Vilnius University Press

Content lists are available at Vilnius University Press

Transformations In Business & Economics
2026, Vol. 25 No 2 (68), pp.235-260.

ISSN 1648–4460 | eISSN 2538-872X
https://doi.org/10.15388/Tibe.2026.25.2.21

To what extent can spreadsheets shape sustainability? A machine learning approach to ESG score prediction

Hussam Musa
E-mail: hussam.musa@umb.sk
ORCID: https://orcid.org/0000-0002-4492-8770
Affiliation: Faculty of Economics, Matej Bel University, Banská Bystrica, Slovakia
ROR: https://ror.org/016e5hy63

Frederik Rech
E-mail: frederikrech@gmail.com
ORCID: https://orcid.org/0000-0002-6392-4951
Affiliation: Faculty of Economics, Matej Bel University, Banská Bystrica, Slovakia
ROR: https://ror.org/016e5hy63

Zdenka Musová
E-mail: zdenka.musova@umb.sk
ORCID: https://orcid.org/0000-0002-1067-8291
Affiliation: Faculty of Economics, Matej Bel University, Banská Bystrica, Slovakia
ROR: https://ror.org/016e5hy63

Janka Grofčíková
E-mail: janka.grofcikova@umb.sk
ORCID: https://orcid.org/0000-0001-5050-4715
Affiliation: Faculty of Economics, Matej Bel University, Banská Bystrica, Slovakia
ROR: https://ror.org/016e5hy63

Annotation. This study examines whether corporate ESG ratings can be predicted using multi-year lagged financial indicators. The objective is to evaluate the extent to which historical financial information explains ESG performance among Slovak manufacturing firms. The analysis uses a sample of 974 Slovak manufacturing firms with ESG ratings for 2023 and financial data from 2018–2022. The primary model applies XGBoost with recursive feature elimination and SHAP analysis, while a PCA-based one-vs-rest XGBoost model is used as a robustness check. The findings show that historical financial indicators provide meaningful but incomplete information about ESG performance. Leverage, liquidity, debt-servicing capacity, firm age, and tax-related indicators emerge as important predictors. Older lagged variables also remain significant, suggesting that ESG outcomes reflect longer-term financial patterns. The alternative model improves prediction, particularly for the Environmental and Governance pillars, whereas the aggregate ESG score remains more difficult to classify. The models predict the middle ESG rating category most reliably, while performance for extreme categories is limited by class imbalance and the ordinal structure of ESG scores. The study concludes that financial data can support ESG prediction, especially at the pillar level, but cannot replace qualitative and ESG-specific information.

Keywords: ESG, financial indicators, machine learning, extreme gradient boosting, manufacturing companies, Slovakia

JEL classification: C1, C4, G30, M21.

Introduction

Environmental, Social, and Governance (ESG) considerations have moved to the forefront of business strategy and reporting worldwide. With heightened stakeholder expectations and stringent regulation requirements, companies have incorporated operations with best practices in governance and sustainability to a higher degree than before (Kubalek, Kudej, 2025). The role of ESG has grown over the

Submitted on March 2026. Accepted on May 2026.

past decade from voluntary best practices to fundamental elements in the management of risks and long-term value creation. Regulatory requirements, including the European Union's Sustainable Finance Disclosure Regulation (SFDR) and the Corporate Sustainability Reporting Directive (CSRD), have accelerated the implementation of ESG practices through requiring transparent and detailed sustainability reporting (Friede *et al.,* 2015; Petrakova, Karacsony; 2024).

The heightened interest in ESG is further fuelled by the promise of better financial performance. A meta-analysis of over 2,000 empirical research papers concluded that approximately 90% had a non-negative relationship between ESG activities and firms' financial performance and that the majority had a positive relationship (Friede *et al.,* 2015). The discovery dispels the traditional notion that profitability and sustainability come with a trade-off and instead suggests that good ESG performance enhances risk management, operational efficiency, and stakeholder trust. Consequently, companies in all industries, including financial and tech to heavy industry, are now adding ESG metrics to models (D'Amato *et al.,* 2022).

In the earlier ESG literature, the predominant focus was on how a firm's ESG performance might affect financial metrics, risk profiles, or broader market valuations. More recently, however, scholars began examining the opposite direction, namely whether it is possible to predict ESG ratings themselves using nothing more than firm-level financial characteristics and external signals. These predictors encompass a range of data sources, including financial statement metrics and sector-specific indicators (D'Amato *et al.,* 2022; Jiang, 2024, Cainelli *et al.* (2020), media sentiment derived through natural language processing techniques (Aue *et al.,* 2025), past ESG ratings (Taskin *et al.,* 2025), as well as integrative models that combine past ESG performance with corporate governance and financial data (Lin, Hsu, 2023). Additionally, the adoption of machine learning algorithms has driven progress in this area by improving predictive accuracy and offering new insights into the interplay between financial health and ESG commitments.

Despite the rapid growth of ESG-related research, important gaps remain in understanding whether ESG outcomes can be inferred from conventional financial information alone. Existing literature has mainly followed two directions. One stream examines how ESG performance affects firm value, financial performance, risk, and stakeholder outcomes (Friede *et al.,* 2015; Eccles *et al.,* 2014; Servaes, Tamayo, 2013). Another stream predicts ESG scores using broader information sets, including financial and market variables, sector-level indicators, media sentiment, past ESG ratings, and combined ESG-financial data (Cainelli *et al.,* 2020; D'Amato *et al.,* 2022; Jiang, 2024; Lin, Hsu, 2023; Aue *et al.,* 2025; Taskin *et al.,* 2025). However, comparatively less attention has been devoted to whether multi-year lagged financial statement indicators alone contain reliable signals about current ESG ratings. This gap is particularly relevant in smaller Central and Eastern European economies, where ESG disclosure practices are still developing and sustainability information may remain incomplete, inconsistent, or affected by symbolic reporting and greenwashing concerns (Petrakova, Karacsony, 2024; Kubalek, Kudej, 2025; Testa *et al.,* 2018). Identifying whether historical financial indicators can support ESG prediction therefore offers useful insights for investors, regulators, lenders, and rating users seeking complementary approaches to ESG assessment when detailed non-financial information is limited.

This study contributes to the ESG prediction literature in three ways. First, while prior ESG prediction studies have often relied on recent financial indicators, market-based variables, past ESG ratings, or textual information, this article focuses specifically on whether multi-year lagged financial data can predict current ESG ratings. By using financial indicators from 2018 to 2022 to predict 2023 ESG scores, the study captures the possibility that ESG performance reflects accumulated financial capacity, managerial discipline, and long-term strategic choices rather than only short-term financial conditions. Second, the analysis is conducted on a sample of Slovak manufacturing firms, a setting in which ESG disclosure and sustainability reporting are still developing, making financial-statement-based prediction particularly relevant. Third, the study combines predictive machine learning with interpretability by applying XGBoost, recursive feature elimination, and SHAP analysis, allowing us not only to evaluate classification performance but also to identify which lagged financial indicators contain the strongest signals for ESG, E, S, and G outcomes.

The remainder of the study is structured as follows. Section 1 reviews the relevant literature on ESG performance and ESG prediction models. Section 2 methodology and analysis. Section 3 presents the empirical results and model evaluation. Section 4 discusses the main findings in relation to prior literature. Section 5 presents the theoretical and practical implications of the study. Section 6 outlines the limitations and directions for future research. Finally, Section concludes the study.

1. Literature Review

ESG implementation is often initiated through a combination of institutional pressures, stakeholder expectations, and strategic alignment. Institutional theory posits that companies undertake ESG practices to satisfy regulation requirements and respond to coercive pressures from government and regulation authorities (Delmas, Toffel, 2004). The pressures affect the internal and external organizational settings and compel companies to conform to societal expectations and industry conventions. Companies are not only driven by compliance with the law but also by reputation and the need to establish legitimacy in the operating context.

Empirical evidence further supports the role played by external drivers in encouraging ESG adoption to a considerable degree. Barko *et al.* (2022) observe the role played by investor activism in serving as a strong impetus and bringing to the forefront the role played by investors in demanding higher transparency and accountability in business operations. Similarly, Servaes, Tamayo (2013) observe the way consumer demand for ethically focused brands provides market-driven incentives to companies to enhance ESG performance. Further, employee retention becomes a strategic imperative with Jones *et al.* (2014) stating that sustainable operations ensure a healthy work culture and thereby minimize turnover and enhance human capital management. Aguinis, Glavas (2012) further observe that the threat posed through regulation in the form of carbon pricing serves to be a strong impetus to companies to undertake integrated ESG approaches. The threat posed through financial and reputational risks forces organizations to introduce sustainability in strategy-making. The process in question is not only not reactive but rather proactive in nature with companies looking to avoid potential liabilities and position themselves well in environmentally conscious markets in the future. Apart from institutional forces, Eccles *et al.* (2014) provide evidence to indicate that companies with long-term commitments to sustainability perform better than less sustainable companies in the long run. Their study identifies a link between the early adoption of ESG operations and improved financial performance and observes that alignment with strategy and sustainability targets leads to long-term profitability and operational efficiency. High-sustainability companies have well-defined stakeholder engagement and better measurement and disclosure mechanisms in place and hence reinforce competitive positioning.

However, some studies caution against assuming that all ESG adoption is inherently genuine or strategically beneficial. Testa *et al.* (2018) explore the phenomenon of “greenwashing,” where companies adopt superficial sustainability practices to appease external stakeholders without implementing substantive changes. Their analysis reveals that while stakeholder pressures can enhance proactive environmental practices, they may also encourage symbolic compliance when companies aim to maintain public perception rather than achieving real environmental impact. In this process, the behavioural characteristics of decision-makers in SMEs play a critical role, as their attitudes toward risk and uncertainty directly determine the strategic resilience and integrity of environmental decisions (Hudakova *et al.,* 2025).

