Service Science and Management 服务科学和管理, 2026, 15(1), 87-97
Published Online January 2026 in Hans. https://www.hanspub.org/journal/ssem
https://doi.org/10.12677/ssem.2026.151012

Hans汉斯

基于加性提升模型的可解释ESG评级研究

朱珂怡，范　宏*

东华大学旭日工商管理学院，上海

收稿日期：2025年12月4日；录用日期：2025年12月28日；发布日期：2025年12月31日

摘　要

在全球可持续发展趋势和负责任投资理念日益受到重视的背景下，环境(Environmental)、社会(Social)与治理(Governance)，即ESG，已成为衡量企业长期价值与综合风险的关键维度。但ESG评级体系中固有的分歧性与不透明性，极大地限制了其在可持续投资和企业战略中的实用价值。为解决ESG评级中长期存在的模型性能与可解释性的平衡难题，本研究提出并应用加性提升机(Additive Boosting Machine, ABM)构建一个内在可解释的ESG评级模型。ABM作为一种先进的广义加性模型，其核心优势在于能够在保持模型完全透明度的同时，通过Boosting算法学习复杂的非线性模式。该模型将复杂的ESG评分预测任务分解为一系列透明、可视化的组成部分——即单一特征的非线性影响函数与关键的成对特征交互效应函数。这使得模型的每一个决策路径和影响因素都可以被精确追溯和直观理解。结果表明，ABM在保持可解释性的前提下，在MSE、RMSE、MAE、R^2、Gini系数和斯皮尔曼系数六大指标上均优于其他模型。该模型不仅成功识别出关键的非线性规律，还揭示了具有实践指导意义的特征交互作用。因此，本研究不仅为ESG评分领域提供了新的解决方案，也为企业管理者、投资者和监管机构提供清晰、可靠决策依据，有助于推动财务回报与可持续发展目标的统一。

关键词

ESG评级，加性提升机制，可解释性

Interpretable ESG Rating Research Based on Additive Boosting Models

Keyi Zhu, Hong Fan*

Glorious Sun School of Business and Management, Donghua University, Shanghai

Received: December 4, 2025; accepted: December 28, 2025; published: December 31, 2025

Abstract

Against the backdrop of the global sustainable development and the growing emphasis on responsible

*通讯作者。

文章引用: 朱珂怡, 范宏. 基于加性提升模型的可解释 ESG 评级研究[J]. 服务科学和管理, 2026, 15(1): 87-97.
DOI: 10.12677/ssem.2026.151012

investment, Environmental, Social, and Governance (ESG) factors have become critical dimensions for assessing corporate long-term value and comprehensive risk. However, the inherent divergence and opacity within existing ESG rating systems significantly limit their practical utility in sustainable investing and corporate strategy. To address the long-standing challenge of balancing model performance and interpretability in ESG ratings, this study proposes and applies an Additive Boosting Machine (ABM) to construct an intrinsically interpretable ESG rating model. As an advanced type of Generalized Additive Model, the ABM's core strength lies in its ability to learn complex non-linear patterns through boosting algorithms while maintaining complete model transparency. The model decomposes the complex task of ESG score prediction into a series of transparent, visualizable components—namely, non-linear shape functions for individual features and key pairwise feature interaction effects. This allows every decision path and contributing factor within the model to be precisely traced and intuitively understood. Our results demonstrate that the ABM outperforms other benchmark models across six key metrics—MSE, RMSE, MAE, R^2, Gini coefficient, and Spearman's rank correlation coefficient—while retaining interpretability. Furthermore, the model not only identifies crucial non-linear relationships but also reveals feature interactions with significant practical implications. Consequently, this research provides a novel solution for the field of ESG scoring and offers clear, reliable decision-support for corporate managers, investors, and regulators, thereby facilitating the alignment of financial returns with sustainable development goals.

Keywords

ESG Rating, Additive Boosting Machine (ABM), Interpretability

1. 引言

在全球可持续发展议程日益深化的背景下，环境、社会与治理(ESG)理念已从早期的企业社会责任(CSR)思想[1]和“三重底线”理论演变为全球资本市场配置与企业战略决策的核心要素。2004 年联合国全球契约组织发布的《在乎者赢》报告，首次系统性地将 ESG 确立为评估企业长期价值的核心分析框架。ESG 表现不再局限于非财务层面的考量，大量实证研究已证实其与企业财务绩效、风险管理和市场估值之间存在紧密联系[2]。卓越的 ESG 表现有助于企业降低资本成本、提升投资效率和促进绿色创新。

然而，尽管 ESG 评级在引导资本配置和促进企业可持续转型方面扮演着关键角色，但当前主流的评级体系却普遍存在两大核心局限，严重削弱了其作为决策依据的有效性与公信力。第一个局限是评级结果的“分歧性”。实证研究表明，不同评级机构对同一家企业的 ESG 评价结果相关性极低[3]，其根源在于各机构在指标的范围界定、数据衡量和权重分配上存在显著的方法论异质性[4]。这种“评级混淆”现象不仅为投资者决策带来了极大困扰，也引发了负面的经济后果。第二个局限是评级过程的“不透明性”。绝大多数评级机构将其详细的方法论体系视为商业机密[5]，导致其内部决策逻辑不透明，外界均难以审视、复现和验证其评分依据，这从根本上损害了评级体系的权威性与问责性。

