Measuring Public Sentiment Towards Nuclear Energy Using Twitter Data

STEPHEN GASS, University of Rochester, Goergen Institute for Data Science, USA
DI WU, University of Southern California, Viterbi School of Engineering, USA
JIEBO LUO, University of Rochester, Department of Computer Science, USA

The public debate over nuclear energy has been smoldering in the United States for many decades. In recent years, increasing concerns over climate change and the reliability of renewable energy sources have elevated the profile of nuclear energy as a zero-emissions energy source despite abiding public concerns over safety and weapons proliferation. This paper presents a general methodology for rapidly and accurately collecting and analyzing the public discourse using the popular social media platform, Twitter. We focus this methodology on the public nuclear energy debate to compare against the vast survey literature available on the subject. Not only do we obtain comparable results to published survey data, but we also show that this methodology can be used to uncover key insights that are not represented in the published survey data. Matching survey data, we find that 67 percent of individual users in our dataset primarily expressed positive sentiments towards nuclear energy, and more finely, that men and older Twitter users express positive sentiments at a higher rate relative to women and younger Twitter users, respectively. Additionally, we match published survey data in showing the key concerns regarding nuclear energy are nuclear waste, accidents, and weapon proliferation. We also find that Twitter accounts belonging to organizations are the predominant advocates of nuclear energy as a potential aid in addressing climate change, while accounts belonging to individuals, particularly those belonging to women and young Twitter users, tweeted more about the threats posed by nuclear weapons and accidents. Our results shine light on the terrain and prospects of the nuclear energy debate in the United States as well as detail the present accessibility of measuring public sentiment cheaply and rapidly using publicly available Twitter data.

CCS Concepts: • **Human-centered computing → Collaborative and social computing**.

Additional Key Words and Phrases: social media, nuclear energy, sentiment classification, topic analysis, Twitter, opinion mining, public sentiment

ACM Reference Format:
Stephen Gass, Di Wu, and Jiebo Luo. 2021. Measuring Public Sentiment Towards Nuclear Energy Using Twitter Data. 1, 1 (March 2021), 14 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Public opinion of nuclear energy has gone through many cycles of interest and disinterest through the decades [7]. While one of the most scalable zero-emissions energy sources, it is not without its downsides. Concerns around nuclear weapons proliferation, environmental contamination, waste disposal and safety have historically held back the industry

Authors' addresses: Stephen Gass, sgass@ur.rochester.edu, University of Rochester, Goergen Institute for Data Science, 1209 Wegmans Hall, University of Rochester, Rochester, New York, USA, 14627; Di Wu, dwu92983@usc.edu, University of Southern California, Viterbi School of Engineering, 3551 Trousdale Pkwy, Los Angeles, California, USA, 90007; Jiebo Luo, jluo@cs.rochester.edu, University of Rochester, Department of Computer Science, 3101 Wegmans Hall, University of Rochester, Rochester, New York, USA, 14627.

[4] [14]. Because of the highly politicized nature of the nuclear energy industry, public support or opposition to nuclear power is often the deciding factor as to whether existing reactors stay online and new nuclear power plants get built [2].

The availability of open-source deep learning models and massive amounts of social media data provide a new, highly-accessible opportunity to rapidly measure public opinion towards politicized topics of this nature. Public opinion survey data regarding nuclear energy have been regularly conducted for decades, which provides an large literature with which to cross exam these new techniques in measuring public sentiment. The nuclear energy debate is also a valuable opportunity for measuring public sentiment because it is a relatively technical and impersonal political issue for which public, online expressions aren't likely to deviate from users' private, offline beliefs. The social media platform, Twitter, provides a particularly unique opportunity for opinion mining because of the highly political and emotional language used on the platform. Tweets have a short character limit, which prevents extensive detailed, technical discussion and incentivizes highly sentimental language [29]. This has the affect of bringing out users' opinions in succinct, sentimental language where this political debate might be too drawn out and technical for sentiment analysis on other online platforms. Additionally, our analysis focuses exclusively on Tweets from users within the United States, and it is common for Twitter users to provide their self-reported location, which we use to filter out Tweets from users outside of the United States.

2 RELATED WORK

Numerous studies have used social media for measuring public sentiment towards energy related topics. For examples, Park [28] measured South Korea's public response (both positive and negative) to major media events related to nuclear energy by analyzing seven years' worth of social media posts from four major platforms. In particular, this paper was able to successfully identify South Korea's public response to the Fukushima accident in 2011, as well as positive news events such as research partnerships and major trade deals. Geotagged Twitter data has been used to measure the Alaskan public's opinion of their current and preferred energy options, such as tidal, solar, and nuclear energy [1]. Likewise, sentiment and topic analysis have been conducted on Twitter data from the U.K. and Spain to identify the public's energy preferences and attitudes towards climate change in general [22].

While numerous studies have documented the socioeconomic and demographic representation limitations regarding mining social media data, particularly for applications regarding public health [30] [25] [10], there have been several machine learning models developed in recent years to combat this problem of representation [9]. While the prediction of sensitive attributes such as demographic information raises a host of ethical concerns around privacy and data ownership, if it is done respectfully, it can play an important role in reducing representation disparities which, left unchecked, might exacerbate existing inequalities. In our methodology, we use the 3M-Inference model, which has been training to make accurate predictions in a variety of popular languages, to predict broad age and gender classification [34]. This model has been successfully combined with sentiment and topic analysis models in other studies to reveal interesting patterns in large Twitter datasets (e.g. [13]).

Finally, there exists a large literature analyzing survey data regarding public opinions on nuclear energy. Most notably, Bisconti [7] summarizes 36 years of survey data in the United States to show how support for nuclear energy has changed over the decades. This work documents the attitudes of different demographics towards nuclear energy, as well as the key issues affecting public opinion, which we use to compare against the results obtained from our methodology.

