

REDDIT POST ANALYSIS R/CLIMATE VS. R/ENERGY

Presentation to Education Publishing Services

Andy Roberts

March 5th, 2021

Agenda

- Motivation
- Climate & Energy Overview
- Subreddit Comparison
- Sentiment Analysis
- Modeling
- Misclassification
- Conclusions & Recommendations

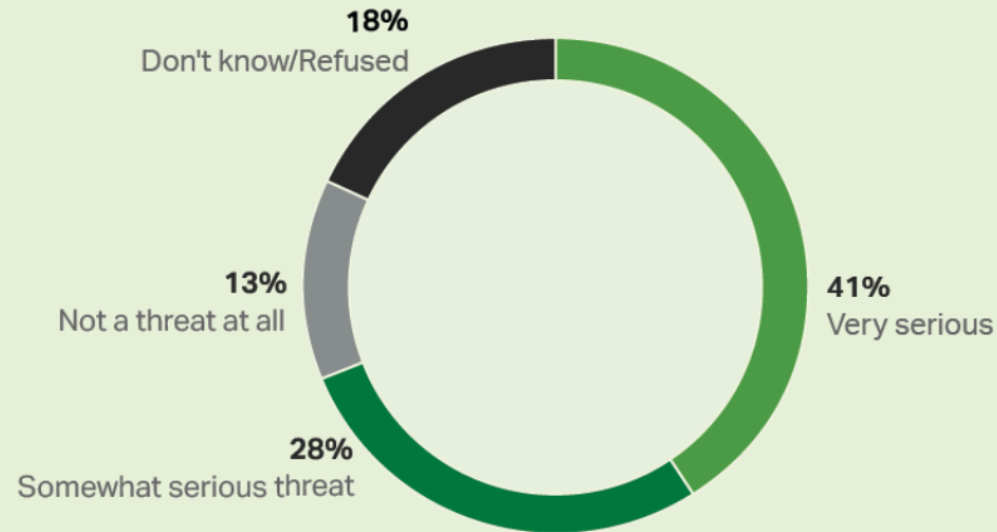
Motivation

- Tradeoffs abound between climate change and energy use
- Discussions often devolve into fraught political arguments
- Education regarding the tradeoffs and implications of these topics is essential for informed debate
- NLP and classification modeling can inform the state of understanding and where to focus education efforts

Views on Climate

Global Perceptions of the Threat of Climate Change

Do you think that climate change is a very serious threat, a somewhat serious threat, or not a threat at all to the people in this country in the next 20 years? If you don't know, please say so.



