Estimation of Corporate Greenhouse Gas Emissions via Machine Learning

Tackling Climate Change with Machine Learning Workshop at ICML 2021
July 23, 2021

You Han, Achintya Gopal, Liwen Ouyang, and Aaron Key

TechAtBloomberg.com

Bloomberg

Background

- Paris Climate Agreement
- Tiny portion of companies disclosing Greenhouse Gas Emissions (GHG)

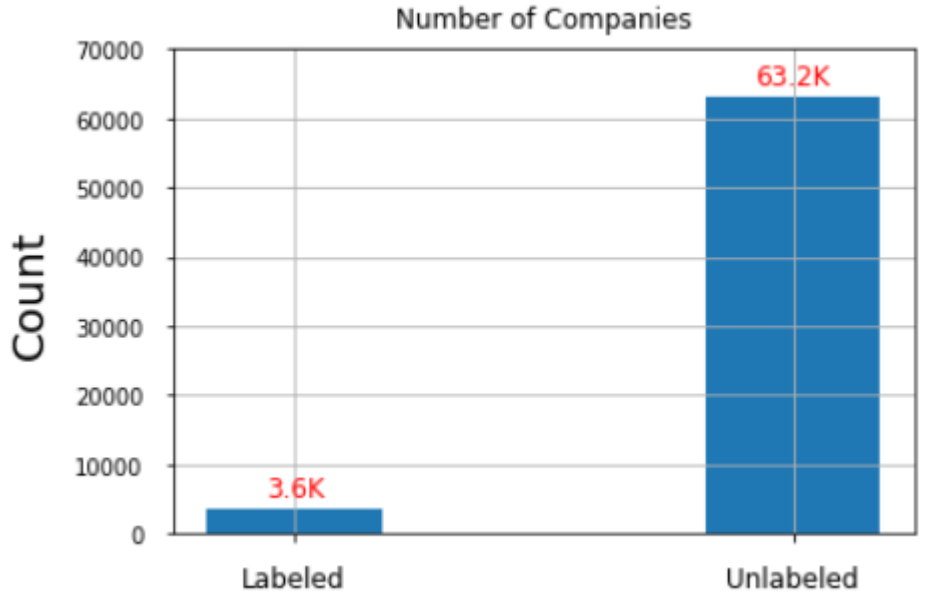
LONG_COMP_NAME	FISCAL_YEAR	GHG_SCOPE_1	GHG_SCOPE_2	GHG_SCOPE_3
Boeing Co/The	2010	NaN	NaN	NaN
Boeing Co/The	2011	NaN	NaN	NaN
Boeing Co/The	2012	NaN	NaN	NaN
Boeing Co/The	2013	672.409	1130.97	326.284
Boeing Co/The	2014	684.535	1167.35	330.693
Boeing Co/The	2015	359.700	1161.00	185.300
Boeing Co/The	2016	178.000	1074.00	NaN
Boeing Co/The	2017	235.000	896.00	NaN
Boeing Co/The	2018	224.000	892.00	NaN
Boeing Co/The	2019	NaN	NaN	NaN
Boeing Co/The	2020	NaN	NaN	NaN