为应对传统主观评价方法的局限性，学术界与业界逐渐转向数据驱动的量化建模方法，以期提升 ESG 评级的客观性与预测能力。然而，这一范式转变在缓解主观性问题的同时，也引发了一个更为棘手的挑战——即模型性能与可解释性的平衡难题[6]。一方面，以梯度提升决策树(GBDT)为代表的高性能集成模型展现出卓越的预测精度[7]，但其复杂的内部机制恰恰加剧了决策的不透明性。另一方面，传统的线性

模型虽具备内在可解释性，但其模型表达能力有限，往往难以捕捉ESG数据中普遍存在的复杂关联动态，导致预测性能不足。

为解决不透明问题，可解释人工智能领域形成了两条主要技术路径。一种是“事后解释”，即使用SHAP [8]或LIME [9]等模型无关工具，在不透明模型预测之后尝试提供近似的归因解释。许多近期的研究也采用了这种“不透明模型 + 事后解释”的模式[10]。一种是构建内生可解释模型，通过对其结构施加特定的约束(例如，线性、单调性、加性等)，使其内在的决策逻辑本身就是透明和易于理解的。其中，广义加性模型(Generalized Additive Models, GAMs) [11]是此类模型的杰出代表。使用者可以直观地通过可视化每个形状函数和交互项，来精确理解每个因素如何独立或协同地影响最终预测[12]。

因此，本研究选择应用加性提升机(Additive Boosting Machine, ABM)作为核心建模框架。ABM是一种先进的、内在可解释的机器学习模型，它通过Boosting算法的迭代学习，能够将复杂的预测函数精确地解构为一系列可加和的、可视化的组件，即单一ESG特征的非线性边际效应以及关键特征之间的成对交互效应。本文的主要研究贡献体现在：第一，在方法论层面，本研究在ESG评级领域引入并验证了一种“事前可解释”的高性能建模范式，正面回应了模型性能与可解释性的平衡难题，为解决当前ESG评级不透明的核心难题提供了可行的、逻辑严谨的方法论新路径。第二，在实证层面，本研究通过实证检验，证明ABM模型能够在保持透明度的前提下，实现与主流模型(如GBDT)相当的预测精度，有力地驳斥了“高性能必然以牺牲可解释性为代价”的传统观念。第三，在洞察层面，本研究利用ABM的内在可解释性，得以精确挖掘和可视化ESG因素与评级之间普遍存在的、复杂的非线性关系和特征交互效应(例如“碳效率”与“雇员增长”的协同作用)。第四，在实践层面，本研究构建的模型不仅是一个预测工具，更是一个兼具诊断与分析功能的透明系统。它能够为企业管理者提供清晰、可量化的ESG改进路径；为投资者提供可审计、可验证的决策依据；同时也为监管机构提供了透明的评估工具，从而为发挥ESG数据在引导可持续资本配置中的核心功能提供更可靠的工具支持。

2. 模型及理论基础

2.1. ABM

ABM的核心思想源于广义加性模型，它假设预测函数$g(x)$可以被分解为一系列独立(或低阶交互)函数项的和。ABM通过Boosting算法来迭代地学习这些函数项。一个包含成对交互的ABM模型，其预测函数$g(x)$的一般形式定义如下：

$$g(x)=\alpha_0+\sum_{j=1}^{P}f_j(x_j)+\sum_{k\neq l}f_{kl}(x_k,x_l) \tag{1}$$

其中，α_0是模型的全局截距项，$f_j(x_j)$是第j个特征x_j的主效应函数。它捕捉了该特征对预测结果的、可能是非线性的独立贡献。$f_{kl}(x_k,x_l)$是特征x_k和x_l之间的成对交互效应函数。它捕捉了这两个特征同时变化时，对预测结果产生的、超越它们各自独立效应之和的协同或拮抗效应。

ABM采用Boosting框架来学习每一个f_j和f_{kl}。Boosting是一种加法模型，它通过迭代地拟合前一轮模型的残差来逐步构建一个强学习器。在ABM中，学习过程通常采用一种特征轮询策略，逐一地更新每一个函数组件f_j和f_{kl}，直到模型收敛。

以更新主效应函数f_j为例，在第t轮迭代中，模型的目标是找到一个更新项h_j来最小化当前残差：首先，计算不包含的模型f_j预测值$g_{-j}(x_i)$。当前残差$r_i^{(t)}$被定义为：

$$r_i^{(t)}=y_i-g_{-j}(x_i)=y_i-\left(\alpha_0+\sum_{k\neq j}f_k(x_{ik})+\sum_{m\neq n}f_{mn}(x_{im},x_{in})\right) \tag{2}$$

