Towards Fine-grained Classification of Climate Change related Social Media Text

Roopal Vaid, Kartikey Pant and **Manish Shrivastava**
International Institute of Information Technology, Hyderabad, India
{roopal.vaid, kartikey.pant}@research.iiit.ac.in,
m.shrivastava@iiit.ac.in.

Abstract

With climate change becoming a cause of concern worldwide, it becomes essential to gauge people's reactions. This can help educate and spread awareness about it and help leaders improve decision-making. This work explores the fine-grained classification and Stance detection of climate change-related social media text. Firstly, we create two datasets, *ClimateStance* and *ClimateEng*, consisting of 3777 tweets each, posted during the 2019 United Nations Framework Convention on Climate Change and comprehensively outline the dataset collection, annotation methodology, and dataset composition. Secondly, we propose the task of Climate Change prevention stance detection based on our proposed *ClimateStance* dataset. Thirdly, we propose a fine-grained classification based on the *ClimateEng* dataset, classifying social media text into five categories: *Disaster*, *Ocean/Water*, *Agriculture/Forestry*, *Politics*, and *General*. We benchmark both the datasets for climate change prevention stance detection and fine-grained classification using state-of-the-art methods in text classification. We also create a Reddit-based dataset for both the tasks, *ClimateReddit*, consisting of 6262 pseudo-labeled comments along with 329 manually annotated comments for the label. We then perform semi-supervised experiments for both the tasks and benchmark their results using the best-performing model for the supervised experiments. Lastly, we provide insights into the *ClimateStance* and *ClimateReddit* using part-of-speech tagging and named-entity recognition.

1 Introduction

The effects of climate change are becoming increasingly apparent, with various natural disasters, including floods, droughts, storms, and fires, increasing in intensity and frequency. The biosphere is changing, endangering the natural resources and agriculture that are essential for our survival. According to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) [1], continued climate change will have severe and irreversible impacts on people and ecosystems worldwide. According to the report, climate change is predicted, with high confidence, to lead to increased intensity and frequency of daily temperature extremes, sea-level rise, ocean acidification, and reduced crop yields. Climate change and its effects have become major causes of concern globally, leading to increased participation in public discourse. They have been the subject of various newspaper articles, scientific papers, blogs, and social media threads.

Although many steps can help control the intensity and effects of climate change, integrating these steps with public policy depends on the vox populi of climate change. There are multiple ways to study and quantify public opinion on climate change. However, traditional methods, including polling, do not take advantage of the growing prevalence and the abundance of public discourse in social media. Twitter is one of the most popular social media and serves as a vital data source for determining public opinion and perception of climate change. Globally, it has more than 200 million daily active users, with an average annual growth of around 20% in the number of active users. Linden (2017) discusses the impact social networks have on developing climate change risk perception, which suggests the importance of understanding discourses in social media for this domain.

One of the primary concerns around climate change is polarization, with social media being one of the key influencers of the same. As has been observed in previous works, climate change skepticism has achieved a higher level of visibility in media than scientific literature (Boykoff and Boykoff, 2004). Nevertheless, it may help mit-

[1] https://www.ipcc.ch/assessment-report/ar5/

Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics Student Research Workshop, pages 434 - 443
May 22-27, 2022 ©2022 Association for Computational Linguistics

igate differences and spread awareness of information related to climate change. Social media also creates an open space for organizations, climate activists, and scientists to reach more people worldwide. *UKCOP26* and *Greenpeace* are a few examples that use social media platforms to share knowledge about current climate conditions and collaborate with artists, activists, politicians, and academic institutions. Hence, the importance of social media outreach in climate change awareness campaigns is immense for effectively reaching a large global audience.

This work proposes climate change prevention stance detection and fine-grained classification of climate-change-related social media text. We release two Twitter-based English datasets [2], *ClimateStance* and *ClimateEng*, consisting of 3777 tweets, manually annotated for both the tasks. Thirdly, we benchmark state-of-the-art text classification models including *BERT*, *RoBERTa* and *DistilBERT* on both the tasks. Fourthly, we create a Reddit-based pseudo-labeled dataset ClimateReddit from the best performing model for ClimateStance and ClimateEng and benchmark its performance based on a smaller manually annotated test dataset. Finally, we perform a linguistic feature-based analysis for both the datasets based on part-of-speech tagging and named entity recognition.

