Machine Learning-Based Examination of ESG Factors in Stock Predictions

Daniel Ślusarczyk, Paweł Tęcza, Daniel Stańkowski

Supervisor: dr inż. Grzegorz Sarwas

Introduction

The increasing emphasis on sustainability and responsible investment has significantly influenced financial markets and corporate decision-making. With environmental, social, and governance criteria (ESG) gaining importance in investment decision-making, questions remain about their actual predictive value. **Can sustainability scores really predict market success - or are they just a green illusion?** The predictive value of environmental scores is examined using machine learning techniques to determine their true impact on stock return forecasting.

Data

- Dataset over 90% of global market capitalization.
- Combined financial and **ESG data**, focusing on **CO_2 emissions**.
- Key carbon-related indicators were manually selected.
- Significant gaps existed in ESG data, especially CO_2 variables.
- Financial data was more complete and enriched the dataset.
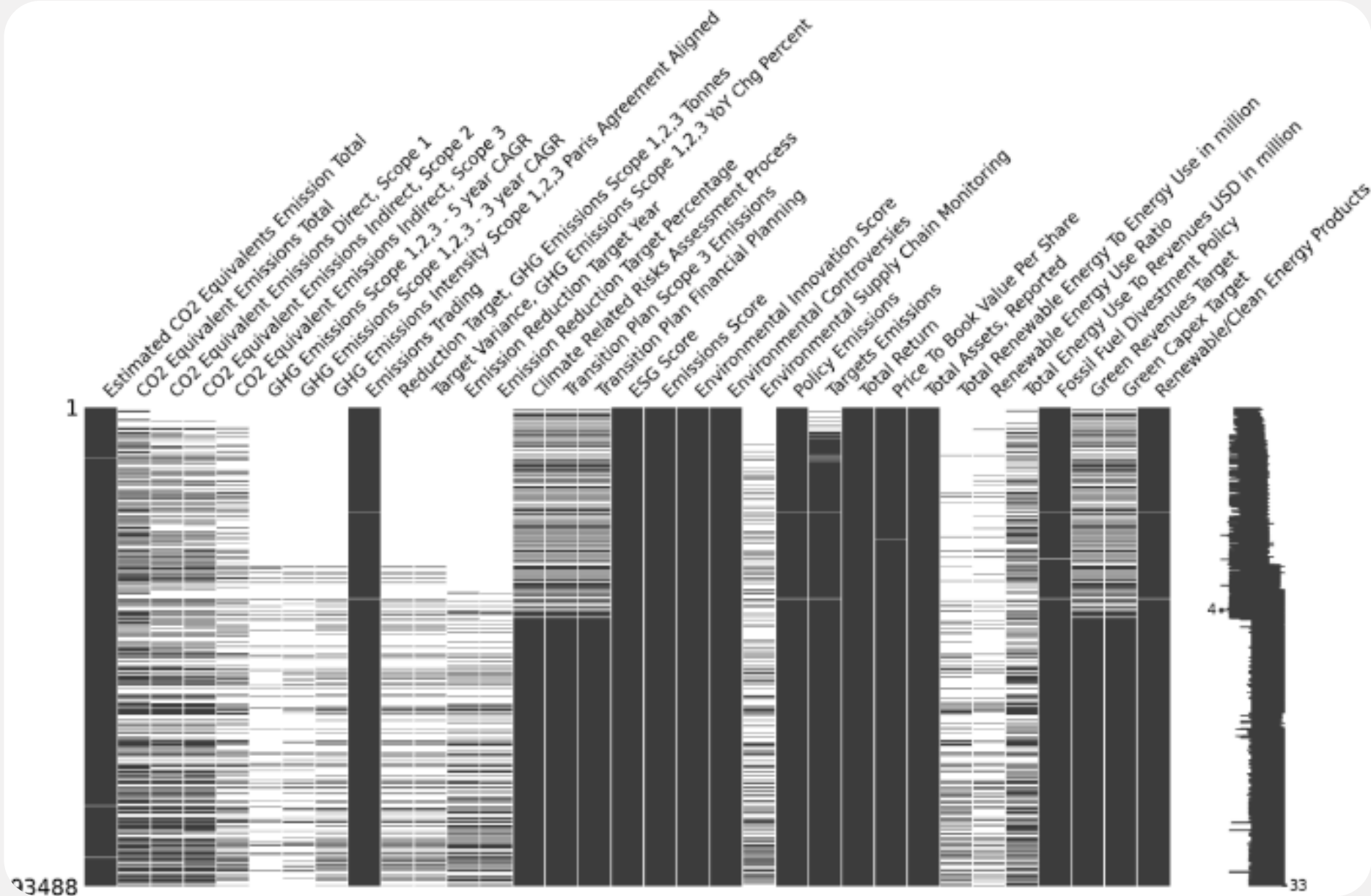
Explainable AI

