Explainable Multimodal Machine Learning for ESG Rating Prediction

ABSTRACT

Environmental, Social, and Governance (ESG) ratings have emerged as key indicators for analysing the sustainability and ethical vision of companies. Yet current ESG rating systems suffer from inconsistency and limited transparency which undermines their reliability. Furthermore, ESG data employs a large corpus of text and various numerical finance data which usually consumes substantial manual effort in score prediction. To address these challenges, we propose an explainable multimodal machine learning based system that combines unstructured corporate disclosures and structured financial indicators.

Our approach integrates annual reports from S&P 500 companies (sourced from SEC.gov using custom scripts) with financial fundamentals and market data (sourced from Yahoo Finance). Textual disclosures are processed using section aware ESG-BERT encoder, while numerical data are modelled with XGBoost. These complementary signals are fused through a simple stacking meta learner.

We evaluate performance using GroupKFold by firm to avoid leakage, reporting R^2, Mean Absolute Error (MAE), and Mean Squared Error (MSE). Results show that numeric data provides a strong predictive backbone, while text features add consistent and modest improvements, specifically on social scores. The stacked model archives $R^2 \approx$ 0.2 (environmental), 0.16 (social), 0.07 (governance), outperforming text only models and modestly improving on numeric only baselines.

Explainability is a first-class objective in this study. SHAP analysis identifies both global and firm level numeric drivers, while stacking weights quantify modality reliance. Moreover, we apply sentence level attribution to surface disclosure snippets most influential in the predictions. This offers interpretable connections between ESG outcomes and textual evidence which is much underexplored in existing studies.

These findings illustrate the feasibility of explainable multimodal learning for automated ESG scoring. Our research establishes both a performance baseline and interpretability tools that can support more transparent and reliable sustainability evaluations.

Keywords: ESG rating, multimodal learning, explainable AI, sustainability, XGBoost, ESG-BERT

Model	Environmental R^2	Social R^2	Governance R^2
ESG-BERT (Text only)	0.0684	0.1489	0.0504
XGBoost (Numeric only)	0.1844	0.1222	0.0556
Stacked meta learner	0.2001	0.1567	0.0728

Table 1: Results comparison

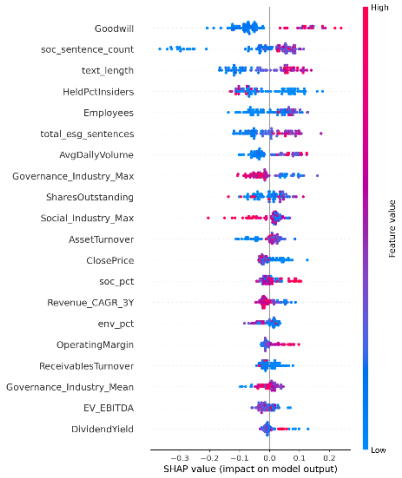

Figure 1: Global SHAP summary – Social score for all companies

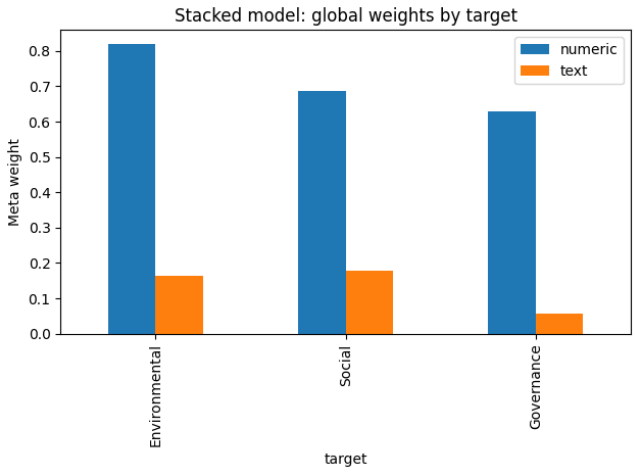

Figure 2: Global weights by target – Stacking meta learner