2 Related Works

Early work in analyzing climate-change-related text in the social media setting is primarily focused on statistical analysis (Kirilenko and Stepchenkova, 2014; Pearce et al., 2013; Kirilenko et al., 2014; Cody et al., 2015). Kirilenko et al. (2014) collected tweets on climate change and global warming in five languages and studied the effect of geography, time, major news events that inspired central topics of discussion over climate change. Pearce et al. (2013) presented the tweet authors and topics associated with the publication of the *IPCC's AR5* on the physical science basis for climate change based on the Tweet's hashtags. Moreover, Kirilenko et al. (2014) performed the analysis on tweets during 2012-2013 to conclude that users are establishing a relationship between temperature anomalies and climate change. On the other hand, Cody et al. (2015) used Hedonometer to determine how collective sentiment differs in response to climate change-related events, news, natural disasters, oil drillings. They conclude that natural disasters and other phenomena related to climate change contributed to a decrease in overall happiness. Although the works mentioned above are immensely helpful in understanding climate change-related discourses in social media, recent advances in natural language processing enable the fine-grained detection of climate-change-related social media text. The advent of contextualized word representation for improving natural language representation for various downstream tasks, including text classification, has been particularly significant.

Recent work in the area employs the techniques of topic modeling (Dahal et al., 2019), and lexicon-based sentiment analysis (Loureiro and Alló, 2020). Dahal et al. (2019) provided an overview of high-impact areas where machine learning and AI can assist the fight against climate change and highlighted climate mitigation and adaptation, as well as meta-level tools that enable other strategies. Loureiro and Alló (2020) analyzed Twitter conversations related to climate change in UK and Spain and employed NLP tools to access the sentiment associated and various emotions evoked by these tweets. They used the lexicon developed by the National Research Center Canada (NRC), denoted as EmoLex (Mohammad and Turney, 2013). Luo et al. (2020) released the Global Warming Stance Detection Dataset, specifically focused on identifying stance on global-warming-related sentences from news articles. Sobhani et al. (2016) released a Twitter dataset for stance detection and further concluded that sentiment features assist in stance classification but are not sufficient on their own. Moreover, Maynard and Bontcheva (2015) release an open-source toolkit for enabling researchers to use Twitter to analyze and understand the engagement of the society regarding climate change.

3 Dataset

This section outlines the dataset creation process for both fine-grained classification of climate-change-related tweets and climate change prevention stance detection. First, we detail the data collection process, which entails scraping, filtering, and preparing the text for annotation. Secondly, we outline the data annotation schema for the fine-grained climate-change-related tweet classification task, along with examples of each of the five categories. We then detail the data annotation schema

[2] https://github.com/serendipity5497/finegrained-climate-change-social-media

Stance	Example
Favor	*"It's time for the American electorate to make #climate change a political do-or-die, up and down the ticket." #ClimateCrisis* *It's not TOO late, but it's late to start reversing climate change.*
Against	*UN not satisfied with hysteria over "Global Warming", "Climate Change". They are seeking language to scare us line "Climate Calamity" to push their fake narrative. #ClimateHoax* *You want to solve climate change, become an electrician.*
Ambiguous	*It's going to be an interesting week in the UK, with elections looming - from a climate change perspective, this is what the major parties are saying* *BBC News - General election 2019: Your questions on climate change and the environment*

Table 1: Examples of Climate Change Prevention Stance Detection Task

for the climate change prevention stance detection task. Finally, we calculate the inter-annotator agreement to evaluate the efficacy of the annotation process.