An ever-growing body of literature has employed machine learning techniques to predict ESG scores using the integration of multiple data points. Cainelli *et al.* (2020) demonstrated that financial fundamentals derived from balance sheet data were successfully employed alongside Random Forest models to generate high-accuracy ESG predictions. Likewise, D'Amato *et al.* (2022) demonstrated that financial statement metrics, such as profitability, liquidity, and solvency ratios, exhibit high predictive power with respect to ESG ratings. Their analysis of firms in the STOXX Europe 600 Index revealed that drivers such as NI_to_Sales were strong predictors of ESG performance. Adding to the findings, Jiang (2024) determined that gradient boosting was the best-performing algorithm to predict ESG risk scores and found key predictors such as Price/Sales, Price/Book, and Market Capitalization. Following up on that line of research, Lin and Hsu (2023) employed several machine learning techniques, Random Forest, Extreme Learning Machines (ELM), Support Vector Machines (SVM), and XGBoost to predict ESG scores among Taiwan's non-financial companies with ELM and XGBoost delivering the best performance in all instances. Advanced machine learning applications also include examining the impact of environmental management training as a key moderating factor in the relationship between corporate profitability and the achievement of SDG 13 – Climate Action (Raza *et al.,* 2025). Notably, Aue *et al.* (2025) took the scope further with predictions derived from news articles using sentiment analysis to extract market signals from an enormous pool of media sources. Moreover, Taskin *et al.* (2025) demonstrated that historical ESG scores turn out to be strong predictors of future ESG performance. Their analysis on Turkey's BIST sustainability index firms revealed that even low-complexity indicators derived only from historical ESG performance were strong predictors of future performance and offered a useful alternative to data-driven models with higher levels of complexity.

2. Methodology and Analysis

2.1 Research Design

The aim of the study is to investigate the predictability of corporate ESG ratings exclusively from firm-level financial features. The study adopts a quantitative predictive research design based on supervised machine learning techniques. A longitudinal framework is employed by combining multi-year lagged financial indicators with 2023 ESG ratings in order to examine whether historical financial conditions contain predictive signals for sustainability outcomes.

The study focuses on a sample of 974 large manufacturing companies in Slovakia, selected according to total assets and number of employees, with matched ESG ratings for 2023. The primary objective is to determine the extent to which aggregate ESG ratings, as well as the Environmental, Social, and Governance pillars separately, can be predicted from historical financial data and to identify which financial indicators and lag structures provide the strongest predictive signals. To achieve these objectives, the study employs eXtreme Gradient Boosting (XGBoost) as the primary machine learning algorithm, combined with Recursive Feature Elimination (RFE) and Shapley Additive Explanations (SHAP) to improve model interpretability.

2.2 Sample and Data Collection

The empirical analysis is based on a final sample of 974 Slovak firms. *Table 1* presents the sectoral composition of the sample and the distribution of companies across the main industry groups. The largest shares are represented by metal manufacturing and metallurgy, mechanical engineering, chemicals and plastics, and the food industry. Financial data covering the period 2018–2022 were retrieved from the Orbis database, a standardized source of firm-level financial statements widely used in corporate finance research. The five-year time horizon enables the construction of lagged financial indicators and provides a longitudinal perspective on firms' economic performance prior to the assignment of ESG ratings. The ESG data employed in the study were obtained from CRIF ANALYTICS and consist of four principal indicators: ESG Score, E Score, S Score, and G Score. Only 2023 ESG ratings were used in order to reflect the most recent sustainability assessment available for each company. Combining ESG indicators with historical financial information allows for an examination of whether past financial conditions are associated with recent ESG evaluations.

Table 1. Sectoral Composition of the Sample

Industry group	**Companies**	**Share (%)**
Metal manufacturing and metallurgy	189	19.40
Mechanical engineering	130	13.40
Chemicals and plastics	122	12.50
Food industry	115	11.80
Electrical engineering	104	10.70
Automotive industry	69	7.08
Clothing and footwear	66	6.78
Wood and paper	55	5.65
Construction	38	3.90
Other manufacturing	33	3.39
Trade and vehicle sales	13	1.33
IT, media, and communications	11	1.13
Healthcare	10	1.03
Transport and logistics	7	0.72
Professional and technical services	5	0.51
Other services	5	0.51
Energy and mining	1	0.10
Agriculture and forestry	1	0.10
Total	**974**	**100.00**

Source: own processing.

2.3 Measures and Variables

The dependent variables consist of ESG Score, E Score, S Score, and G Score. All ESG indicators are measured on a five-point ordinal scale, where 1 indicates maximum adequacy (strong ESG performance) and 5 indicates minimum adequacy. The ESG Score represents a composite indicator aggregating environmental, social, and governance dimensions, while the E, S, and G scores capture pillar-specific sustainability performance. *Table 2* summarizes the definitions of the ESG indicators.

Table 2. Overview of ESG Scores

Variable	Description
ESG Score	Synthetic indicator assessing the entity based on its E, S, and G factors. In the Environmental dimension, this score excludes Scope 3 emissions and physical climate risk.
E Score	Measures the extent of the entity's alignment with environmental subfactors, including emissions, energy efficiency, water consumption, waste generation, and biodiversity/ecosystems.
S Score	Gauges the entity's social practices, focusing on community engagement, employee relations, customer relationships, and broader societal concerns such as human rights, poverty, and famine.
G Score	Evaluates the entity's corporate governance with emphasis on ethical aspects, risk management strategies, inclusivity, and transparency in oversight and decision-making.

Note: All ESG indicators are measured on a five-point ordinal scale, where 1 indicates maximum adequacy and 5 indicates minimum adequacy. Thus, lower values represent stronger ESG performance. The aggregate ESG score is a provider-generated synthetic indicator based on CRIF ANALYTICS methodology. The study does not impose equal weights on the E, S, and G pillars because the provider documentation does not disclose a simple equal-weight aggregation rule. The E pillar includes emissions, energy efficiency, water use, waste production, and biodiversity; the S pillar includes community and society, employee relations, customer relations, human rights, and poverty-related aspects; and the G pillar includes ethical considerations, strategy and risk management, inclusiveness, and transparency.

Source: own processing based on CRIF ANALYTICS documentation.

The explanatory variables consist of firm-level financial indicators derived from the Orbis database. These variables capture multiple dimensions of corporate financial performance, including profitability, liquidity, leverage, operational efficiency, solvency, cash-flow generation, and capital structure. Examples include EBITDA, ROA, liquidity ratios, debt-service coverage, leverage ratios, and turnover indicators. Firm age was also included as a demographic control variable. Each financial indicator was measured annually between 2018 and 2022, producing multiple lagged versions of each predictor. *Table 3* provides a detailed overview of all financial variables included in the analysis.

Table 3. Overview of Corporate Financial Indicators

Variable(s)	Description
Accounting Cash Flow	Cash flow derived from accounting statements (i.e., net income plus non-cash expenses), reflecting the firm's internal financing capacity.
Age	Age of the firm in years as of 2022.
Asset Turnover Period, Asset Turnover	Measures of how efficiently a firm uses its assets to generate revenue, reported as either a period (days) or a ratio (revenue ÷ total assets).
COGS	Cost of goods sold, representing direct production costs (materials, labor).
Debt Repayment Period, Debt Repayment to Revenue Period, Trade Debt Repayment Period	Indicators of the time (in days or years) required to repay outstanding debt relative to revenue or other financial metrics.
Debt to Equity Ratio	Compares total liabilities to shareholder equity, reflecting a company's financial risk and capital structure strategy.
Deleveraging Ratio, Short Term Flow Deleveraging	Ratios depicting how quickly a firm can reduce debt levels through operational earnings or net cash flow.
EBIT	Earnings before interest and taxes, indicating operating profitability prior to financing costs and taxes.

Table 3 *(continuation)*. Overview of Corporate Financial Indicators

EBITDA	Earnings before interest, taxes, depreciation, and amortization, providing insight into cash-based operational performance.
Effective Tax Rate	Actual tax rate (total tax expense ÷ pre-tax earnings), measuring the firm's compliance burden.
Financial Accounts to Assets	Proportion of financial accounts (e.g., cash, marketable securities) relative to total assets, indicating a firm's liquid asset position.
Financial Leverage	Leverage ratio reflecting the firm's total assets relative to its equity base, gauging financial risk.
Gross Margin, EBITDA Margin, Operating Margin, Profit Margin, Accounting Cash Flow Margin	Various profitability measures—gross margin, EBITDA margin, operating margin, profit margin, and accounting cash flow margin—each expressed as a percentage of revenue.
Gross Operating Cash Flow, NOPAT	Gross operating cash flow indicates total cash generated pre-financing activities; NOPAT (net operating profit after tax) measures after-tax operational returns.
Interest Coverage, Debt Service Coverage	Coverage ratios quantifying the firm's ability to service interest and overall debt payments from its operating earnings or net cash flow.
Interest Coverage from Net Cash Flow, Debt Repayment Period from Net Cash Flow	Versions of coverage ratios assessing interest and debt repayment feasibility specifically from net cash flow rather than EBITDA or operating income.
Inventory Turnover, Inventory Turnover Period	Frequency (or days) of inventory turnover, reflecting how quickly inventory is converted into sales.
Liquidity Ratio 1, Liquidity Ratio 2, Liquidity Ratio 3	Classic liquidity measures (current, quick, and cash ratios) evaluating the firm's capacity to meet short-term liabilities with available current assets.
Markup, Financial Efficiency	Markup indicates the percentage increase over cost to set prices; financial efficiency measures overall resource utilization effectiveness.
Net Cash Flow	Net change in cash after accounting for all operating, investing, and financing activities.
Net Debt	Total debt minus cash and cash equivalents, providing a snapshot of leverage.
Net Debt to EBITDA, Liabilities to EBITDA	Ratios comparing debt or total liabilities to EBITDA, illustrating leverage and solvency status.
Non-Current Asset Turnover, Current Asset Turnover	Ratios indicating how efficiently a firm utilizes its fixed (non-current) and current assets to drive revenue.
Operating Leverage	Ratio illustrating how fixed costs compare to variable costs, affecting earnings sensitivity to changes in revenue.
Personnel Cost Coverage, Personnel and Depreciation Coverage	Coverage of personnel expenses or combined personnel and depreciation costs by operating income, gauging margin adequacy.
Personnel Cost to Value Added	Proportion of personnel expenses to the firm's value added, reflecting labor cost intensity.
Receivables Turnover Period, Short-Term Receivables Collection Period, Trade Receivables Collection Period	Metrics assessing collection efficiency on receivables, measured either as a turnover ratio or the average collection period in days.
Revenue Total, Revenue Adjusted	Gross revenue (unadjusted or adjusted) recognized in a given year (2018–2022), reflecting total sales minus returns and allowances.
ROE, ROA, ROA_EBIT, ROC_EBIT, Long-Term Capital Return EBIT, ROI	Performance ratios relating net income, EBIT, or total capital returns to equity or assets, capturing various dimensions of profitability and return on investment.
Total Debt Ratio, Self Financing Ratio, Long-Term Debt Ratio, Short-Term Debt Ratio	Measures of capital structure, indicating the proportion of total assets financed by debt, the extent of internal (equity) financing, and the split between long-term and short-term debt, respectively.
Trade Illiquidity, Total Illiquidity	Indicators of the firm's liquidity constraints, typically measuring the ease of converting assets to cash to meet short-term obligations.

