ESG ratings explainability through Machine Learning techniques

Alessandro Del Vitto*, Daniele Marazzina*, Davide Stocco*
*Department of Mathematics, Politecnico di Milano
Ispra, 13-14 July 2023

The problem with ESG Ratings

ESG ratings are quantitative assessments of companies' performance and commitment from the Environmental (E), Social (S), and Governance (G) points of view.
In the last two decades ESG ratings have become very popular tools in the financial industry. However, the **reliability**, **consistency**, and **effectiveness** of such instruments **are still under discussion**.

- Berg et al. (2022) document the presence of great divergences between the ratings issued by different data providers. (1)
- Billio et al. (2021) analyze the consequences of the divergences showing that the confusion disperses the effects of investors' preferences. (2)
- Bams and van der Kroft (2022) highlight information asymmetries and question possible ESG rating inflation phenomenon incentivized by the assessment systems. (3)

Our study concerns and focuses on the **lack of transparency problem**:

ESG scores are issued by rating agencies exploiting **proprietary algorithms** that are not publicly disclosed due to insufficient regulation and transparency.
It is impossible for investors and policymakers **to effectively assess the reliability of the scoring processes** exploited by data providers.

Refinitiv ESG Scores

- Refinitiv computes and collects more than 630 sustainability metrics, but only 186 of these variables are considered in the scoring process.
- The three pillar scores are computed independently and then they are linearly combined to obtain a final overall percentile score.
- Refinitiv computes the E and S scores relatively to the company's **industry sector**, while it determines the G score relatively to the firm's **country of incorporation**.

We lack two fundamental information:
1. Which of the 630+ features are actively relevant in the scoring process.
2. Which are the precise weights of the aggregation rule.

Methodology

DATA
We consider a snapshot of Refinitiv's 2022 ratings to avoid model inconsistencies (see 4).
We consider the Financial, the Manufacturing and the Information industries for the E and S scores. We consider the EU, the USA and China as countries of incorporation for the G score.

METHODOLOGY
1. We pre-process the available ESG information.
2. We **replicate the proprietary issuance model** with ML regression techniques.
3. We **explain the replication models** exploiting ML interpretability techniques.
   3a) **Local** Interpretability: interpret and motivate individual companies' ESG ratings.
   3b) **Global** Interpretability: understand and explain, in general, the inner workings of the proprietary grey-box issuance model.

MODELS
We consider both white-box and black-box regression models, since the former are easier to interpret while the latter are, in general, more powerful.
1. **White Box Models** (WB): Linear regression, Lasso and Ridge, Decision Tree, KNearestNeighbors.
2. **Black Box Models** (BB): Random Forests, AdaBoost ensemble models, Artificial Neural Networks.

TRAINING
We find the best hyperparameters combinations with a *GridSearch* algorithm.
We optimize the model's RMSE thanks to a 10-Fold *Cross-Validation* procedure.
We value the models measuring their RMSE on a held-out set of never-seen samples.

Interpretability

While the interpretation of WB models is straightforward, BB models need *ad hoc* techniques.
We choose to explain BB models with the **Shapley Values technique** taken from cooperative game theory through the SHAP python library (SHapley Additive exPlainations).

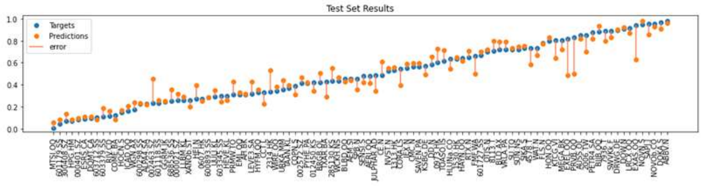
LOCAL INTERPRETABILITY
The Shapley Values technique allows us to **estimate each feature value's impact** on the company's output score.

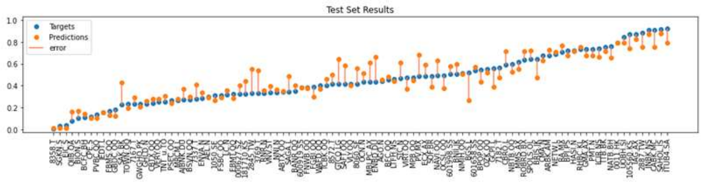
GLOBAL INTERPRETABILITY
The Shapley Values technique allows us to obtain a **global ordering of the model's features** comparing their mean Shapley Values on the test set samples.

Regression Results

- Results show that it is possible to predict E, S, and G scores with satisfactory accuracy.
- We notice the presence of unlearnable noise which is unevenly spread across the three pillars.
- The E score is the easiest one to predict, maybe thanks to a more well-defined scope.

	I	II	III
Environmental	0.078	0.084	0.071
Social	0.089	0.096	0.090
Governance	0.075	0.103	0.092





Interpretability Results

- It is possible to explain thanks to the Shapley Values technique even BB replication models.
- We can meaningfully **interpret and motivate individual companies ESG scores**.
- We can estimate which are the **most relevant features** in Refinitiv's proprietary model and how much they explain of the final score.
- The interpretation is consistent between WB and BB models (Ridge vs ANN).
- The most relevant features we have found reflect differences in how different sectors are managed by the proprietary algorithm.

Conclusions & Remarks

We contribute to the existing ESG literature in the attempt **to contrast the problem of lack of transparency** with a **new ML approach**.

In particular, we apply ML **regression** and **interpretability** techniques in the **specific case of the Refinitiv data provider** to **find**, **understand** and **explain** the proprietary model used to issue sustainability scores.

In general, we propose a viable tool for investors and policymakers to effectively assess the reliability of the scoring systems exploited by rating agencies.

- Our methodology can be applied whenever the granular ESG data exploited in the scoring process are available.
- The results can be extremely useful to investors and policymakers to make better and more informed decisions.
- Companies can exploit these techniques to understand and improve more effectively their sustainability ratings.
- By repeating the procedure for other rating agencies, we can further investigate the divergences between the scoring systems employed by different data providers.

References

1. Florian Berg, Julian K¨olbel, and Roberto Rigobon. Aggregate confusion: The divergence of ESG ratings. 2022
2. Monica Billio, Michele Costola, Iva Hristova, Carmelo Latino, and Loriana Pelizzon. Inside the ESG ratings: (dis)agreement and performance. 2021
3. Dennis Bams and Bram van der Kroft. Divestment, information asymmetries, and inflated ESG ratings. 2022
4. Florian Berg, Kornelia Fabisik, and Zacharias Sautner. Is history repeating itself? The (un)predictable past of ESG ratings. 2021
5. Valeria D'Amato, Rita D'Ecclesia, and Susanna Levantesi. Esg score prediction through random forest algorithm. 2022

alessandro.delvitto@mail.polimi.it
daniele.marazzina@polimi.it
davide.stocco@polimi.it