3.1 Data Collection

3.1.1 Twitter Data Collection

Using the Twitter Application Programming Interface (API) [3], we collected a sample of tweets between 1st December 2019 and 14th December 2019 as the UN Climate Change Conference COP 25 was held from 2 – 13 December 2019. To accommodate different time zones, we start collecting data one day before the conference and collect it until one day after the conference. In total, we collected 378772 tweets along with their metadata. In order to extract climate-change-related tweets from this dataset, we constructed a list of keywords relevant to the concerns regarding climate change - *Climate Change*, *Global Warming*, *Warming Planet*. Apart from these keywords, we also collect tweets containing the following hashtags *#climatechange*, *#climateaction*, *#globalwarming*,*#fossilfree*, *#climatehoax*, *#climatetaxfraud*. After removing non-English tweets, we were left with 263041 tweets. We used Twitter ID deduplication to remove overlapping redundant tweets from multiple hashtags or keywords. Further, we deduplicate tweets based on tweet text to remove duplicates leaving us with 243781 tweets. Lastly, for performing the human annotation process, we sampled 3777 tweets.

[3] https://developer.twitter.com/en/docs/twitter-api

3.1.2 Reddit Data Collection

We use Pushshift (Baumgartner et al., 2020) for extracting Reddit comments related to climate change. For this purpose, we use four subreddits that engage in climate change discourse, namely: r/climate, r/Climateskeptics, r/ClimateActionPlan, r/climatechange. Through this method, we extracted 6591 comments in total. We then preprocess these comments to remove hyperlinks and markdown symbols representing stylized text (i.e., bold and italic). Finally, we split the dataset into two parts: 6262 comments for creating the pseudo-labeled dataset and 329 comments for manual annotation for benchmarking the pseudo-labeled dataset.

3.2 ClimateStance: Climate Change Prevention Stance Detection

We use the term stance as a broad concept covering sentiment, evaluation, appraisal, or attitude and its associated information that is stance target and further use this to evaluate the stance. Similar to Sobhani et al. (2016) we use favor, against and ambiguous labels. We categorize each tweet into one of the three categories in terms of its stance towards climate change prevention:

- **Favor**: Expressions of opinion, action, concern against the climate change phenomenon.
- **Against**: Expressions of distance, ignorance towards signs of climate change, extreme climates, and the opposition of climate change policies or actions taken by the governing bodies.

Class	Examples
Disaster	Take a swim in the charcoal, kids - Sydney beach today (Malabar) #NSWfires #ClimateChange #AustraliaFires Too late to act on fires after they start - need to stop them by acting on climate change #qana
Ocean/Water	IUCN report: Oceans losing oxygen at rapid rate due to #CLIMATE change, #POLLUTION Good news, when climate change melts all of Greenland's ice and deflects the gulf stream away from Europe, you'll get all the snow you could ever ask for!
Agriculture/Forestry	Our most important mountains are under threat—thanks to climate change Extensive livestock farming in rainfed pastures and grazing land could mitigate climate change while being more humane and just.
Politics	It's astonishing that battling climate change is politicized, mainly because it hurts republicans right in their pockets. Vote and vote for someone sees the importance of mitigating climate change by any necessary means
General	Everything is due to climate change? This sounds like a propaganda Am committed and will expect all of the below and more - including tackling homelessness and climate change issues with determination

Table 2: Examples of Fine-grained Classification Task

- **Ambiguous**: Do not express any clear stance towards climate change. Tweets with sarcastic tones were also marked as ambiguous.

3.3 ClimateEng: Fine-grained Classification

The collected data was then manually annotated on the following categories: Disaster, Ocean/Water, Agriculture/forestry, Politics, General.

3.3.1 Disaster

This category contains tweets related to various climate-change-influenced natural disasters, including wildfires, floods, hurricanes, and droughts. These references entail:

- References containing opinions about specific instances of natural disasters.
- Information regarding specific instances of natural disasters.

3.3.2 Ocean/Water

This category contains tweets that are:

- References to the effects of climate change on biodiversity on ocean, seas, and other water bodies.
- References to water-based activities that accelerate climate change.
- References to how biodiversity on land adapts to the effects of climate change.

3.3.3 Agriculture/forestry

This category contains tweets that are:

- References to the effects of climate change on biodiversity on land, crop yields.
- References to activities including deforestation and fossil fuel burning accelerating climate change.
- References to how biodiversity on land is adapting itself to the effects of climate change.