Note: Each variable is measured annually from 2018 through 2022 (e.g., EBIT for 2018, EBIT for 2019, etc.) and is aggregated into the unified dataset.

Source: own processing.

Aside from the ESG measures, financial variables between 2018 and 2022 were retrieved from the Orbis database, a reliable source of standardized firm-level financial statements (*Table 3*). Combining the ESG indicators with a firm's historical financial performance enables an analysis of the relationship between past firm-level financial performance and recent appraisals in terms of sustainability (Cainelli *et al.,* 2020; Jiang, 2024; Lin, Hsu, 2023).

2.4 Data Preprocessing

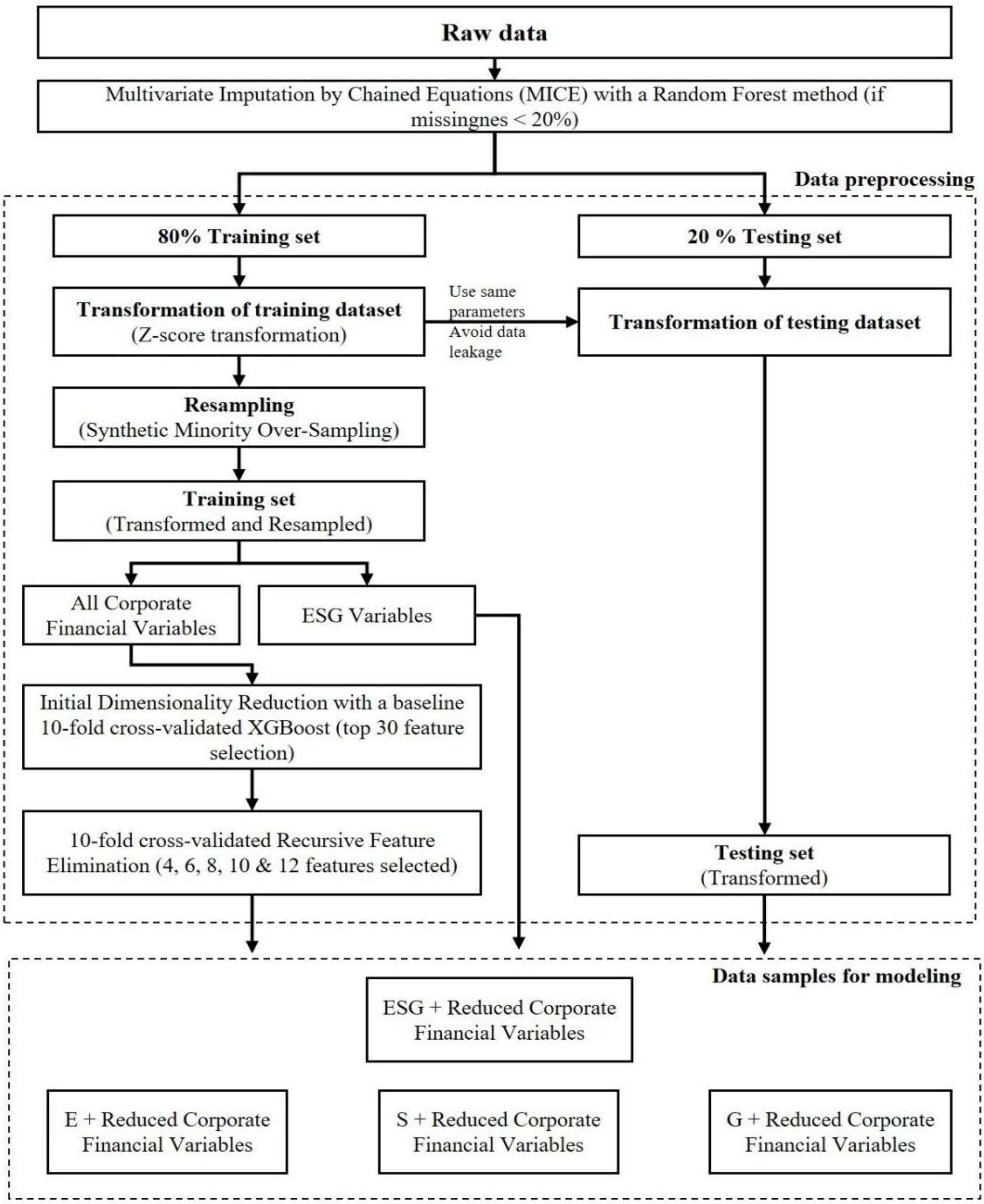
Several preprocessing procedures were implemented to prepare the dataset for machine learning analysis. First, the data were divided into training (80%) and testing (20%) subsets using a fixed random seed to ensure reproducibility (James *et al.,* 2013). Missing observations were subsequently addressed using Multivariate Imputation by Chained Equations (MICE) with a Random Forest estimation procedure in order to preserve underlying relationships among variables. To reduce the influence of extreme outliers, winsorization was applied by truncating observations below the 2nd percentile and above the 99th percentile (Ghosh, Vogt, 2012). Predictor variables were then standardized using Z-score normalization:

$$Z=\frac{X-\mu}{\sigma} \qquad \text{Eq. 1}$$

where μ and σ represent the mean and standard deviation calculated exclusively from the training dataset.

Restricting standardization parameters to the training subset prevents data leakage and improves the reliability of model evaluation (Kuhn, Johnson, 2013). *Figure 1* summarizes the overall data preprocessing and modelling workflow.

*Figure 1 i*llustrates the overall data pre-processing and modelling process with an initial dimensionality reduction via a baseline XGBoost model trained using 10-fold cross-validation. The baseline model computes feature importance using the "gain" metric that measures the incremental predictive accuracy gained per feature (Chen, Guestrin, 2016). We retain the top 30 features with cumulative gain contribution; the initial filtering reduces the dimensionality of the dataset and retains variables most relevant to predicting ESG score.



Source: own processing.

Figure 1. **Data Preprocessing and Dimensionality Reduction**

2.5 Feature Selection and Dimensionality Reduction

To reduce dimensionality and retain the most informative predictors, an initial XGBoost model was estimated using 10-fold cross-validation. Feature importance was evaluated using the gain metric, which measures the incremental predictive contribution of each variable (Chen, Guestrin, 2016). Based on cumulative gain contribution, the top 30 predictors were retained for further analysis. These predictors were subsequently refined using Recursive Feature Elimination (RFE) (Guyon, Elisseeff, 2003). In each iteration, the model was repeatedly trained while the least important predictors were removed until an optimal subset was identified. The procedure was validated through repeated 10-fold cross-validation to minimize overfitting and improve generalization performance. Following the strategy proposed by Papíková, Papík (2023), final feature subsets containing 4, 6, 8, 10, and 12 predictors were evaluated in order to balance model complexity and explanatory power.

2.6 Data Analysis Techniques

Because ESG ratings represent ordinal categorical outcomes, the study applies a classification framework. Model training employed repeated 10-fold cross-validation to improve the robustness of predictive estimates (Kuhn, 2008). To address class imbalance among ESG categories, the Synthetic Minority Over-sampling Technique (SMOTE) was incorporated within each cross-validation fold (Chawla *et al.,* 2002). SMOTE synthetically augments minority-class examples, aiding the model in accurately distinguishing across all ESG score categories (Rech *et al.,* 2025).

The primary predictive algorithm used in the study is eXtreme Gradient Boosting (XGBoost), which constructs an ensemble of decision trees in an additive manner (Chen, Guestrin, 2016). The prediction for observation is defined as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \qquad \text{Eq. 2}$$

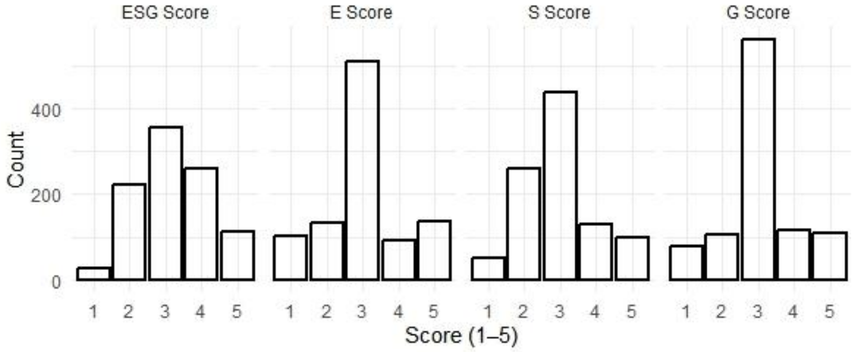
where each f_k represents an individual regression tree and x_i denotes the feature vector for observation i. Hyperparameter tuning was conducted using random search, which enables efficient exploration of the parameter space compared to exhaustive grid search (Bergstra, Bengio, 2012).