接着需要拟合残差，模型的目标是找到一个函数 f_j 来拟合这些残差，同时施加平滑性约束以防止过拟合。这通过最小化一个带正则化的损失函数来实现：

$$\min_{f_j}\sum_{i=1}^{N}\left(r_i^{(t)}-f_j\left(x_{ij}\right)\right)^2+\lambda\Omega\left(f_j\right) \tag{3}$$

其中 $\Omega\left(f_j\right)$ 是一个正则化项，λ 是控制平滑强度的超参数。在实践中，连续特征 x_j 首先被分箱，模型学习的是每个箱子对应的权重 β_k。函数 $f_j\left(x_j\right)$ 被建模为：

$$f_j\left(x_j\right)=\sum_{k=1}^{K}\beta_k\cdot I\left(x_j\in\mathcal{B}_k\right) \tag{4}$$

其中 $\mathcal{B}_k$ 是第 k 个箱，$I(\cdot)$ 是指示函数。通过对残差 $r^{(t)}$ 进行带岭回归来估计这些权重 β_k，并以一个较小的学习率 η 更新 f_j。交互项 f_{kl} 的学习过程与主效应类似，但它是在一个二维的特征空间上拟合残差。模型会迭代轮询所有自动检测出的特征对，直到 α_0、所有的 f_j 和 f_{kl} 都达到收敛。

2.2. ABM 的可解释性原理

ABM 模型的一个核心优势在于其内在的可解释性，这种可解释性源于其构建的透明加性结构。这使得本研究能够从全局层面理解模型的整体行为，从局部层面剖析对单个样本的预测归因，并对特征间的交互效应进行直观的可视化。

在全局解释层面，ABM 允许评估每个 ESG 特征对最终评分的总体影响。首先，可以计算每个特征的特征重要性。这通过量化其在 ABM 模型中主效应函数 $f_j\left(x_j\right)$ 的贡献大小来实现。一种常用的计算方法是该函数在所有训练样本 $x_{ij}\left(i=1,\ldots,N\right)$ 上的平均绝对值与整体均值的偏差：

$$I\left(x_j\right)=\frac{1}{N}\sum_{i=1}^{N}\left|f_j\left(x_{ij}\right)-E\left[f_j\left(x_j\right)\right]\right| \tag{5}$$

其中 $E\left[f_j\left(x_j\right)\right]$ 是 f_j 在特征 x_j 的经验分布上的期望。通过对所有特征的重要性进行排序，可以识别出对 ESG 评分最具决定性的驱动因素。其次，可以绘制每个特征 x_j 的全局形态函数 $f_j\left(x_j\right)$，即 x_j 与其对应边际贡献 $f_j\left(x_j\right)$ 之间的关系图。这可以直观地揭示当特征 x_j 的值变化时，预测 ESG 评分的平均变化趋势，这对于理解特定 ESG 因素与最终评分之间是否存在非线性关系至关重要。

在特征交互分析层面，ABM 利用可视化技术来检验显著的成对效应。学习到的交互效应函数 $f_{kl}\left(x_k,x_l\right)$ 可以生成二维热力图，从而清晰地揭示特征对之间的非加性关系，避免了在 ESG 评估中过度简化的风险。通过上述全局及交互效应两个层面的综合解释机制，ABM 从一个单纯的预测工具，转变为一个兼具分析功能的透明的框架。

3. 基于 ABM 算法的 ESG 评级模型

当前 ESG 评级研究的核心痛点在于：复杂的机器学习模型以牺牲透明度为代价换取高性能，而简单的可解释模型又难以捕捉复杂的 ESG 动态。本研究的内容正是为了解决这一难题，致力于构建一个兼具高预测精度与完全可解释性的 ESG 评级框架。ABM 作为一种“事前可解释”模型，其核心内容是将复杂的预测分解为可加和、可独立分析的组件。本章将重点介绍如何将该模型具体应用于 ESG 评级任务，包括所采用的数据处理流程、实验框架以及模型的训练与评估机制。

为实证检验 ABM 模型在 ESG 评级任务中的有效性，本研究构建了一个系统化的实验流程。该流程的核心是利用 ABM 模型，针对企业的“综合 ESG 总分”构建一个统一的预测模型。本研究的实验流程如下：

(1) 数据准备和指标选取开始。鉴于原始 ESG 数据(如 Refinitiv)包含大量底层指标，为减轻多重共线

性和噪声干扰，本研究采用了一套系统化的特征选择方法，包括剔除高缺失值和低方差特征、移除强相关特征，并最终采用 Lasso 回归(L1 正则化)为该评级任务自动筛选出最具信息量的核心指标子集。

(2) 模型构建。针对“综合 ESG 总分”这一预测目标，本研究构建了一个 ABM 模型。为保证模型评估的客观性与可靠性，本研究首先将完整的 ESG 数据集严格划分为三个互不重叠的子集，其中训练集，用于模型的迭代学习；验证集，不参与训练，专门用于模型调优；测试集用于评估模型的最终泛化性能。