WORLD RISK POLL, 2019

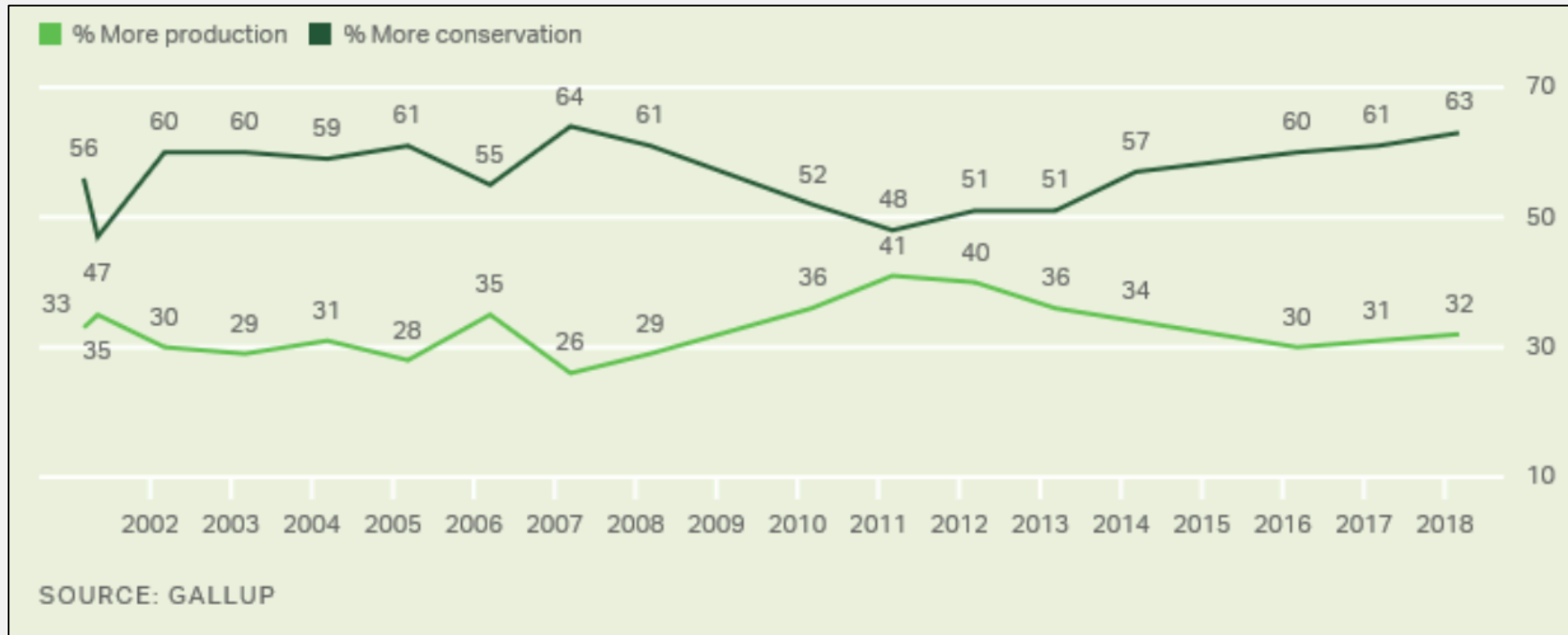
<https://news.gallup.com/opinion/gallup/321635/world-risk-poll-reveals-global-threat-climate-change.aspx>

Perception of climate risk:

- Majority view as threat
- Significant proportion unaware of risk

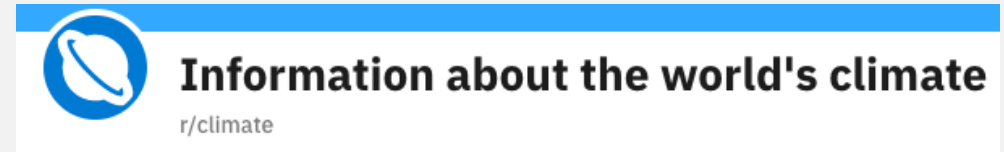
Views on Energy

Preferred approach to solving nation's energy problems:
What's the right strategy?



Reddit Overview

- Community-based social news aggregation, web content rating, and discussion website
- 1,200 posts scraped from r/climate and r/energy over the previous 12 months

