Hertie School

Stance Detection on Climate Science Tweets

Augustine Malija

Thesis for Master of Public Policy
Class of 2021
Supervised by Dr. Slava Jankin
Hertie School
Finn Muller-Hansen
Prof. Dr. Jan Minx
The Mercator Research Institute on Global Commons and Climate Change (MCC)

14th June 2021

Abstract

Stance detection is an exciting research methodology in the policy realm since it seeks to understand positions people take towards a cross-cutting phenomena. This methodology is particularly useful in studying social media data in form of tweets. Detecting stance towards climate science tweets is of particular importance in facilitating efforts on mitigating climate change. Understanding the divide in stance tweeter users take towards climate science has been the main motivation for this study. Previous research has established that compared to continental Europe, the controversy in delivery of climate change information to United States citizenry leads for it to be less informed (Jang 2015) on climate change. Much of this controversy is driven by party lined media. Whereas democrat oriented traditional media share pro climate science information, republican oriented media share information on the flip side. This leads to, in principle, a clear division (Jason 2017) in tweet stances towards climate science in the United States. This study expands further by deploying Support Vector Machines as a baseline and a Bidirectional Encoder Representations from Transformers as a deep learning component to study such stances. These state of the art models achieve promising results that foster the understanding of such a divide in stances and drivers towards them. They can also be adopted to understand stances tweet users take towards other cross-cutting discussion topics. However, the models need improvement when other axes are incorporated in studying stances people take towards climate change.
[*].

[*] Notebooks and code for the project can be accessed here: Thesis Github Repository

Contents

Contents

1 Executive Summary

Summary

chapter.1section.2.1section.2.2chapter.3section.3.1section.3.2section.3.3section.3.4c

Since the beginning of the discipline, social sciences research has been studying social phenomena in order to inform policy making. Qualitative and quantitative research methods have always made sure that public opinion is represented through data both qualitative and quantitative. All across time, the evolution of social science research methods have always made sure that both primary and secondary data kept representing the public's views. Recent developments in machine learning (Franco 2020) and natural language processing (Boutilier 2020) have offered an opportunity to have an even wider representation in form of being able to partake very large datasets.

A combination of natural language processing and machine learning offers an insurmountable opportunity to perform analysis from text. To understand the state of a phenomena, it is important to know what people both in position of power and the public think about it. Corpuses from parliamentary speeches serve as one of the sources to understand opinion from an authoritative point of view. But to facilitate democracy, it is important to understand opinion form the public's point of view.

Technological advancement has simplified the process of obtaining data. The emergence of social media has offered a platform where such data can be extracted. Social media such as Tweeter and Facebook (Munzert 2020) are playing a big role as data platforms. The advantages of extracting data from social media over preceding methods are that respondents won't be biased to withhold information and data obtained is willingly posted by users. A social media researcher does not have to invest ingenuity in building rapport to respondents for the sake of obtaining the accurate information. In itself, social media data is highly genuine to be reliable for social sciences research purposes. In comparison, Tweeter has been the media that stood out because the extent to which data is willingly posted is much more fluid than other media. Above all, it is a favorable (Kristin 2019) platform to discuss matters relating to public policy where both people in position of influence and the public get to debate on pressing public policy issues.

Natural Language Processing and Machine Learning methods are instrumental in studying such debates. A variety of approaches such as keyword studies, sentiment analysis and stance detection exist in studying social media debates. While keywords are useful in understanding the ecosystem of political frames to the public, sentiment analysis offers the opportunity to study how people feel about certain topics and rank them on a scale from positive to negative. Although ranking people's sentiment about a topic informs on how best to improve policy, it is telling to understand stances they take towards such topics. Stance detection studies the position people take towards a policy issue. A person's position towards a policy issue, tells you how firm they are holding such a grip. Such positions could be supporting, refuting or not taking any side towards a topic in question.

This paper studies the stance people take towards climate science. It applies stance detection to gauge whether people accept, deny or take a neutral position when they tweet about climate science. It also discerns from climate policy related tweets. Accepting climate policy implies supporting climate science while denying climate policy does not mean denying climate science. Based on the target feature – climate science – tweets are sequentially annotated subject to three categories "in favor", "against" and "neutral". At the same time, tweets that are related to climate policy are identified and set aside. These will be used for future research in studying stances people take when they tweet about climate policy. Category classification depends on annotation rules developed.

To initially evaluate the data, Support Vector Machines (SVMs) function as the baseline. Even though the model performs well with other data, strikingly suboptimal prediction results are found. Much of this is explained by a few data points used compared to experimental datasets. With the existent model, predictions were made on unclassified tweets that led to communicated results. A deep learning component using Bidirectional Encoder Representations from Transformers was used to asses the accuracy of predictions made by the baseline model.

A weak but positive effect of users taking a neutral stance is observed to nudge users denying climate science towards supporting mitigation efforts. As it appears, such deniers are ones that have been swayed by deniers with authority such as prominent right wing politicians. In tandem, this positive effect is also contributed with major events happening in the climate change calendar such as the Global Week for the Future and the warmest month on record. Keenly, practicality and a bipartisan divide turned out to be major drivers for users to be climate science deniers. This practicality manifests through rigidity and pragmatism towards reception of climate science information. Moreover, users following the right wing politicians turned out to be deniers compared to ones following left wing politicians. On the other hand, a strong trust in climate science institutions was noted as major driver for users who take a stance in favor of climate science.

2 Related Work

Summary

2.1 Stance Detection Methodology

A following synopsis of related studies offers theoretical underpinning and shed light into the progression of this study.

Dilek Kucuk and Fazli Can determine stance on tweets posted about two opponent football clubs in Turkey – Galatasaray and Fenerbahce. They achieve their objective both through the use of explicit names of clubs and implicitly on tweets mentioning respective management or footballers themselves.

The paper is selected as a baseline because of its employment of SVM-based techniques in detecting stance from football tweet data.

One million tweets about popular Turkish Sports clubs – Galatasaray (Target 1) and Fenerbahce (Target 2) – published between August 18 and September 6, 2015 were compiled. Only Favor and Against Stance classes were considered. The Club stood as a stance target either in cases where the tweets mentioned the sports clubs in their entirety or in cases where management or some footballers were praised or criticized. From these annotations, 175 tweets were in favor of Target 1 while another 175 tweets were against it. On the other hand, 175 tweets were in favor of target 2 while another 175 tweets were against it. Therefore, the stance annotated dataset contains 700 tweets.

The target identification criteria comprised tweet features based on unigrams, bigrams, hashtags, external links, emoticons and named entities. Through the use of Natural Language Processing, SVMs for target 1 and two were formed. These SVMs were then trained and tested on the annotated datasets. Two SVM Classifiers – one for each target – using unigrams as features were trained. Prior to the extraction of unigrams, automated pre-processing was used to filter-out stopwords. A SVM implementation employing the SMO algorithm to train a classifier with a linear kernel was used. It provided results of the two classifiers using the metrics of precision, recall and F-Measure. To observe the contribution of hashtags on stance detection, the existence of hashtags in tweets was used as an additional feature to unigrams.

It proposes the use of n-gram based features of SVM classification for a favorable performance on stance detection problems. This is because using bigrams as the only feature leads to poor results. Conversely, unigram features lead to superior results compared to bigram features.

Aseel Addawood et al uses stance classification to study tweets that were related to the spring 2016 debate over FBI's request that apple decrypt a user's iPhone. They studied the public's stance between advocates for individual privacy and advocates for national security. Their choice of stance detection serves to understand positions users take on top of just determining if their opinion is positive or negative. They explore whether classifying stance in an ideological debate can determine how frequently each position and what attitudes users express. Using Support Vector Machines (SVM) approach they labeled tweets by both the topic of discussion and user's stance towards that topic. Data was annotated based on a user's position towards individual right to privacy and national security. Their SVM results for all features of the three classifiers using a 10-fold cross validation attain a precision, recall and an F1 score of 93.2 percent. On the other hand, their SVM results for the three classifiers using 10-fold cross validation attains a precision, recall and an F1 score of 83.8 percent.

An analysis of twitter data collected over six weeks before the June 2016 Brexit referendum is collected to understand the relationship between twitter mood and referendum outcome as well as identifying the most influential twitter users in the pro-Brexit camps. A total of 4.5 million tweets from almost one million users posting about Brexit between May 12th and June 24th 2016 were collected. A machine learning text classification model was developed to detect the stance people take when tweeting about Brexit. To conclude on tweet stances, Grcar et al balances prevailing tweets of a user to come to an establishment if it is pro, against or neutral. If most of a user's prevailing tweets are neutral, the stance for such a user is regarded as neural. They establish that the most influential twitter users are not the most productive, that the pro-brexit camp was more influential and had a larger impact in the campaign compared to the opponents and that the pro-brexit camp was more polarized than the contra.