Black Box + **XAI techniques** = **Interpretable model**

SHAP analysis showed that financial variables were the most influential, while ESG factors had minimal impact.
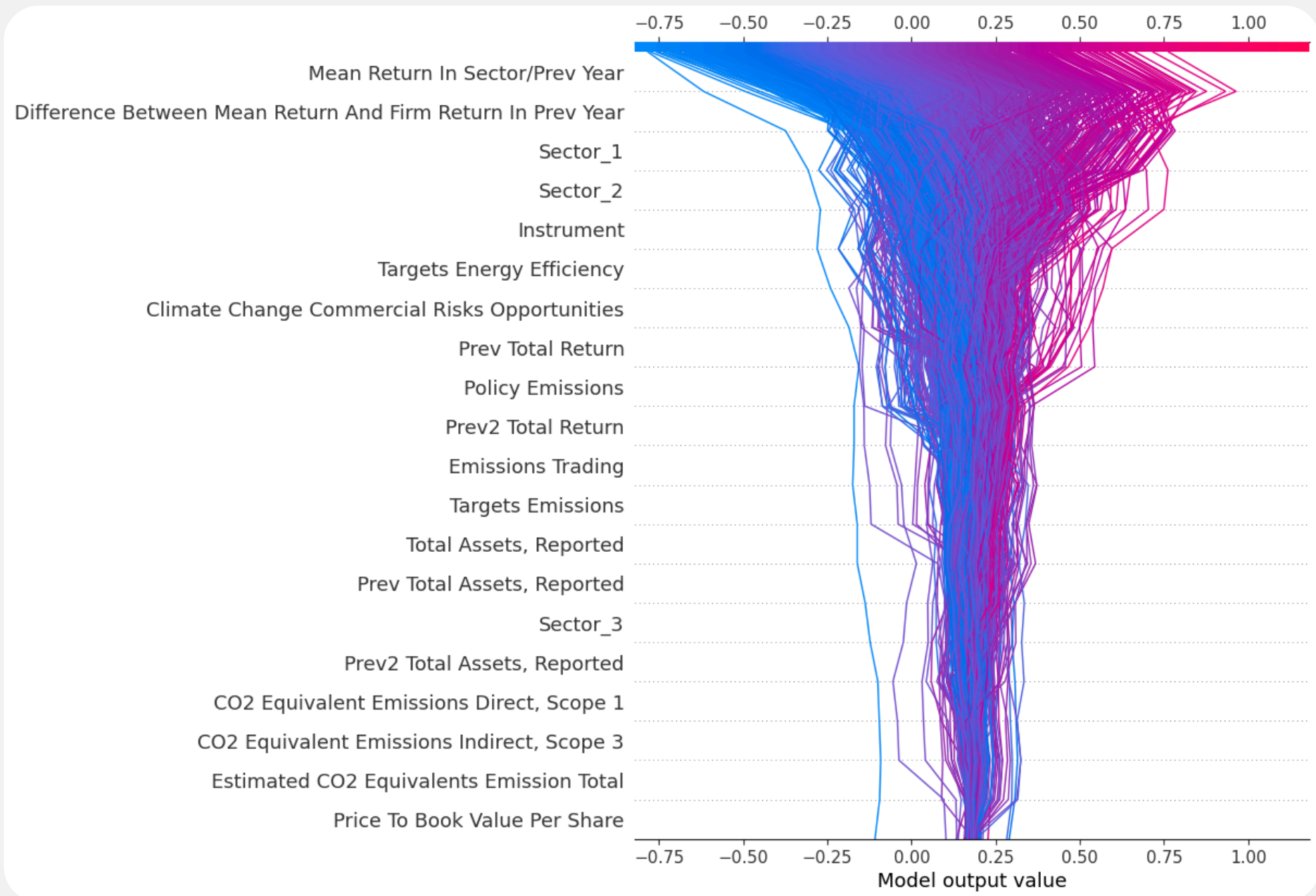

Interpretable Models

To assess the impact of ESG factors on stock return predictions, we employed interpretable models including **XGBoost** and **decision trees**. The results, summarized in the adjacent table, highlight that ESG inclusion does not consistently improve model accuracy, and financial indicators remain more reliable predictors.