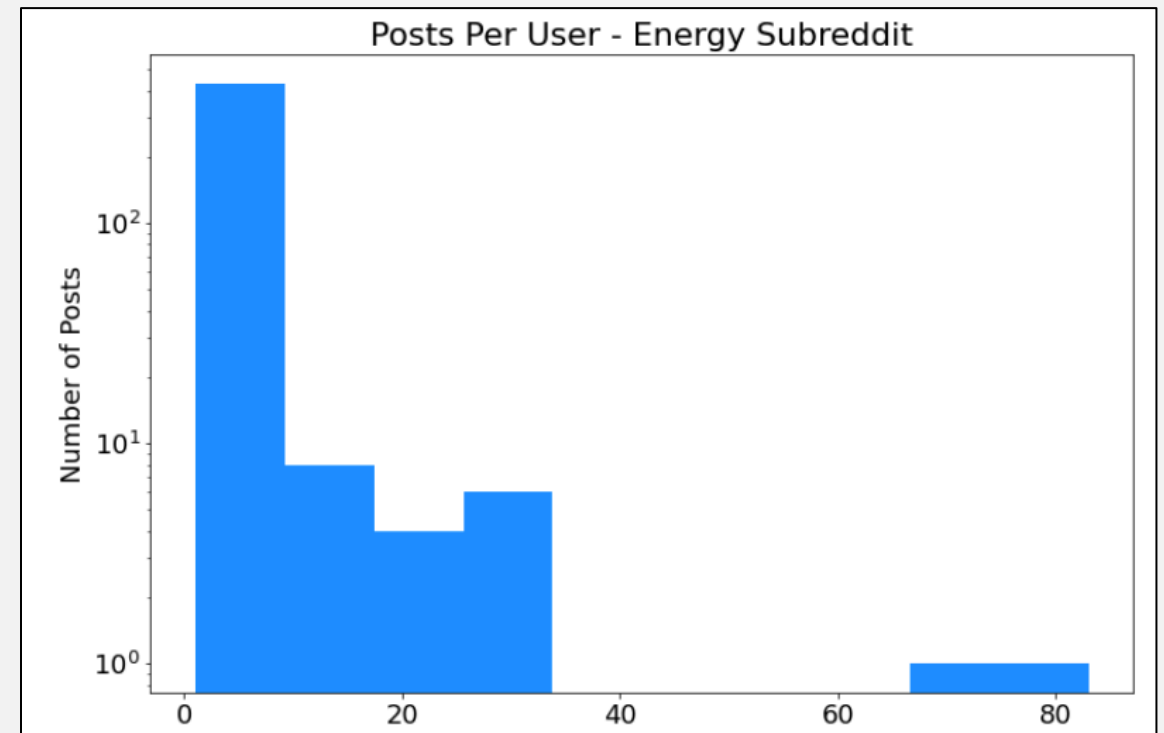
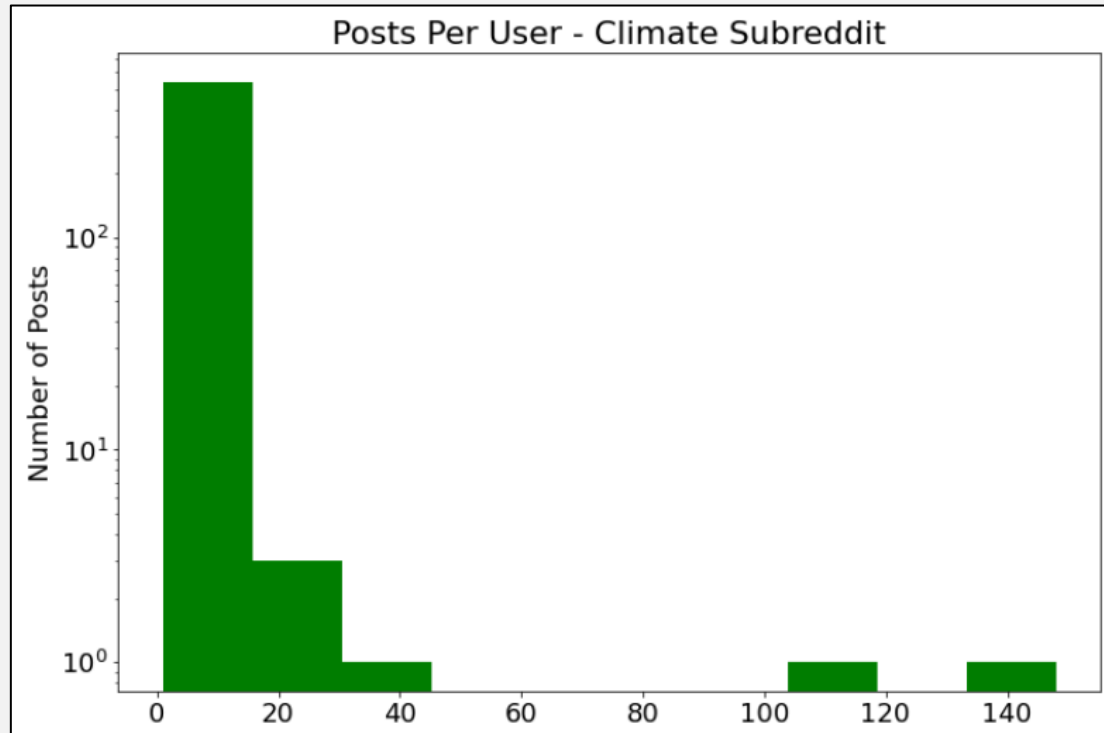