(3) 模型的训练与优化。ABM 的训练是一个迭代的 Boosting 过程，模型在训练集上通过“特征轮询”策略逐步拟合损失函数。为防止模型过度迭代而导致的“过拟合”，本研究在训练过程中引入了早停机制。具体而言，模型在训练时会持续监控其在验证集上的性能 MSE 指标，一旦验证集上的性能在连续多轮内不再提升，训练将自动停止，并选用在验证集上性能最佳的模型。同时，ABM 模型的性能还受一系列关键超参数的控制，如学习率、最大分箱数和交互项数量等。本研究采用随机搜索等方法，在预定义的超参数空间内进行探索，对于每一组超参数组合，模型均按照包含早停机制的流程进行训练，并最终选择在验证集上性能最优的那一组超参数，作为 ABM 模型的最终配置。

通过这一完整的实验流程，本研究构建了一个能够客观、透明地评估企业 ESG 表现，并同时具备高预测精度和泛化能力的 ABM 评级模型。

4. 性能实验与结果分析

4.1. 数据描述与处理

本研究的实证分析所采用的数据来源于国际领先的路孚特(Refinitiv) Eikon ESG 数据库。该数据库覆盖了全球主要上市公司，提供了全面且细颗粒度的 ESG 指标。本研究选取了 2002 年至 2023 年的年度数据，并以企业的“综合 ESG 总分”作为预测目标。在数据预处理阶段，本研究首先对原始数据进行了严格的清洗，包括剔除重复样本、缺失值比例超过 70%的特征以及相关性绝对值大于 0.95 的冗余特征。在此基础上，本研究采用 Lasso 回归进行二次特征选择，为“综合 ESG 总分”这一预测任务自动筛选出最具预测价值的核心指标子集。最终获得 89,066 个样本个数，包含 79 个特征，关于 ESG 评分数据集的详细信息见表 1。最终，本研究将处理后的数据集按照训练集:验证集:测试集 = 70%:15%:15%的比例进行严格划分，以确保模型评估的客观性与可靠性。

Table 1. Description of the dataset
表 1. 数据集信息

数据集	数据量	特征数
e_data	89,066	25
s_data	108,756	28
g_data	108,808	26
esg_data	89,066	79

4.2. 对比模型及性能评估指标

为全面评估 ABM 模型的性能，本研究选取了一系列具有代表性的机器学习模型作为对比基准。这些模型覆盖了从简单到复杂的不同算法范式，主要包括：(1) 作为基础性能参照的经典线性模型，如普通最小二乘线性回归(LR)和广义线性模型(GLM)；(2) 其他广义加性模型(GAMs)，如 LinearGAM、ExpectileGAM、GammaGAM，以对比 ABM 的 Boosting 实现在捕捉非线性关系上的优势；(3) 高性能的

集成学习模型(Ensemble Learning)，如梯度提升决策树(GBDT)和自适应提升(AdaBoost)，它们被视为不透明模型中的性能标杆，是检验 ABM 是否在不牺牲性能的情况下实现可解释性的关键；(4) 经典的决策树(DT)模型；(5) 其他先进的可解释性与混合模型，如 GBDT-PL、DeepGBM、EGM、NAM。

本研究采用多维度的性能指标来综合评估所有模型。这些指标可分为三类：第一类是预测误差指标，包括均方误差(MSE)、均方根误差(RMSE)和平均绝对误差(MAE)，用于衡量预测值与真实值的偏离程度，其值越小越好；第二类是拟合优度指标，即判定系数(R-squared)，用于衡量模型对数据总变异的解释能力，其值越接近 1 越好；第三类是排序能力指标，包括归一化基尼系数(Normalized Gini, N-Gini)和斯皮尔曼等级相关系数(Spearman Correlation, SPCC)，用于评估模型对企业 ESG 表现排序的准确性，其值越高越好。

4.3. 实验结果分析

(1) 综合性能对比

表 2 总结了 ABM 模型在 ESG 评级测试集上六个关键性能评价指标的详细结果。在测试集上的实证结果清晰地表明，本研究采用的 ABM 模型在该数据集上的表现优异，在六个评价指标上取得最佳结果。最关键的是，ABM 是在保持模型结构完全透明、“事前可解释”的前提下取得这一高性能的，这是其独特的优势所在。

Table 2. Comparison of model performance
表 2. 性能对比表

Algorithm	MSE	RMSE	MAE	R2	N-Gini	SPCC
GBDT	0.0081	0.09	0.0716	0.7877	0.8883	0.8856
DecisionTree	0.0144	0.1200	0.0878	0.6223	0.8108	0.8060
LinearRegression	0.0084	0.0919	0.0727	0.7786	0.8832	0.8804
LinearGAM	0.0082	0.0904	0.0716	0.7856	0.8875	0.8848
ExpectileGAM	0.0082	0.0904	0.0716	0.7856	0.8874	0.8847
GammaGAM	0.0101	0.1005	0.0790	0.7351	0.8789	0.8762
AdaBoost	0.0123	0.1107	0.0906	0.6786	0.8573	0.8519
GLM	0.0088	0.0940	0.0749	0.7684	0.8785	0.8755
GBDT-PL	0.0081	0.0900	0.0716	0.7877	0.8883	0.8856
DeepGBM	0.0097	0.0985	0.0766	0.7452	0.8711	0.8685
EGM	0.0112	0.1056	0.0865	0.7073	0.8796	0.8766
NAM	0.0107	0.1034	0.0825	0.7195	0.8515	0.8486
ABM	**0.0073**	**0.0853**	**0.0675**	**0.8092**	**0.9002**	**0.8978**