Conforti et al studies stance from articulated datasets such as news articles connecting the deep mutual influence existing between social media and news sources. STANce Detection and Evidence Retriaval (STANDER) is introduced that collects news articles in English discussing mergers and acquisitions from high reputation sources. Stance is classified based on four categories; Support, Refute, Comment and Unrelated. An MLP leveraging sentence BERT embedding (BertEmb), MLP leveraging Universal Sentence Encorder's sentence embeddings and BiLSTM over Glove Embeddings methodologies are employed. A drop in performance is observed when considering only one class against the rest. Interestingly, a gain in stance classification is observed when BertEmb is jointly trained to perform both stance Detection and Evidence Retrival.

Building from the SemEval 2016 dataset, Dey et al uses a traditional machine learning approach SVM to detect stance from a limited annotated dataset of 2,914 labeled tweets. Their study is done through two approaches whereby the first approach classifies tweets into neutral and non-neutral (favor/against). The non-neutral tweets are then classified to favor and against classes. This gives the study an average F-Score of 74.44 that outperforms their baseline literature. It also yields triumphant F-Score improvements of 5.46 and 5.29 for phase one and two respectively.

2.2 Other Methodologies

In a study by S.Mo Jang and P. Sol Hart, 5.7 million tweets mentioning either "climate change" or "global warming" emanating from four English speaking countries (US, UK, Canada and Australia) from July 1st 2012 to June 30th 2014 were accessed from Topsy. Topsy is a third party firm licensed by twitter that provides open access to actual tweets and metadata aligned to about 100 million active accounts. Retweets were also included because they show the perception extent for the importance of original tweets. To answer questions on geographical and partisan differences, tweets were filtered based on location and democratic-republican coded states.

A guiding assumption was that any tweet mentioning either "climate change" or "global warming" and "real" or "fact" at the same time was regarded to be engaging in legitimate climate change discussions. This is to identify uniqueness of public rhetoric that represents single themes of a more complex issue.

Top 500 tweets that were retweeted mentioning either climate change or global warming were looked into first. Through the analysis of such tweets, terms and phrases that commonly represented particular themes were gathered. Moreover, a series of word clouds showing the most used words in content was visually arranged so that the size of a word corresponded to the frequency of its appearance in a text. This procedure aided in identifying important keywords that reflect important themes – keywords with less than 1000 tweets per year and ones with too much noise were discarded.

Five climate change themes were selected as elaborated in this list; (1) real themes included the term "real" or "fact" (2) hoax themes included the term "hoax" or "lie" or "fraud" (3) impact themes included the term "impact" or "impacts" or "threat" or "threats", or "consequences", or "effects" or "affect" or "affects" or "disaster". (4) cause themes included the term "cause" or "causes" or "fuel" or "carbon" or CO2 or "human" and (5) action themes included the terms "action" or "act" or "stop" or "fight" or "policy" or "policies".

It found that tweets reflect much of the controversy observed in traditional media. The US presented a higher ratio of real themes than the rest of the countries. The same applied to hoax themes. The US registered a lower ratio of cause themes than the rest. This pattern was commonly observed concerning impact and action themes except that impact themes in the US and Australia were not significantly different.

There are systematic differences between red and blue states. Red states include real and hoax more than blue ones. In contrast, blue states focused more on impact and action than red states. But they are all equal in cause themes. Red states turned out to be using global warming more than blue states. But the use global warming reduced when tweets discussed climate change in terms of its impact.

It appears that all countries used climate change more frequently than they used global warming. Although American users showed preference of using global warming than climate change among the countries. Moreover, all countries were more likely to use the term global warming than climate change when tweets related to hoax themes.

Before diving into data and analysis Cody et al's study is backed by the following theoretical propositions. First is that climate change is anthropogenic – a result of human activity. They posit that most of what the general public learns comes from social media and twitter is a dominant source. They rely on findings that twitter activity positively correlates with proximity to the storm and physical damage. Also that individuals are likely to form close connections and groups on twitter if they have been affected by a natural disaster.

Cody et al collected 1.5 million tweets mentioning the word "climate" at least once from twitter's gardenhose API over the period of 6 years ranging from 14th September to July

14th 2014. They also include retweets to ensure a higher weighting of messages.

Their methodology is underpinned by the application of a hedonometer – an instrument that uses sentiment scores to calculate the level of happiness score for a large volume of text. 10,222 words of the most frequently used English words from four corpora were assigned happiness ratings using Amazon's Mechanical Turk online marketplace.

Each word was rated based on a rating from 1 – least happy to 9 – most happy subject to how the word made the raters feel. This rating was based on the time scales of day, week and month. Words earning scores between 4 and 6 are classified as neutral and thus were omitted from the study because of the ambiguity of describing them. The word climate earned a score of 5.8 and thus was not included at the time of calculating average happiness. For comparison, they also calculate average happiness for 5 related climate keywords.

To keep an eye on tweets including the word climate but not related to climate, a manual coding of a sample of 1500 tweets determined that 93.5 percent of the tweets were about climate change. This offered assurance that the removal of non-climate tweets did not alter results of the overall happiness score.

Cody et al's methodology relates to the Jang et al in the sense that it picked tweets with the word climate to form a dataset. It also included retweets to add on its stock of dataset. It diverges from Jang through offering an alternative approach of applying the hedonometer to calculate average happiness scores of words.

Williams et al's theoretical underpinning emanates from the argument that studying social media reveals social network structures in the context of online debates which occur to be highly polarized. Social network structures shape opinion and behaviour since they tend to have a strong influence on an individual's perception to phenomena like climate change. In this regard, many individual characteristics such as smoking and political opinion are congregated in social networks. Such homophily arises due to preferential connection to similar individuals and from peer influence. Influences such as musical preferences, health related behavior and emotional transfer are the primary triggers for causes of this homophily. Thus, such clustering determines an individual's adoption of a new attitude or behavior. For example, belonging to a certain online group leads to members of such a group having a similar linguistic construct when communicating within the group.

In tandem with Jang, twitter API was used to collect messages between 13th January 2013 and 30th May 2013 that included "climate globalwarming climatechange" representing tweets that discussed climate change while "agw" and climaterealists representing users that expressed skeptic views about climate change. Friend and follower connections were collected forming a bulk of 590,608 tweets from 179,180 users. To ensure data quality, 100 tweets for each hashtag were sampled and manually assessed for their relevance to the topic of climate change.

Follower, retweet and mention networks were constructed and analyzed using the NetworkXv1.8.1 module for Python. They were filtered and visualized as directed graphs using the ForceAtlas2 force directed layout algorithm in a sense that closely connected users were located near each other. A panel of three climate science researchers classified the most active users based on their expressed views towards climate change into "sceptic", "activist", "neutral" or "unknown" categories and "ambiguous" for users who the panel did not reach consensus.

High frequency of edges connecting users with similar or differing views was used as a measure for homophily. Furthermore, the Louvian method was used to algorithmically sense user communities – people interacting with each other more frequently than they do with others – based on social network. These community level interactions were visualized using nodes while edges were used to represent residual interactions.

Kim et al's theoretical arguments are clustered around user authority structures and homophily. They communicate that 90 percent of tweets containing @ - a form of referring a specific user – are conversational in context, making it opportune to study conversational aspects of tweeter. They argue that tagging and hashtags play a role of topic and community identification that extends a reach to users that are not connected. Notably, only 1 percent and 9 percent of users are the most active and highly active respectively while the rest 90percent only share very few tweets – making the 10 percent evangelists that reach the mass through hashtags. They go on to posit that homogenous minds tend to tweet among themselves more often compared to their contrasting minds. This has proven so when pro-life and pro-choice members within abortion groups tweeted to each other.

Their dataset comprised of 152,893 English tweets with an acronym IPCC between 17th September and 8th of October 2013 extracted from twitter's API. Author names, usernames and hashtags mentioned in the tweets were filtered from the dataset. Attention was paid to the most active and most frequent tweets in order to shy away the impact of spam tweets and noise that might arise during the course of analysis. To study conversational aspects of the messages, tweets mentioning usernames with the @ sign were extracted. They were converted to a network with Webometric Analyst then analyzed using Force Atlas to arrive at positions of nodes and layouts of the network. Attention was paid to frequently mentioned connections out of which a threshold of ten or more connections determined the most active networks. These were further checked for their relevance to climate change.

Kristin Demetrious's theory acknowledges twitter as a space for opinion leadership, demographically skewed and a platform to open and close public debate. Being a source for opinion leadership, twitter is a living barometer for political sentiment where elite users influence mainstream media that in turn report issues for a wider public. In comparison to other media platforms, twitter is the favorite to the young, urbanized, new users and well educated demographic groups. It outweighs others in terms of media content such as news, global reach and preference for forum/issue discussion. Public debates are amplified by twitter's subtle power to influence the unconscious assimilation of public discourses by readers – intertextuality. This occurs through weight, narratives and discourses carried in specific texts/quotes comprised in tweets.