The final tuned XGBoost model was evaluated using the hold-out testing dataset. Predictive performance was assessed through confusion matrices, classification accuracy, Cohen’s Kappa, recall, and related performance metrics. These measures provide evidence regarding the extent to which historical financial indicators can distinguish between ESG rating categories.

3. Results

3.1 Main Modelling Approach

As a starting point, *Figure 2* provides the distributions of the overall ESG score and its three components, Environmental (E), Social (S), and Governance (G), for the 974 firms in the sample. Not surprisingly, the scores of each pillar cluster in the middle of the 1–5 scale, suggesting that most firms have moderate levels of compliance with sustainability standards in all dimensions and neither excel nor fail in any one dimension. The E pillar has a very tight concentration in the vicinity of score 3, suggesting relatively homogeneous performance in emissions management, energy efficiency, and resource utilization. The S and G pillars have higher dispersion between scores 2 and 4, suggesting higher heterogeneity in firms' labour policies and community relations and in firms' governance and adherence to ethics. The overall ESG rating has a similar middle-skewed shape but with a heavier tail at higher scores, suggesting that a significant percentage of firms have high levels of compliance with ESG standards and others cluster near average levels of performance.



Source: own processing.

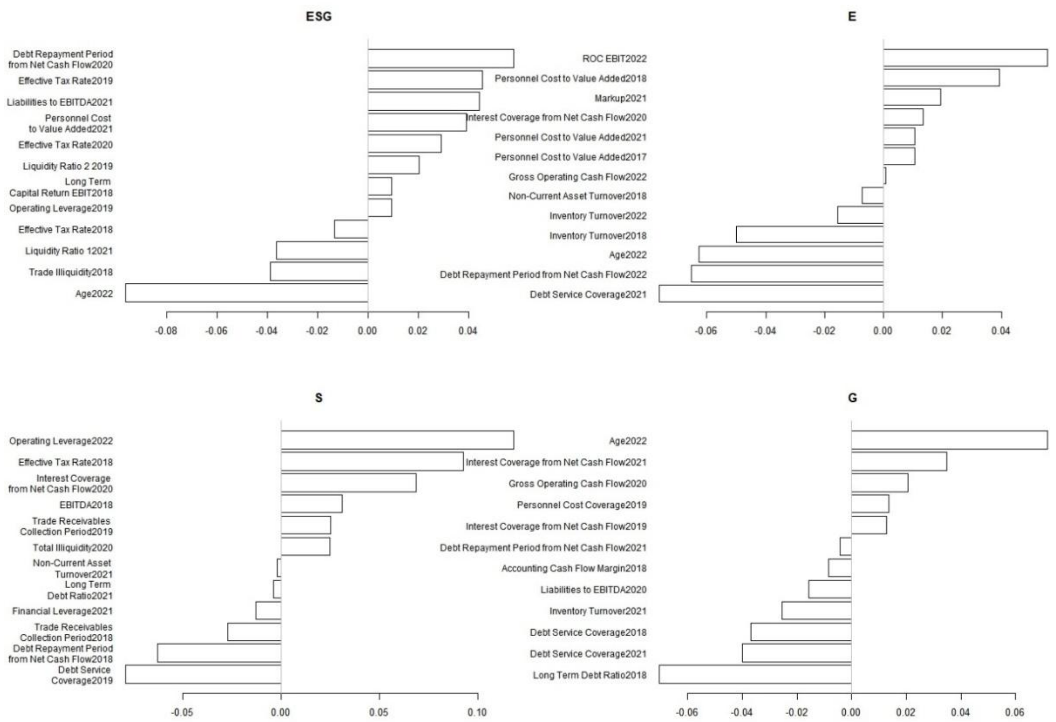
Figure 2. **Distribution of ESG and Pillar Scores (E, S, G)**

After an initial screening of the full variable set, only the 30 best predictors were retained based on their feature importance scores. Building on this reduced pool, we then performed 10-fold cross-validated RFE, wherein the model systematically re-evaluated subsets by removing two features at each iteration. The final data configurations resulting from RFE correspond to subsets containing 4, 6, 8, 10, and 12 features, respectively. Detailed outputs for this process, including metrics such as Accuracy and Kappa, are provided in the Appendices, separately for the overall ESG score and each pillar (E, S, G).

Figure 3 displays four adjacent bar charts, each representing the top-ranking features based on their SHAP values for different models: one for the overall ESG score and one each for the individual dimensions. In general, a positive SHAP value indicates that a feature contributes positively to the predicted score relative to the baseline, whereas a negative SHAP value indicates that the feature reduces the predicted score. Higher absolute SHAP values denote a greater influence on the model's predictions, thereby highlighting the most critical predictors for each ESG dimension.

In each dimension, indebtedness ratios emerge as the foremost drivers exerting downward pressure on the scores. This means that as these ratios increase, they tend to lower the firm's predicted ESG rating, signalling a financial structure that might be overly leveraged and potentially risky, thereby reflecting negatively on the firm's overall sustainability performance. Similarly, liquidity ratios, when showing negative SHAP values, indicate that weaker liquidity can detract from a firm’s score by suggesting potential issues with short-term financial stability, which in turn may deter investors and stakeholders who value robust sustainability practices.

Age stands out as another highly significant contributor: for the overall ESG score it exhibits the largest absolute negative influence, while for the E dimension it ranks as the third largest negative driver. The negative association between firm age and ESG/E performance may reflect challenges older firms face in adapting legacy systems to modern sustainability standards. Newer firms often integrate these principles into their foundational operations, aligning with current regulatory and market expectations. While this trend suggests a structural advantage for younger companies, the relationship is likely influenced by additional factors.



Source: own processing.

Figure 3. **Top 12 Features Based on SHAP for Each ESG Dimension**

On the positive side, effective debt management ratios, such as interest coverage, debt repayment, and liabilities to EBITDA, generally boost the scores, indicating that firms with a strong handle on managing their debt obligations are rewarded with higher sustainability ratings. A higher effective tax rate also contributes positively, possibly reflecting responsible financial practices or stringent compliance measures that align with broader governance and sustainability objectives. For the E dimension specifically, operational efficiency variables are particularly influential, meaning that companies optimizing their production processes and resource use tend to be rated more favourably.

The prominence of debt variables with 3- to 5-year lags, and the broader reliance on older predictors rather than recent data, suggests ESG performance is shaped by long-term financial and operational trends, not short-term adjustments. Debt decisions likely exert delayed but compounding effects: high leverage from several years prior may constrain sustainability investments or erode governance credibility over time. Similarly, older operational metrics reflect enduring practices that underpin ESG resilience. The scarcity of 1-year lag variables implies immediate changes may lack the sustained impact required for ESG assessments, which prioritize consistency. This pattern highlights that ESG excellence hinges on multi-year commitments to financial discipline and resource efficiency, with historical decisions casting long shadows over future ratings.

Table 4. Comparative Performance of Different Feature Subsets across ESG Dimensions

Number of features	ESG	E	S	G
Four	0.1934 (-0.0106)	**0.2928 (0.0307)**	**0.2818 (0.0623)**	0.2762 (-0.0041)
Six	0.2155 (0.0019)	0.2928 (0.0292)	0.2652 (0.0135)	0.2818 (-0.0339)
Eight	0.2431 (0.016)	0.2376 (-0.0401)	0.2652 (0.0186)	0.3425 (0.0346)
Ten	**0.2928 (0.0466)**	0.2707 (-0.008)	0.2486 (-0.0019)	**0.3591 (0.0438)**
Twelve	0.2818 (0.0326)	0.2873 (0.0181)	0.2707 (0.003)	0.3536 (0.0236)

Note: The first value in each cell indicates the accuracy metric, while the value in parentheses represents the corresponding Kappa statistic.

Source: own calculations.

The results showed that while four key variables were sufficient for optimal classification in E and S dimensions, a broader feature set of ten variables was necessary for ESG and G dimension and the overall ESG classification (see *Table 5*). Most importantly, the low accuracy of our model is not necessarily problematic given the inherent challenges of predicting ESG ratings solely from balance sheet data. ESG ratings encompass a range of qualitative factors, market perceptions, and industry-specific practices that are not fully captured by quantitative financial metrics. As such, the modest accuracy figures underscore the complexity of the task rather than indicating a failure of the model. This outcome is in line with expectations and highlights the limitations of using balance sheet data as the sole predictor for ESG performance. Instead, the model's value lies in the insights it offers regarding how long-term financial indicators, such as indebtedness, liquidity, and age, contain informative signals related to sustainability outcomes, even if these relationships cannot fully account for the nuanced nature of ESG assessments. The variables identified as most relevant for each dimension are summarized in *Table 5*.

Table 5. Assessment of Model Overfitting

Number of features	ESG	E	S	G
Test				
Four	0.1934	0.2928	0.2818	0.2762
Six	0.2155	0.2928	0.2652	0.2818
Eight	0.2431	0.2376	0.2652	0.3425
Ten	0.2928	0.2707	0.2486	0.3591
Twelve	0.2818	0.2873	0.2707	0.3536
Train				
Four	0.3807	0.6634	0.8152	0.8221
Six	0.4497	0.8634	0.9145	0.9228
Eight	0.5986	0.949	0.9352	0.9324
Ten	0.9159	0.9255	0.9421	0.9724
Twelve	0.9338	0.9724	0.9448	0.9779

Source: own calculations.