3.3.4 Politics

This category contains tweets that are related to:

- Quotes of different world leaders on the topic of climate change.
- References about actions taken by institutions like UN to spread awareness about the increasing concerns about climate change.
- References to policies being put in place like Newgreendeal, COP25.

3.3.5 General

This category contains tweets that are:

- References of people discussing and spreading awareness about climate change without a specific focus like ocean, water.
- References of climate change affecting suburban lives.

3.4 Semi-supervised Experiments

We also create *ClimateReddit* dataset to perform experiments with semi-supervised learning for the task of stance detection and Fine-grained Classification for a Reddit-based dataset. Semi-supervised learning is often used for utilizing a large amount of unlabeled data to improve the predictive performance of models across various machine learning tasks (Blum and Mitchell, 2000; Chapelle et al., 2006). For our semi-supervised experiments, we use the method of pseudo-labeling. In this method, we first train a "teacher" model based on our Twitter-based annotated datasets, namely, *ClimateEng* and *ClimateStance*. We then use this model to predict the labels for the un-annotated Reddit dataset and create a pseudo-labeled dataset from the predictions. We denote this pseudo-labeled dataset of Reddit comments along with its predicted stance and fine-grained climate-based classification labels as *ClimateReddit*.

3.5 Inter-annotator Agreement

Two human annotators with a linguistic background and proficiency in English conducted the annotation of the dataset to classify the tweets according to the schemas mentioned above. We selected a sample annotation set consisting of 100 tweets per class from all across the dataset. Throughout the annotation process, these sample annotation sets served as the reference baseline of each category.

We also analyze the disagreements between the two annotators on both the fine-grained classification task and the stance detection task. The use of sarcasm in the tweets led to disagreements in many such cases, particularly in the case of stance detection. To accurately capture the stance for those cases, we marked them to be ambiguous. Moreover, the implicit bias of the annotators towards specific entities also led to disagreements between the annotators. We tried our best to select the more objective answer from those labels for creating our corpus.

We calculated the Inter-Annotator Agreement (IAA) to validate the annotation quality. For both annotation tasks, we compute the IAA between the two annotation sets of 3777 tweets using Cohen's Kappa coefficient (Fleiss and Cohen, 1973). We obtained the Cohen Kappa scores of 0.817 and 0.739 for the *ClimateStance* and the *ClimateEng* respectively. Moreover, we also calculate the Cohen Kappa score to be 0.850 for the fine-grained classification task and 0.864 for the stance detection task between the two annotation sets for the manually annotated test split of the *ClimateReddit* dataset. These denote that the quality of the annotations and the presented datasets are significantly productive.

4 Methodology

This section briefly describes the various state-of-the-art models that we used for our benchmarking experiments.

4.1 FastText

FastText (Joulin et al., 2017) is an open-source library for efficient learning of word representations and sentence classification. It allows training both supervised and unsupervised word and sentence representations, also supporting training using both continuous bag-of-words and skip-gram techniques. Since *FastText* uses character n-grams while generating embeddings, it can create representations for words that do not appear in the training corpus. Moreover, *FastText* is capable of achieving good predictive performance efficiently without a pre-trained corpus.

4.2 BERT

BERT released by Devlin et al. (2019) is a bidirectionally trained language model. It exploits a novel technique called Masked LM (MLM) Masking processing text in both directions and using the full context of the sentence, i.e., words to both left and right of the masked word, to predict the masked word. It relies on the Transformer model, which works by performing a small, constant number of steps applied to understand relationships between all words in a sentence, regardless of their respective position, using an attention mechanism. In terms of the type of training data used, it can be classified into cased and uncased variants, based on the letter casing of the training data. We use the Base cased and Large cased variants for our benchmarking experiments.

