2025). An optimized machine learning framework for predicting and interpreting corporate ESG greenwashing behaviour. *PLoS One*, 20(3), e0316287. doi: [10.1371/journal.pone.0316287](https://doi.org/10.1371/journal.pone.0316287).
- Zhang, D. (2022). Green financial system regulation shock and greenwashing behaviors: Evidence from Chinese firms. *Energy Economics*, 111, 106064. doi: [10.1016/j.eneco.2022.106064](https://doi.org/10.1016/j.eneco.2022.106064).
- Zhang, D. (2023). Subsidy expiration and greenwashing decision: Is there a role of bankruptcy risk? *Energy Economics*, 118, 106530. doi: [10.1016/j.eneco.2023.106530](https://doi.org/10.1016/j.eneco.2023.106530).
- Zhang, D., Meng, L., & Zhang, J. (2023a). Environmental subsidy disruption, skill premiums and ESG performance. *International Review of Financial Analysis*, 90, 102862. doi: [10.1016/j.irfa.2023.102862](https://doi.org/10.1016/j.irfa.2023.102862).
- Zhang, J. W., Wang, R. Z., & Jiang, Y. (2023b). Digital economy and tunneling behavior of major shareholders: A new perspective of technological governance. *Business and Management Journal*, 45(04), 64–81. doi: [10.19616/j.cnki.bmj.2023.04.004](https://doi.org/10.19616/j.cnki.bmj.2023.04.004).
- Zhang, W., Qin, C., & Zhang, W. (2023c). Top management team characteristics, technological innovation and firm’s greenwashing: Evidence from China’s heavy-polluting industries. *Technological Forecasting and Social Change*, 191, 122522. doi: [10.1016/j.techfore.2023.122522](https://doi.org/10.1016/j.techfore.2023.122522).
- Zhang, C. J., Yang, Y., & Wang, W. T. (2024). A review of corporate greenwash research: Drivers, consequences and governance. *Finance and Accounting Monthly*, 45(04), 28–33. doi: [10.19641/j.cnki.42-1290/f.2024.04.004](https://doi.org/10.19641/j.cnki.42-1290/f.2024.04.004).
- Zhao, T., & Wang, H. (2024). The industry peer effect of enterprise ESG performance: The moderating effect of customer concentration. *International Review of Economics & Finance*, 92, 1499–1525. doi: [10.1016/j.iref.2024.03.018](https://doi.org/10.1016/j.iref.2024.03.018).
- Zhou, K., Jin, X., Li, X., & Tao, Y. (2024). Enhancing sustainable development through effective disclosure: Corporate environmental performance and readability. *Corporate Social Responsibility and Environmental Management*, 31(1), 274–291. doi: [10.1002/csr.2569](https://doi.org/10.1002/csr.2569).

Further reading

- Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1), 685–725. doi: [10.1146/annurev-economics-080217-053433](https://doi.org/10.1146/annurev-economics-080217-053433).
- Castree, N., Hulme, M., & Proctor, J. D. (2018). *Companion to environmental studies*. London: Routledge.
- Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, 41(8), 1381–1411. doi: [10.1002/smj.3152](https://doi.org/10.1002/smj.3152).
- Cicchello, A. F., Cotugno, M., & Foroni, C. (2023). Does competition affect ESG controversies? Evidence from the banking industry. *Finance Research Letters*, 55, 103972. doi: [10.1016/j.frl.2023.103972](https://doi.org/10.1016/j.frl.2023.103972).

- Du, X., Jian, W., Zeng, Q., & Chang, Y. (2018). Do auditors applaud corporate environmental performance? Evidence from China. *Journal of Business Ethics*, 151(4), 1049–1080. doi: [10.1007/s10551-016-3223-6](https://doi.org/10.1007/s10551-016-3223-6).
- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, 24(24), 395–419. doi: [10.1146/annurev-polisci-053119-015921](https://doi.org/10.1146/annurev-polisci-053119-015921).
- Hafezi, M., & Zolfagharinia, H. (2018). Green product development and environmental performance: Investigating the role of government regulations. *International Journal of Production Economics*, 204, 395–410. doi: [10.1016/j.ijpe.2018.08.012](https://doi.org/10.1016/j.ijpe.2018.08.012).
- Hastie, T. J., & Pregibon, D. (2017). Generalized linear models. In *Statistical models in S* (pp. 195–247). Routledge.
- He, F., Guo, X., & Yue, P. (2024). Media coverage and corporate ESG performance: Evidence from China. *International Review of Financial Analysis*, 91, 103003. doi: [10.1016/j.irfa.2023.103003](https://doi.org/10.1016/j.irfa.2023.103003).
- Hu, J., Fang, Q., & Long, W. B. (2023d). Carbon emission regulation, corporate emission reduction incentive and total factor productivity: A natural experiment based on China's carbon emission trading system. *Economic Research Journal*, 58(4), 77–94. doi: [10.1007/s10668-022-02552-8](https://doi.org/10.1007/s10668-022-02552-8).
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *The Journal of Economic Perspectives*, 31(2), 87–106. doi: [10.1257/jep.31.2.87](https://doi.org/10.1257/jep.31.2.87).
- Tan, J., & Tang, Y. (2016). Donate money, but whose? An empirical study of ultimate control rights, agency problems, and corporate philanthropy in China. *Journal of Business Ethics*, 134(4), 593–610. doi: [10.1007/s10551-014-2386-2](https://doi.org/10.1007/s10551-014-2386-2).
- Tidhar, R., & Eisenhardt, K. M. (2020). Get rich or die trying. . . finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal*, 41(7), 1245–1273. doi: [10.1002/smj.3142](https://doi.org/10.1002/smj.3142).
- Valizade, D., Schulz, F., & Nicoara, C. (2024). Towards a paradigm shift: How can machine learning extend the boundaries of quantitative management scholarship?. *British Journal of Management*, 35(1), 99–114. doi: [10.1111/1467-8551.12678](https://doi.org/10.1111/1467-8551.12678).
- Xiang, X., Liu, C., & Yang, M. (2022). Who is financing corporate green innovation?. *International Review of Economics and Finance*, 78, 321–337. doi: [10.1016/j.iref.2021.12.011](https://doi.org/10.1016/j.iref.2021.12.011).

Corresponding author

Miao Yu can be contacted at: yumiao@dufe.edu.cn