Estimating Corporate Scope 1 Emissions Using Tree-Based Machine Learning Methods

Maida Hadziosmanovic

Concordia University, Montreal, Canada

Elham Kheradmand

Université de Montréal, Montreal, Canada

Introduction

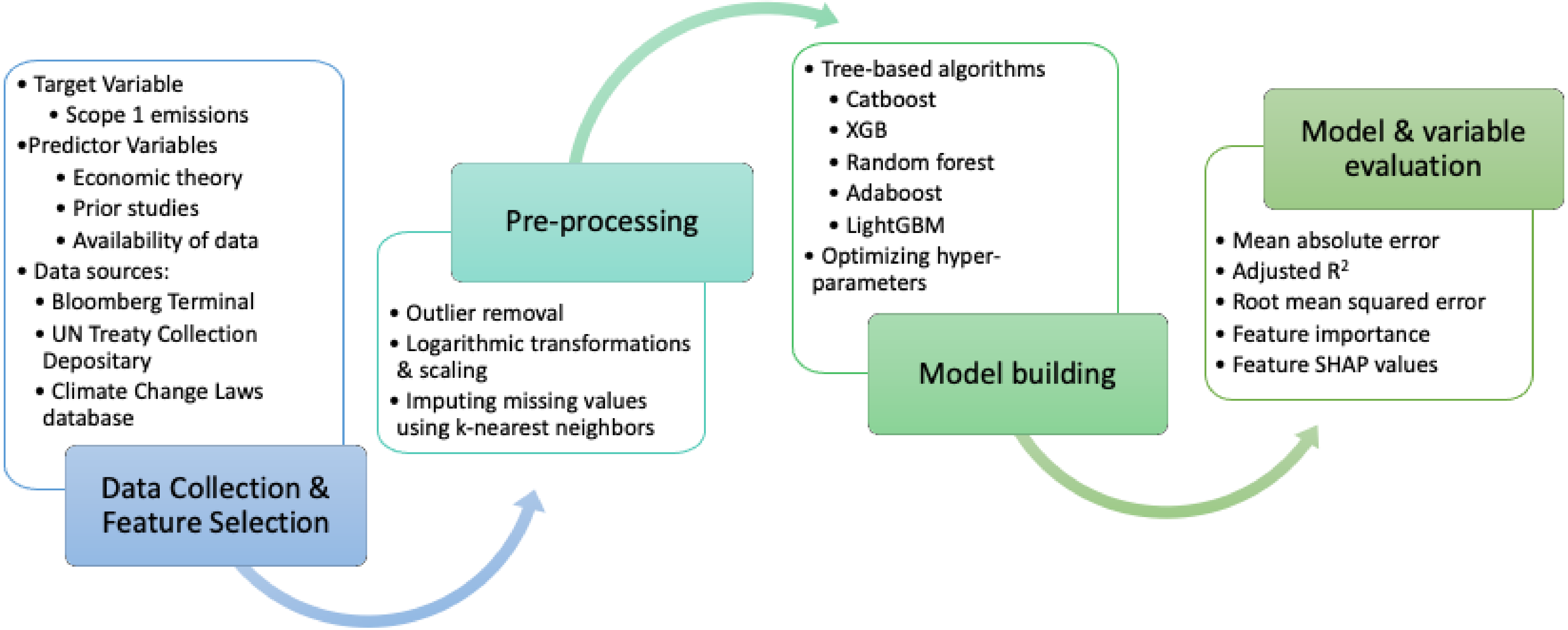
- <5% of public companies disclose their **direct (Scope 1)** GHG emissions[1]
- Difficult to **reconcile company-level emissions with national & global GHG estimates** – *essential for assessing decarbonization efforts*[2]
- There is a need to track & corroborate emissions **target** and **reduction** claims[3]

Related Works

- GHG estimation **models can fill the gap** in corporate emissions data
- Three machine learning (ML) models from the literature[4,5,6]
- **Tree-based** algorithms have shown the best results
- Limitations of these ML models:
  - Large set of features rendering model complex & difficult to replicate
  - Feature data is not easily accessible or available
  - Estimate other scopes of emissions (e.g., scopes 2, 3)
- ML techniques for estimating corporate emissions are in the **early stages**

Objective

- Fill in the corpporate emissions data gap by:
  - training a series of models based on **decision trees** for the **estimation** of company-level **Scope 1 emissions**