4.3 RoBERTa

RoBERTa (Liu et al., 2019) is *BERT*-based contextualized word embedding that uses modified key hyperparameters, simpler pre-training objectives, and a different size of training data. Unlike *BERT*, *RoBERTa* does not use the next sentence prediction training objective while using dynamic masking for changing the masked token during training epochs. It uses a larger batch-training size and ten times the training data when compared to *BERT*. These improvements enable *RoBERTa* to obtain significant gains in the predictive performance in various downstream tasks, including *GLUE* (Wang et al., 2018) for text classification. Similar to *BERT*, it also comes in two variants in terms of transformer architecture: Base and Large. Unlike *BERT*, it only comes in the cased variant in terms of the type of training data used. We benchmark both Base and large variants of *RoBERTa*.

4.4 DistilBERT

DistilBERT (Sanh et al., 2019) is a distilled version of *BERT* that uses 40% fewer parameters and is 60% faster while retaining the majority of its predictive performance. It does not use token-type embeddings while removing the pooler in its architecture, reducing the number of layers compared to *BERT* by half. *DistilBERT* uses a composite loss combining distillation, cosine-distance, and language modeling losses to leverage the inductive biases learned by undistilled models during pre-training. In terms of the type of training data used, it can be classified into two variants:- *cased* and *uncased*. We use the cased version of *DistilBERT* for our benchmarking experiments.

5 Experiments

5.1 Experimental Settings

5.1.1 Supervised Experimental Setting

We evaluate our models on a held-out test dataset for all experiments that consist of 10% of the total dataset. For validation purposes, we split the training dataset was further divided in 8 : 1 training:validation split. We use *F1*, *Precision*, *Recall*, and *Accuracy* for evaluating the models. We use the macro variant of the *F1*, *Precision*, and *Recall* which treats all classes equally by taking an unweighted arithmetic mean of all per-class scores.

We use *FastText*'s recently open-sourced automatic hyperparameter optimization functionality and run 100 trials of optimization. For *BERT*, *RoBERTa* and *DistilBERT*, we fine-tune with a learning rate of $1 * 10 - 5$, batch size of 12, and a maximum sequence length of 128 tokens. We validate the models for up to five epochs using the validation dataset and report the best-performing model in our results.

5.1.2 Semi-Supervised Experimental Setting

For generating pseudo-labels and performing the benchmarking experiments, we use the best-performing model in terms of *F1* score for both tasks of stance detection and fine-grained classification. We use the same methodology for training the models as explained in Subsection 5.1.1.

We use all splits of the Twitter-based datasets, namely *ClimateStance* and *ClimateEng*, for their respective tasks, for training the generating the pseudo-labels from the Reddit dataset. For validation, we re-split the dataset into a 9 : 1 split. Now, upon pseudo-labeling, we use the aggregated dataset consisting of both Twitter and Reddit text and re-split the dataset again into a 9 : 1 split for validation. For all our evaluation experiments, we use the same manually annotated dataset split of *ClimateReddit* as the test dataset.

5.2 Experimental Results

5.2.1 Supervised Experiments

From Table 3 which illustrates the results of the climate change prevention stance detection experiment, we observe *RoBERTa-Base* outperform all models in *F1* with a score of 0.510. In contrast, *RoBERTa-Large* outperforms all models in *Accuracy* and *Recall* with *Accuracy* 82.54% and 0.507 *recall* score. *BERT-LARGE* achieved the best *precision* score of 0.530.

Model / Metric	F1	Accuracy	Precision	Recall
FastText	0.343	79.63%	0.503	0.354
BERT-Base	0.464	77.51%	0.507	0.446
BERT-Large	0.489	77.78%	**0.530**	0.470
RoBERTa-Base	**0.510**	81.22%	0.528	0.502
RoBERTa-Large	0.489	**82.54%**	0.473	**0.507**
DistilBERT	0.448	79.37%	0.497	0.430

Table 3: Results for the Stance Detection using *ClimateStance* dataset

Table 4 illustrates the results of the fine-grained-classification experiment. For this task, we observe *RoBERTa-Large* to outperform all models in *F1*, *Accuracy*, and *Precision*, obtaining an *F1* score of 0.735, *accuracy* of 83.07%, and *Precision* of 0.738

in the experiments. At the same time, *RoBERTa-Base* was able to achieve a better *Recall* score of 0.756.

