Artificial Intelligence - Bachelor Thesis

Classification of Conversations: Distinguishing between Opposing Climate Change Communities on Reddit

July 2, 2021

by Chieling Yueh
Student number: 6565255

Supervisor: Anna Wegmann
Second supervisor: Rianne van Lambalgen
Credits: 7.5 ECTS

Utrecht University
Faculty of Humanities
BSc Artificial Intelligence
c.l.yueh@students.uu.nl

Abstract

Gaining a full understanding on the discussion on climate change is of importance since it can help counteract climate skepticism and implement effective climate policies. This thesis investigates how two opposing Reddit communities, r/climate and r/climateskeptics, differ in the way they talk about climate change. We develop a logistic regression model, which uses textual features to classify which community a discussion belongs to. Our results show that conversations including words related to certainty (*obviously*, *undeniable*), second person pronouns (*you, yourself*), third person singular pronouns (*he*, *she*) and third person plural pronouns (*they*, *them*) are more likely to be part of the r/climateskeptics community, whereas conversations including first person plural pronouns (*we*, *us*), words related to gratitude (*thanks*, *appreciate*) and greeting (*hi*, *hello*) are more likely to to be part of the r/climate community. Our findings may improve climate change-related chatbots to better distinguish between individuals that are climate skeptics and those that are not. By doing so, the chatbot can have personalised answers depending on the values of the user.

Keywords: online communities, climate change discussions, climate skepticism

Contents

1 Introduction

The Intergovernmental Panel on Climate Change (IPCC) is an organisation of the United Nations that evaluates the risks of climate change by reviewing published scientific literature. In 2018, the IPCC published a special report "Global warming of 1.5°C" [33]. This report shows the impact of global warming at 1.5°C and 2°C, and how the global warming risks are much higher if we exceed 1.5°C. More importantly, the IPCC [33] made clear that immediate and drastic action is needed to limit global warming to 1.5°C to lessen the global warming effects.

There is a strong consensus on climate change among scientists that humans are the main cause of global warming [20, 50]. Based on 11944 abstracts of scientific papers between 1991 and 2011, it was found that 97% supports the scientific consensus on climate change [20]. More recently, it was reported that the scientific consensus on climate change in 2019 had even grown to 100% [50]. Despite the strong scientific consensus, the polarisation around the consensus of climate change in the public remains. In 2020 around 12% of inhabitants in the United States did not believe that climate change is happening and around 32% did not believe that climate change is caused by humans [66].

The polarisation around climate change can be amplified by *echo chambers* [64]. Echo chambers arise when participants only expose themselves with their own points of view [64, 27, 18]. Internet users tend to form communities that have the same views [43]. For example, van Eck et al. [64] found the effects of echo chambers in climate change blogs, in which users on those blogs only consumed content that was in line with their view. It has been suggested that online echo chambers should be exposed more to opposing viewpoints to counteract this polarisation [60, 64].

However, in a recent study, it was reported that exposing echo chambers with opposing views did not counteract the polarisation [45]. In this study, they investigated the impact of opposing views in an online community on Reddit, namely the subreddit r/climateskeptics. The r/climateskeptics users who encountered opposing views within r/climateskeptics showed more activity in this subreddit within the first hour than the users who did not encounter opposing views within r/climateskeptics [45]. The authors explained this by the so-called *identity defence*, which means that individuals tend to defend their views when they encounter a threat to their view [45]. The users that were exposed to opposing views defended their stance on climate change by posting content that reinforces their views again, which can eventually cause more polarisation around climate change [45].

Previous research on online climate change discourse focused on various social media platforms. On Twitter, it was found that individuals that do not believe in climate change were more hostile to supporters of the consensus of climate change than vice versa [63]. Further, Pearce et al. [46] found that Twitter users mostly interact with users that have similar views on climate change. On Facebook, it was found that deniers of climate change used misrepresentations of peer-reviewed scientific articles to reinforce their views [12]. Matthews [39] found that 27% of the users on the *Air Vent* blog were supporters of the consensus of climate change at first, but became skeptical about the consensus of climate change.