The study's data comprises a spiked Australian twitter activity for the month of February 2017 scrapping tweets with clean coal as a key term during discussions about the AEFL campaign on clean coal. It also included tweets and retweets that exemplified intertextuality. This data was accessed from Tracking Infrastructure for Social Media Analysis – a project that tracks Australian twitter activity in real time to form a large deposit. Using tableau, the data was profiled and select periods of times were singled either due to high levels of engagement or key events in the course of the AEFL campaign. These revealed hashtags that were used to understand key themes discussing the issue of coal.

3 Data Sources

Summary

3.1 Data Acquisition

Twitter offers access to data for research purposes either directly or through specialized institutions. As a thesis practice partner to Hertie School, The Mercator Research Institute on Global Commons and Climate Change secured tweets subject to this study from Tweepy. Tweepy is a specialized institution that helps researchers access data from social media to be used for research purposes. Tweets pertaining climate change from December 2018 to April 2021 were accessed. A total of 10 million tweets were collected. But only fifty thousand tweets were subjected to the study. This was for the purpose of making sure the model works swiftly without having to run indefinitely. Tweets that contained "Climate Change" were singled out. Tweet extraction included filtering English tweets, URLs, hashtags, users, other symbols mentioned in the tweets and retweets that pertained to climate change.

3.2 Data Annotation

A test-driven and iterative approach was used to develop annotation guidelines. A sample of fifty tweets were annotated on a test basis in order to determine the basis for which guidelines will be developed. Guideline development was backed up with the understanding that there exist an asymmetry between implicit positions about climate change and climate policy. Supporting climate policy means implicit acceptance of climate science where as being against does not mean being a climate skeptic. Rather, being a climate skeptic implies being against climate policy. In this regard, there exists a clear distinction between climate science and climate policy. Corollary, a stance on either can – but must not – imply a stance on the other. A stage-wise approach was used to develop annotation guidelines. First, guidelines on climate science were developed while ones on climate policy were generally highlighted. A sample of first 659 tweets were annotated based on climate science guidelines. The rest 341 tweets were annotated based on both climate science and climate policy guidelines.

Thus, two annotation classifications "Climate science" and "Climate policy" came into being with three positions "In-favor", "against" and "neutral". The following are the guidelines used with their narrations and examples.

3.3 Annotation Guidelines

- Climate skepticism

  Claims climate change is not caused by human activity: Primarily asserts that changes in climate are natural and have been there since the start of planet earth.

  - Example tweet: *"I have a math and science degree and man made climate change is a hoax. So ashamed of the supporter of climate is living in Colorado what the* *liberals and no progress progressives have done to my state is pathetic."*

Climate impacts will not be very serious: Primarily refutes calls for immediate action to be taken to combat climate change. They challenge the establishment that planet earth is likely to extinct if no action is taken.

- Example tweet: *"This is because of climate change. If you can taste and smell, then nothing serious!"*

Claims scientific basis is exaggerated: Their core for this refute is that climate scientists are tweaking data and knowledge to make people believe that planet earth is facing climate challenges that threaten humanity's existence.

- Example tweet: *"Sign the petition: Demand the Koch network stop funding junk science to spread doubt about climate change. Take action here: https://t.co/McwwOtejhl"*

Claims advocates for climate policy are acting in bad faith: Primarily asserts that they perpertuate acting on things they advocate against.

- Example tweet: *"Then it's best not to question why the climate change activists Obamas just bought a multi-million dollar house on the east coast coastline. Seems irresponsible since it's going to be underwater soon. Keeping your head buried in the sand is probably best. Be well."*

Claims climate policy / climate policy advocacy is part of a conspiracy: Primarily assert that climate science is regressive and intends to reverse planet earth back to stone age.

- Example tweet: *"Now the little shit will write a book. Copies will by bought with American dollars by the lemmings called Democrats and this little shit will get rich and buy everything that causes the bogus climate change. Go to China Greta and they will lock you up and throw away the key."*

Claims local weather conditions disprove climate change: Hold a stance that weather modification is confused with climate change and that human effort to combat climate change makes that modification even worse.

- Example tweet:*"The seasons have moved back a month in the last few years imo. The joys of climate change"*

- Neutral Stance

  Against climate policy without expressing outright skepticism

  - Example tweet: *"@EvanFor2020 @LeoCTweets @ninaturner Maybe the 45,000 Americans who die every year because they can't afford basic healthcare have something to do with it. But by your logic, no progressive policy is important unless it's climate change."*

  Procedural tweets, news tweets / information without positional content (reporting climate impacts as climate impacts indicates positional content):

  - Example tweet:*"Few industries, if any, feel the direct impacts of climate change more than agriculture. Find out which 5 crop regions are most impacted via the @GEOSYS*

2019 Climate Change and Crop Report: https://t.co/UUO9MdCF9J worldagritech climate cropreport"

- Accepts climate science

  Warns of / highlights the impacts of climate change: Primarily backed by climate science knowledge either explicitly or implicitly.

  - Example tweet:*"@ericlo @AOC @People4Bernie Climate change has always been man made. Just ask the Neanderthals who melted the glaciers with their carbon bleeding SUVS"*

  Supports climate policy: Happens in scenarios such as; they understand climate policy, they retweet or draw information from an authoritative user who is well versed with climate policy or have mere knowledge but approves any climate policy there is.

  - Example tweet: *"Anyone else with a more than passing interest in climate change policy amp; politics - which *should* be every human on the planet - follow @UNFCCC coverage of COP25 this morning. ClimateEmergency ClimateCrisis ClimateChange"*

  Criticises skeptics: Based on either improvements in local climatic conditions or their knowledge on climate change, they refute propositions climate change deniers put forth.

  - Example tweet:*"I cannot believe Rochester, NY is supposed to get a foot of snow on November 11th fuck climate change deniers"*

- Supports Climate Policy

  Shows support for climate policy in general or a specific climate policy: They are either conversant with climate policies or possess mere knowledge on climate policy but they are in strong support of such policies.

  - Example tweet: *"An iceberg elephant?! We hope not. But seriously, we too were hoping for more details on climate change policy. This is going to be a climate election so @OxfamAustralia will keep on pushing both sides for higher ambition and a solid plan to get there budgetreply2019 auspol"*

- Neutral Stance Describes policy without taking a stance: Generally interested in sharing climate policy to create awareness.

  - Example tweet: *"We hope you're as excited as we are for the 6th mydtrainingseries! Here's a summary of one of our speakers, Ms. Marhaini, Principal Assistant Secretary of the Climate Change Policy Division of MESTECC. Our event is free amp; open to public. Link here: https://t.co/QefuGcKHdt https://t.co/SHCbITjj4V"*

- Against climate policy

  Argues against climate policy: Grounded with or without climate policy denial feedback mechanisms, they challenge climate policy in part or in whole.

  - Example tweet:*"@GlobalTV i get it. Climate change caused the very poor Forest management practices of California Oregon and Washington. Is there any chance you guys can present balance reporting some day?"*

3.4 Annotation Challenges

Having the annotation process as a progressive undertaking led to initial annotation issues. As annotation rules were being developed, annotation of approximately first six hundred tweets was challenging. Partly because these tweets functioned as ones to review in order to understand the nature of tweets and develop annotation rules. As a result, some of tweets that would fall on the neutral stance due to their nature of information sharing were annotated to either in favor or against classes. To avoid confusing the classifier, these tweets were omitted from the corpus that was subjected to the classifier. This led to a thousand three hundred out of two thousand annotated tweets being subjected to the classifier.

There were tweets that did not directly talk about climate change, many of them just mentioning climate change. For ones that took a stance in favor of climate change – even though they were not targeted towards it – were regarded as taking an "in-favor" position. At the same time, they were labeled as being irrelevant to the target – climate science and/or policy. Conversely, ones that were against several other topics and mentioning climate change as a subsequent topic were regarded as taking an "against" position. They too were classified as being irrelevant to the target. Furthermore, Ones that did not seem to take any position were classified as neutral.