Model / Metric	F1	Accuracy	Precision	Recall
FastText	0.638	73.55%	0.730	0.594
BERT-Base	0.696	78.84%	0.697	0.701
BERT-Large	0.695	78.31%	0.730	0.689
RoBERTa-Base	0.734	80.16%	0.725	**0.756**
RoBERTa-Large	**0.735**	**83.07%**	**0.738**	0.742
DistilBERT	0.694	77.51%	0.695	0.713

Table 4: Results for the Fine-grained Classification using *ClimateEng* dataset

Apart from these, *DistilBERT* and *FastText* also perform competitively while being trained significantly faster than the others. *DistilBERT* obtains an *F1* score of 0.448 in the Climate Change Prevention Stance Detection task and an *F1* of 0.694 in the fine-grained classification task. In contrast, *FastText* obtains an *F1* score of 0.343 in the Climate Change Prevention Stance Detection and an *F1* of 0.638 in the fine-grained classification task.

5.2.2 Semi-Supervised Experiments

For this experiment, we use the best performing models in terms of *F1* score for the Climate Change Prevention Stance Detection task using *ClimateStance* (*RoBERTa-Base*) and Fine-grained Classification task using *ClimateEng* dataset (*RoBERTa-Large*).

Training Data	F1	Accuracy	Precision	Recall
ClimateEng	0.775	88.15%	0.800	0.769
ClimateEng + Pseudo-labelled Reddit Data	0.834	90.27%	0.850	0.823

Table 5: Results for the Semi-Supervised Fine-grained Classification Task

From Table 5, for the task of fine-grained classification, we find that *RoBERTa-Large* trained with all splits of *ClimateEng* performs significantly well in the fine-grained classification task for ClimateReddit dataset, obtaining an F1 of 0.775 and an accuracy of 88.15%. Moreover, using the pseudo-labeled Reddit dataset for training along with *ClimateEng*, we find an even higher F1 of 0.834 and an accuracy of 90.27%.

Training Data	F1	Accuracy	Precision	Recall
ClimateStance	0.343	60.79%	0.403	0.387
ClimateStance + Pseudo-labelled Reddit Data	0.311	60.49%	0.396	0.369

Table 6: Results for the Semi-Supervised Stance Detection Task

In contrast, as illustrated in Table 6, the predictive performance of *RoBERTa-Base* reduces sharply for the task of Stance detection in the semi-supervised setting. It obtains an F1 score of 0.343 and an accuracy of 60.7% when only trained with the *ClimateStance* dataset. Upon adding the additional Reddit-based pseudo-labeled corpus for the Stance detection, we find the model's performance to dip even further, reaching an F1 score of 0.311 and an accuracy of 60.49%. This drop can be attributed to the significant imbalance in class distribution as highlighted in Subsection 6.1.

6 Discussion

6.1 Dataset Composition

In the annotated *ClimateStance* dataset, we observe the primary stance to be in *favor* with a count of 2990 (79.16%), i.e., in conclusion, most discussions showed concern and proposed actions to mitigate climate change. Further, we observed the *ambiguous stance* state with no clear stance on climate change 414 (10.96%) times. In contrast, the tweets against and with confusion towards climate change, i.e., those having a *against stance* state, occurred 373 (9.87%) times.

In the annotated *ClimateEng* dataset, we found the popularity of *General* tweets with a count of 2159 (57.16%) followed by *Politics* class with a count of 1045 (27.67%), which sheds light on how different governing bodies are acting against climate change and citizens' expectations from the governing parties for climate change mitigation. We observed *Ocean/Water* class has a count of 204 (5.40%) as we see the signs of climate change, including rising shorelines and melting glaciers. The *Agriculture/Forestry* class consisted of 197 (5.21%) tweets due to the rising effects of climate change on agricultural practices and biodiversity. We also observed that disastrous events around the globe did follow an increase in discussions regarding climate change and global warming; in the dataset, we were able to capture 172 (4.55%) tweets that could be classified as *Disaster*.