(2) 全局可解释性结果

如图 1 所示，对于预测 ESG 评分的模型，其最重要的前三个特征依次为 CorpAward、CO2RevRt 和 WaterRev。这一结果清晰地表明，企业的整体 ESG 表现并非由某一个支柱的绩效所主导，而是由一组跨越 E、S、G 三个维度的、最具代表性的关键事件和绩效指标共同决定。最重要的特征是来自社会维度的

特征 CorpAward，其具体含义为“公司是否因其社会、道德、社区或环境活动/表现而获得过外部重要奖项或认证”。它表明，来自独立第三方的、对企业整体可持续发展实践的正面认可和荣誉，是预测其最终综合 ESG 评分的最强信号。值得注意的是，这一特征的重要性可能受到公司规模和行业的影响。通常，大型企业或消费品行业的公司拥有更多资源进行品牌建设，也更倾向于参与各类评奖活动以提升公众形象，这可能导致该特征在这些公司样本中呈现出更高的预测权重。在未来的研究中，通过分层分析检验该特征重要性的稳健性将是一个有价值的方向。紧随其后的是两个来自环境维度的效率指标：CO2RevRt (单位营收的二氧化碳及当量排放)和 WaterRev (单位营收的总取水量)。这两个指标的重要性凸显了企业的运营生态效率是其整体可持续发展能力的核心基石。CO2RevRt 衡量企业的碳效率，直接关联全球最重要的环境议题——气候变化；而 WaterRev 则衡量水资源利用效率，应对日益严峻的全球水资源挑战。模型将这两个效率指标置于高位，表明评估体系高度重视企业能否以更少的环境资源消耗创造更多的经济价值，即实现经济增长与环境影响的“解耦”。这一发现也与当前全球对资源密集型行业(如制造业、能源业)环境风险的普遍担忧相符。对于这些行业的企业而言，运营效率的提升不仅是成本控制问题，更是其 ESG 评级和长期生存能力的核心。

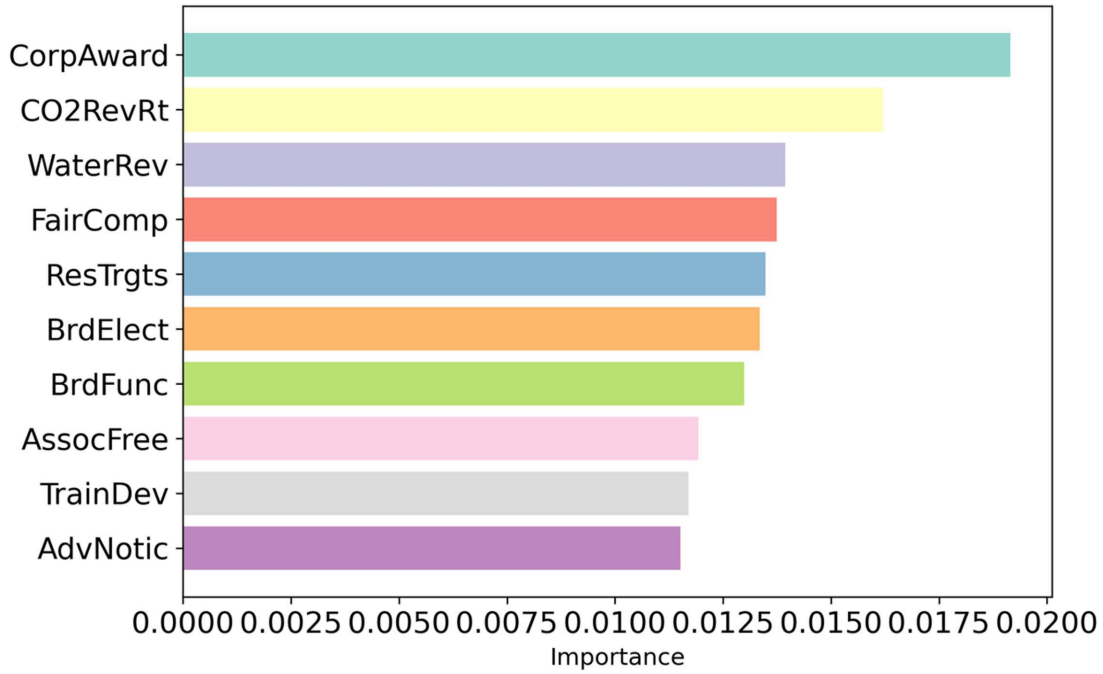
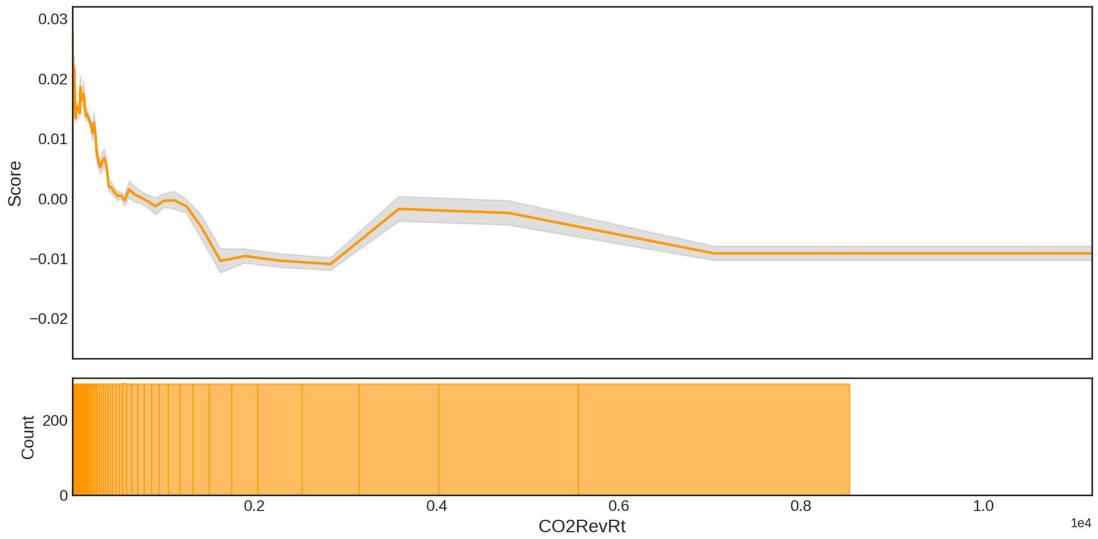