- 85.1k members



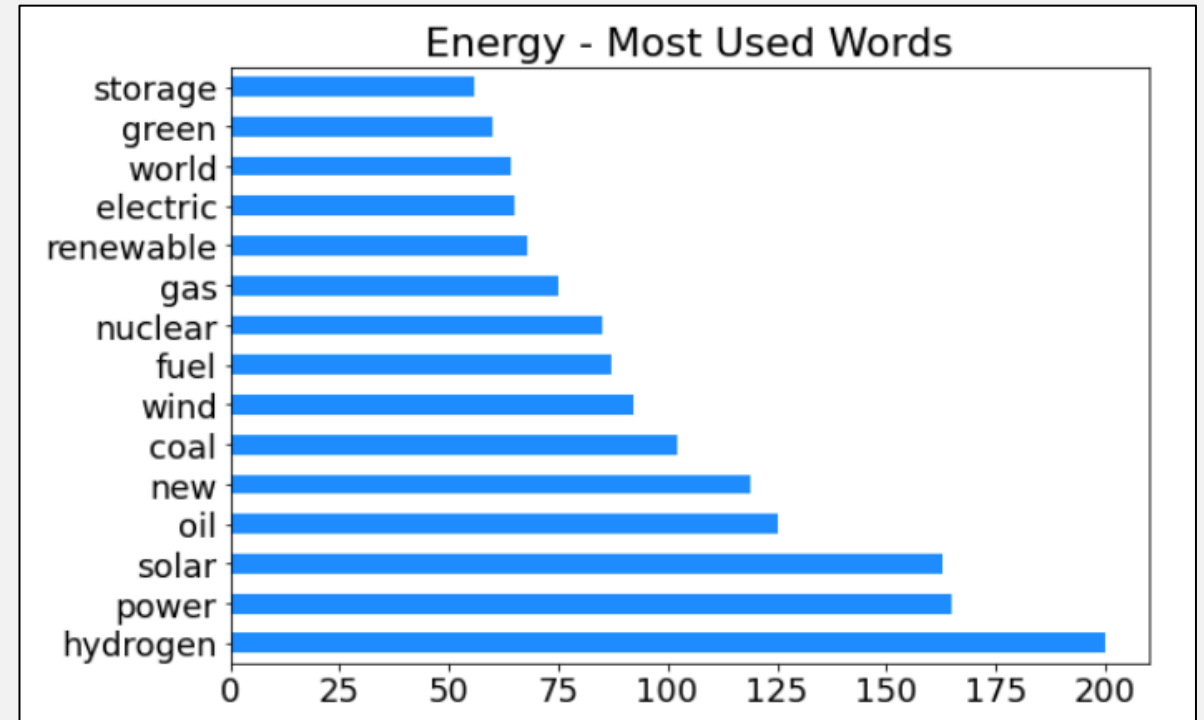