3 METHODOLOGY

We first set out to find the key issues that contribute to people's opinion of nuclear energy. The age and gender of Twitter users who tweeted about topics related to nuclear energy are inferred from the users' profiles and used to form five major user clusters. Topic analysis is then applied to these user clusters to reveal the main priorities of different groups. This is followed by three-class sentiment analysis of the users' Tweets in an effort to measure the overall approval of nuclear energy as well as the approval rate for various demographic classifications.

3.1 Data Collection

Tweets related to nuclear energy from users within the United States were collected from the Twitter API using the tweepy python wrapper[1] by a series of keyword searches using the phrases "nuclear energy", "nuclear power", "nuclear waste", "nuclear radiation", "nuclear reactor", "nuclear accident", and "radiation pollution". Naturally, the topics of nuclear energy and nuclear weapons exhibit significant overlap, so to prevent the oversampling of Tweets related solely to nuclear weapons, we used the above keywords in conjunction such that all relavent Tweets returned by the API must have both key words in their text, e.g. both "nuclear" and "energy". As discussed in detail below, our dataset does contain significant discussion of nuclear weapons, but by conducting the key word search in this way, we keep the discussion of such topics solely in relation to the subject of nuclear energy. Each tweet returned from an API query includes a user's screen name, self-reported "real" name, self-reported location, and profile picture which are each collected and used for subsequent analysis. Note that the accounts of both organizations and individual Twitter users are returned undifferentiated by the API, so it is necessary to make predictions as to the organizational status of Twitter users for a valid analysis of public sentiment. Additionally, we use the recent search feature of the API to retrieve time-constrained Tweets for a period of several months at the end of 2020, resulting in a dataset comprising over 50,000 nuclear energy-related Tweets from about 25,000 unique user accounts within the United States during this time period.

3.2 Text Preprocessing

To analyze the Tweets' text, we develop a text preprocessing pipeline similar to that of Baziotis et al. [5]. This pipeline includes performing sentiment-aware tokenization, spell correction, word normalization, word segmentation (for splitting hashtags) and word annotation. The challenging part of tokenization is to correctly split the word without separating expressions or words that should be kept intact. This is particularly important in texts from social networks, with "creative" writing and expressions like emoticons, hashtags and so on. Thus, we implement a social tokenizer geared specifically towards Twitter. This tokenizer is capable of recognizing Twitter markup, emoticons, emojis, expressions such as dates, times, currencies, acronyms, censored words (e.g. f**k), words with emphasis (e.g. a *great* time) and more. After tokenization, we use the Viterbi algorithm to perform spell correction and word segmentation with word statistics (unigrams and bigrams) computed from our unlabeled dataset to obtain word probabilities. Finally, the text is converted to lowercase, and all URLs, emails and mentioned usernames are removed to retain the natural language elements from the text data.

3.3 Latent Dirichlet Allocation

We employ Latent Dirichlet Allocation (LDA), a generative, unsupervised, probabilistic model that is commonly used to reveal latent topics in text corpora [8], to identify the key concerns that people have regarding nuclear energy. Each

[1] https://git.io/JvAjh

Tweet in the text corpus is modeled as a bag of words with an occurrence count for each token. In our analysis, we not only look at individual tokens but also highly correlated groups of words. Therefore, bigrams and trigrams are also added to our corpus. These documents and occurrence counts are then used to generate a probability distribution over a pre-specified number of topics. Each topic, therefore, is the probability distribution over each token in the corpus. The highest probability tokens within each topic can then be used to give the topic a human-interpretable name.

3.4 M3-Inference

Opinions on nuclear energy have been well documented to depend with various demographic factors [26]. To generate demographic predictions for users in our dataset, we use the deep learning model, M3-Inference (Multimodal, Multilingual, and Multi-attribute) to predict the age category, gender, and organizational status of each user account. M3-Inference was designed specifically to improve demographic representation in social media opinion mining by combining multilingual demographic inference with post-stratification to training the model on a representative training sample. While the three predicted attributes of age, gender, and organizational status are not comprehensive of a user's demographics identity, they are nonetheless fundamental in understanding some of the primary demographics affecting nuclear energy opinion while being mindful of the user's privacy.

The M3-Inference model is a multi-attribute deep learning system which takes a user's biography, screen name, user name, and profile image as inputs to make demographic classifications [34]. The four age categories are "<= 18", "19-29", "30-39", and ">= 40"; gender classifications are simply "male" and "female"; and organizational statuses are "isorg" and "notorg", the latter of which designates that the account is for an individual rather than an organization. For demographic classifications in English, this model has a macro-F1 score for age, gender, and organizational status of 0.42, 0.92, and 0.90, respectively. The relatively low performance for age classification indicates that conclusions made regarding age classification should be taken with a grain of salt. And while it's not the main focus of this paper, it must be noted that potential bias in gender definitions when creating annotated training datasets for deep learning model pose a potential limitation to this approach of machine-learning inference and is a serious concern regarding the unscrupulous application of facial recognition algorithms more broadly [31].

3.5 Transformers for Sentiment Analysis

To identify the sentiment for each Tweet, we use the transformer models that have shown state-of-the-art performance on natural language processing tasks. This paper compares five models that hold the best performance on the SST-2 dataset for text classification including BERT, RoBERTa, XLNet, ELECTRA and BART. We use pretrained models from the huggingface transformers library [2] and fine tune them by adding an additional output layer to the pretrained models. A more detailed introduction for each model will be presented in the sentiment analysis section below. We use BERT as a baseline model and compare its performance with four other models across three annotated datasets designed for Tweets sentiment classification tasks. We use the best performing model for sentiment analysis of nuclear energy-related Tweets. To validate the performance of our chosen model on our nuclear energy dataset, a random sampling of 100 Tweets is hand labeled and used to calculate the accuracy of the chosen model.