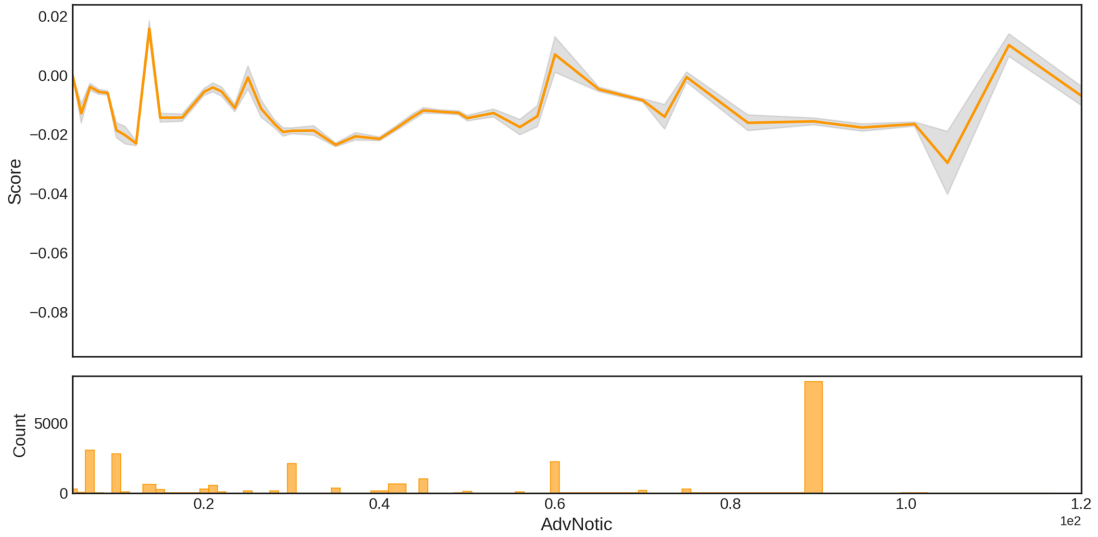
Figure 1. Plot for global model interpretability
图 1. 全局可解释性图

(3) 单效应特征效应分析

图 2(a)表明企业的二氧化碳减排率(CO2RevRt)对整体 ESG 绩效有强烈的负面影响，随着企业碳排放强度的增加，其对整体 ESG 评分的贡献值持续下降。这有力地证实，企业的碳效率不仅是衡量其环境表现的核心，更是决定其整体可持续发展形象和综合 ESG 评价的关键负向驱动因素。图 2(b)表明公司给予股东提交提案的窗口期较长，截止日离大会较近会对 ESG 得分产生不利影响。一个较短的提前通知期要求，意味着公司更能保障股东的权利，鼓励股东积极参与公司事务，这本身就是稳健、透明和负责任的公司治理的强烈信号，因此对整体 ESG 形象有显著的正面提升作用。