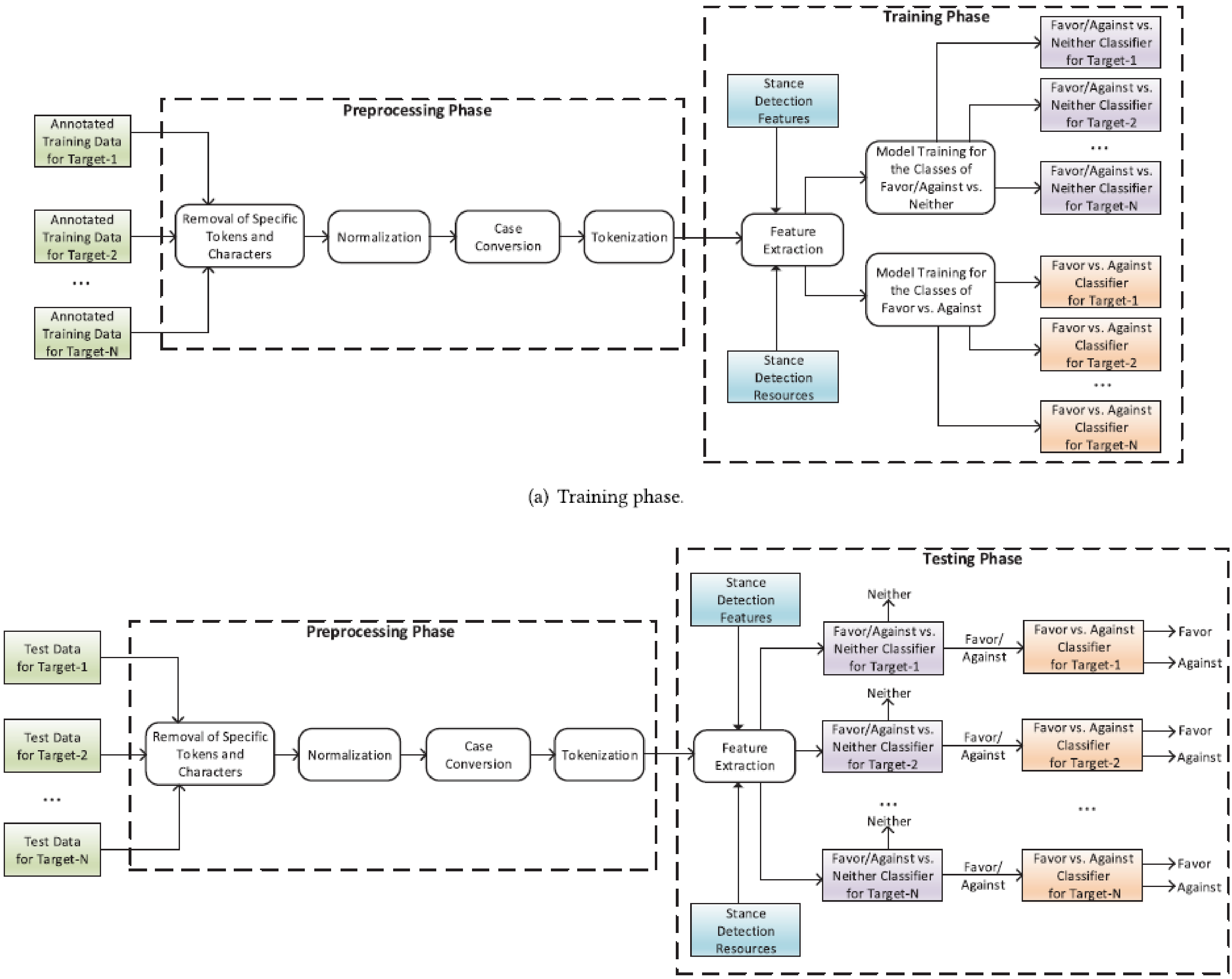
Another challenge was tweets that only mentioned other user accounts, included only links without any words accompanying them and ones that just said "Climate Change". In their nature, these tweets do not say anything about their stance towards climate change. They were therefore classified as neutral. While ones that only had mentions and links were classified as irrelevant.

4 Methodology

Summary

4.1 Approach

To proceed with this research, tweet data will be annotated into three stance classes: In Favor, Against and Neutral forming three different classes. Having three classes makes this a multiclass problem. Through the use of appropriate regular expressions the corpus will undergo pre-processing phase that includes removal of specific tokens and characters, normalization, case conversion and tokenization. With the use of Scikit Learn's multiclass classfier, the processed tweets will then be subjected to training and testing phases as summarized in the following figure.



(a) Training phase.

(b) Testing phase.

Figure 4.1: The Architecture of a Generic Stance Detection System

The two phases of this pipelined classification scheme are geared towards two objectives. The first phase determines relevancy when inputs are classified as having a stance: (Favor or Against class) or not(Neither class). In the second phase, inputs that were classified as having a stance are further classified as favor or against towards the stance target.

The explicit stance target will be "Climate Science". This means tweets explicitly mentioning climate science will be analyzed to see if they are taking a stance in favor, against or neutral.

A tweet such as ClimateAction superhero" implicitly takes a stance that is in favor of climate science. A tweet going by "Wetlands are a secret power against climate change" explicitly takes a stance in favor of the target - Climate science. An example of tweet against would be "It is our internal issue, don't speak GretaGoToSchool". This is implicitly takes a stance against climate science.

4.2 Baseline Model

Functioning as a baseline, Support Vector Machines (SVMs) are used as a type of supervised learning classification algorithm. The main reason for using them is due to their reputation in attaining brilliant results. The advantages of SVMs include; their effectiveness in high dimensional spaces, where the number of dimensions is greater than the number of samples, they are memory efficient because they use a subset of training points in the decision function. They are also versatile in the sense that different common kernels can be specified in the decision function but also allows for specifying custom kernels. SVM's downsides include; the risk of over-fitting if the number of features is greater than the number of samples. They also do not directly provide probability estimates - they have to be calculated using cross validation and probabilities.

It is tantamount to pre-process tweets through removing unwanted features so that they come of the form ready for performing analysis. The removal of unwanted features leads to clean textual data that can easily be transformed to numeric arrays ready to be subjected to machine learning techniques. Performing this we hone an opportunity for detecting stance towards the target. Corollary text preprocessing offers the opportunity for the model's optimality to be tested.

The following is an exemplar of a raw tweet extracted from the corpus of tweets under study.

"*Recourse, Greenpeace, Earthlife Africa and Centre for Financial Accountability unvail contradictions in @IMFNews policy advice and climate change-related macroeconomic risks including subsidies and other investment incentives for fossil fuels and coal: https://t.co/WL9ujibrvO*"

Pre-processing of these tweets was then initialized with removing all unnecessary white spaces in tweet sentences, special characters, removing mentions, substituting multiple spaces with a single systematic space, removing prefixes, removing single characters from the start of tweets, hashtags, punctuations, numbers (e.g.,@, *). All single characters that would jeopardize the transformation of features to arrays were removed. This also included ones that were found at the beginning of tweets. There being more than one white space between words, they were cleared to remain with a single standard space. To have features much clearer, all prefixes were removed. This led to having clean tweets as stipulated in the clean version of the aforementioned tweet.

"*recourse greenpeace earthlife africa and centre for financial accountability unvail contradictions in policy advice and climate change related macroeconomic risks including subsidies and other investment incentives for fossil fuels and coal*"

To generate features, the most popular method – TF-IDF (Term Frequency – Inverse Document) was used. TF-IDF builds a vocabulary of words that assign a unique integer to each of the words reflecting how common or rare a word is in the entire document set. The following formulas manifests what is going on under the hood of TF-IDF where t represents the term, d the document and D a corpus.

TF(t) = tf(t,d)

IDF(t) = $\log(|D| \div (1 + |d : t \sum d|))$

tf - IDF(t) = TF * IDF

From these formulas, equation one represents the term frequency of word t, equation 2 represents the inverse document frequency and equation three the TF-IDF score calculation of term for term t.

The following parameters were specified in order to have the TF-IDF pre-processor working efficiently. Words with frequencies higher than 90 percent from a cluster of pre-processed data were ignored. In the same respect, records with a lower frequency than 2 were ignored. Lower and upper boundaries of the range of n-values for different word grams were unigrams and bigrams. This is because having both of them offered a high enough accuracy, a perfect ROC AUC score, and a precise standard error from our model's cross-validation test. We also specified the strength of our regularization parameter "C" to be 1.

Both Gaussian and Linear kernels were specified to take care of irregularities that may have arisen from the dataset as the machine learning model runs. With a Gaussian kernel, a kernel coefficient "gamma" of zero was specified. Class weight was set to "balanced" to automatically adjust weights/probabilities inversely proportional to class frequencies. A linear kernel needed not to be specified further since it regards the dataset as a regular one with no imbalances.

Since this is purely a multi-class problem, a pipeline that applies a TF-IDF vectorizer taking into account the specified parameters was built. The pipeline deploys a OneVsRest classier of SVC that takes into account probabilities/weights. With these in check, the processed dataset was separated to training and test sets where we have 80 percent of the data as a training set while 20 percent of it as a test set. A Grid Search is then used to do an exhaustive search over parameter values where all processors are used. To be concise, the verbosity of messages is controlled by only allowing computation time and parameter candidates for each fold to be displayed. The use of Grid search is deemed paramount since it picks the best performing kernel out of the Gaussian and linear. The output it produces is one for the best-performing kernel. In this regard, there was no need to worry about being explicit of a single kernel among the two. A classifier was then built by fitting training sets from which test sets were able to be predicted.