[2]https://git.io/JfUEh

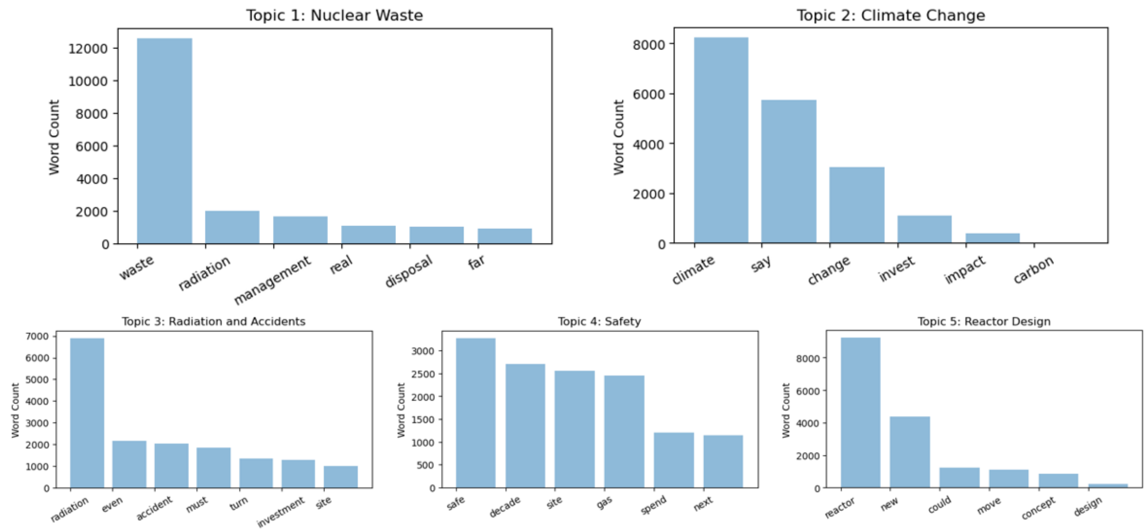

Fig. 1. Word Frequency of Top Ten Topic Keywords.

3.6 FP-Growth

Connections between users' demographics and their sentiment towards nuclear energy are mined using frequent patterns mining. We divide our users into two classes based on the sentiment classification of their Tweets and then mine the user demographic labels within each class of users. We use FP-Growth, proposed by Han et al. [15] to extract these frequent patterns. This algorithm has been shown to be efficient and scalable for mining both long and short frequent patterns, and is about an order of magnitude faster than the Apriori algorithm as well as faster than some recently reported frequent pattern mining methods. Our results are obtained using a support confidence of 15 percent.

4 EXPERIMENTS

4.1 Key Topic Analysis

The number of topics used in LDA analysis is determined by optimizing for the coherence score of the LDA model on our nuclear energy-related Tweet dataset. The coherence score for this dataset has a maximum at 70 topics, resulting in the relatively high coherence score of 0.57. The top five topics resulting from this model are then given human-interpretable labels according to key words and an interpretation of sampled Tweets in each topic. Figure 1 presents each of these topics with their top six key words and corresponding word counts. This figure shows that the most common subtopic in discussions of nuclear energy on Twitter is Nuclear Waste. We found 9,149 Tweets in our dataset belong to this topic which is about 18% of the dataset. This is followed by discussions of Climate Change (9%), Radiation and Accidents (6%), Safety (6%), and Reactor Design (5%). Sample Tweets from these five topics are presented in Figure 2.

Surveys of public opinion towards nuclear energy show that negative opinions tend to be much more focused than positive opinions, so it is to be expected that the most common topic would be a negative one like the discussion of nuclear waste. This analysis also matches survey results that show accidents, waste, and safety to be the general public's primary areas of concern around nuclear energy [7].

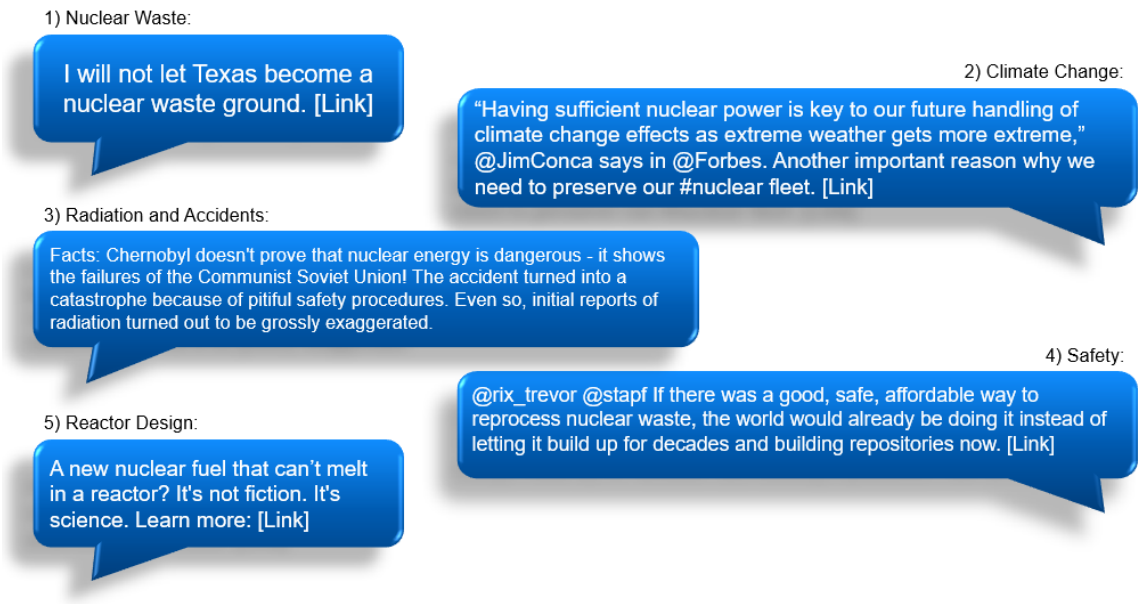

Fig. 2. Sample Tweets from the top 5 topics in the entire dataset (Links removed for brevity).

Interestingly, there is quite a bit of detailed discussion of nuclear reactors in our dataset as evidenced by Topic 5. This reflects the nature of the public discourse around nuclear energy as a technical and technologically oriented topic of discussion as well as illustrating the value of Twitter as a social media platform with regard to the ease with which users can share news articles about current research and events.