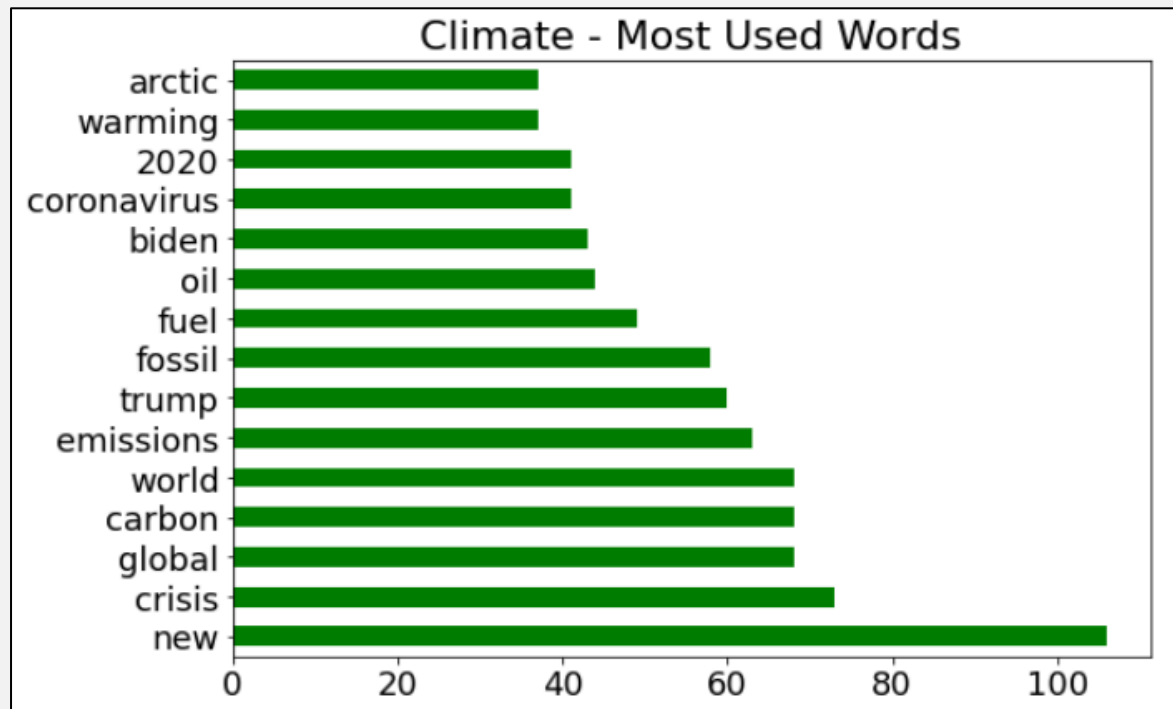
- 108k members