RESULTS FOR INTERPRETABLE MODELS

Model	Dataset variant	MAE	
		without ESG	with ESG
XGboost	Entire dataset	14.84	**14.63**
	Energy sector only (test set 2022/23)	28.06	**27.12**
Decision Tree	Entire dataset	**19.36**	20.09
	Entire dataset (with missing value)	39.85	**38.60**

Neural Networks

- Usage of **MLP networks** to model **c**omplex, non-linear relationships between financial and ESG variables.
- Applied classic supervised learning instead of time-series forecasting due to limited and irregular data.
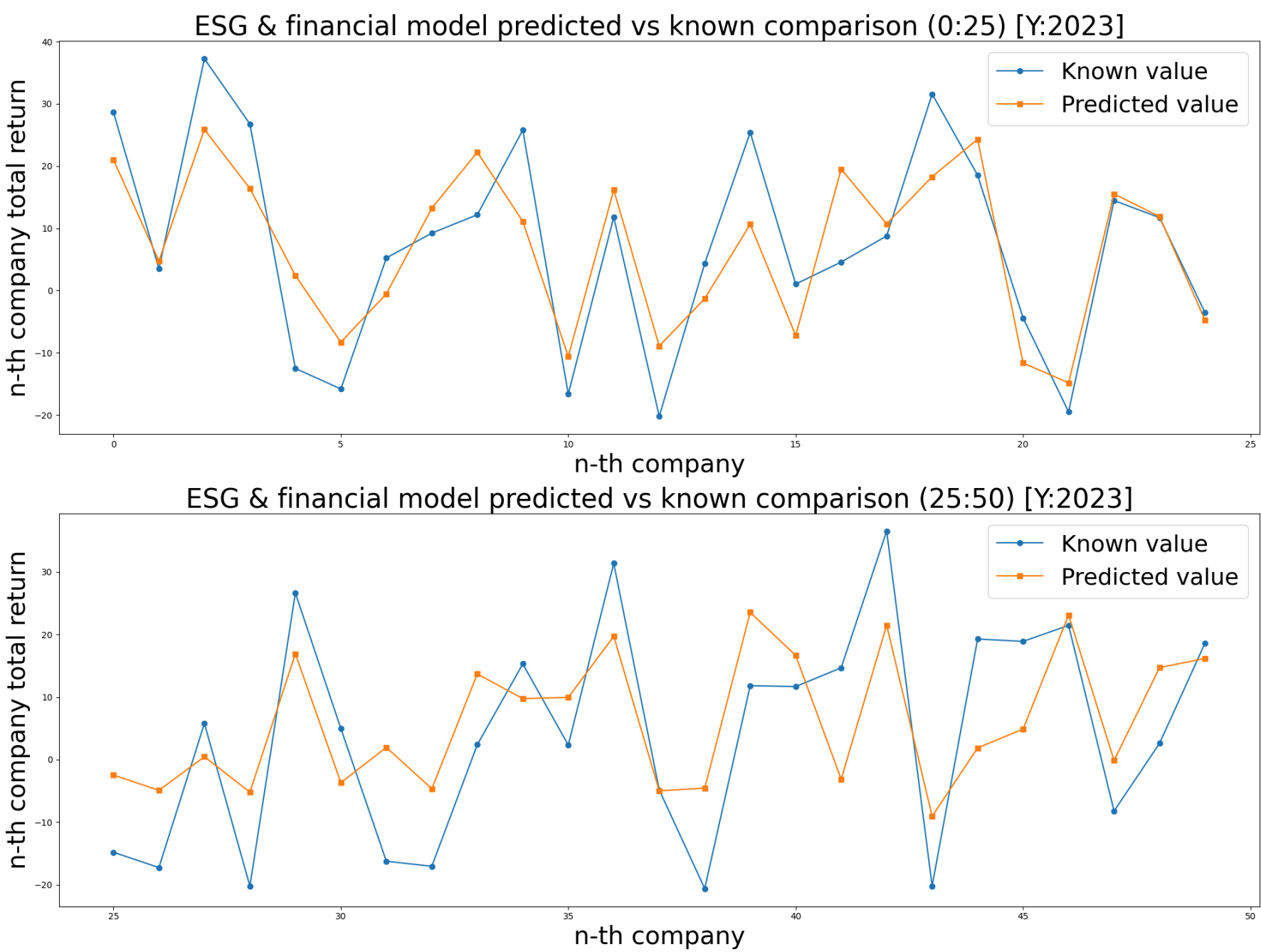
MEAN ERRORS PER MODEL VARIANT

Model Variant	Mean Errors	
	RMSE	*MAE*
ESG & financial	13.11	11.09
Financial only	12.43	9.76
Mostly ESG	15.65	13.15

Conclusion

- **The model based solely on financial data outperformed both the combined and ESG-focused models in stock predictions.**
- Financial variables remained the strongest predictors, consistently outperforming models that included ESG data.
- Significant data gaps and inconsistencies in ESG disclosures reduced model reliability and accuracy.
- Enhancing ESG data quality and applying alternative modeling approaches are necessary to better assess the financial relevance of ESG factors.