4.2 Topic Analysis of User Clusters

Demographic predictions of users in our dataset resulted in a gender distribution that is 69.1 percent male and 30.9 percent female which suggests that males tweet about nuclear energy much more than female Twitter users. While Twitter as a whole has been documented to have similar gender representation disparities, it is difficult to quantify just how much more skewed the topic of nuclear energy is towards male Twitter users than Twitter as a whole when taking into account the accuracy of these prediction methods [3] [24]. The predicted age distribution was 9.5 percent “$\leq$ 18”, 19.7 percent “19-29”, 19.7 percent “30-39”, and 51.1 percent “$\geq$ 40”. Notice that this age distribution is heavily skewed towards older Twitter users, particularly compared to Twitter users’ age distribution as a whole [23]. Finally, the organizational status prediction indicated that only 11.7 percent of our users were organization accounts, however, Tweets from organizations accounted for roughly 20 percent of all Tweets in our dataset, potentially revealing organizations’ aspirations to have a high impact on the public discourse around nuclear energy on Twitter. This is not particularly surprising considering one would expect an organization dedicated to nuclear energy or environmental issues to be more vocal than the average Twitter user regarding nuclear energy.

These demographic class predictions are combined with other user attributes, such as number of followers, to form user clusters using the K-prototype algorithm. The optimal number of clusters is determined using the elbow method based on within-cluster variance, which results in a total of five clusters for our dataset of users. These five clusters

Table 1. Topic Analysis by User Clusters.

	Whole Dataset	Men ≥ 40	Women ≥ 40	Men and Women 30-39	Men 19-29	Organizations
Topic 1	Waste	Weapons	News Events	Accidents	Weapons	Reactor Design
Topic 2	Climate Change	Radiation	Weapons	Potential Problems	Utility Details	Future Plans
Topic 3	Accidents	Reactor Design	Energy Needs	Reactor Design	Energy Needs	Climate Change
Topic 4	Safety	Waste	Security	Weapons	Renewables	Utility Details
Topic 5	Reactor Design	Safety	Waste	Renewables	Reactor Design	Waste

each roughly comprise between 3,000 and 7,500 users and are given the following labels, in no particular order, based on some of their majority user attributes: “Men over 40", “Women over 40", “Men and Women 30-39", “Men 19-29", and “Organizations”. Again, these human-interpretable labels are not a strict feature filter of users, but rather they represent clustered groups of Twitter users based on a more comprehensive use of the users’ profile data. We find that users with similar profile features, such as self-reported location, are more likely to share similar topic patterns than strict filters on predicted demographic characteristics as measured by the topic analysis’ coherence scores.

Accordingly, by dividing our dataset of nuclear energy-related Tweets according to these user clusters, we can apply topic analysis to obtain the top five most popular subtopics for each user cluster. This result can be seen in Table 1. Topic analysis for Tweets in each of these clusters achieved a comparably high coherence score to that of the overall dataset, each between 0.55 and 0.65. There are several interesting patterns emerging from this analysis. First, the topic of “weapons” does not enter into the top five topics of the dataset as a whole, but it is, however, a major topic for four of the five user clusters. This relative obscurity of “weapons” discussions in the dataset as a whole is likely due to the fact that Organizations tend not to discuss issues around nuclear weapons and, instead, tweet about nuclear energy in relation to climate change. Conversely, notice that while the topic of “climate change" is the second most common topic overall, it does not appear in the top five topics for any of the user clusters except the Organizations cluster. The implication here, that the general public does not primarily think of nuclear energy in relation to climate change, is also backed by survey data [7].

Additionally, the topic of “nuclear waste" appears to be a major topic among three of the five user clusters. Interestingly, nuclear waste does not appear as a major topic for the clusters of users under 39 years old. Instead, these two clusters of users tend to tweet about nuclear energy in relation to accidents and weapons. Because the 2011 accident at Fukashima is the only major nuclear accident that has occurred during the lifetime of these users, this result seems to indicate the weight with which that single accident continues to affect public opinion in the US today. This is in line with survey data which show a longitudinal decline in approval of nuclear energy starting around 2011 for which it has never recovered. Although, whether Fukashima is the ultimate cause of this observation in the literature is unclear [33].

Finally, the focus of the cluster labeled "Women ≥ 40" on the topic of “News Events" is an interesting appearance in these results and refers the major news events such as the Iran nuclear negotiations and President Trump’s sudden firing of three federal agency heads including the chairman of the Federal Energy Regulatory Commission. This result

Table 2. Performance Comparison Between Classifiers.

	Twitter Airline		Apple Sentiment		SentEval2017		Average Score	
	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
BERT	0.805	0.814	0.752	0.758	0.745	0.745	0.767	0.772
ROBERTA	0.816	0.827	0.756	0.769	0.749	0.752	0.774	0.783
XLNET	0.836	0.838	0.758	0.774	0.737	0.741	0.777	0.784
BART	0.831	0.835	**0.799**	**0.801**	0.725	0.725	0.786	0.787
ELECTRA	**0.852**	**0.856**	0.781	0.782	**0.791**	**0.793**	**0.808**	**0.810**

demonstrates one of the main benefits of restricting the time frame in the API queries to retrieve recent Tweets relating to current events for topic and sentiment analysis. Conducting data collection and topic analysis in this way is also necessary in informing the interpretation of the sentiment analysis results.

4.3 Sentiment Analysis

To expand the scope of our analysis, we then train and evaluate a range of sentiment analysis models to select the best performing model for conducting a three-class sentiment classification (positive, neutral, and negative) for Tweets in our dataset. We can then compute the sentiment score for each user who tweeted about nuclear energy based on the results of our chosen classifier. Each sentiment is represented by an integer (Positive: 1, Neutral: 0, Negative: -1), and the sentiment score for each user is, therefore, the average of the sentiment predictions of all the Tweets belonging to each user. A positive classification is given to users with a positive average sentiment score and vice versa.