Subreddit Comparison



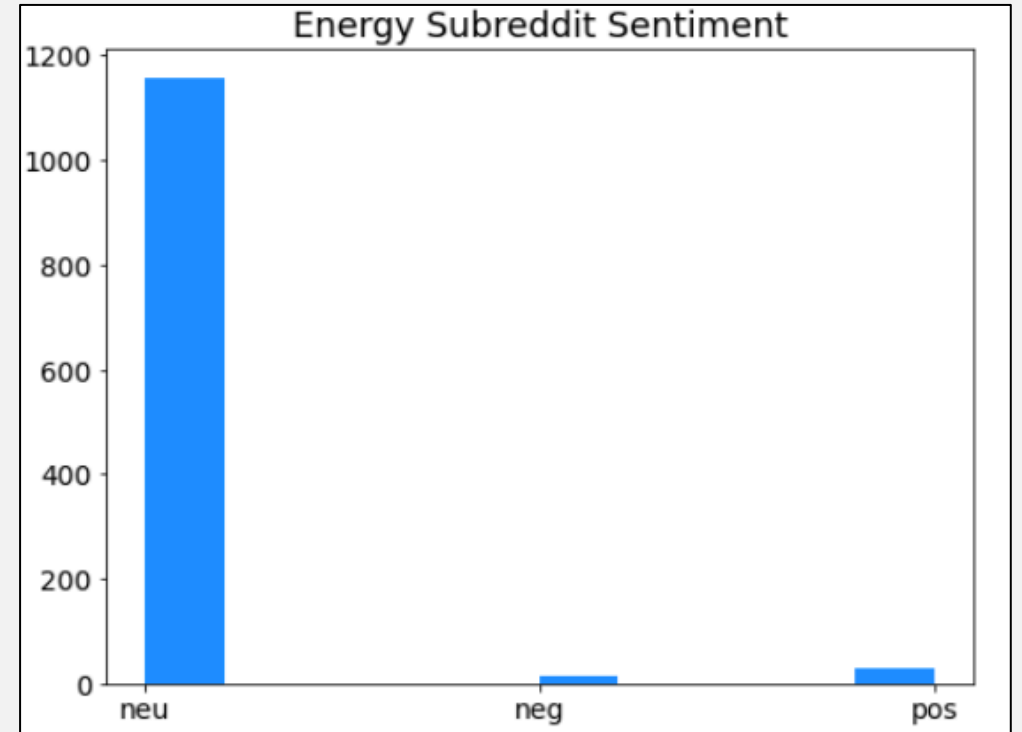
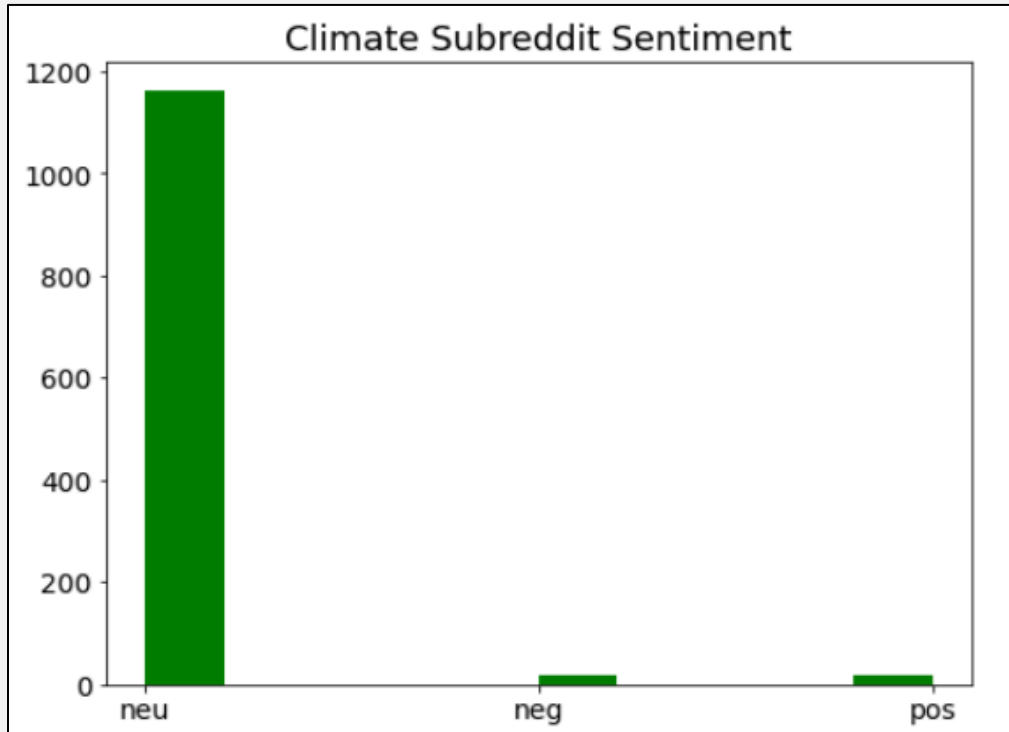
- Climate subreddit has users who tend to either post rarely or very often
- Energy subreddit shows more even distribution of posts per user

Subreddit Comparison



- Most common word occurrences reflect trending topics
- These terms provide insight into general interest in climate and energy topics