Class	Part-of-Speech						Named Entities				
	PROPN	VERB	NOUN	ADJ	PRON	ADV	PERSON	GPE	MONEY	ORG	DATE
Favor	4.22	3.69	7.76	2.00	1.39	1.15	0.29	0.29	0.39	0.99	0.26
Against	3.22	3.71	7.18	2.26	1.76	1.41	0.30	0.17	0.20	0.71	0.24
Ambiguous	3.55	3.01	6.47	1.81	1.47	1.16	0.31	0.20	0.35	0.82	0.24

Table 7: Mean Value of Part-of-Speech tags and Named Entities in the *ClimateStance* dataset per Class.

Class	Part-of-Speech						Named Entities				
	PROPN	VERB	NOUN	ADJ	PRON	ADV	PERSON	GPE	MONEY	ORG	DATE
General	3.58	3.33	7.08	1.92	1.47	1.13	0.28	0.16	0.35	0.85	0.23
Politics	4.75	4.26	8.14	2.21	1.68	1.34	0.38	0.41	0.39	1.12	0.28
Ocean/ Water	4.94	3.17	7.96	1.78	0.82	0.86	0.22	0.40	0.37	0.97	0.32
Agriculture/ Forestry	4.10	3.51	8.73	1.92	0.72	0.95	0.16	0.22	0.47	0.95	0.19
Disaster	4.46	3.81	8.24	2.23	1.09	1.34	0.27	0.61	0.42	0.95	0.35

Table 8: Mean Value of Part-of-Speech tags and Named Entities in the *ClimateEng* dataset per Class.

In the *ClimateReddit* dataset consisting of 6591 Reddit comments, we observe the primary stance to be in favor with a count of 6269 (95.11%). Further, we observed the *against* stance 251 (3.80%) times and those having a *ambiguous* stance, occurred 71 (1.08%) times. Moreover, upon observing in terms of the fine-grained labels, we found 4699 (71.29%) comments to lie in the *General* category. The next most frequent category was *Politics*, having 1197 (18.16%) comments. The next three categories of comments had a fairly equivalent number of occurrences having 243 (3.69%), 227 (3.44%), and 225 (3.41%) comments for *Ocean/Water*, *Agriculture/Forestry*, and *Disaster* respectively.

6.2 Linguistic Feature Analysis

We compare our annotated features with various linguistics features including part-of-speech (POS) and named entities (NE) on *ClimateStance* and *ClimateEng* datasets. To perform this analysis, we exploit SpaCy [4], an open-source library for advanced natural language processing. We use the *en_core_web_sm* for extracting the part-of-the-speech tagging and performing named-entity recognition from all 3777 tweets.

Table 7 illustrates the results for the part-of-speech tagging and named entity recognition for the *ClimateStance* dataset. We observe that tweets in *favor* stance use proper nouns and nouns the most when compared to other classes. In contrast, tweets with stance *against* displayed a higher use of adjectives, pronouns, and adverbs. While observing NEs, we found the highest occurrence of GPE, MONEY, and ORG tagged NEs in tweets with in *favor* stance. The arguments to support this observation could be stated as in favor stance towards climate change would lead to concern and demand action against climate change. Organizations (ORG) and geopolitical entities (GPE) would be required to make significant changes to bring a systematic change that could slow down climate change. Moreover, the economy needs to adapt to the changing climate, which might be the reason for using entities with the MONEY tag in tweets having stance in *favor* of climate change.

Table 8 illustrates the results for the part-of-speech tagging and named entity recognition for the *ClimateEng* dataset. Tweets classified as *Disaster* had the majority of GPE NEs as well as DATE NEs. We believe this could be due to the localization of disastrous events and tweets holding the political body of the geography for action for mitigation and relief work. Tweets classified as *General* observed the least mention of MONEY NEs. In contrast, we see a higher count of the MONEY NEs in *Agriculture* and *Disaster* classes, which might be due to the cost associated with agricultural industries and disaster mitigation and relief organizations to adapt to the climate change effects witnessed during a disaster. We also observe the most leading mention of ORG NEs in *Politics* class. This observation could be due to references of actions needed to be adopted or are adopted by different organizations to mitigate climate change.