4.3.1 Transformer models used for sentiment classification. In this section, we compare five pretrained transformer models that have achieved state-of-the-art performance in natural language processing tasks across multiple datasets. They have achieved similar performance on SST-2, a Stanford Sentiment Treebank (SST) dataset with binary labels (positive and negative). To find the model that performs best for our classification task, we train and evaluate each of our candidate models across three annotated datasets. We then use the model that achieves the best average performance for analysis of our nuclear energy dataset and evalute it on a small, hand-labeled sample from that dataset.

- BERT [12] was designed to pretrain deep bidirectional representations of tokens from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create highly accurate models for a wide range of tasks, such as question answering and language inference, without substantial task-specific modifications to its architecture. The large version of Bert achieved a 0.932 F1 score on SST-2.
- BART [18] was trained by corrupting text with an arbitrary noising function, and learning a model to reconstruct the original text. BART uses a standard Transformer-based neural machine translation architecture which, despite its simplicity, can be seen as generalizing BERT (due to the bidirectional encoder), GPT (with the left-to-right decoder). The major advantage of this set up is the flexibility of the arbitrary transformations that can be applied to the original text, including changing its length. As a result, they reported a 0.966 F1 score of their large version of BART on SST-2.
- RoBerta [20] is an improved recipe for training BERT models. The large RoBERTa model achieved a 0.964 F1 score on SST-2.

- XLNet [35] is a permuted language model. In their study, Yang et al. point out that BERT neglects dependency between the masked positions and suffers from a pretrain-finetune discrepancy. Therefore, they proposed XLNet, a generalized autoregressive pretraining method that enables learning bidirectional contexts by maximizing the expected likelihood over all permutations of the factorization order and overcomes the limitations of BERT thanks to its autoregressive formulation. In addition, Li et al. [19] observed the target-position-aware representation and relative position encoding features of XLNet, leading to a better benchmark score at the cost of 1.2× arithmetic operations and 1.5× execution time on a modern CPU. As a result, the large XLNet model achieved a 0.97 F1 score on SST-2.
- ELECTRA [11], like XLNet, was also designed to solve the input noise limitation of BERT. Instead of masking the input, they proposed an alternative approach that corrupts it by replacing some tokens with plausible alternatives sampled from a small generator network. Then, instead of training a model that predicts the original identities of the corrupted tokens, they trained a discriminative model that predicts whether each token in the corrupted input was replaced by a generator sample or not. Thorough experiments demonstrate this new pre-training task is more efficient than MLM because the task is defined over all input tokens rather than just the small subset that was masked out. As a result, ELECTRA achieved a 0.971 F1 score on SST-2.

4.3.2 Training and Evaluation. To fine tune these models for sentiment classification, we add an additional output layer on top of the pretrained models' output. Since we use the large version of each of the models, we add a linear layer with dimensions of [1024, 256, 3] with the ReLU activation function. We use the AdamW optimizer [21] to optimize the loss function. Because this is a multi-class classification task, we use the cross-entropy as the metrics to optimize for loss function. Moreover, we apply advanced fine tuning techniques (Discriminative fine tuning, Slanted triangular learning rate, Gradual unfreezing) proposed by Howard et al. [16] for task-specific features (i.e., learning word embeddings, learning context embeddings, and learning sentiment outputs). As suggested by Howard's work, the combination of all these techniques allow our classifier to achieve state-of-art performance for our sentiment classification task.

We train our models on three different datasets (SemEval-2017 Task 4A, Twitter US Airline Sentiment and Apple Computers Twitter Sentiment), each of which is annotated with one of three sentiments: negative, neutral, and positive. After fine tuning our classifier, all of our models outperform the top performers of SemEval-2017 and Twitter US Airline Sentiment (Leaderboard of Apple Computer Twitter Sentiment is not available). These results are shown in Table 2. In particular, the pretrained ELECTRA model achieves an impressive weighted macro F1 of 0.81 score across the three datasets. We then train the ELECTRA model on 20,000 randomly selected samples from the combination of all three datasets. The model achieves 0.79 accuracy on the combined dataset. The confusion matrix is shown on the left in Figure 3. Lastly, we hand label 100 samples from our dataset to evaluate the fine-tuned ELECTRA model, which is able to correctly predict 74 percent of our samples. Since we are making predictions of datasets containing three classes, the lower accuracy compared to the results from [11] is reasonable. The confusion matrix for our ELECTRA model's performance on the hand-labeled samples is given on the right in Figure 3.

4.3.3 Sentiment results. We observe broad public approval expressed from Tweets related to nuclear energy. Out of 25,212 users, 67 percent of the users expressed positive attitudes towards nuclear energy, compared to just 25 percent of users who predominately expressed negativity towards nuclear power. This result reflects published survey data showing that those who favored nuclear energy outnumbered those who opposed by 61 percent to 26 percent [7]. Moreover, people who posted positive Tweets are more outspoken compared to negative users. On average, positive users tweeted about nuclear energy 20 percent more than negative users.

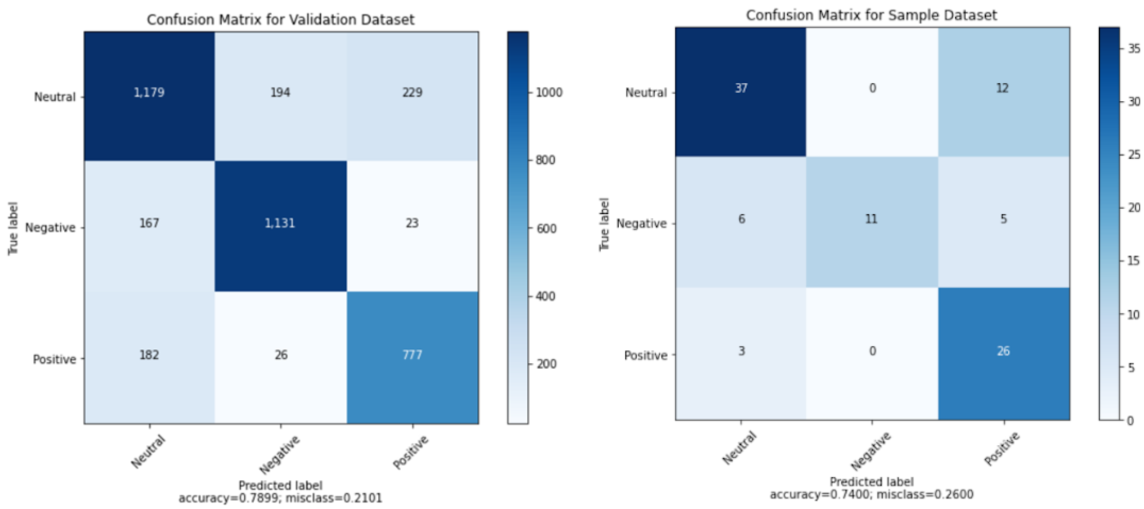

Fig. 3. Confusion matrices for the combined validation dataset (left) and hand-labeled nuclear energy Tweets (right).