Sentiment Analysis



	Positive	Neutral	Negative
Climate	1.5%	96.8%	1.6%
Energy	2.5%	96.3%	1.1%

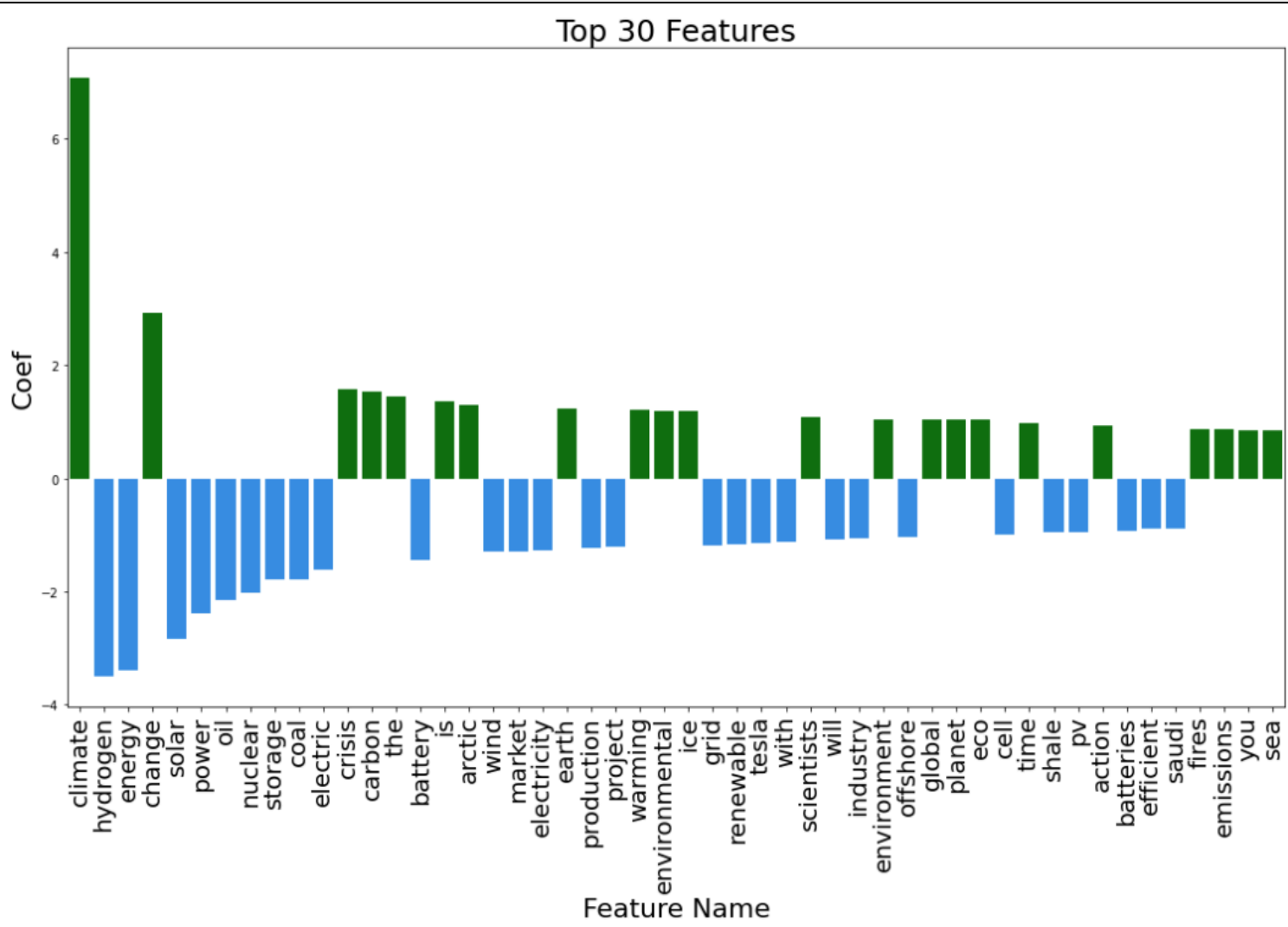
Classification Model

Grid Search Parameters	
max_df	0.9
max_features	2000
min_df	3
ngram_range	(1,1)
stop_words	None