This thesis focuses on the platform Reddit, which is a popular social media platform [2]. Users on Reddit can share content by posting stories, links, images and videos, comment on these posts and up- or downvote comments in more than 100.000 communities [54]. The already mentioned study of Oswald and Bright [45] investigated how a Reddit community, r/climateskeptics, that rejects climate change, reacts to opposing viewpoints. We aim to build further on this research to grasp a better understanding on the discussions around climate change on Reddit, by investigating how two opposing communities differ in the way they talk. The communities, i.e. subreddits, that are used are: r/climate and r/climateskeptics. The first one is a community that supports the consensus on climate change, and the latter is skeptical about the consensus on climate change.

Individuals who are skeptical about the consensus on climate change are also known as the climate skeptics[1]. Climate skepticism can have negative consequences. To illustrate, climate skepticism can have large impacts on the beliefs and preferences of individuals in relation to climate change policies [1]. Eventually this can be harmful, when these beliefs and preferences push into the mainstream [45], since climate skepticism can delay climate change mitigation and adaptation policies [37]. Therefore, it is important to gain insights of the online climate change discourse and how these views arise, to counteract climate skepticism. This can benefit climate policy makers to implement effective climate policies [45] by informing them about the public opinions around climate change.

Furthermore, the results from this study are useful for Artificial Intelligence applications such as chatbots. Chatbots are programs that communicate with users and mimic human conversations [22]. A climate change-related chatbot has already been developed by Toniuc and Groza [62]. This chatbot, called Climebot, advocates for the scientific consensus on climate change [62]. We aim to investigate how conversations between two opposing communities differ from each other by developing a logistic regression model. The features used for this model can be relevant for a chatbot like Climebot, since it could support them to distinguish between users that are skeptical about climate change and those that are not. In this way, chatbots can have different answers depending on if it is having a conversation with a climate skeptic or not. Knowledge about the preferences or values of an user benefits Climebot since it stimulates the conversation flow and the engagement of the user [15].

In short, the goal of this study is to investigate how two opposing Reddit communities differ in the way they talk about climate change. We investigate this by comparing two opposing subreddits, namely r/climate and r/climateskeptics. We develop a logistic regression model using textual features to classify which community a discussion belongs to. We show that the r/climate community uses more first person plural pronouns (*we*, *us*), words related to gratitude (*thanks*, *appreciate*) and greeting (*hi*, *hello*). Whereas the r/climateskeptics community uses more second person pronouns (*you*, *yourself*), third person singular pronouns (*he*, *she*), third person plural pronouns (*they*, *them*) and words related to certainty (*obviously*, *undeniable*).

1.1 Research Question

The central question in this thesis is:

- **How do two opposing online communities on Reddit differ in the way they talk about climate change?**
  i.e. how do the subreddits r/climateskeptics and r/climate differ in the way they talk?

1.2 Hypotheses

In this section, we formulate hypotheses that are based on former studies. These hypotheses help us to create features for the logistic regression model and answer the research question.

An online community on climate skepticism can be seen as a *counter public space*. Counter public spaces are spaces where people come together and challenge the mainstream public [11]. It

[1] The term climate skepticism has been criticised for giving scientific skepticism a bad name [32]. The criticism on the term climate skepticism resulted into alternative labels (e.g climate deniers or climate contrarians [32, 44, 65, 7]) to describe individuals who challenge the consensus on climate change [32]. For simplicity, we use the term climate skeptics.

appears that climate change skeptics not only attack people with opposing views, but also attack each other in their own online counter public spaces [35]. Kaiser [35] stated that this could be because there is less of a common ground in climate skepticism. For example, there are climate skeptics who deny climate change altogether, but there are also climate skeptics who are only doubting some aspects of climate science [30]. Disagreements could lead to conflicts between individuals, which could lead to impoliteness. Therefore, we formulate the first hypothesis:

1. The subreddit r/climateskeptics uses less politeness cues in their conversations, than the subreddit r/climate that supports the consensus on climate change.

The use of negations (*no, not, never*) has been brought together in the context of *inhibition* [61]. Inhibition is defined as: "the process of restraining one's impulses or behavior, either consciously or unconsciously, due to factors such as lack of confidence, fear of consequences, or moral qualms." [6]. Psychological inhibition can be a potential response to the impact of climate change [25]. An example of inhibition as a response to the impact of climate change is that supporters of the consensus of climate change cannot properly express the impacts of climate change [25]. Therefore, we formulate the second hypothesis:

2. The subreddit r/climate, who supports the consensus on climate change, uses more negation cues in their conversations, than the subreddit r/climateskeptics that do not believe in climate science.