This analysis of linguistic features can be fur-

[4] https://spacy.io/

ther extended to entail a study on the correlation of these features alongside fine-grained labels and stance labels created in *ClimateStance* and *ClimateEng* dataset. The study may lead to interesting sociolinguistic findings while helping out in general understanding of how we use language in a social setting while writing climate-related short-form text. Moreover, this study may also help with information retrieval (Li et al., 2022) based on the named entities alongside our created labels.

7 Conclusion

In this work, we proposed the task of predicting Stance in social media texts related to climate change. We further proposed the task of categorizing these texts into five categories. We benchmarked the datasets using state-of-the-art contextualized word embeddings and provided baselines for both the proposed tasks. We observed that *RoBERTa-Large* outperforms all other models in three of the four evaluation metrics for the fine-grained classification task, obtaining an *F1* of 0.735. Moreover, we also observed that *RoBERTa-Base* obtained the best *F1* score in the Stance detection task with a 0.510 *F1* score. We further extend this work to the semi-supervised setting and use pseudo-labeling to predict for the Stance detection and fine-grained classification tasks in a Reddit-based dataset. This work can be further expanded to analyze people's reactions to climate change in multi-modal and multilingual settings to get a broader understanding.

Ethical Considerations

This paper uses data obtained from the Twitter Developer API [5] and freely available social media data from the Reddit platform using the Pushshift API (Baumgartner et al., 2020). Moreover, we only provide the Tweet ID in the annotated datasets along with a data preparation script in accordance with the Twitter Terms of Service. We also compensated the human annotators with a stipend more than the minimum wage in India.

Acknowledgements

We would like to thank the annotators Kumar Abhishek and Shreya Chandorkar for their immensely useful contribution to this work. We would like to thank the anonymous reviewers for providing critical suggestions.

[5] https://developer.twitter.com/en/docs/twitter-api

References

Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *ICWSM*.

Avrim Blum and Tom Mitchell. 2000. Combining labeled and unlabeled data with co-training. *Proceedings of the Annual ACM Conference on Computational Learning Theory*.

M. Boykoff and Jules Boykoff. 2004. Balance as bias: global warming and the us prestige press. *Global Environmental Change-human and Policy Dimensions*, 14:125–136.

Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien, editors. 2006. *Semi-Supervised Learning*. The MIT Press.

Emily M. Cody, A. J. Reagan, Lewis Mitchell, P. Dodds, and C. Danforth. 2015. Climate change sentiment on twitter: An unsolicited public opinion poll. *PLoS ONE*, 10.

Biraj Dahal, S. Kumar, and Zhenlong Li. 2019. Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, 9:1–20.

J. Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL*.

J. Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33:613 – 619.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *EACL*.

A. Kirilenko, T. Molodtsova, and S. Stepchenkova. 2014. People as sensors: Mass media and local temperature influence climate change discussion on twitter. *Global Environmental Change-human and Policy Dimensions*, 30:92–100.

Andrei P. Kirilenko and Svetlana O. Stepchenkova. 2014. Public microblogging on climate change: One year of twitter worldwide. *Global Environmental Change*.

Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. 2022. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):50–70.

S. Linden. 2017. Determinants and measurement of climate change risk perception, worry, and concern. *Social Science Research Network*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, M. Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692.

M. Loureiro and M. Alló. 2020. Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the u.k. and spain. *Energy Policy*, 143:111490.

Yiwei Luo, D. Card, and Dan Jurafsky. 2020. Desmog: Detecting stance in media on global warming. *ArXiv*, abs/2010.15149.

D. Maynard and Kalina Bontcheva. 2015. Understanding climate change tweets: an open source toolkit for social media analysis. In *EnviroInfo/ICT4S*.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29.

W. Pearce, K. Holmberg, I. Hellsten, and B. Nerlich. 2013. Climate change on twitter: topics, communities and conversations about the 2013 ipcc report.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.01108.

Parinaz Sobhani, Saif M. Mohammad, and Svetlana Kiritchenko. 2016. Detecting stance in tweets and analyzing its interaction with sentiment. In **SEMEVAL*.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2018. Glue: A multi-task benchmark and analysis platform for natural language understanding. *ArXiv*, abs/1804.07461.