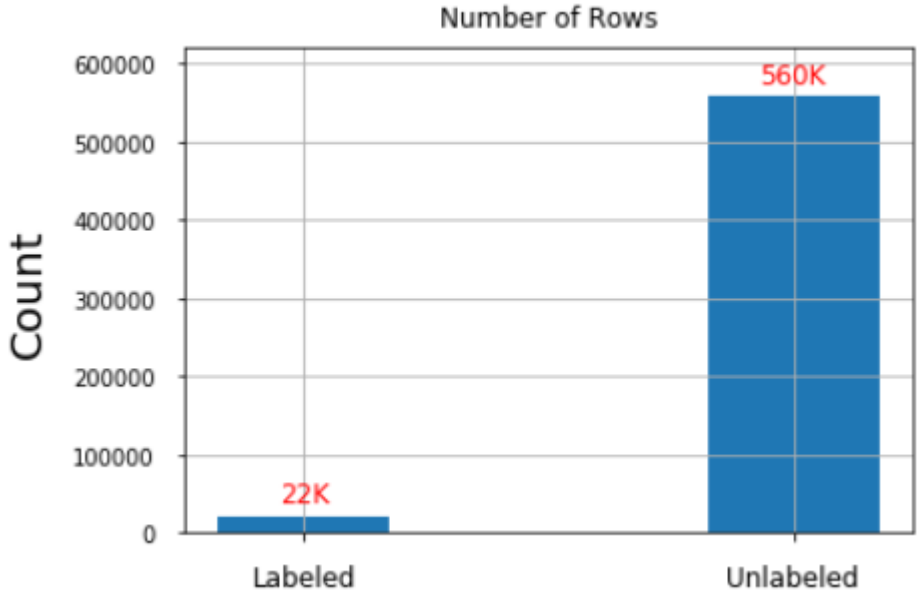
LONG_COMP_NAME	FISCAL_YEAR	GHG_SCOPE_1	GHG_SCOPE_2	GHG_SCOPE_3
Rosneft Oil Co PJSC	2010	NaN	NaN	NaN
Rosneft Oil Co PJSC	2011	NaN	NaN	NaN
Rosneft Oil Co PJSC	2012	NaN	NaN	NaN
Rosneft Oil Co PJSC	2013	NaN	NaN	NaN
Rosneft Oil Co PJSC	2014	NaN	NaN	NaN
Rosneft Oil Co PJSC	2015	NaN	NaN	NaN
Rosneft Oil Co PJSC	2016	NaN	NaN	NaN
Rosneft Oil Co PJSC	2017	NaN	NaN	NaN
Rosneft Oil Co PJSC	2018	NaN	NaN	NaN
Rosneft Oil Co PJSC	2019	NaN	NaN	NaN
Rosneft Oil Co PJSC	2020	NaN	NaN	NaN