Although the models achieve relatively high accuracy on the training set, they generalize poorly to the holdout sample, with test accuracy levels typically hovering around 20–30%. The train-test gap suggests that the sparse interpretable model is prone to overfitting, especially in the S and G dimensions. However, this does not imply that financial indicators lack predictive content. Rather, it indicates that a small set of directly interpretable variables cannot fully recover the five-class ESG rating structure. This motivates the alternative modelling approach in Section 3.2, where a broader PCA-based representation of the financial data is used as a robustness check.

Table 6. Correct Classification Accuracy of Individual ESG Scores

		1	2	3	4	5
ESG	Four	0.125	0.2142	0.2542	0.08	0.2727
	Six	0.375	0.2143	0.1864	0.18	0.3182
	Eight	0.375	0.2619	0.2203	0.22	0.2727
	Ten	**0.125**	**0.2619**	**0.3729**	**0.3**	**0.1818**
	Twelve	0.375	0.2381	0.3559	0.28	0.1364
E	Four	**0.3333**	**0.1053**	**0.3889**	**0.1579**	**0.1724**
	Six	0.2083	0.1053	0.3889	0.1579	0.2759
	Eight	0.125	0.0526	0.3222	0.1579	0.2414
	Ten	0.125	0.0526	0.4	0.1579	0.2069
	Twelve	0.125	0.1579	0.4	0.1579	0.2414
S	Four	**0.0667**	**0.2885**	**0.3582**	**0.1481**	**0.35**
	Six	0.0667	0.2308	0.4179	0.1111	0.2
	Eight	0	0.2308	0.4179	0.1852	0.15
	Ten	0	0.2885	0.3433	0.1481	0.15
	Twelve	0	0.2885	0.4179	0.1481	0.1
G	Four	0.2667	0.1923	0.3398	0.15	0.1765
	Six	0.0667	0.1538	0.4078	0.1	0.1177
	Eight	0.1333	0.1538	0.4757	0.15	0.2353
	Ten	**0.2**	**0.1154**	**0.5340**	**0.1**	**0.1177**
	Twelve	0.2	0.1154	0.5243	0.1	0.1177

Note: The highlighted rows indicate the feature subset that achieved the highest overall classification performance.

Source: own calculations.

Table 6 reveals that the best-predicted ratings tend to occur in the mid-range for the Environmental, Social, and Governance dimensions, with score 3 consistently achieving the highest accuracy across several feature subsets. In contrast, extreme ratings, especially the lowest (scores 4 and 5) and highest (score 1) categories, tend to be predicted less accurately. For the overall ESG score, the picture is more mixed, with certain feature subsets performing better at the extremes while others struggle with mid-range values. These findings imply that while balance sheet data can effectively capture moderate sustainability performance, it falls short in reliably predicting companies at the performance extremes.

3.2 Alternative Modeling Approach

As a robustness analysis, we estimate an alternative PCA-based (Principal Component Analysis) one-vs-rest XGBoost model. This specification differs from the main XGBoost, RFE, and SHAP framework in two important ways. First, instead of relying on a small subset of selected original financial variables, the full set of lagged financial indicators is transformed into principal components, with the retained components explaining approximately 90% of the variance in the original predictor set. Second, we evaluate the model under several experimental train-test partitions, namely 60/40, 70/30, 80/20, and 90/10, to assess whether the results are sensitive to the choice of sample split. For each ESG outcome, separate one-vs-rest XGBoost classifiers are estimated for the five rating categories. Final classification in *Tables 7 and 8* is based on the highest raw predicted probability, while Appendix E and Appendix F report the recall-weighted version of the same model.

Table 7. PCA-Based One-vs-Rest XGBoost Model (Raw-Probability Ensemble)

Target	Best split	Train N	Test N	PCA components	PCA variance explained (%)	Accuracy	Macro recall	Recall 1	Recall 2	Recall 3	Recall 4	Recall 5
E	90/10	816	90	71	90.1	0.456	0.283	0.091	0.273	0.723	0.250	0.077
S	70/30	636	270	69	90.0	0.370	0.203	0.000	0.307	0.620	0.000	0.087
G	70/30	635	271	67	90.1	0.480	0.253	0.080	0.118	0.747	0.222	0.097
ESG	70/30	636	270	69	90.3	0.311	0.232	0.000	0.254	0.366	0.371	0.167

Note: Final classes are assigned using the highest raw one-vs-rest predicted probability. PCA components indicate the number of retained principal components required to explain approximately 90 percent of the variance in the financial predictor set. Accuracy measures overall correct classification. Macro recall is computed as unweighted averages across the five ESG rating classes. Recall 1 to Recall 5 reports class-specific recall for each ESG score category.

Source: own calculations.

Compared with the main results, the alternative model confirms that historical financial indicators contain predictive information about ESG ratings, but it also shows that full multiclass prediction remains difficult. The raw-probability model reaches accuracies of 0.456 for E, 0.370 for S, 0.480 for G, and 0.311 for ESG in the best split specifications reported in *Table 7*. These results are generally stronger than the sparse main model for several dimensions, but they remain lower than what would normally be expected from simpler binary classification tasks. This difference is not surprising, because the model is required to assign firms across five ordinal rating categories rather than distinguish between two broad groups.

The class-level results show that the alternative model predicts the individual ESG pillars fairly well, especially the Environmental and Governance dimensions. In *Table 7*, recall for the third rating category reaches 0.723 for E, 0.620 for S, and 0.747 for G, which is strong given the five-class ordinal structure of the dependent variable. The stronger performance for E and G suggests that financial statement data are more closely linked to environmental and governance outcomes, possibly because these dimensions reflect operational efficiency, investment capacity, financial discipline, risk management, and transparency. The Social pillar is also predictable, but to a lesser extent, likely because social performance depends more on qualitative factors such as labour relations, employee welfare, community engagement, and customer policies. By contrast, the composite ESG score is harder to predict, with score 3 recall-reaching only 0.366 in *Table 7*. This suggests that aggregation may dilute pillar-specific financial signals. A firm may perform well in governance but less well in environmental or social dimensions, making the overall ESG score less directly connected to any single financial pattern. This supports the decision to model E, S, and G separately and shows that lagged financial indicators are especially informative for E and G, while the aggregate ESG score remains more ambiguous.

Table 8. Split-Sensitivity Results for the PCA-Based One-vs-Rest XGBoost Model (Raw-Probability Ensemble)

Target	Split	Train N	Test N	PCA components (variance explained, %)	Accuracy	Macro recall	Recall 1	Recall 2	Recall 3	Recall 4	Recall 5
E	60/40	544	362	67 (90.2)	0.431	0.273	0.152	0.163	0.644	0.185	0.220
E	70/30	635	271	69 (90.2)	0.410	0.250	0.231	0.194	0.628	0.038	0.158
E	80/20	725	181	68 (90.1)	0.436	0.251	0.062	0.179	0.702	0.118	0.192
E	90/10	816	90	71 (90.1)	0.456	0.283	0.091	0.273	0.723	0.250	0.077
S	60/40	545	361	67 (90.2)	0.296	0.191	0.048	0.333	0.420	0.020	0.133
S	70/30	636	270	69 (90)	0.370	0.203	0.000	0.307	0.620	0.000	0.087
S	80/20	726	180	70 (90)	0.328	0.195	0.000	0.451	0.420	0.000	0.105
S	90/10	817	89	74 (90.3)	0.326	0.175	0.000	0.423	0.450	0.000	0.000
G	60/40	545	361	64 (90.3)	0.490	0.219	0.129	0.024	0.783	0.075	0.081
G	70/30	635	271	67 (90.1)	0.480	0.253	0.080	0.118	0.747	0.222	0.097
G	80/20	726	180	70 (90.1)	0.489	0.244	0.167	0.091	0.784	0.133	0.043
G	90/10	816	90	73 (90.2)	0.467	0.199	0.000	0.167	0.736	0.091	0.000
ESG	60/40	546	360	66 (90.3)	0.289	0.213	0.000	0.259	0.351	0.330	0.125
ESG	70/30	636	270	69 (90.3)	0.311	0.232	0.000	0.254	0.366	0.371	0.167
ESG	80/20	726	180	71 (90.1)	0.328	0.222	0.000	0.302	0.403	0.404	0.000
ESG	90/10	817	89	72 (90.1)	0.236	0.174	0.000	0.250	0.303	0.217	0.100

Source: own calculations.

Table 8 shows that these patterns are not limited to one specific train-test partition. The E and G models remain comparatively stable across the 60/40, 70/30, 80/20, and 90/10 splits, with E accuracy ranging from 0.410 to 0.456 and G accuracy ranging from 0.467 to 0.490. This suggests that the financial signal behind the Environmental and Governance dimensions is relatively robust across alternative experimental settings. The Social pillar is less stable but still reaches its best performance under the 70/30 split, while the composite ESG score remains the weakest and most sensitive target. This again indicates that the aggregation of E, S, and G into a single composite score may obscure pillar-specific financial signals.

The appendix results provide an additional robustness perspective. Appendix E and Appendix F report the recall-weighted version of the PCA-based one-vs-rest XGBoost model. This version improves overall accuracy for several targets, reaching 0.492 for E, 0.382 for S, 0.550 for G, and 0.348 for ESG. More importantly, Appendix F shows that the recall-weighted model substantially improves correct classification of the third rating category. For the E pillar, recall for score 3 reaches 0.926 under the 80/20 split and 0.936 under the 90/10 split. For the G pillar, score 3 recall is even stronger and remains above 0.94 across all splits, reaching 0.981 under the 60/40 split. These are strong results for a five-class ordinal classification problem and indicate that the model can identify the dominant middle ESG category with high reliability. At the same time, the weaker recall for the extreme categories confirms that the main limitation is not the absence of a financial signal, but the difficulty of learning sparsely populated rating classes, even after synthetic resampling through SMOTE.