4.3 Preliminary Results

It is quite striking to learn how challenging this particular classification problem is from the preliminary results of the SVM classifier. 67 percent of tweets that were classified as in favor were correctly predicted while 9 percent and 24 percent were predicted as against and neutral respectively. Particularly striking is that 72 percent of tweets that were classified as against were predicted as in favor while 17 percent were correctly predicted and 11 percent were predicted as neutral. 41 percent of tweets that were classified as neutral were predicted as in favor, 4 percent predicted as against and 55 percent were correctly predicted.

To understand why this is the case, misclassified tweets falling off confusion matrix diagonal were retrieved. Particular characteristics of these tweets are manifested with words they contain. Most of these words are used by deniers of climate science, for example, hoax as shown in the following example.

"*climate change can believe that people really do call that hoax did they not go to school*"

Even though the tweet contains the word hoax, it takes a stance in favor of climate change. It was also noticed that some tweets contained strong words such as pathetic and horror that are imaginable to be used in a climate science denier sense.

"*this is horror on very grand scale anyone who says there is no climate change must be completely mad*"

On top of words, there are tweets that contain phrases that negate climate science, even though they are in favor of climate science. As it will be seen in the following example, the tweet contains a phrase "don't believe in climate change" even though it takes a stance in favor of climate science.

"*even if you for whatever reason don believe in climate change it the wildest thing to me that people can watch this girl and somehow get outraged anyone angry out there willing to walk me through this what gives*"

"*you re the one that that linked patrick moore m just debunking him again those names you keep throwing at me moore bastrati happer lindzen are the of people or scientist that don believe in climate change believe the majority not the minority*"

The same situation applies to tweets that have been classified as against but predicted as in favor. Tweets with climate progressive words such as "greenpeace", "environmental protection", "global cooperation", "sustainable" and "left" have been predicted as in favor even though their stances were against.

"*greenpeace is fraud climate change is not dangerous at ppm not even at*"

The main reason explaining this is the downside of SVMs assigning fixed weights to tokens. In this regard, the model ends up categorizing tweets with the same words to the same category even though they might be taking opposing stances.

To have a robust classifier, there evolved need to specify a range of parameters from which the grid search functionality will pick the best specifications to feed the classifier. A trial and error approach was used to determine the best combination of parameters that will lead to better results. This entailed requiring the pipeline to ignore terms with frequency higher than the range between 0.5 and 0.7, ignoring lower frequencies than three and five documents. A range of unigrams and bigrams, gamma ranging from negative one to positive one and a regularization parameter of one and two were specified.

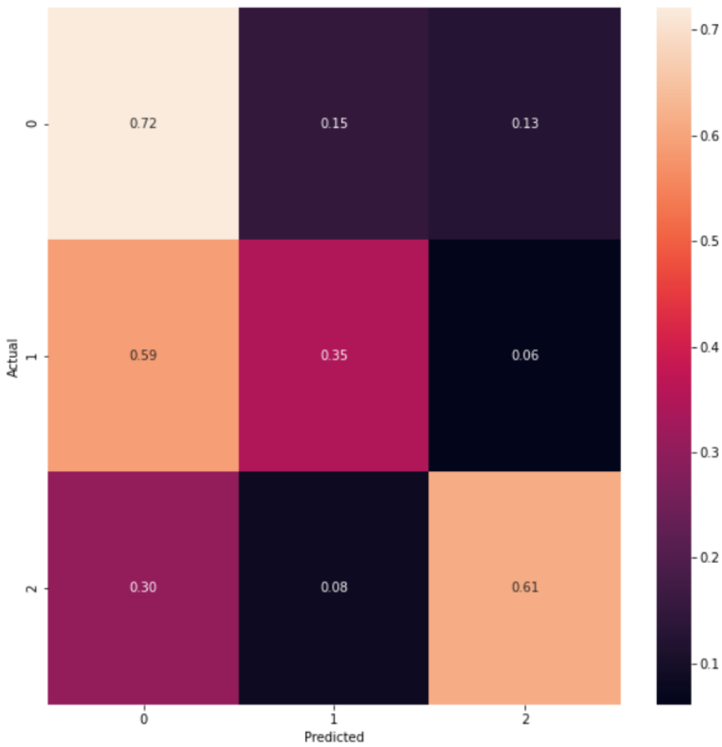


Figure 4.2: Normalized Distribution of Tweets according to Stances

From the confusion matrix, 72 percent of tweets that were classified as in favor were correctly predicted while 15 percent and 13 percent were predicted as against and neutral respectively. Even intriguing are the predictions for tweets that were classified as against. Only 35 percent of tweets that were classified as against have been correctly predicted while a bothersome 59 percent and 6 percent were classified as in favor and neutral respectively. A curious question is why is that the case? This is primarily driven by a few annotated data points used to predict stances. In comparison to excellently performing experimentation results, 1300 annotated data points were not enough in light of the best parameter specification. Moreover, 61 percent of tweets that were classified as neutral were correctly predicted while 30 percent and 6 percent were predicted as in favor and against respectively.

With these results, the model was then applied to predict stances of the fifty thousand unclassified tweets. The results of which are presented on the Findings section.

4.4 Evaluation of the Baseline Model

Getting a clear idea of how the model works is assuring. Two methods for evaluating the model were used. These are calculating the ROC AUC score and using cross-validation evaluation score. Unlike precision, recall, and F1 scores, they do not form part of the training and testing process. This is because if they were, they would return a perfect evaluation score suggesting that the model, both optimal and sub-optimal, does a good job. In the event of a sub-optimal model, this would lead to model over-fitting. Thus, it was essential to have evaluation techniques that did not form part of the training and testing process.

With the ROC AUC score, the area under the receiver operating characteristics is being computed from prediction scores. With its ease of interpretation, it served as the first evaluation technique. Having a score above 0.5 means that the model performs suitably well. Regardless of model sub optimality, it seems to be performing well with this evaluation technique. The ROC AUC score obtained is 0.77 which suggests the model performs well.

Cross-validation was used to determine the level of model precision. By running cross-validation, scores for standard deviation can be obtained—the lower the standard deviation, the more precise the model. The model turned out to be exact enough to form a baseline since standard deviation score obtained was 0.03278, which is relatively low, suggesting that the model's precision is quite satisfactory. With such a precise standard deviation, the model can apply to an even larger dataset.

5 Experimentation

Summary

With satisfactory evaluation results, it deemed necessary to test the model on a separate dataset. A SemEval dataset by Davidson et al (2017) on hate speech comprising 24,802 tweets was paramount. Its tweets are categorized as hateful, offensive or neither. This dataset was already manually annotated by CrowdFlower workers. The inter-annotator agreement of three annotators provided by CrowdFlower was 92 percent. Labels to each tweet were assigned based on majority decision that was easily achieved provided a high inter-annotator agreement. In light of automatic detection of hate speech tweets that were labeled offensive were dropped.

With this dataset, a precision, recall and accuracy of 93 percent were achieved. The classifier also appeared to do very well. 77 percent of tweets that were classified as hate speech were correctly predicted while 23 percent were predicted as neutral. On the other hand, 98 percent of tweets that were classified as neither were correctly predicted while only two percent were predicted as hate speech.

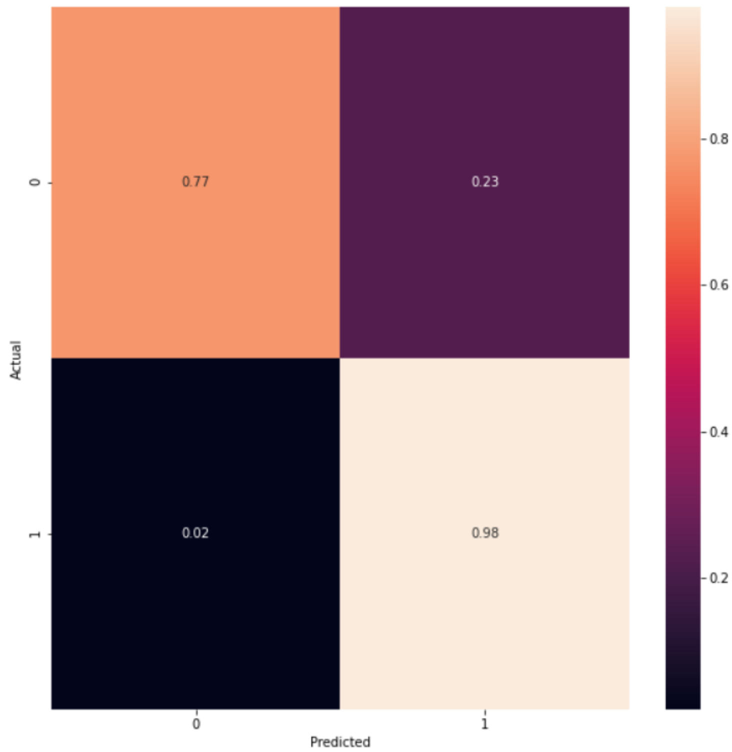


Figure 5.1: A Visual Representation of BERT Model Architecture

What explains the miniscule misclassification in these tweets is in part the aforementioned weakness of SVMs and in part tweets that are hard for humans to annotate. Equally likely, the evaluation scores for this model indicate that it is doing well with ROC_AUC score 0.877 and a cross-validation standard deviation of 0.001687.