TechAtBloomberg.com

Bloomberg

Imbalanced Labeled and Unlabeled Data

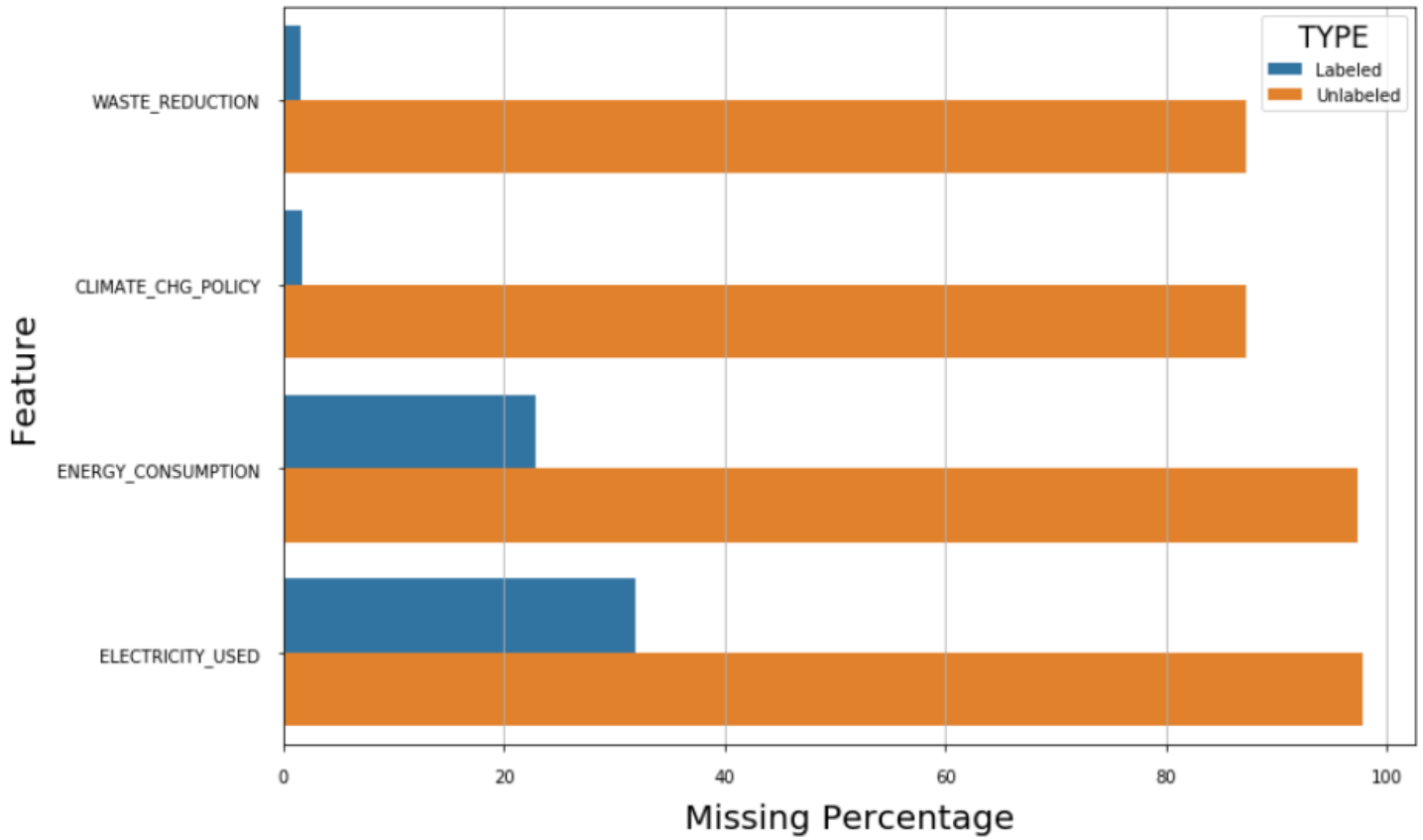




TechAtBloomberg.com

Bloomberg

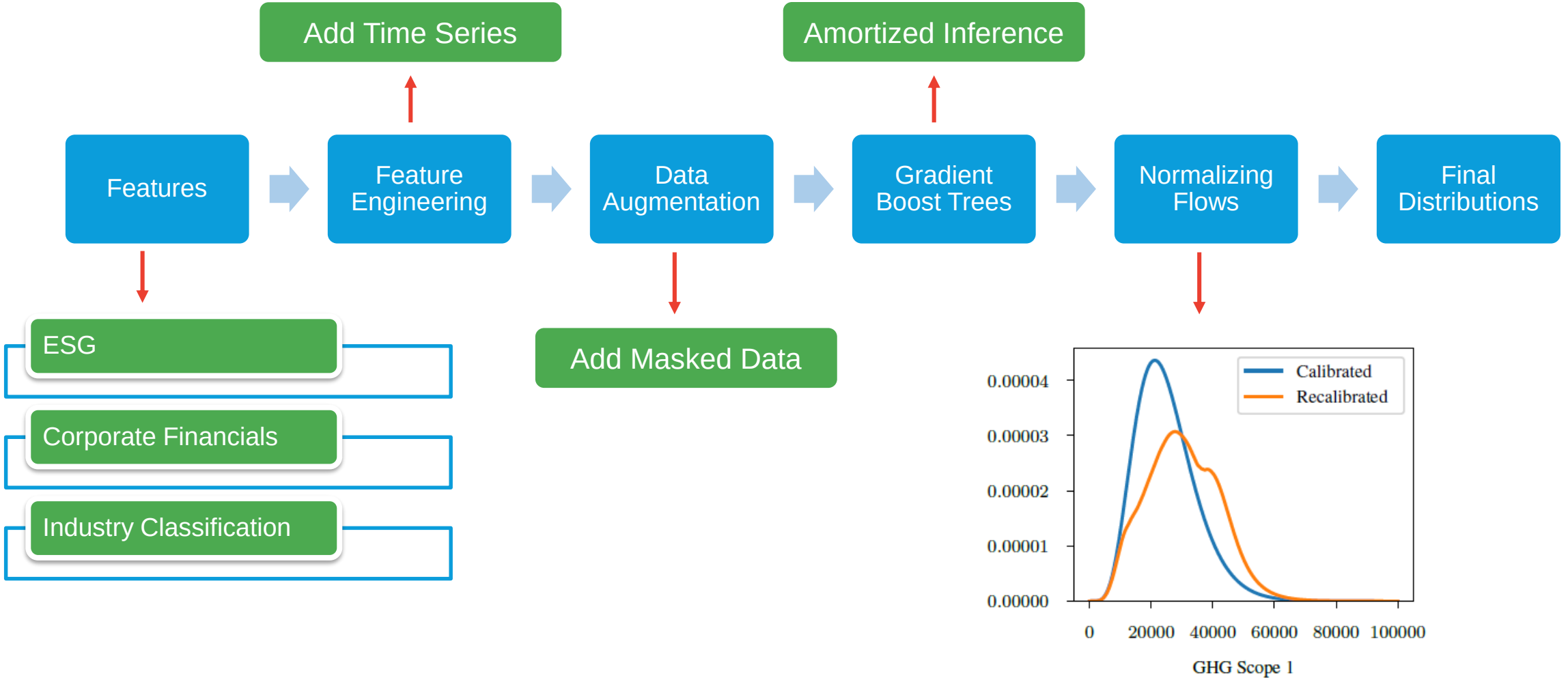
High and Different Feature Missing Rates