4. Discussion

The results show that historical financial indicators contain meaningful information about ESG ratings, but this signal is stronger for the individual ESG pillars than for the aggregate ESG score. This is consistent with prior studies showing that financial fundamentals can support ESG prediction (Cainelli *et al.,* 2020; D'Amato *et al.,* 2022; Jiang, 2024), but our results add that the predictive value is not evenly distributed across ESG dimensions. The alternative PCA-based model performs particularly well for the Environmental and Governance pillars, suggesting that these dimensions are more closely connected to observable financial conditions such as operational efficiency, investment capacity, debt management, risk control, and financial discipline. By contrast, the composite ESG score is harder to predict, likely because aggregation combines different pillar-level signals and may dilute the direct link between financial data and specific ESG outcomes.

The main XGBoost, RFE, and SHAP results further show that leverage, liquidity, debt-servicing capacity, firm age, and tax-related indicators are among the most informative predictors. This supports the idea that ESG performance is shaped not only by current firm characteristics but also by accumulated financial capacity and long-term managerial discipline. In this respect, the study extends prior ESG prediction studies that often rely on contemporaneous financial variables, market indicators, or past ESG scores (D'Amato *et al.,* 2022; Jiang, 2024; Taskin *et al.,* 2025).

At the same time, the results confirm the limits of predicting ESG ratings from financial statements alone. The five-class ordinal structure of the ESG scores makes the task more difficult than a binary classification problem, since each rating category may contain firms with heterogeneous underlying sustainability quality. This is especially visible in the class-level results. The models predict the middle category much more reliably than the extremes. In the recall-weighted specification, recall for score 3 exceeds 90% for E and G in several partitions, showing that the model can identify the dominant middle ESG group with high reliability.

The weaker performance for extreme ESG classes reflects the extreme class imbalance in the data. Most firms are concentrated in the middle rating category, so the model learns this group from a larger number of real observations. By contrast, the extreme classes contain fewer firms and rely more heavily on synthetic resampling through SMOTE, which cannot fully replace real observations. Overall, the findings suggest that financial data provide a useful but incomplete signal. They are especially informative for E and G performance, but qualitative information, ESG disclosures, governance narratives, and past ESG scores remain necessary to fully capture the multidimensional nature of ESG ratings.

Additionally, the findings highlight the importance of temporal persistence in ESG-related financial behaviour. The fact that several highly ranked predictors originate from older lag structures suggests that ESG outcomes are influenced by cumulative organizational processes rather than by isolated short-term managerial actions. This observation is theoretically important because it implies that sustainability performance develops gradually through long-term investment patterns, financing decisions, and operational stability. Firms do not appear capable of substantially altering ESG outcomes through temporary financial adjustments immediately before ESG assessment periods. Instead, ESG performance reflects embedded organizational characteristics and long-term strategic consistency. This finding aligns with the broader sustainability literature emphasizing path dependency and long-term capability building in corporate sustainability transitions.

The results also contribute to the growing debate regarding the relationship between financial resilience and sustainability orientation. The consistent importance of leverage and debt-servicing variables suggests that financially constrained firms may face difficulties implementing ESG-oriented investments, particularly those requiring significant upfront capital expenditures. Environmental improvements, governance reforms, compliance systems, and social responsibility initiatives often require sustained financial resources and organizational slack. Consequently, firms with stronger balance-sheet stability may possess a structural advantage in developing sustainability capabilities. This supports the argument that ESG performance is partly connected to the broader concept of organizational resilience, where financially stable firms are better positioned to absorb the costs and uncertainties associated with sustainability transformation.

Another important observation concerns the distinction between interpretability and predictive performance in ESG modelling. The sparse RFE-based models provide clearer interpretability by identifying a limited number of influential financial variables, yet they exhibit weaker out-of-sample generalization. In contrast, the PCA-based one-vs-rest XGBoost specification achieves stronger predictive accuracy by incorporating broader latent financial structures, although at the cost of reduced direct interpretability. This trade-off illustrates a common challenge in machine learning applications within finance and sustainability research. While highly interpretable models are valuable for understanding causal and managerial implications, broader latent-factor approaches may better capture the multidimensional interactions embedded in corporate financial data. The coexistence of these two modelling perspectives strengthens the robustness of the study's conclusions because both approaches independently confirm that meaningful ESG-related signals are embedded within historical accounting information.

The findings may also reflect institutional characteristics specific to the Slovak and broader Central and Eastern European context. In emerging ESG environments, sustainability reporting frameworks and disclosure standards may still be less mature and less standardized than in Western European markets. Under such conditions, financial statement indicators may indirectly capture aspects of organizational quality, managerial professionalism, and operational discipline that are more strongly correlated with ESG assessments. This could partly explain why Environmental and Governance pillars exhibit relatively stronger predictability in the sample. In more mature ESG reporting environments, by contrast, non-financial disclosures may contain richer and more direct sustainability information, potentially reducing the relative importance of accounting-based predictors.

At the same time, the moderate overall predictive accuracy observed across models reinforces concerns raised in the literature regarding ESG rating heterogeneity and methodological inconsistency among providers. ESG ratings are influenced not only by measurable firm characteristics but also by subjective weighting systems, data availability, sector adjustments, and qualitative evaluations. As a result, some unexplained variation likely reflects differences in rating construction rather than purely firm-level sustainability performance. This interpretation is consistent with studies arguing that ESG scores frequently contain substantial measurement divergence across providers (Berg *et al.,* 2022). Therefore, the inability of financial variables to perfectly predict ESG ratings should not necessarily be interpreted as model weakness alone, but also as evidence of the inherently complex and partly subjective nature of ESG evaluation systems.

Finally, the study demonstrates the practical usefulness of interpretable machine learning methods in sustainability research. Traditional statistical approaches often struggle to capture nonlinear relationships and interaction effects among financial indicators, whereas machine learning techniques such as XGBoost can model these complex dependencies more effectively. At the same time, the integration of SHAP analysis ensures that the predictive process remains transparent and economically interpretable. This combination is particularly valuable in ESG research, where stakeholders increasingly demand both predictive capability and explainability. The results therefore support the growing role of explainable artificial intelligence (XAI) approaches in corporate sustainability analysis and ESG risk assessment.

5. Theoretical and Practical Implications

5.1 Theoretical Implications

This study contributes to ESG prediction research by showing that historical financial indicators contain useful but incomplete information about ESG ratings. Prior studies show that financial fundamentals can help explain ESG scores (Cainelli *et al.,* 2020; D'Amato *et al.,* 2022; Jiang, 2024; Lin, Hsu, 2023). This study extends that literature by using multi-year lagged financial indicators and by showing that ESG outcomes are not only related to recent financial conditions, but also to longer-term financial discipline, liquidity, leverage, debt-servicing capacity, and firm maturity.

The findings also show that ESG should not be treated as one uniform outcome. The Environmental and Governance pillars are more predictable from financial statement information than the aggregate ESG score. This suggests that pillar-level ESG ratings may reflect different financial and organizational foundations. It also supports the view that ESG ratings combine both firm-level sustainability characteristics and rating-provider-specific measurement choices (Berg *et al.,* 2022; Chatterji *et al.,* 2016). Therefore, the study adds to theory by showing that financial data can explain part of ESG performance, but cannot fully capture the qualitative and multidimensional nature of ESG assessment.

5.2 Practical Implications

The findings have practical value for investors, lenders, rating users, and ESG rating agencies. Financial statements can serve as an early screening tool when ESG disclosures are limited, delayed, or difficult to compare. This is especially relevant in smaller economies and manufacturing sectors, where detailed sustainability reporting may still be developing. This implication is consistent with prior ESG prediction studies showing that financial fundamentals and firm-level information can help predict ESG ratings using machine learning methods (D'Amato *et al.,* 2022; Lin, Hsu, 2023).

For ESG rating agencies, the results suggest that historical financial indicators can be useful as complementary inputs in rating validation and quality control. Variables related to leverage, liquidity, debt-servicing capacity, firm age, and tax-related behaviour may help identify firms whose ESG ratings appear inconsistent with their long-term financial profile. This does not mean that financial data should determine ESG ratings. Rather, accounting information can provide an additional consistency check, especially when direct sustainability data are incomplete.

However, the results also show that financial indicators should complement ESG-specific information rather than replace it. The models perform better for middle ESG categories than for extreme categories, which means that financial data are less reliable for identifying the best and worst ESG performers. For practitioners and rating agencies, this implies that balance sheet data can support preliminary ESG assessment, but final evaluation should still include sustainability reports, governance information, environmental disclosures, certifications, controversy data, and qualitative evidence.

6. Limitations and Future Research

This study has several limitations. First, ESG ratings are available only for one year and the sample is limited to manufacturing firms in Slovakia. This design is suitable for examining whether historical financial indicators contain ESG-related signals in a focused setting, but it limits the ability to test longer ESG dynamics, cross-sector differences, and broader international generalizability. Future research should extend the analysis to multiple ESG years, additional industries, and other countries, especially in Central and Eastern Europe where ESG disclosure practices are still developing.

Second, the study relies on ESG scores from one rating provider. ESG ratings are provider-specific assessments and may reflect methodological choices, data availability, sectoral assumptions, and weighting decisions that are not fully observable to researchers. This means that the model may partly learn the provider's rating logic rather than only the underlying sustainability quality of each firm. Future studies should compare ratings from multiple ESG providers to test whether financial indicators predict ESG performance consistently across alternative rating methodologies.