Another dataset subject to experimentation is the Kaggle dataset named "Twitter Climate Change Sentiment Dataset". Updated one year ago, this version 1 dataset comprises of 43943 tweets pertaining to climate change collected between April 27, 2015 and February 21, 2018. Its collection was funded by Canada Foundation for Innovation. These tweets have been annotated and the annotation has been agreed upon by three reviewers. Similar to this study's labelling, four labels are subject for this dataset; News – if the tweet links to factual news about climate change, Pro – if the tweet supports the belief of man-made climate change, Neutral – if the tweet neither supports nor refutes the belief of man-made climate change. Anti –if the tweet does not believe in man-made climate change.

Using a progressive approach, this study's model records improvements in performance as batches of the dataset keep increasing. Starting with a random selection of 2,000 data points from the dataset, an accuracy of 64 percent is recorded. Improvements in accuracy are observed to increase to 65, 67 and 70 percent when data points are increased to 5000, 6000 and 10000 respectively.

Table 5.1: Progressive Evolution in Accuracy Scores

Iterations	Samples	Accuracy
Iteration 1	2000	0.64
Iteration 2	5000	0.65
Iteration 3	6000	0.67
Iteration 4	10000	0.70

All this suggests that, the model is deemed to do well when subjected to an even larger dataset. Out of a stock of 10 million tweets, 50000 were randomly selected. After taking them through the pre-processing phase, they were subjected to the classifier for prediction into three of the established classes.

6 Deep Learning

Summary

chapter.1chapter.2section.2.1section.2.2chapter.3section.3.1section.3.2section.3.3se

To assess the accuracy of predictions from the SVM model, a further step is taken to develop a deep learning model named Bidirectional Encoder Representations from Transformers (BERT). This model takes care of word contexts and thus avoid generalizations in classifications. Through computing the probability of a word based on a combination of words it has seen previously, BERT learns to predict the probability of a sequence of words. A Hugging face flavor to the model is applied with the use of Pytorch as a language.

BERT is a Transformer-based machine learning technique for natural language processing developed by Google in 2019. In turn, Transformers are encoder-decoder architectures developed in 2017. The encoder consists of a set of encoding layers that processes the input iteratively one layer after another and the decoder consists of a set of decoding layers that does the same thing to the output of the encoder. The introduction of Transformers, and specifically BERT, has led to the replacement of previous models such as LSTM-based Recurrent Neuro Networks.

As opposed to the majority of the architectures that have preceded BERT, this algorithm does not make use of directional models, which read the text input sequentially, either left-to-right or right-to-left. BERT encoders read the entire sequence of words at once, making it bi-directional—simultaneously left-to-right and right-to-left. This particular construction feature allows the BERT model to learn the context of a word based on all of its surroundings, regardless of position in relation to the word. The general representation of the Transformer architecture can be seen in the figure below:

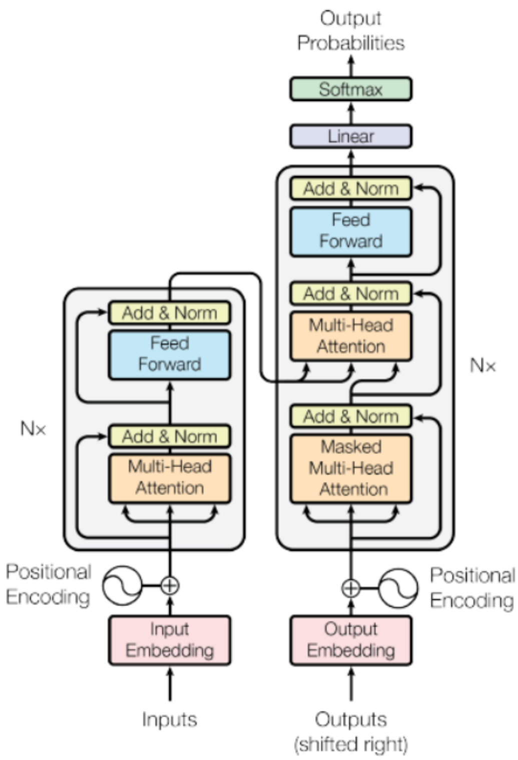


Figure 6.1: A Visual Representation of BERT Model Architecture

In general, a BERT base model comprises a 12 transformer blocks encoder, 12 self-attention heads, and a hidden size of 768. It takes a sequential input of no more than 512 tokens and outputs the sequence representation. One or two segments are attached to the sequence where the first token is comprised with a special classification embedding (CLS) and another one for separating segments (SEP). It is bidirectional in the sense that it considers left and right contexts when learning representations of texts.

In text classification, BERT utilizes the final hidden state h of the CLS token to represent the whole sequence as seen in the following equation.

p($c \leqslant h$) = softmax(Wh)

To predict the probability of label c, a softmax classifier gets added to the top of BERT. W is the matrix for task-specific parameter. All parameters from BERT as well as W are jointly fine-tuned by maximizing the log probability of the correct label.

As a start, a configuration for CUDA execution was done to prepare a computing device for leveraging onboard GPU. In particular, this served to accelerate the deep learning process. A dataset and a data loader functioning as variables to be used during fine tuning and training phase of the model were then prepared.

This entailed creating a dataset class and a data loader itself. The dataset class defines how text gets preprocessed before being sent to the neuro network. The dataset class was defined to accept a dataframe as an input and generate tokenized output from preprocessed tweets that can be used with DistilBERT Model for training. During pre-training, DistilBERT reduces the size of BERT by 40 percent while retaining 99 percent of its capabilities. It also makes inference 60 percent faster. An encode plus method of the tokenizer is used to perform tokenization that generates output in the form of ids and attention mask. It also incorporates a column of the dataset that comprises the three labels assigned to the tweets, namely, in-favor, against and neutral that were coded as 1,2 and 3 respectively. It goes on further to split the dataset into two, one for training comprising 80 percent and another one for validation that serves for evaluating the model's performance.

Dataloader's definition serves to creating training and validation aspects that load data to the neuronetwork in a structured manner. This serves to control the amount of data loaded to the neuronetwork since data cannot be loaded at once from the dataset to the memory. To achieve this structure, parameters such as "batch size" and "max len" are used.

To enhance the fine-tuning process, a neuro network and a loss function optimizer are created. Through a DistillBERTClass, the creation of a neuro network is done. It encompases a DistilBERT Language model with a dropout and a linear layer that will feed the dataset. Final output layers from the feeding will be compared to encoded categories to determine the accuracy of the model's prediction. A further instance of the network refered to as model was then created to be used for training and saving the model for further inference.

A Loss function and an optimizer are further created. Whereas the loss function is used to calculate the difference between the output created by the model and the actual output, an optimizer is used to update weights of the neuro network to improve its performance.

An actual fine-tuning happens with the creation of a training function. In this function; a data loader passes data to the model based specified size of batches – a specified batch is 2, model outputs are and actual categories are compared to calculate the losses that are used to optimize the weights of the neuros in the neuro network and get printed out after every 5000 steps.

To test how good the model performs, the validation stage deems paramount. At this stage, the unseen data or the testing dataset gets passed to the model to check how well the model performs on unseen data. At this stage, weights of the model are not updated. It's only the final output that is compared to the actual value from which the accuracy of the model is calculated.

For experimentation, this model was subjected to the dataset on hate speech that that was also applied to experiment the baseline model. The following are its satisfactory results.

Table 6.1: Results from the Deep Learning Model.

Metrics	Training	Validation
Loss	0.37	0.28
Accuracy	86.6	90.80

With confidence in the model, a data set containing predicted labels was subjected to the model. The main aim was to test how good the baseline model works and how well the data set will blend to the model. Contrary to specifications on the baseline model, a number of steps the classifier needs to look at the data set before outputting metrics was reduced from 5000 to 30 steps. This was because with 5000 steps in a epoch, the model seemed to run indefinitely without producing metrics. In part, this can be explained with a slightly higher number of data points compared to the approximately 44000 data points in the experimentation stage. The training stage produces satisfactory metrics. Training loss has significantly declined from 1.125 to 0.69 while training accuracy has increased from 25 percent to 71.5 percent. This serves to indicate that the predictions made from the baseline have a satisfactory level of accuracy and can be relied upon. However, metrics from the validation stage trigger a curiosity towards the data. Converse to the training phase, validation loss has seen a constant increase. Equally, validation accuracy has gone overboard. In part, this misbehaviour can be explained with the aforementioned contextual words in tweets that can mislead a classifier.