TechAtBloomberg.com

Bloomberg

Model Design



TechAtBloomberg.com

Bloomberg

Performance Evaluation

	Simple	Baseline	Recalib
Unmasked Data	0.1620	**0.3705**	0.2730
Masked Data	0.1620	0.1830	**0.2648**

Table 1. Precision Comparison

	Simple	Baseline	Recalib
Unmasked Data	8854.89	9184.00	**6657.00**
Masked Data	8854.89	9690.46	**6744.02**

Table 2. RMSE Comparison

Performance Evaluation

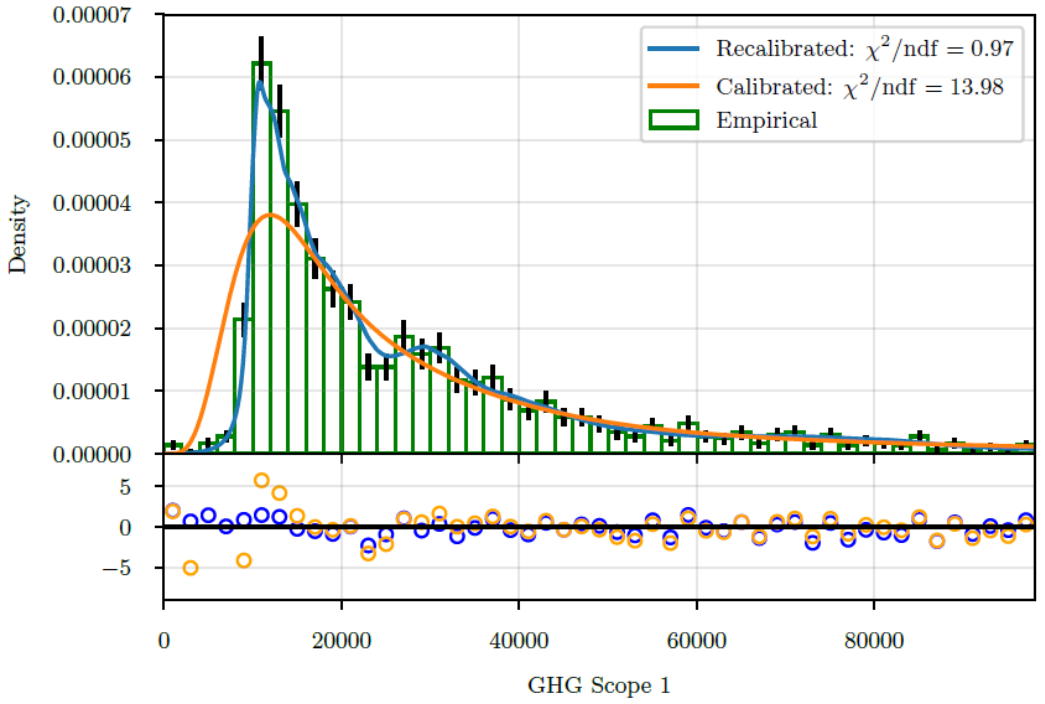


Figure 3. Comparison of the predicted mixture distribution for before and after recalibration versus observations. The lower panel shows the distance between the predicted density and the observed density, normalized by the standard error of the counts in the bin.

TechAtBloomberg.com

Bloomberg

Conclusions and Future Works

- Conclusions
  — An amortized inference model is proposed to estimate distributions of companies' GHG emissions
  — The proposed model outperforms the baseline models

- Future Work
  — Scope 3
  — Supply chain and factory data

TechAtBloomberg.com

Bloomberg

Thank you!

https://www.bloomberg.com/careers

Questions?

TechAtBloomberg.com

Bloomberg