	Train	Test
Model Score	0.92	0.85

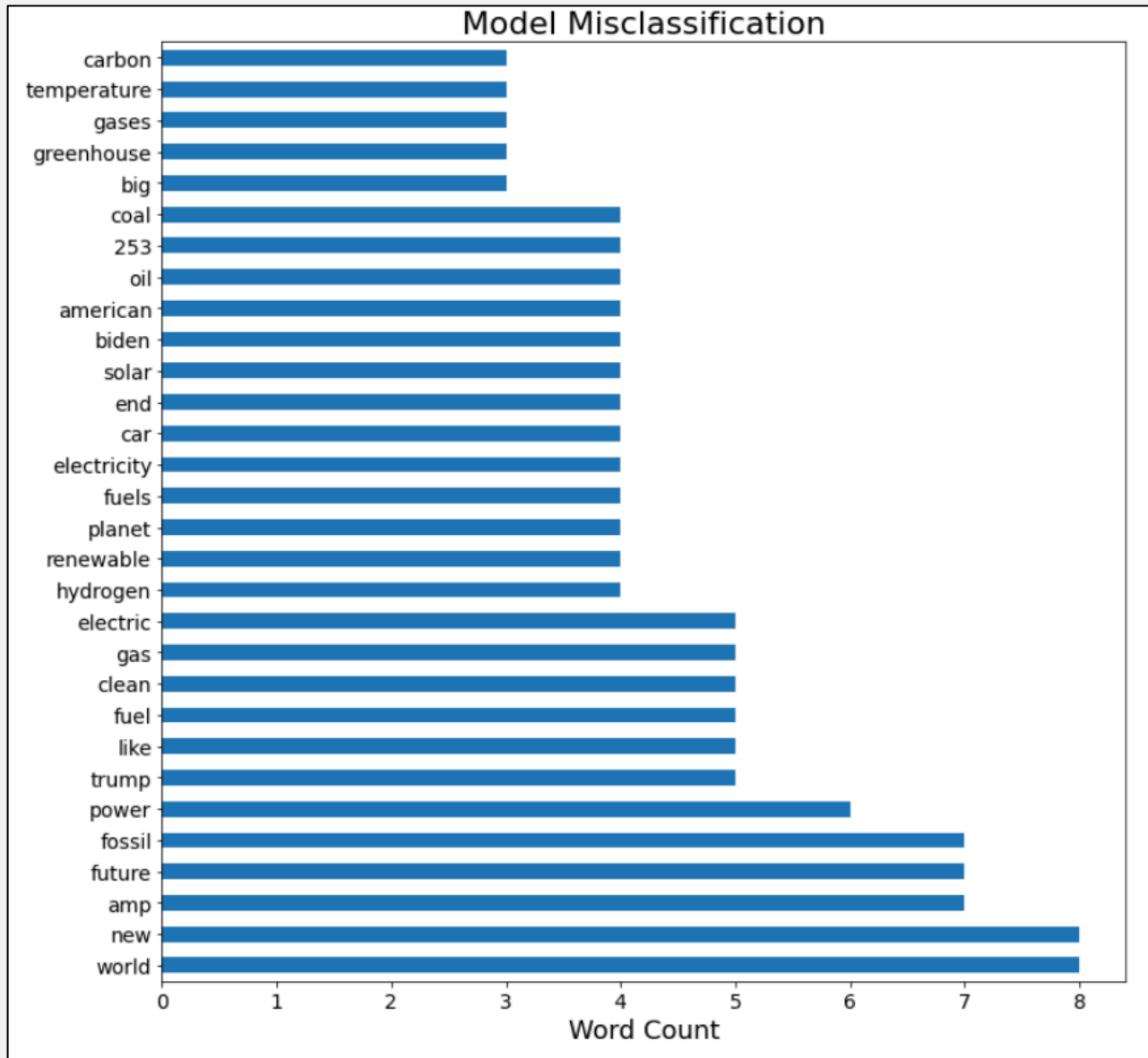
- Production model utilized TF-IDF & logistic regression
- Relatively overfit, yet still provides reliable predictions

Feature Importance



- Demonstrates most impactful features
- ‘Climate’ most influential, yet energy-related terms hold high overall effect

Misclassification



- Most commonly misclassified terms tend to be ambiguous
- Energy-related terms tend to be associated with misclassification
- Misclassified terms hold potential for guiding education

Conclusions and Recommendations

- Web-scraping of Reddit post from r/climate and r/energy identifies commonly used terms, their sentiment, and predictive power
- TF-IDF transformer and logistic regression estimator provided most viable model for predicting whether text belonged to either subreddit
- Analysis provided herein will provide better understanding of knowledge may be lacking in the general public regarding climate change and energy use