In interest of further research, there seems to be a need to pay a closer look at the data set. For example, some broken tweets were noted. Much of this break has been contributed by characters going by "amp". As a result, tweets containing with this character did not have a smooth connection in context.

7 Results and Policy Implications

Summary

7.1 Findings

Descriptive statistics of predicted stances shows an encouraging trend with a distribution that is favorable towards climate change mitigation efforts. 26,516 tweets were predicted as being in favor of climate science while 8,225 and 15,259 were predicted as against and neutral respectively. Interestingly, a parallel evolution is observed in the monthly trend of climate science stances between December 2018 and May 2021.

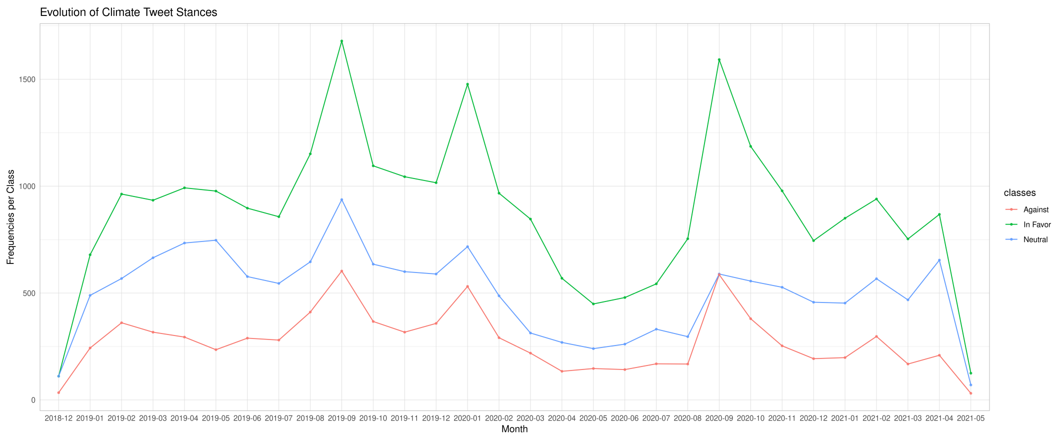


Figure 7.1: Timely Evolution of Climate Science Stances

As stances in favor of climate science fluctuate, ones against and neutral follow suit. However, a very weak positive influence of neutral stances to climate science deniers is manifested, particularly from January to May 2019 and September to November 2020. It occurs that the tendency of users sharing information on climate science, with time, nudges deniers towards supporting mitigation efforts. A comparatively weak spike on against stances to neutral stances during the month of March 2021 together with the aforementioned are symbolic.

Three stark spikes in all stances comes to being in the months of September 2019, January 2020 and September 2020. These are due to spikes in tweet activity during climate events that happened during these times. September 2019 was the month when global climate strikes happened prior and after the United Nations Climate Summit. Dubbed as the Global Week for Future, over 4 million people gathered to protest in 4,500 locations across 150 countries. Shortly after the strikes, Australian bush fires that happened in December 2019 triggered much tweet activity where people demanded immediate climate action to be taken. Equally, devastating floods happened in Indonesia during the same month. Consequently, September 2020 was reported to be the warmest month on record worldwide. Such a rise in temperature is due to a good number of climate disasters that have happened.

As it stands, not only users sharing updated factual climate science information that nudge deniers to supporters but also critical events in the climate change timeline. As aforementioned, much of the weak nudge are observed in months towards the Global Week for Future and after the warmest month on record. In themselves, these are an indication that some of deniers haven't been engaging intellectually with climate science information. This factual information has dawned on them through assimilation on a practical basis.

With such a broader picture, it is worthwhile to understand what drives users to take their stances – particularly "in favor" and against.

A strong trust in institutions is established as a unifying driver for tweeter users tweeting in favor of climate science. This trust is observed to materialize in many ways. First is when some tweeters appreciate and acknowledge pro-climate science users for nudging them towards green behavior. In itself, a feeling of zenith that some mindsets have been transformed for the better of climate change mitigation efforts. Second, such users have shown much concern on the future impacts of climate change in both environmental and financial dimensions. Much of this concern is drawn from information shared by climate change authoritative institutions such as the IPCC. Third, most users manifest the recognition that current changes in climate have not been witnessed since the beginning of time. In synergy, they demand immediate action to be taken in all angles of implementation from policy to activism to curb a potential extinction, should nothing be done. A following tweet offers a snapshot;

"but even when we push our perspective to the earliest days of the roman empire we cannot discern any event that is remotely equivalent either in degree or extent to the warming over the last few decades"

Some users have tweeted to draw inspiration from the young generation that is concerned about climate change. Such an inspiration comes from discussions they have during daily activities such as conversations on a dinner table. Children draw this concern from what they learn in school – information that is driven by factual findings from climate change authoritative institutions.

"love it how my yr s are so passionate about climate change that we had off piece discussion about how we should not be travelling to mars and should stay and clean up our planet first feel like am winning as teacher today"

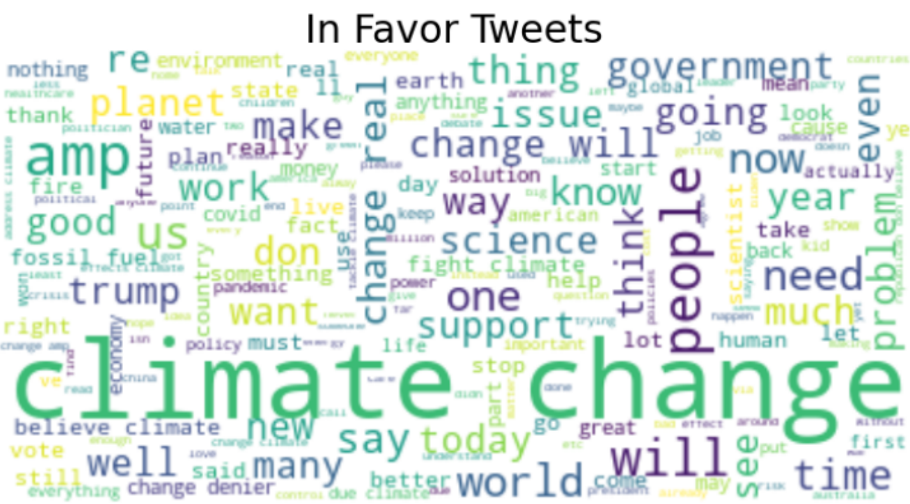


Figure 7.2: Drivers for Climate Science

As it can be visualized, words such as want, need and now are pronounced to indicate the need for immediate action in order to chart a foreseeable future. In tandem, exemplar words such as government, real, science and support serve to indicate a strong trust in institutions beheld by supporters. Unlike previous studies, this trust in institutions cuts across continents since it is observed from users all over the world. This is mainly because this research's data set comprises all English speaking countries and this research does not seek to study geographic differences.

Two stark drivers are observed to contribute towards users taking a stance against climate science – practicality and bipartisan division.

Practicality arises through rigidity and pragmatism. Rigidity fundamentally lies under the assumption that climate changes have been there since the beginning of time. In this regard, all changes to the climate we are witnessing are natural – a phenomena that calls for no mitigating action to be taken. Corollary, some users even go extreme to not caring at all about efforts to mitigate climate change. Pragmatism manifests when users depict inability to connect climate change theory and information to real world events. This disconnect happens in various forms but the following are visible;deniers blame supporters for being too reliant on authoritative information from institutions such as the IPCC without challenging such evidence. Other deniers believe that climate change statistics are tweaked to mislead the public. Finally, some claim that climate change statistics go overboard and they are nothing relatable to past events like the second world war.