Another important insight from this analysis is that different user groups express very different sentiments towards nuclear power. The results of FP-Growth, shown in Table 3, reveal that older male users, organizations and accounts with relatively many followers are the most likely demographics to express positive sentiment when tweeting about nuclear energy. In contrast, women and younger users are more likely to express negative sentiments towards nuclear energy.

Among positive users, {'male', '≥40'} is the most frequent 2-itemset with a frequency of 5376 occurrences which means about 30 percent of positive users are male who are older than 40 years old. Based on the frequent patterns that we mined, the number of positive Tweets from {'male', '≥40'} are three times more than the number of negative Tweets from this demographic. More interestingly, out of roughly 16,200 positive users, 2,350 are organizations, accounting for 15 percent of the positive users in our dataset. In contrast, out of 6,200 negative users, only 470 are from organization accounts (about 7 percent). This means organizations appear roughly twice as frequently among positive users than negative users, and pro-nuclear organizations are more numerous overall. Lastly, follower count appears to be heavily correlated with positive attitudes towards nuclear energy with follower counts in the top qaurtile often appearing among frequent patterns for positive users and follower counts in the bottom quartile appearing among the frequent patterns of negative users. This interesting pattern suggests that high-profile Twitter users tend to be more positive towards nuclear energy.

To further illustrate this finding, we plot the ratio of positive to negative users for each demographic class in Figure 4. This reaffirms the gender and age incongruity in attitudes towards nuclear energy in our dataset and matches published analyses of survey data reporting that men are more likely to have a favorable opinion of nuclear energy than women and that favorability of nuclear energy increases with the age of the participant [7].

Our analysis also reveals a large disparity between how outspoken pro-nuclear versus anti-nuclear organizations are on social media. This result manifests as a positive to negative Tweet ratio, shown in Figure 4, greater than 1.0 for organizations. While some in the media have speculated for some time that many anti-nuclear organizations might be reconsidering their long-held stance that nuclear energy is not "green" in the face of the increasing threats posed by

Table 3. Partial Frequent Patterns and their Support Counts

Features Patterns		Support Count
"positive"	"male"	10171
"positive"	"followers_top25%"	9001
"positive"	"male" "≥ 40"	5376
"positive"	"followers_25-50%"	4114
"positive"	"followers_50-75%"	4070
"positive"	"followers_50-75%"	3995
"positive"	"followers_bottom25%"	3927
"positive"	"female"	3820
"positive"	"isorg"	2350
"negative"	"followers_bottom25%"	1640

climate change, our findings can't support such claims. We simply present this result as a point of consideration that many anti-nuclear individuals and organizations might find interesting.

By filtering individual user locations by state, our analysis reveals that California is in-fact a "pro-nuclear" state. Individual users reporting to be living in California had a positive to negative sentiment ratio of 2.53, which is a significant majority. In particular, this result could shed light on the direction of the current debate over whether to close the Diablo Canyon nuclear plant. In 2016, regulators voted to shut down the nuclear plan due to a desire to incorporate more renewable energy into the electric grid. However, this decision has received some political backlash and could be reversed if legislatures choose to classify nuclear energy as "green energy" [32]. California has some of the most ambitious renewable energy goals in the country, aiming to get 60 percent of its electricity from renewables by 2030 and raising that to 100 percent by 2045. The motion to close down Diablo Canyon, the last operational nuclear plant in California which single-handedly generated 9 percent of California's electricity and 20 percent of its clean, carbon-free electricity, has generated a vibrant political discussion around the role of nuclear energy in addressing climate change [27]. While our results do not directly relate to the closure of Diablo Canyon, our methodology could be used in future research to assist in analyzing the ongoing debate over the closure of the last nuclear power plant in California.

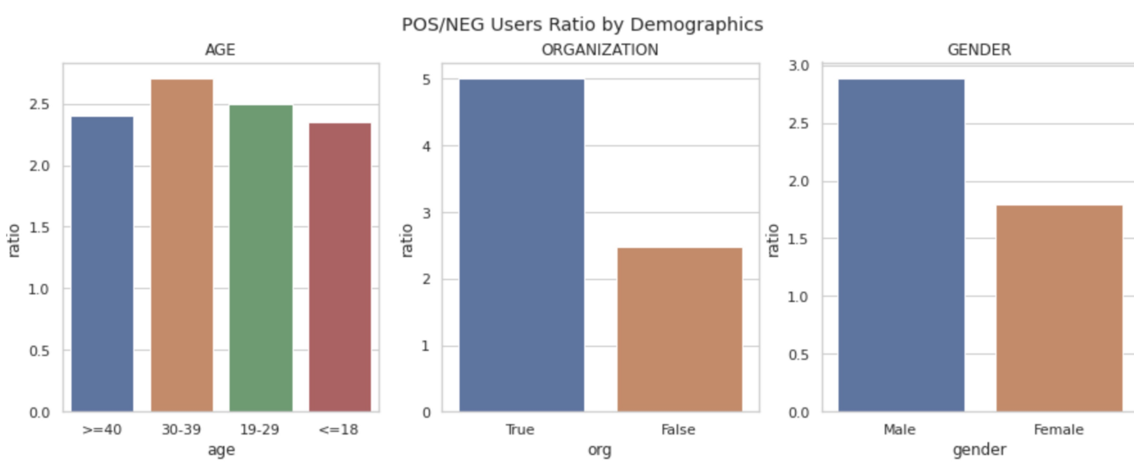

Fig. 4. Ratio between pos/neg users by demographics.

