Predicting Corporate Climate Impact Using Twitter Activity

CS 539 – Jules Cazaubiel, Nicholas Tourtillott

Agenda

1st

Introduction

To give insight into the problem, our motivation, and our proposed solution

2nd

Methods

To give insight into our datasets and the methods we used to process them

3rd

Models

To give insight our model, its performance, and our attempts to iterate

4th

Conclusions

To give insight into our limitations, key findings, and proposed next steps

The Climate Crisis

- We are in an anthropogenic climate crisis
    - Caused by exploitation and immense levels of carbon emissions
- Rising temperatures are destabilizing ecosystems on a global scale
    - If trends remain unchanged average temperatures may raise by 5℃ by the end of the century
- Extreme weather is causing unprecedented damage to society

Corporate Caused

71% of global emissions

Come from just 100 corporations world wide– according to the Carbon Majors Report in 2017.

Why Do We Need Our Model?

- This corporate greed is unsustainable and will lead to:
    - Mass extinctions
    - Famines
    - Floods
    - Emergent diseases
    - And worse
- Corporations are not apt to take responsibility for their actions
    - Infact, they often feign environmental activism despite their true impact

There needs to be a way to hold corporations accountable whether or not they decide to disclose their true impact

Datasets

Carbon Disclosure Project + Twitter Data

We generated a novel dataset which combines multiple sources:

- A report from the Carbon Disclosure Project which grades corporations (A-E scale) on environmental impact
- Corporate Twitter feeds since they are often used for:
  - outreach
  - marketing
  - & activism campaigns

Methods

Twitter Web Scraping - Snscrape

Manually get company usernames

snscrape --jsonl --max-results 750 --since 2011-01-01 twitter-search 'from: Nike' > Nike.json"

Snscrape command
- Company username
- Maximum number of tweets to get (750)
- Lower date limit

→

Individual JSON files (one per company) with tweet information

→

Central tweet csv
- Username
- Content
- Date of creation

➔ Easily readable file to use for the next part of our project

Natural Language Processing

- Character Removal
- Stopword Removal
- Lemmatization

@The quick brown #fox jumps over the lazy dog.