Third, the ESG indicators are modelled as ordinal rating categories. Firms within the same rating class may still differ substantially in their actual sustainability practices, while extreme rating classes contain fewer observations and are harder to predict. Although SMOTE helps reduce class imbalance, it cannot fully replace real observations in underrepresented ESG categories. Future research could apply ordinal classification methods, class-weighted learning, and alternative resampling strategies to better account for the ordered and imbalanced structure of ESG ratings.

Finally, this study intentionally focuses on financial statement indicators to test whether historical accounting data alone can support ESG prediction. The results show that such data provide useful but incomplete information. Future research should extend the analysis to multiple years, additional industries, and other countries. It should also combine financial indicators with non-financial data, such as sustainability disclosures, governance reports, media sentiment, regulatory information, and historical ESG scores, to improve both predictive accuracy and interpretability.

Conclusion

This study examined whether corporate ESG ratings can be predicted using multi-year lagged financial indicators in a sample of Slovak manufacturing firms. By combining financial data from 2018–2022 with 2023 ESG ratings and applying machine learning techniques, the study investigated whether historical financial conditions contain predictive signals for sustainability outcomes.

The findings demonstrate that financial indicators provide meaningful, although incomplete information about ESG performance. Leverage, liquidity, debt-servicing capacity, firm age, and tax-related indicators emerged as important predictors, suggesting that ESG outcomes are associated not only with current firm characteristics, but also with longer-term financial discipline and operational stability. The results further indicate that the Environmental and Governance pillars are more predictable from financial statement information than the aggregate ESG score, highlighting important differences across ESG dimensions.

At the same time, the study confirms that financial data alone cannot fully explain the multidimensional nature of ESG ratings. While historical accounting information can support ESG assessment and early screening, qualitative disclosures, governance practices, sustainability reporting, and broader non-financial information remain essential for capturing the complexity of corporate sustainability performance.

Overall, the study contributes to the growing literature on ESG prediction by showing that multi-year financial indicators can serve as valuable complementary signals of ESG positioning, particularly in environments where ESG disclosure remains limited. The findings also highlight the potential of interpretable machine learning methods to improve understanding of the relationship between financial performance and sustainability outcomes.

Literature

Aguinis, H., Glavas, A. (2012), "What We Know and Don't Know About Corporate Social Responsibility: A Review and Research Agenda", *Journal of Management*, Vol. 38, No 4, pp.932-968

Aue, T., Jatowt, A., Faerber, M. (2025), "Predicting Company ESG Ratings from News Articles Using Multivariate Timeseries Analysis", *The Web Conference Companion Proceedings (Pre-Print),* available at, https://doi.org/10.1145/3701716.3717509, referred on, 28/04/2025.

Barko, T., Cremers, M., Renneboog, L. (2022), "Shareholder Engagement on Environmental, Social, and Governance Performance", *Journal of Business Ethics*, Vol. 180, No 2, pp.777–812.

Bergstra, J., Bengio, Y. (2012), "Random search for hyper-parameter optimization", *Journal of Machine Learning Research*, Vol. 13, pp.281-305.

Berg, F., J. F. Ko, R. Rigobon (2022), "Aggregate Confusion: The Divergence of ESG Ratings." *Review of Finance*, Vol. 26, No 6, pp.1315–1344. https://doi.org/10.1093/rof/rfac033.

Cainelli, G., D'Amato, A., Mazzanti, M. (2020), "Resource efficient eco-innovations for a circular economy: Evidence from EU firms", *Research Policy*, Vol. 49, No 1, pp.103827

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002), "SMOTE: Synthetic Minority Over-Sampling Technique", Journal of Artificial Intelligence Research, Vol. 16, pp.321–357.

Chen, T., Guestrin, C. (2016), "XGBoost: A Scalable Tree Boosting System", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785–794, available at, https://doi.org/10.1145/2939672.2939785, referred on 06/02/2025.

D'Amato, V., D'Ecclesia, R., Levantesi, S. (2022), "ESG Score Prediction through Random Forest Algorithm", *Computational Management Science*, Vol. 19, No 2, pp.347–373.

Delmas, M., Toffel, M.W. (2004), "Stakeholders and Environmental Management Practices: An Institutional Framework", *Business Strategy and the Environment*, Vol. 13, No 4, pp.209–222.

Eccles, R.G., Ioannou, I., Serafeim, G. (2014), "The Impact of Corporate Sustainability on Organizational Processes and Performance", *Management Science*, Vol. 60, No 11, pp.2835–2857.

Friedman, J.H. (2001), "Greedy Function Approximation: A Gradient Boosting Machine", *Annals of Statistics*, Vol. 29, No 5, pp.1189–1232.

Ghosh, D., Vogt, A. (2012), "Outliers: An Evaluation of Methodologies", *Proceedings of the Joint Statistical Meetings*, pp.3455–3460.

Guyon, I., Elisseeff, A. (2003), "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research*, Vol. 3, pp.1157–1182.

Hudakova, M., Gabryśova, M., Zerebakova, A. (2025), "Assessment of Behavioural Characteristics of Owners' and Managers' Decision-Making in the SME Segment of V4 Countries", *Journal of Business Sectors*, Vol. 3, No 2, pp.103–112. https://doi.org/10.62222/WUDU1009.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013), *An Introduction to Statistical Learning: With Applications in R*, Springer, New York.

Jiang, X. (2024), "Predicting Corporate ESG Scores Using Machine Learning: A Comparative Study", *Advances in Economics, Management and Political Sciences*, Vol. 118, No 1, pp.1–8.

Jones, D.A., Willness, C.R., Madey, S. (2014), "Why Are Job Seekers Attracted by Corporate Social Performance? Experimental and Field Tests of Three Signal-Based Mechanisms", *Academy of Management Journal*, Vol. 57, No 2, pp.383–404.

Kubalek, J., Kudej, M. (2025), "Greenwashing Resistance in SMEs: The Role of ESG Perceptions and Practices", *Journal of Business Sectors*, Vol. 3, No 1, pp.104–111.

Kuhn, M. (2008), "Building Predictive Models in R Using the caret Package", *Journal of Statistical Software*, Vol. 28, No 5, pp.1–26.

Kuhn, M., Johnson, K. (2013), *Applied Predictive Modeling*, Springer, New York.

Lin, H.Y., Hsu, B.W. (2023), "Empirical Study of ESG Score Prediction through Machine Learning—A Case of Non-Financial Companies in Taiwan", *Sustainability*, Vol. 15, No 19, Article 14376.

Papíková, L., Papík, M. (2023), “Application of Intellectual Capital in SME Bankruptcy”, *Applied Economics*, Vol. 56, No 55, pp.6574–6590.

Petrakova, Z., Karacsony, P. (2024), “Sustainability of SMEs and Their Determinants: Comparative Analysis from V4 Countries”, *Journal of Business Sectors*, Vol. 2, No 2, pp.50–61.

Raza, H., Kovacs, S., Khan, M.H. (2025), “Profitability and SDG 13–Climate Action: The Moderating Role of Environmental Management Training: A Machine Learning Approach”, *Journal of Business Sectors*, Vol. 3, No 1, pp.1–12. https://doi.org/10.62222/BVCH4721.

Rech, F., Isaboke, C., Xu, H. (2025), “Surviving the Pandemic: Financial Distress Prediction for Slovak SME Manufacturers”, *Journal of Business Sectors*, Vol. 3, No 1, pp.41–51.

Servaes, H., Tamayo, A. (2013), “The Impact of Corporate Social Responsibility on Firm Value: The Role of Customer Awareness”, *Management Science*, Vol. 59, No 5, pp.1045–1061.

Sokolova, M., Lapalme, G. (2009), “A Systematic Analysis of Performance Measures for Classification Tasks”, *Information Processing and Management*, Vol. 45, No 4, pp.427–437.

Taskin, D., Sariyer, G., Acar, E., Cagli, E.C. (2025), “Do Past ESG Scores Efficiently Predict Future ESG Performance?”, *Research in International Business and Finance*, Vol. 74, Article 102706.

Testa, F., Boiral, O., Iraldo, F. (2018), “Internalization of Environmental Practices and Institutional Complexity: Can Stakeholder Pressures Encourage Greenwashing?”, *Journal of Business Ethics*, Vol. 147, No 2, pp.287–307.

Acknowledgements. The paper has been supported by the Scientific Grant Agency of Slovak Republic under project VEGA No. 1/0479/23 *Research of circular consumer behaviour in the context of STP marketing model*, and VEGA No. 1/0120/25 *Research of paradigms and determinant processes and ESG implementation in the context of the required financial performance of companies and changes resulting from the CSRD directive*. This paper has been supported by the Scientific Grant Agency of Slovak Republic under contract No. APVV-23–0018. The authors would like to express their gratitude to the Scientific Grant Agency of The Ministry of Education, Research, Development and Youth of the Slovak Republic for financial support of this research and publication.

Detailed Description of the Use of Artificial Intelligence Tools: Additionally, we acknowledge the use of language model tools such as ChatGPT, which were employed to enhance the grammar, clarity, and readability of certain sections of this text. These tools provided supplementary assistance in ensuring the linguistic quality of the manuscript while maintaining the integrity of the original content.