(a) CO2RevRt 的效应曲线



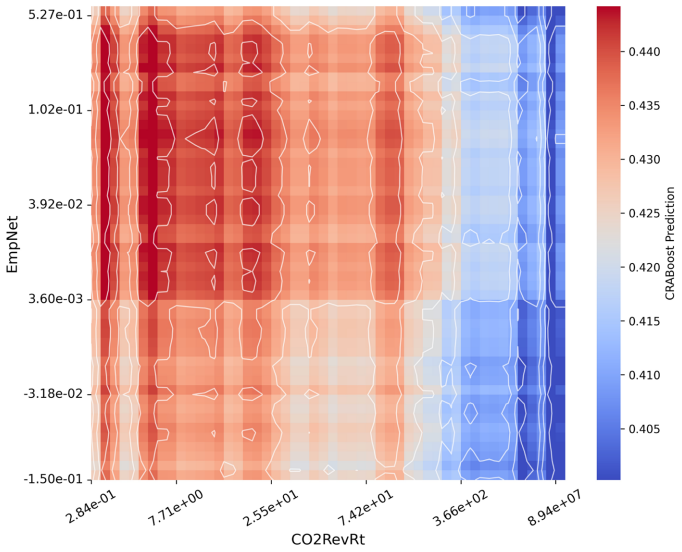
(b) AdvNotic 的效应曲线

Figure 2. Key single-feature effect curves
图 2. 关键单特征效应曲线

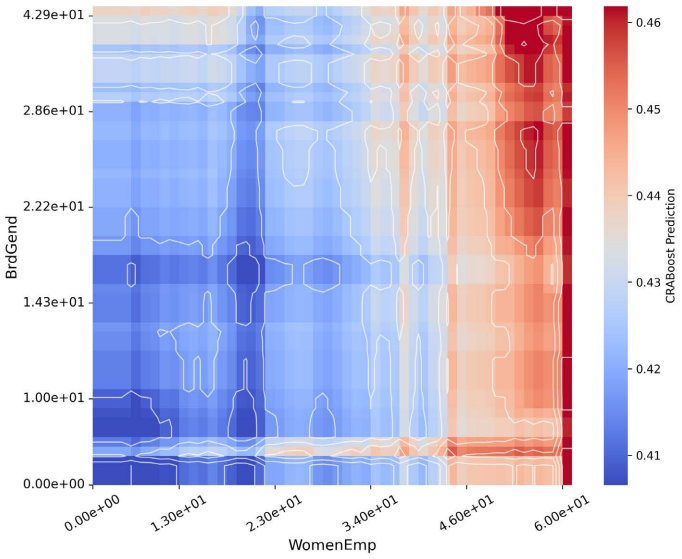
(4) 特征交互效应分析

如图 3(a)所示，该图展示了特征 CO2RevRt 和 EmpNet 之间的交互效应对 ESG 评分的边际贡献。图中颜色最暖、表示最强正向交互效应的区域，清晰地出现在左上角。该区域对应于企业同时实现高 EmpNet (即快速的岗位增长)和低 CO2RevRt (即卓越的碳排放效率)的情况。当一个企业的快速扩张并非以牺牲环境为代价，而是伴随着高效的低碳运营时，其获得的整体 ESG 评价提升，要大于“创造就业”和“低碳高效”这两个单项优点贡献的简单加总。这种协同效应揭示了高质量发展的内涵：增长的可持续性比增长本身更为关键。然而，需要警惕潜在的混杂因素，例如，技术驱动型的高增长企业可能天然具备更高的能源效率，这使得“低碳”与“增长”同时发生。尽管如此，该交互作用清晰地为企业指明了一条路径：在追求业务扩张时，必须将环境影响管理置于战略核心地位。同时本研究发现当 CO2RevRt 的值足

够高时，对模型起主要影响效果，即普遍呈现负面效应。如图 3(b)所示，该图展示了特征 WomenEmp 和 BrdGend 之间的交互效应对 ESG 评分的边际贡献。这揭示了模型对“系统性性别平等(Systemic Gender Equality)”的高度认可。当一家公司不仅在员工队伍中实现了较高的性别均衡，其最高决策层——董事会——也同样展现出良好的性别多元化时，模型认为这是一种从上至下、贯穿始终的、真实的 D&I 承诺。这一发现支持了“关键少数”理论(critical mass theory)，即当女性在决策层达到一定比例后，其影响力才能得到实质性发挥，从而推动整个组织层面的性别平等文化。模型捕捉到的这种交互效应，超越了对单一指标的孤立考察，印证了在评估企业 D&I 实践时，需同时考量其员工层面的广度与治理层面的深度。



(a) CO2RevRt 和 EmpNet 的交互效应



(b) WomenEmp 和 BrdGend 的交互效应

Figure 3. Key feature interaction effects diagram
图 3. 关键特征交互效应图

5. 局限性与未来展望

5.1. 研究局限性

尽管本研究提出的基于 ABM 的可解释 ESG 评级模型在性能和透明度上取得了显著成果，但仍存在若干局限性：

(1) 数据来源的单一性：本研究的数据完全来源于路孚特(Refinitiv)数据库。虽然该数据源具有较高的权威性和覆盖度，但未能纳入其他主流评级机构(如 MSCI、Sustainalytics)的数据进行交叉验证。鉴于不同评级机构间存在“评级分歧”，依赖单一数据源可能无法完全捕捉 ESG 表现的全貌。