5 LIMITATION OF SENTIMENTAL CLASSIFICATION

In this work, we utilize large language models to conduct sentiment analysis on Tweets. Language models such as Electra classify Tweets based on the hidden representation extracted from sentence embeddings. Therefore, due to the nature of language model, we suspect some of the positive comments might be falsely classified as negative comments. Below are some examples of a false negative Tweets. In fact, the sentiment classifier did a good job on phrases such as "a lot of misconceptions" which, indeed, reflects strong negative sentiment. We notice false classifications tend to occur within double negative sentence and rhetorical questions. Therefore, we should expect the overall ratio of positive comments to be even higher than current results.

Example of false negative Tweets:

1. "There are a lot of misconceptions about nuclear energy."
2. "If you care so much about the climate why are you trying to close California's last nuclear power plant?"
3. "This was one of the hardest things i've listened to. Loved to hear Setsuko Thurlow's story and that she got to see the results of her activism. My dream is that we get rid of all nuclear weapons, process the material and use it to create energy."

Despite such misclassification, we note that these language models have shown excellent performance across three diverse datasets (Section 4.3) and should provide sufficiently accurate sentiment signals to facilitate our data analytics.

6 CONCLUSION AND FUTURE WORK

The motivation of this study is to highlight the accessibility of measuring and interpreting individual political concerns at a rapid pace and low cost. We aim to identify the key issues affecting overall public sentiment towards nuclear energy using nothing more than publicly available Twitter data. In doing this, we are able to not only reproduce several key results from large survey analyses but also produce several new finding not available in survey data. In particular, our results demonstrate the affect that organizations have on shaping the major topics of discussion in the public discourse. In our dataset, it is largely the Twitter accounts of organizations that are advocating nuclear energy as a potential zero-emissions energy source to combat climate change while individual users tend to be more concerned about nuclear waste, weapons, and accidents. In addition, our sentiment analysis shows a disparity between pro-nuclear and anti-nuclear organizations in which pro-nuclear organizations are not only more numerous on Twitter, but are also much more vocal in their advocacy of nuclear energy as a potential aid in addressing climate change.

By conducting a comparative analysis, we find ELECTRA to have the best performance for sentiment classification of Tweets and used it to measure an overall approval rate of nuclear energy similar to published results. Based on these sentiment predictions, we are able to mine the frequent patterns of predicted demographic attributes within each sentiment class. The results of this analysis match published survey results indicating that older males have a more positive opinion towards nuclear energy, while female and younger users are more likely to be less approving of nuclear energy. While this paper mainly focuses on latent topics of subgroups based on demographic information, future research in this area could be extended to incorporate occupational inference [17] and the users' level of knowledge about the topic, which has been shown to correlate with favorable opinions of nuclear energy [6].

One geographic trend that this paper is able to highlight is the high rate of positive sentiment that Californians show towards nuclear energy, which is particularly relevant given the current debate over the closure of the Diablo Canyon nuclear plant. While this is only a preliminary finding in this respect, we show that the California public is potentially much more favorable towards nuclear energy than the regulators controlling the destiny of the power plant. Future research with specific emphasis on Diablo Canyon is needed to strengthen this finding. Our work can be extended by collecting more precisely location data with geo-tagged information as well as focusing on collecting a more demographically and temporally representative sample of sentiments from the target population. In general, a more representative dataset could show more interesting trends across many geographic areas and across time at a much finer granularity than the current survey literature.

The success of our analysis in matching published survey results validates the strength of our methodology and social media's relevance in measuring the reported political opinions of the broader public. The methodology presented in this paper is general enough to be extended to any political topic of comparable importance and can provide vital insight into issues that are not well represented in the survey literature. Ultimately, this paper emphasizes the accessibility of using social media data to accurately understand political sentiment on individual political issues.

REFERENCES

[1] Moloud Abdar, Mohammad Ehsan Basiri, Junjun Yin, Mahmoud Habibnezhad, Guangqing Chi, Shahla Nemati, and Somayeh Asadi. 2020. Energy choices in Alaska: Mining people's perception and attitudes from geotagged tweets. *Renewable and Sustainable Energy Reviews* 124 (2020), 109781. https://doi.org/10.1016/j.rser.2020.109781

[2] A. Abdulla, P. Vaishnav, B. Sergi, and D.G. Victor. 2019. Limits to deployment of nuclear power for decarbonization: Insights from public opinion. *Energy Policy* 129 (2019), 1339–1346. https://doi.org/10.1016/j.enpol.2019.03.039

[3] Pablo Barbera and Gonzalo Rivero. 2015. Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review* 33, 6 (2015), 712–729. https://doi.org/10.1177/0894439314558836

[4] Jonathon Baron and Stephen Herzog. 2020. Public opinion on nuclear energy and nuclear weapons: The attitudinal nexus in the United States. *Energy Research & Social Science* 68 (2020), 101567. https://doi.org/10.1016/j.erss.2020.101567

[5] Christos Baziotis, Nikos Pelekis, and Christos Doulkeridis. 2017. DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis. 747–754. https://doi.org/10.18653/v1/S17-2126

[6] Ann Stouffer Bisconti. 2018. Changing public attitudes toward nuclear energy. *Progress in Nuclear Energy* 102 (2018), 103 – 113. https://doi.org/10.1016/j.pnucene.2017.07.002 Shippingport 60th Anniversary: A Time to Take Stock of Nuclear Energy's Status.

[7] Ann Stouffer Bisconti. 2020. Factors Affecting Public Opinion of Nuclear Energy in the United States. In *Reference Module in Earth Systems and Environmental Sciences*. Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.12338-3

[8] David M. Blei, Andrew Y. Ng, Michael I. Jordan, and John Lafferty. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3 (2003), 2003.