KIEK SKAIČIUOKLĖS GALI FORMUOTI TVARUMĄ? ESG BALŲ PROGNOZAVIMAS TAIKANT MAŠININIO MOKYMOSI METODĄ

Hussam Musa, Frederik Rech, Zdenka Musová, Janka Grofčíková

Santrauka. Straipsnyje nagrinėjama, ar įmonių ESG (aplinkosaugos, socialinės atsakomybės ir valdysenos) reitingus galima prognozuoti naudojant kelių ankstesnių metų finansinius rodiklius. Tyrimo tikslas – įvertinti, kiek istorinė finansinė informacija paaiškina ESG veiklos rezultatus Slovakijos gamybos įmonėse. Analizė atlikta pasitelkus 974 Slovakijos gamybos įmonių imtį, 2023 m. ESG reitingus ir 2018–2022 m. finansinius duomenis. Pagrindiniame modelyje taikytas XGBoost algoritmas, rekursyvus požymių eliminavimas (*Recursive Feature Elimination*, RFE) ir SHAP analizė. Rezultatų patikimumui įvertinti pasitelktas PCA pagrįstas one-vs-rest XGBoost modelis. Tyrimo rezultatai atskleidė, kad istoriniai finansiniai rodikliai reikšmingi, tačiau jų nepakanka ESG veiklos rezultatams prognozuoti. Svarbiausi prognozuojantys veiksniai: finansinis svertas, likvidumas, gebėjimas aptarnauti skolą, įmonės amžius, su mokesčiais susiję rodikliai. Taip pat nustatyta, kad reikšmingi išlieka ir senesnių laikotarpių rodikliai. Tai leidžia teigti, kad ESG rezultatai atspindi ilgalaikes finansines tendencijas.

Alternatyvus modelis pagerino prognozavimo tikslumą, ypač aplinkosaugos (*Environmental*) ir valdysenos (*Governance*) dimensijose, tačiau bendro ESG balo prognozavimas išliko sudėtingesnis. Modeliai patikimiausiai prognozavo vidutinę ESG reitingų kategoriją, o kraštutinių kategorijų prognozavimo tikslumą ribojo klasių disbalansas ir ESG vertinimų ordinalinė struktūra. Daroma išvada, kad finansiniai duomenys gali būti naudingi prognozuojant ESG rezultatus, ypač atskirų ESG dimensijų lygmeniu, tačiau jie negali visiškai pakeisti kokybinės ir specifinės ESG informacijos.

Reikšminiai žodžiai: ESG; finansiniai rodikliai; mašininis mokymasis; ekstremalus gradientinis stiprinimas (Extreme Gradient Boosting, XGBoost); gamybos įmonės; Slovakija.

Appendix A.

Recursive Feature Selection (ESG Pillar)

Number of features	Accuracy	Kappa	Accuracy SD	Kappa SD	Selected
2	0.2288	0.033137	0.05662	0.07201	
4	0.2207	-0.002726	0.02929	0.03830	
6	0.2827	0.068937	0.03694	0.04843	
8	0.2830	0.062000	0.05153	0.06226	
10	0.2940	0.069951	0.06691	0.07858	
12	0.3144	0.094968	0.05573	0.06134	
14	0.2869	0.049546	0.04229	0.05043	
16	0.2979	0.057405	0.04052	0.05004	
18	0.3062	0.074448	0.04322	0.05195	
20	0.3050	0.067409	0.04128	0.04664	
22	0.2868	0.045784	0.03388	0.04258	
24	0.3075	0.070644	0.02528	0.03167	
26	0.3062	0.068917	0.04331	0.05007	
28	0.3173	0.083712	0.04055	0.05762	
30	**0.3296**	**0.096024**	**0.02859**	**0.03853**	*

Note: The asterisk (*) in the "Selected" column indicates the model configuration chosen for variable count 30.
Source: own calculations.

Appendix B.

Recursive Feature Selection (Environmental Pillar, E)

Number of features	Accuracy	Kappa	Accuracy SD	Kappa SD	Selected
2	0.2246	0.03600	0.06252	0.07622	
4	0.2786	0.08130	0.05913	0.06813	
6	0.2783	0.06929	0.04458	0.04651	
8	0.2867	0.06324	0.06252	0.07307	
10	0.2810	0.05219	0.05831	0.07077	
12	0.2852	0.04363	0.04698	0.05405	
14	0.2795	0.03633	0.06130	0.07945	
16	0.3307	0.07491	0.04806	0.07739	
18	0.3235	0.07905	0.05856	0.08121	
20	0.3457	0.08973	0.04764	0.07407	
22	0.3296	0.07396	0.05115	0.05588	
24	0.3516	0.09130	0.03995	0.05752	
26	0.3458	0.09116	0.05415	0.07712	
28	**0.3792**	**0.11844**	**0.04488**	**0.06740**	*
30	0.3724	0.11278	0.04350	0.06412	

Note: The asterisk (*) in the "Selected" column indicates the model configuration chosen for variable count 28.
Source: own calculations

Appendix C

Recursive Feature Selection (Social Pillar, S)

Number of features	Accuracy	Kappa	Accuracy SD	Kappa SD	Selected
2	0.2055	0.01694	0.03885	0.04858	
4	0.2330	0.02587	0.04616	0.04733	
6	0.2607	0.04607	0.04298	0.05039	
8	0.2815	0.04497	0.03624	0.04435	
10	0.2360	0.00668	0.06618	0.07987	
12	0.2676	0.02770	0.04996	0.04958	
14	0.2760	0.03202	0.05155	0.06740	
16	0.2800	0.02483	0.04746	0.06373	
18	0.2964	0.05015	0.04806	0.05011	
20	0.2978	0.04048	0.05754	0.06982	
22	0.3076	0.04824	0.04999	0.06306	
24	0.3035	0.04440	0.04009	0.05975	
26	0.3174	0.05374	0.05228	0.07099	
28	0.3049	0.04142	0.04284	0.06010	
30	**0.3228**	**0.06250**	**0.05468**	**0.06922**	*

Note: The asterisk (*) in the "Selected" column indicates the model configuration chosen for variable count 30.
Source: own calculations

Appendix D

Recursive Feature Selection (Governance Pillar, G)

Number of features	Accuracy	Kappa	Accuracy SD	Kappa SD	Selected
2	0.2081	0.00299	0.04623	0.03553	
4	0.2386	0.04193	0.05919	0.06811	
6	0.2660	0.04129	0.05797	0.06214	
8	0.2800	0.05100	0.08005	0.07194	
10	0.3114	0.06335	0.06961	0.07123	
12	0.3308	0.06660	0.06273	0.06389	
14	0.3337	0.05784	0.05553	0.07408	
16	0.3324	0.06670	0.04889	0.05025	
18	0.3449	0.06419	0.04671	0.05576	
20	0.3421	0.05671	0.06735	0.08183	
22	0.3504	0.05271	0.06302	0.06729	
24	0.3408	0.06067	0.05745	0.05775	
26	0.3407	0.04880	0.05775	0.05698	
28	0.3562	0.05323	0.09536	0.10430	
30	**0.3671**	**0.07444**	**0.05511**	**0.05836**	*

Note: The asterisk (*) in the "Selected" column indicates the model configuration chosen for variable count 30.
Source: own calculations

Appendix E

PCA-Based One-vs-Rest XGBoost Model (Recall-Weighted Ensemble)

Target	Best split	Train N	Test N	PCA components	PCA variance explained (%)	Accuracy	Macro recall	Recall 1	Recall 2	Recall 3	Recall 4	Recall 5
E	60/40	544	362	67	90.2	0.492	0.221	0.091	0.020	0.856	0.000	0.136
S	60/40	545	361	67	90.2	0.382	0.196	0.000	0.333	0.648	0.000	0.000
G	80/20	726	180	70	90.1	0.550	0.201	0.000	0.000	0.961	0.000	0.043
ESG	70/30	636	270	69	90.3	0.348	0.245	0.000	0.190	0.564	0.271	0.200

Note: Final classes are assigned using one-vs-rest probabilities weighted by each binary model's cross-validated recall. PCA components indicate the number of retained principal components required to explain approximately 90 percent of the variance in the financial predictor set. Accuracy measures overall correct classification. Macro recall is computed as unweighted averages across the five ESG rating classes. Recall 1 to Recall 5 reports class-specific recall for each ESG score category.
Source: own calculations.

Appendix F

Split-Sensitivity Results for the PCA-Based One-vs-Rest XGBoost Model (Recall-Weighted Ensemble)

Target	Split	Train N	Test N	PCA components (variance explained, %)	Accuracy	Macro recall	Recall 1	Recall 2	Recall 3	Recall 4	Recall 5
E	60/40	544	362	67 (90.2)	0.492	0.221	0.091	0.020	0.856	0	0.136
E	70/30	635	271	69 (90.2)	0.487	0.196	0.039	0	0.890	0	0.053
E	80/20	725	181	68 (90.1)	0.503	0.221	0.063	0	0.926	0	0.115
E	90/10	816	90	71 (90.1)	0.511	0.224	0	0.182	0.936	0	0
S	60/40	545	361	67 (90.2)	0.382	0.196	0	0.333	0.648	0	0
S	70/30	636	270	69 (90)	0.430	0.208	0	0.213	0.826	0	0
S	80/20	726	180	70 (90)	0.433	0.211	0	0.255	0.803	0	0
S	90/10	817	89	74 (90.3)	0.416	0.201	0	0.231	0.775	0	0
G	60/40	545	361	64 (90.3)	0.576	0.196	0	0	0.981	0	0
G	70/30	635	271	67 (90.1)	0.557	0.201	0	0.029	0.974	0	0
G	80/20	726	180	70 (90.1)	0.550	0.201	0	0	0.961	0	0.044
G	90/10	816	90	73 (90.2)	0.556	0.189	0	0	0.943	0	0
ESG	60/40	546	360	66 (90.3)	0.294	0.193	0	0.198	0.485	0.255	0.025
ESG	70/30	636	270	69 (90.3)	0.348	0.245	0	0.191	0.564	0.271	0.200
ESG	80/20	726	180	71 (90.1)	0.344	0.216	0	0.256	0.612	0.213	0
ESG	90/10	817	89	72 (90.1)	0.281	0.178	0	0.200	0.515	0.174	0

Source: own calculations