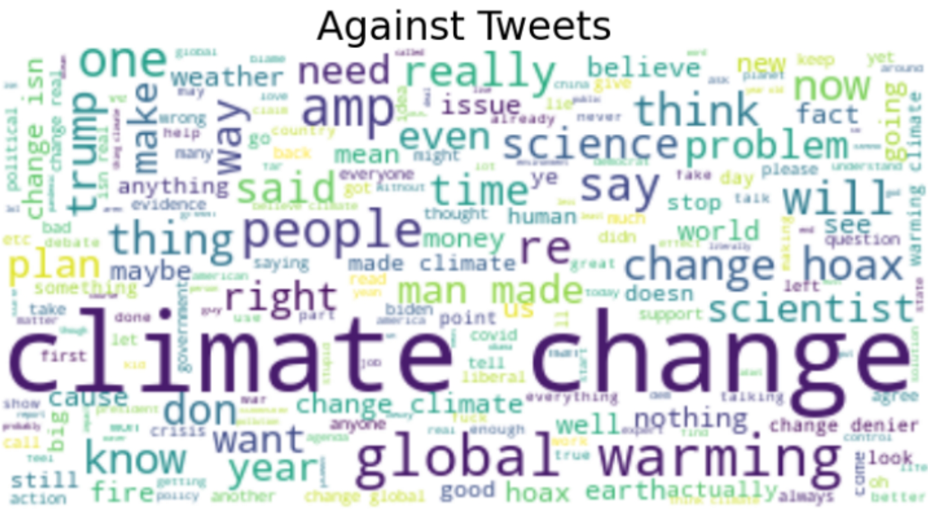


Figure 7.3: Drivers for Climate Science

A bipartisan line between left and right deemed a tangible determinant. Most of right wing users took a stance against climate science – and challenged left wing users. The same discussion is observed on the other side of the pendulum. Much of this division is attributed to political ideologies left and right parties hold towards climate change. Tantamount to (Jang 2015), the stance taken by right wing users is attributed to information they receive from traditional media. Right wing media such as Fox News tends to offer information that nudges its users against climate science. The opposite applies to left wing traditional media. This contradictory exchange in information breeds much of this division and fosters fierce debates in many ways. This is also visible from the word cloud through pronounced words such as hoax, trump and global warming that deniers tend to use mostly. This division is also observed through users referring political leaders. Trump is visibly pronounced compared to Biden. This indicates that deniers draw much of their climate stance from Donald Trump's authority – a climate change denier – compared to Joe Biden. Contrary to previous studies, much of a shift towards the use of the word Climate Change in place of Global Warming is observed in deniers. This indicates a sign of progress in nudging deniers towards taking a stance in favor of climate science.

7.2 Policy Implications

In light of the practical nature of deniers, climate policy can offer incentive to activists in getting more creative on delivering climate science facts to the general public. Aforementioned, some deniers have been nudged towards being supportive of climate science due to major events in the climate timeline. Creativity in delivering facts is likely to change perceptions of deniers that are being dissuaded from supporting mitigation efforts by denier users with tweet–authority. In principle, their foundations for being swayed to denial do not hold tangible theoretical arguments to battle with climate science facts and a growing influence from decorated activists. Creative activists can then penetrate climate facts through mainstream, social, physical and institutional media to have a wider reach. For deniers encountering information that brings climate facts to life be it on the road, while listening or on screen insidiously nudges their subconscious in favor of mitigating efforts. Creativity can also amplify the current visible strides in educating the younger generation on climate change to clustering a significant pool of climate science supporters in the long-run. Having a child igniting green conversations during family hangouts not only enhances a green lifestyle but also a green mindset to parents regardless of their stances beforehand.

To enhance a fair bipartisan debate, there is need to regulate output from mainstream media and think tanks that goes overboard in contradicting established credible facts on climate science. Regulating these two axis of information delivery is critical since they are primary sources of information before it is fed to social, physical and institutional media. Much of what is shared on secondary media is derived from primary media output. Setting stage for such a fair exchange in climate information is likely to reduce the extent of users being misled by authoritative climate deniers, be it politicians or normal users. This leaves room for a constructive debate on climate science. Much of what the Intergovernmental Panel on Climate Change (IPCC) is doing tallies to recommendation but it could go further to monitoring mainstream media.

7.3 Conclusion

This research has given insight into drivers for users taking stances in favor and against climate science. It has particularly shone light on what exactly drive deniers into taking such stances. While the baseline model has shown that it is possible to detect stances of users from unclassified tweets, the deep learning model has reinforced how good the baseline model performs. Given a few data points due to sub-optimal metric results prior predictions, the baseline model has shown how good it performs with the experimentations.

Although further development and expansion of the methodology are essential to improve its efficacy and reliability, the success of these models constructed from publicly available data illustrate that they can serve as a viable alternative in detecting stances users take on other cross-cutting issues.

7.4 Areas for Further Research

In its scope, this research focused on climate science tweets. But whilst annotating, tweets pertaining to climate policy were noted. Although they were not as massive as ones for climate science, they hold a promise into researching climate policy. A tailored improvement in methodology to gain insight in these tweets deems paramount. These can be through having more labels for climate policy or performing a comparison between outcomes for climate science and climate policy. Building on proposed recommendations, climate policy tweets are likely to yield significant insight into improving climate policy.

It will be interesting to look at the geographical dynamics in terms of the variation in institutional trust. Such a distribution will shine light into the context of how climate policy is implemented. Moreover, it will also highlight on the effectiveness of climate legislation in respective regions. Researching these will be telling points for improving public administration responsible for a just climate.

8 Bibliography

Summary

chapter.1chapter.2section.2.1section.2.2chapter.3section.3.1section.3.2section.3.3se

[1] Hywel.T.P Williams, James R. Mc Murray, Tim Kurz, and F.Hugo Lambert. *Network analysis reveals open forums and echo chambers in social media discussions of climate change*. Elsevier,2015

[2] S. Mo Jang and P. Sol Hart. *Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data*. Elsevier, 2015

[3] Emily. M. Cody, Andrew. J. Reagan, Lewis Mitchell, Peter Sheridan Dodds and Christopher M. Danforth. *Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll*. Plos One, 2015

[4] Warren Pearce, Kim Holmberg, Lina Hellsten and Brigitte Nerlich. *Climate Change on Twitter: Topics, Communities and Conversations about the 2013 IPCC Working Group 1 Report*. Plos One, 2014

[5] Kristin Demetrious. *Twitter and the Struggle to Transform the Object: A Study of Clean Coal in the 2017 Australian Energy Policy Public Debate*. Journal of Public Interest Communications, 2019

[6] Dilek Kucuk and Fazli Can *Stance Detection: A Survey*. ACM Comput. Surv. 53, 1, Article 12 (February 2020), 37 pages.

[7] Dilek Kucuk and Fazli Can *Stance Detection on Tweets: An SVM-based Approach*. arXiv preprint arXiv:1803.08910, 2018

[8] Aseel Addawood, Jodi Schneider and Masooda Bashir *Stance Classification of Twitter Debates: The Encryption Debate as A Use Case*. University of Illinois at Urbana Champaign, 2017

[9] Kazhuparambil, S and Kaushik, A *Cooking is all about people: Comment Classificatrion on Cookery Channels Using BERT and Classification Models* Preprints 2020, 2020060223 (doi: 10.20944/preprints202006.0223.v1

[10] Maslej-Kresnakova, V, Sarnovsky, M, Butka, P and Machova K *Comparison of Deep Learning Models and Various Text Pre-Processing Techniques for the Toxic Comments Classification* Technical University of Košice, 2020

[11] Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. *Deep Learning Based Text Classification: A Comprehensive Review. 1*, January 2020.

[12] Costanza Conforti, Jakob Berndt, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Toxvaerd and Nigel Collier *STANDER: An Expert-Annotated Dataset for News Stance Detection and Evidence Retrieval*, Association for Computational Linguistics, November 2020.

[13] Miha Grcar, Darko Cherepnalkoski, Igor Mozetic and Petra Kralj Novak *Stance and influence of Twitter users regarding the Brexit referendum*, Computational Social Networks, 2017

[14] Kuntal Dey, Ritvik Shrivastava and Saroj Kaushik *Twitter Stance Detection - A Subjectivity and Sentiment Polarity Inspired Two-Phase Approach* Indian Institute of Technology, 2017.

[15] Davidson et al *Automated Hate Speech Detection and the Problem of Offensive Language*

roceed- ings of ICWSM 2017 2017

[16] Robert. G. Boutilier and Kyle Bahr *A Natural Language Processing Approach to Social License Management* MDPI, 2020

[17] Giovanni Di Franco and Michele Santuro *Machine learning, artifcial neural networks and social research* Sapienza University of Rome, 2020

[18] Simon Munzert, Andrew. M. Guess, Pablo Barbera and JungHwan Yang *The consequences of online partisan media* National Academy of Sciences, 2021

[19] Jason.T.Carmichael, Robert. J.Brule, Joanna. K. Huckster *The great divide: understanding the role of media and other drivers of the partisan divide in public concern over climate change in the USA, 2001–2014* Springer Science and Business Media, 2017

9 Statement of Authorship

Summary

chapter.1chapter.2section.2.1section.2.2chapter.3section.3.1section.3

I hereby confirm and certify that this master thesis is my own work. All ideas and language of others are acknowledged in the text. All references and verbatim extracts are properly quoted and all other sources of information are specifically and clearly designated.

DATE	**NAME**
Berlin, 14.06.2021	Augustine Malija

I hereby confirm and certify that this master thesis is my own work. All ideas and language of others are acknowledged in the text. All references and verbatim extracts are properly quoted and all other sources of information are specifically and clearly designated.

DATE
Berlin, 14.06.2021

NAME
Augustine Malija