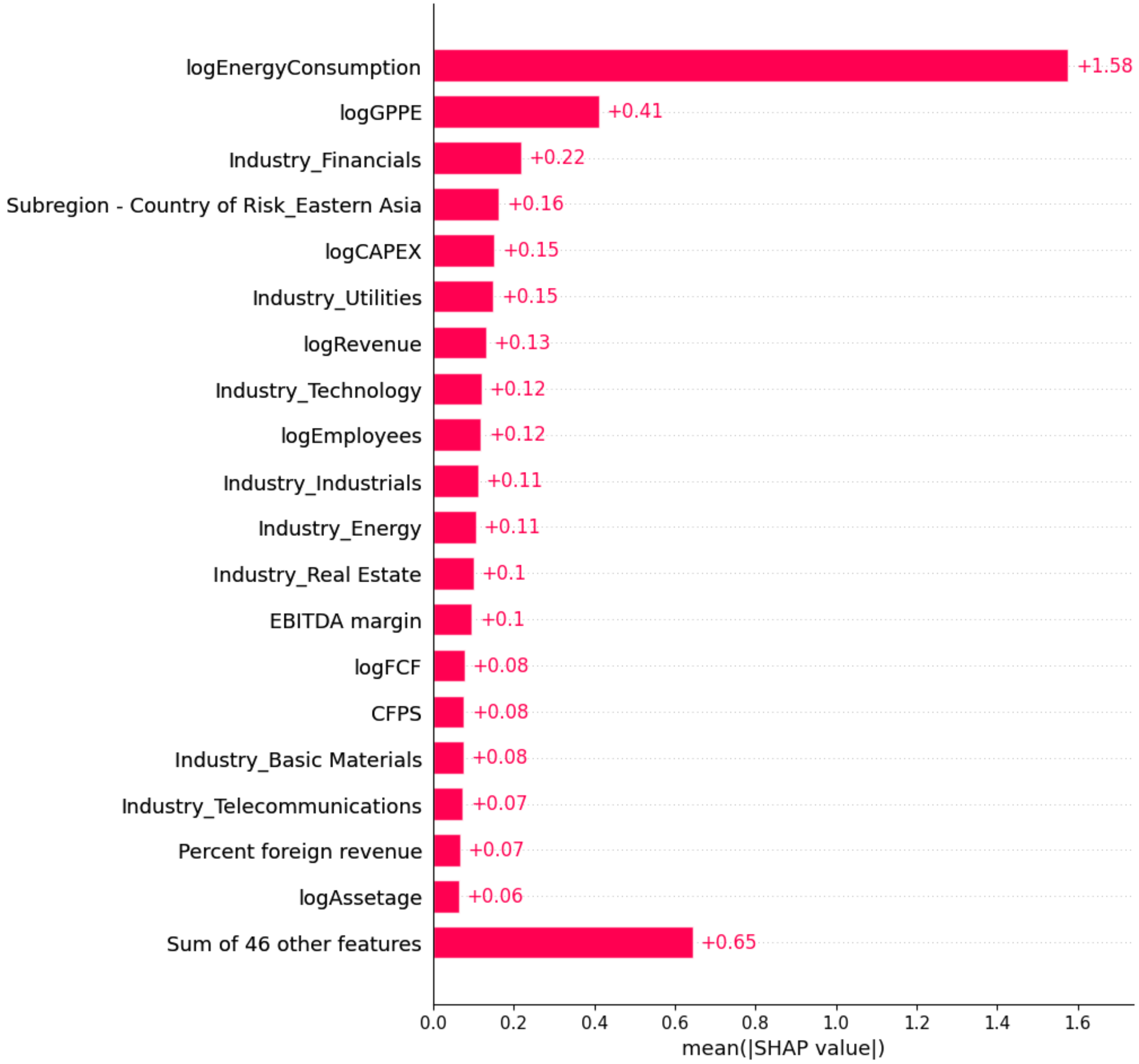


Data & Methods

Results

Model	RMSE	MSE	MAE	MAPE	Adjusted R2	MAE (Nguyen et al. 2021)	MAE improvement against benchmark model (%)
Catboost-1	1.43	2.03	0.96	0.32	0.81	n/a	6.80%
Catboost-2	1.41	1.99	0.96	0.29	0.82	n/a	6.80%
XGB	1.30	1.69	0.83	0.29	0.84	1.03	19.42%
Random Forest	1.32	1.74	0.87	0.30	0.84	1.03	15.53%
Adaboost	1.94	3.77	1.38	0.36	0.661	n/a	-33.98%
LightGBM	1.31	1.73	0.86	0.30	0.84	n/a	16.50%

Results



logEnergyConsumption +1.58
logGPPE +0.41
Industry_Financials +0.22
Subregion - Country of Risk_Eastern Asia +0.16
logCAPEX +0.15
Industry_Utilities +0.15
logRevenue +0.13
Industry_Technology +0.12
logEmployees +0.12
Industry_Industrials +0.11
Industry_Energy +0.11
Industry_Real Estate +0.1
EBITDA margin +0.1
logFCF +0.08
CFPS +0.08
Industry_Basic Materials +0.08
Industry_Telecommunications +0.07
Percent foreign revenue +0.07
logAssetage +0.06
Sum of 46 other features +0.65
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
mean(|SHAP value|)

Discussion

- Result show **significant improvement** in accuracy of our XGBoost model compared to benchmark model
- We show that Scope 1 emissions can be estimated with models of **lower complexity & greater computational efficiency**
- Model can be used for **data gap-filling** – allows for better GHG accounting & tracking

"*what gets measured, gets managed*"

References

1. Hadziosmanovic, M., Lloyd, S. M., Bjørn, A., Paquin, R. L., Mengis, N., & Matthews, H. D. (2022). Using cumulative carbon budgets and corporate carbon disclosure to inform ambitious corporate emissions targets and long-term mitigation pathways. Journal of Industrial Ecology. 1-13. https://doi.org/10.1111/jiec.13322
2. Luers, A., Yona, L., Field, C. B., Jackson, R. B., Mach, K. J., Cashore, B. W., ... & Joppa, L. (2022). Make greenhouse-gas accounting reliable—build interoperable systems. Nature.
3. NewClimate Institute. (2022). Corporate Climate Responsibility Monitor. https://newclimate.org/sites/default/files/2022-06/CorporateClimateResponsibilityMonitor2022.pdf
4. Han, Y., Gopal, A., Ouyang, L., & Key, A. (2021). Estimation of Corporate Greenhouse Gas Emissions via Machine Learning. arXiv preprint arXiv:2109.04318.
5. Serafeim, G., & Velez Caicedo, G. (2022). Machine Learning Models for Prediction of Scope 3 Carbon Emissions. Available at SSRN.
6. Nguyen, Q., Diaz-Rainey, I., and Kuruppuarachchi, D. Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. Energy Economics, 95:105129, 2021. ISSN 0140-9883. doi: https://doi:org/10:1016/j:eneco:2021:105129