the quick brown fox jumps over the lazy dog

quick brown fox jumps over lazy dog

quick brown fox jump over lazy dog

Term Frequency Inverse Document Frequency

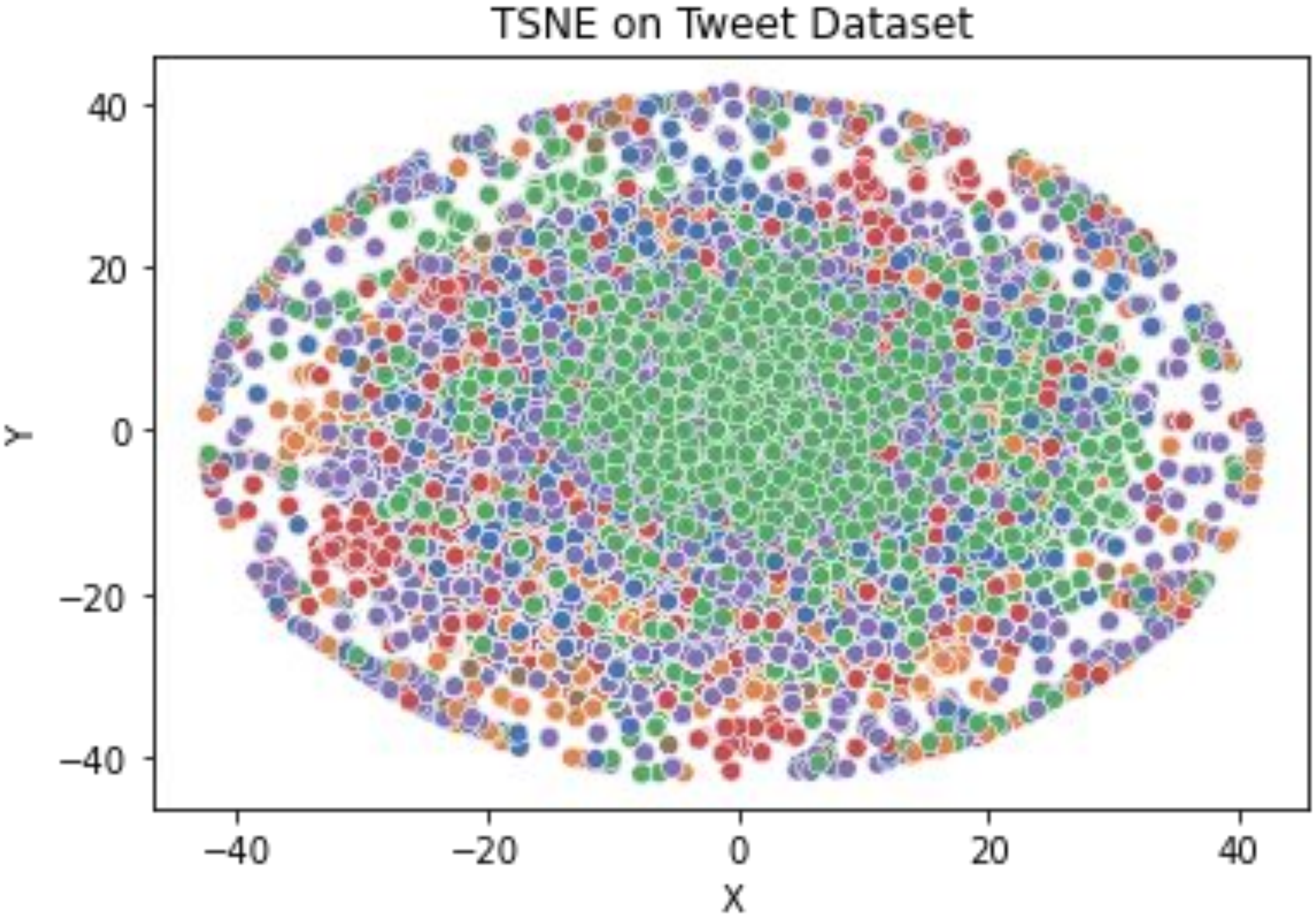
Weights each word

$$\mathrm{tf}(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$$

$$\mathrm{idf}(t,D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

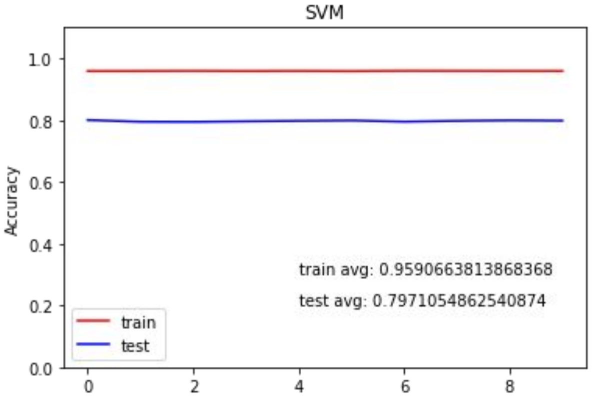
Dimensionality Reduction using TSNE

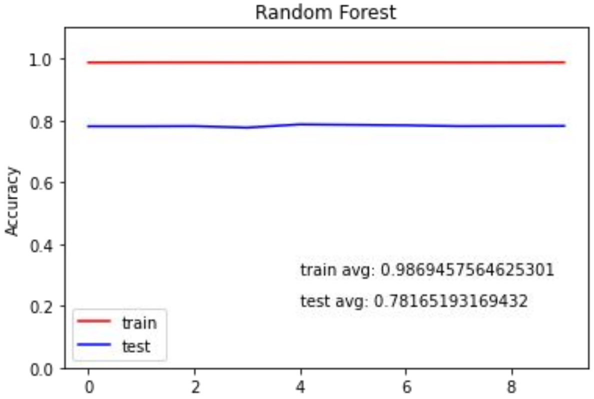
- No distinct clustering of tweets by company grade