[9] Nina Cesare, Christan Grant, Quynh Nguyen, Hedwig Lee, and Elaine O. Nsoesie. 2018. How well can machine learning predict demographics of social media users? arXiv:1702.01807 [cs.SI]

[10] Nina Cesare, Christan Grant, and Elaine Nsoesie. 2019. Understanding Demographic Bias and Representation in Social Media Health Data. In *Companion Publication of the 10th ACM Conference on Web Science* (Boston, Massachusetts, USA) *(WebSci '19)*. Association for Computing Machinery, New York, NY, USA, 7–9. https://doi.org/10.1145/3328413.3328415

[11] Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. arXiv:2003.10555 [cs.CL]

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs.CL]

[13] Viet Duong, Phu Pham, Tongyu Yang, Yu Wang, and Jiebo Luo. 2020. The Ivory Tower Lost: How College Students Respond Differently than the General Public to the COVID-19 Pandemic. arXiv:2004.09968 [cs.SI]

[14] Kuhika Guptaa, Matthew C. Nowlinb, Joseph T. Ripbergera, Hank C. Jenkins-Smitha, and Carol L.Silvaa. 2019. Tracking the nuclear 'mood' in the United States: Introducing a long term measure of public opinion about nuclear energy using aggregate survey data. *Energy Policy* 133 (2019), 110888. https://doi.org/10.1016/j.enpol.2019.110888

[15] Jiawei Han, Jian Pei, and Yiwen Yin. 2000. Mining Frequent Patterns without Candidate Generation. *SIGMOD Rec.* 29, 2 (May 2000), 1–12. https://doi.org/10.1145/335191.335372

[16] Jeremy Howard and Sebastian Ruder. 2018. Universal Language Model Fine-tuning for Text Classification. arXiv:1801.06146 [cs.CL]

[17] Tianran Hu, Haoyuan Xiao, Thuy-vy Thi Nguyen, and Jiebo Luo. 2017. What the Language You Tweet Says About Your Occupation. *CoRR* abs/1701.06233 (2017). arXiv:1701.06233 http://arxiv.org/abs/1701.06233

[18] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. arXiv:1910.13461 [cs.CL]

[19] Hailong Li, Jaewan Choi, Sunjung Lee, and Jungho Ahn. 2020. Comparing BERT and XLNet from the Perspective of Computational Characteristics. *AInternational Conference on Electronics, Information, and Communication (ICEIC)*, 1–4. https://doi.org/10.1109/ICEIC49074.2020.9051081

[20] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692 [cs.CL]

[21] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. arXiv:1711.05101 [cs.LG]

[22] Maria L. Loureiro and Maria Allo. 2020. Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the U.K. and Spain. *Energy Policy* 143 (2020), 111490. https://doi.org/10.1016/j.enpol.2020.111490

[23] Jonathan Mellon and Christopher Prosser. 2017. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics* 4, 3 (2017), 2053168017720008. https://doi.org/10.1177/2053168017720008

[24] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and James Rosenquist. 2011. Understanding the Demographics of Twitter Users. In *ICWSM*.

[25] Dhiraj Murthy, Alexander Gross, and Alexander Pensavalle. 2015. Urban Social Media Demographics: An Exploration of Twitter Use in Major American Cities. *Journal of Computer-Mediated Communication* 21, 1 (11 2015), 33–49. https://doi.org/10.1111/jcc4.12144 arXiv:https://academic.oup.com/jcmc/article-pdf/21/1/33/22316402/jjcmcom0033.pdf

[26] Viet Phuong Nguyen and Man-Sung Yim. 2018. Examination of different socioeconomic factors that contribute to the public acceptance of nuclear energy. *Nuclear Engineering and Technology* 50, 5 (2018), 767–772. https://doi.org/10.1016/j.net.2018.02.005

[27] Rob Nikolewski. 2018. *Regulators vote to shut down Diablo Canyon, California's last nuclear power plant.* Retrieved March 25, 2021 from https://www.latimes.com/business/la-fi-diablo-canyon-nuclear-20180111-story.html

[28] Eunil Park. 2019. Positive or negative? Public perceptions of nuclear energy in South Korea: Evidence from Big Data. *Nuclear Engineering and Technology* 51, 2 (2019), 626 – 630. https://doi.org/10.1016/j.net.2018.10.025

[29] Rene Pfitzner, Antonios Garas, and Frank Schweitzer. 2012. Emotional divergence influences information spreading. In *AAAI ICWSM 2012*. 2–5.

[30] Samuel Scarpino, James Scott, Rosalind Eggo, Nedialko Dimitrov, and Lauren Meyers. 2016. Data Blindspots: High-Tech Disease Surveillance Misses the Poor. *Online Journal of Public Health Informatics* 8 (03 2016). https://doi.org/10.5210/ojphi.v8i1.6451

[31] Morgan Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed Brubaker. 2020. How We've Taught Algorithms to See Identity: Constructing Race and Gender in Image Databases for Facial Analysis. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 058 (May 2020), 35 pages. https://doi.org/10.1145/3392866

[32] Michael Shellenberger. 2019. *Why California May Go Nuclear.* Retrieved March 2, 2021 from https://www.forbes.com/sites/michaelshellenberger/2019/09/03/why-california-may-go-nuclear/?sh=5c498afaae02

[33] Anmol Soni. 2018. Out of sight, out of mind? Investigating the longitudinal impact of the Fukushima nuclear accident on public opinion in the United States. *Energy Policy* 122 (2018), 169–175. https://doi.org/10.1016/j.enpol.2018.07.024

[34] Zijian Wang, Scott Hale, David Ifeoluwa Adelani, Przemyslaw Grabowicz, Timo Hartman, Fabian Flöck, and David Jurgens. 2019. Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. *The World Wide Web Conference* (5 2019). https://doi.org/10.1145/3308558.3313684

[35] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ Salakhutdinov, and Quoc Le. 2019. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/dc6a7e655d7e5840e66733e9ee67cc69-Paper.pdf