(2) 时间动态性的忽略：本研究采用截面数据构建模型，将不同年份的样本汇集处理，未能充分考虑企业 ESG 表现随时间演变的动态特征和路径依赖性。一个更理想的模型应能捕捉到企业 ESG 策略改进或恶化的轨迹。

(3) 内生性与因果推断的挑战：尽管模型揭示了关键特征与 ESG 得分之间的强相关性和交互作用，但作为一种预测模型，其本质上仍是相关性分析，难以做出严格的因果推断。例如，获得奖项(CorpAward)与高 ESG 得分之间可能存在双向因果关系。

5.2. 未来展望

针对上述局限，未来的研究可在以下方向深化拓展：

(1) 构建多源融合的评级框架：整合来自多个评级机构、政府监管报告、非政府组织评估等多源数据，利用 ABM 等模型分析不同来源评级分歧的驱动因素，构建一个更全面、更稳健的评级体系。

(2) 引入时序模型：将 ABM 与时间序列分析方法(如 LSTM、GRU)相结合，构建动态可解释模型，以捕捉企业 ESG 表现的演变趋势，并预测其未来走向。

(3) 探索非结构化数据：利用先进的自然语言处理(NLP)技术，将新闻舆论、社交媒体讨论、企业社会责任报告文本等非结构化数据中蕴含的情感、主题和争议点量化为特征，融入模型，以捕捉更实时、更动态的 ESG 信号。

通过上述改进，未来的研究有望构建出更加精准、透明且具备动态诊断能力的 ESG 评估工具，为实现可持续金融目标提供更强大的支持。

6. 总结

在全球可持续发展趋势下，ESG 评级在投资者决策、企业战略制定以及监管政策的形成中的重要性愈加凸显。本文提出的基于加性提升模型的可解释 ESG 评级模型，旨在应对传统 ESG 评分方法在无法平衡性能和可解释性的局限性。实验结果表明，ABM 模型在 ESG 数据集上的表现均优于传统统计方法和其他机器学习模型，尤其在 MSE、MAE 等关键指标上展现出显著优势。这一成果不仅验证了 ABM 模型的有效性，也为 ESG 评分领域提供了新的思路和方法，在实际应用中能够为投资者决策以及企业政策制定提供更为精准的支持。

未来的研究可以进一步探索 ABM 在其他 ESG 数据集、企业风险评估等场景中的应用。其次，可以尝试将更丰富的非结构化数据，如实时新闻、社交媒体文本，通过先进的自然语言处理技术融入模型，以捕捉更动态的 ESG 信号。

参考文献

[1] Carroll, A.B. (1991) The Pyramid of Corporate Social Responsibility: Toward the Moral Management of Organizational Stakeholders. *Business Horizons*, **34**, 39-48. https://doi.org/10.1016/0007-6813(91)90005-g

[2] Friede, G., Busch, T. and Bassen, A. (2015) ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies. *Journal of Sustainable Finance & Investment*, **5**, 210-233. https://doi.org/10.1080/20430795.2015.1118917

[3] Berg, F., Kölbel, J.F. and Rigobon, R. (2022) Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance*, **26**, 1315-1344. https://doi.org/10.1093/rof/rfac033

[4] Christensen, D.M., Serafeim, G. and Sikochi, A. (2021) Why Is Corporate Virtue in the Eye of the Beholder? The Case of ESG Ratings. *The Accounting Review*, **97**, 147-175. https://doi.org/10.2308/tar-2019-0506

[5] Dell'Erba, M. and Doronzo, M. (2023) Sustainability Gatekeepers: ESG Ratings and Data Providers. Social Science Research Network. https://papers.ssrn.com/abstract=4470672

[6] Rudin, C. (2019) Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, **1**, 206-215. https://doi.org/10.1038/s42256-019-0048-x

[7] Drempetic, S., Klein, C. and Zwergel, B. (2019) The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings under Review. *Journal of Business Ethics*, **167**, 333-360. https://doi.org/10.1007/s10551-019-04164-1

[8] Lundberg, S.M. and Lee, S.I. (2017) A Unified Approach to Interpreting Model Predictions. arXiv: 1705.07874

[9] Ribeiro, M.T., Singh, S. and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the* 22*nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 1135-1144. https://doi.org/10.1145/2939672.2939778

[10] Zhang, M., Shen, Q., Zhao, Z., Wang, S. and Huang, G.Q. (2025) Optimizing ESG Reporting: Innovating with E-BERT Models in Nature Language Processing. *Expert Systems with Applications*, **265**, Article ID: 125931. https://doi.org/10.1016/j.eswa.2024.125931

[11] Hastie, T.J. (1992) Generalized Additive Models. In: Chambers, J.M. and Hastie, T.J., Eds., *Statistical Models in S*, Routledge, 59.

[12] Nori, H., Jenkins, S., Koch, P., *et al.* (2019) InterpretML: A Unified Framework for Machine Learning Interpretability. arXiv: 1909.09223v1 https://arxiv.org/abs/1909.09223v1