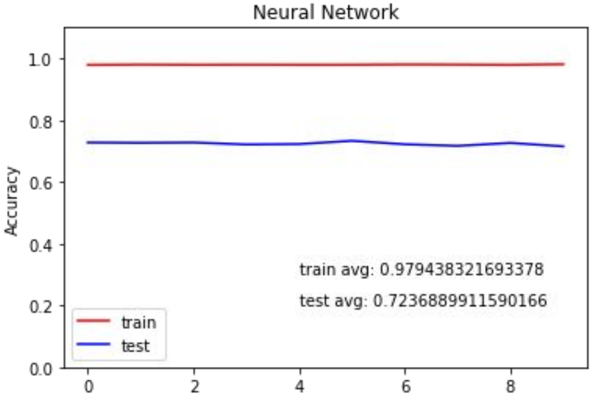


Models

Classifiers used







Predict environmental impact grade (A-E)

→ SVM and Random Forest perform the best, with similar performances.

Conclusions

Limitations

- Our environmental impact predictions are fairly out of date
  - The dataset was from 2013
  - Modern versions were not freely available
- Our method heavily relies on high levels of corporate activity on social media
  - While this is ubiquitous in our culture currently some companies may fly under the radar

Key Insights

- Corporate climate impact is discernible from social media activity
  - Provides new avenues to investigate corporate climate impacts regardless of their disclosure of the facts
- Our best model can predict an impact grade with an accuracy of 79.7%
  - Meaning we can provide journalists, researchers, and others an effective way to hold corporations accountable

Next Steps

- Retrain the model with a modern equivalent to the 2013 CDP dataset
  - Likely would not decrease model accuracy
  - Would give more weight to the current predictions
- Journalists and scientists should apply this model to investigate corporations and hold them accountable for the damage they cause

Sources

Abadi, Mart'in, Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … others. (2016). Tensorflow: A system for large-scale machine learning. In *12th $USENIX$ Symposium on Operating Systems Design and Implementation ($OSDI$ 16)* (pp. 265–283).

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc."

Cdp. (2016, June 1). *2013 - global 500 emissions and response status: CDP Open Data Portal*. 2013 - Global 500 Emissions and Response Status | CDP Open Data Portal. Retrieved December 1, 2021, from https://data.cdp.net/Companies/2013-Global-500-Emissions-and-Response-Status/marp-zazk.

Dahlman, R. L. and L. A. (2021, March 15). *Climate change: Global temperature*. Climate Change: Global Temperature | NOAA Climate.gov. Retrieved December 1, 2021, from https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature.

Griffin, P. (2017). Carbon majors report 2017. Retrieved December 1, 2021, from https://b8f65cb373b1b7b15feb-c70d8ead6ced550b4d987d7c03fcdd1d.ssl.cf3.rackcdn.com/cms/reports/documents/000/002/327/original/Carbon-Majors-Report-2017.pdf?1499691240.

JustAnotherArchivist. (2018). snscrape: A social networking service scraper in Python. Retrieved December 1, 2021, from https://github.com/JustAnotherArchivist/snscrape.

Roesslein, J. (2020). Tweepy: Twitter for Python! *URL: Https://Github.Com/Tweepy/Tweepy*.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (1970, January 1). *Scikit-Learn: Machine learning in Python*. Journal of Machine Learning Research. Retrieved December 1, 2021, from https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

Stohl, C., Etter, M., Banghart, S., & Woo, D. J. (2015). Social Media Policies: Implications for contemporary notions of corporate social responsibility. *Journal of Business Ethics*, *142*(3), 413–436. https://doi.org/10.1007/s10551-015-2743-9