Measuring firms' ESG performance using NLP

Yuqi Zhang[1], Sunny Son[2], Shuming Zhang[2]

[1]Center for Data Science, New York University

[2]Leonard N. Stern School of Business, New York University

Overview

- ESG (environmental, social, and corporate governance) becomes an increasingly important factor in evaluating firms' performance
- Financial institutes developed ESG metrics based on financial expertise and more inclusive documents
- Individual investors instead have less information to refer to and prefer sharing ESG-related information in individual investor communities.
- Different behaviors create information disparity compared to the institutes.

- In our research, we contributed to
  - 1. develop a preliminary ESG-Reddit dataset
  - 2. finetune FinBERT with ESG-related social media context (FinBERT-ESG, referred to as FinBERT)
  - 3. provide an approach to analyze how components of a sentence contribute to its ESG score

Objectives

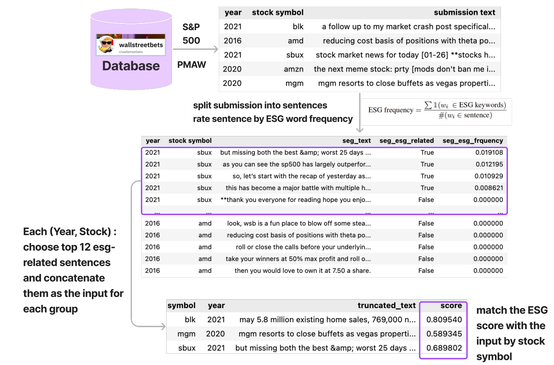
- To verify the reliability of ESG information on public communities on ESG performance evaluation
- To compare the adaptability of the learned representation from FinBERT on classification task and regression task with regard to ESG grade/score
- To better understand the associated impact words have against any given rating

Experiment Design

- Obtain pretrained representation from FinBERT (Huang et al., 2022)
- The pretrained FinBERT was frozen and only the classifier layer was trained
- Transform the model into Auto Classes of Hugging-
- Face and implement the Pipeline for SHAP

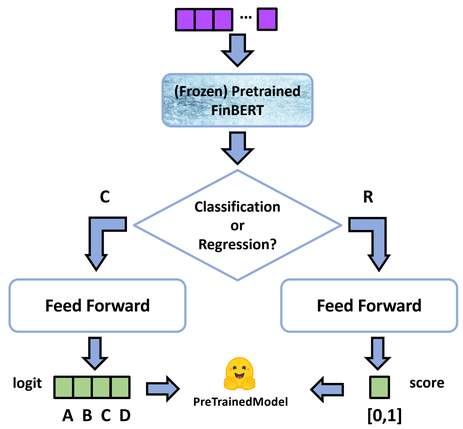
Data Collection

- Refinitiv ESG score is used as the gold label for its high data quality and widely covered range of companies
- "r/wallstreetbets" is a representative individual investor community with a high density of users and stock information including that of ESG

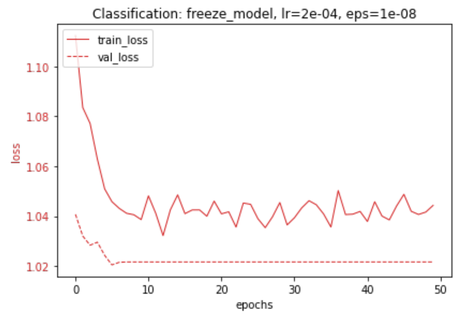


- Table 1: Distribution of the length of input sequences for each (year, company) before and after the top-12 selection

Quantile	Before	After
0.5	1211.5	261.5
0.75	3679.75	373
0.95	20387.8	715.7
0.99	73435.22	1613.68



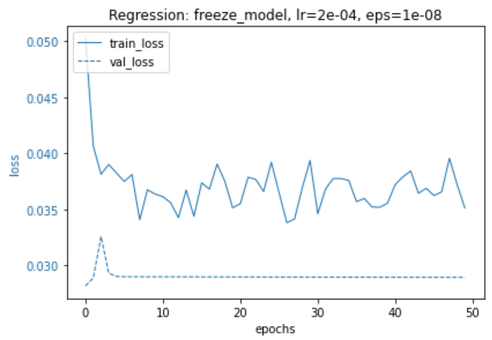
Result



- The classification task achieved the accuracy of **52.66%**, which is significantly higher than random guess. And the MSE of the regression task was **0.03.**

- **Both metrics implied the reliability of the ESG information in public communities, at least in "r/wallstreetbets", for its capacity to predict relatively authoritative grades/scores.**

- **The classification task was better learned** since the validation loss of the regression model has no significant decrease



Analysis

- SHAP by Lundberg and Lee (2017), attempts to distribute weights of importance (contribution to the final result) using combined cooperative game theory.

- Can we determine there's a positive correlation between SHAP value and grade for positive words, and for negatively-associated words, the correlation is the opposite?
- This was generally the case for higher grades (A, B) with positive words and lower grades (C, D) with negative words, but for other combinations, the results were more ambiguous.

Reflection

- Verified reliability of the ESG information at least in "r/wallstreetbets" for ESG performance evaluation
- Through leveraging "weakly supervised learning," in our case fine-tuning a pretrained model while freezing the pretrained layers (Yu et al., 2020), with the shown accuracy we were able to reconcile systematic disparities in information access between two parties: the already prominent corporate party and institutional investors