Individuals tend to draw attention to themselves when they are in emotional pain [61]. It has been found that such individuals use more first person singular pronouns (*I, me, mine*) [61]. For example, Rude et al. [55] found that depressed individuals used more first person singular pronouns than other individuals. Since the r/climate community is a community that supports the consensus on climate change, it might be that the r/climate community experiences emotional pain due to the climate change-related losses, such as the loss of species and ecosystems [21]. So, it might be interesting to investigate whether the r/climate community uses more first person singular pronouns, than the r/climateskeptics community. Hence, the third hypothesis:

3. The subreddit r/climate that supports the consensus on climate change uses more first person singular pronouns cues in their conversations than the subreddit r/climateskeptics that do not believe in climate science.

Accordingly, the emotional responses such as sadness and anger can also arise due to climate-related losses [21]. Therefore, our fourth and fifth hypothesis:

4. The online community, r/climate, who supports the consensus on climate change uses more sadness cues in their conversations than the online community, r/climateskeptics, that do not believe in climate science.

5. The online community, r/climate, who supports the consensus on climate change uses more anger cues in their conversations than the online community, r/climateskeptics, that do not believe in climate science.

Pennebaker et al. [47] found that responding defensively in conversations could correlate to an increased anxiety. As it might be that climate skeptics do not have a clear set of common beliefs [35], it could be that the subreddit r/climateskeptics respond more defensive in their conversations. Thus, we formulate the sixth hypothesis:

6. The subreddit r/climateskeptics uses more anxiety cues in their conversations than the opposing subreddit r/climate.

1.3 Related Work

We aim to investigate how two opposing online communities on Reddit talk about climate change. This section will focus further on the related literature.

Online conversations. Social media platforms like Twitter, Reddit and Facebook are internet forums that are an important part of the World Wide Web [8]. Conversations on such platform have become important in our daily life. Former research found that people on Twitter have conversations to develop a common ground and feel connected [70]. Choi et al. [16] focused on finding patterns in conversations on Reddit based on volume, responsiveness and virality of the conversations. Choudhury et al. [17] investigated aspects of online conversations that make them interesting, in terms of encouraging user participation. They found that interesting conversations are often associated with active discussions about themes that reflect external events. We continue to focus on the online conversations.

Conversation classification. Multiple conversation classifiers has been developed. Such as, predicting whether online conversations will lead to failure [69], whether online conversations will have positive outcomes [9] and whether a conflict in an online conversation would arise, using conversational receptiveness, e.g. the willingness to engage with opposing views [67]. Others developed classifiers to classify conversations into different discourse acts [68] and to predict user participation in online political conversations [57]. Our study develops a classifier to predict whether a conversation is part of the subreddit r/climate or r/climateskeptics.

Climate change conversations. There has been research on climate change conversations, both online and offline. Schweizer et al. [56] investigated how to engage people in climate conversations in offline settings. Grundmann and Krishnamurthy [29] studied different news articles to find how climate change is framed differently in countries and how this explains the differences in climate policies. Further, in online conversations on Twitter, it was found that the term *global warming* was used more than the term *climate change* by climate skeptics [34]. We aim to research the climate change conversations on Reddit, by investigating how two opposing communities talk differently about climate change.

Politeness strategies. The logistic regression model in this paper uses the proposed politeness strategies features by Danescu-Niculescu-Mizil et al. [24]. They developed a classifier to predict whether an utterance is polite or impolite. These features have also been used to develop a framework that detects early signs of conversational failure [69]. Failure in their framework means that participants attack each other in the conversation [69].

Personal Pronouns The use of personal pronouns has been correlated to psychological processes. For example, Simmons et al. [58] studied the use of personal pronouns in interactions between spouses. They showed that first person singular pronouns correlated to problem solutions and satisfaction in their marriage, while the use of second-person pronouns correlated more to negative interactions [58]. Another study has been done on the use of pronouns and depression. It was found that participants that are depressed use "I" more than non-depressed participants [55]. We use the different personal pronouns as features for our logistic regression model.

1.4 Outline

In this chapter, we gave an introduction, formulated the research question and the hypotheses, and presented related work. Chapter 2 contains information about the Reddit Corpus data, logistic regression model, the features for this model and how the model and features are evaluated. Chapter 3 presents the results and analysis of our model. Our most interesting findings are presented in Chapter 4, in which the limitations of this study are also discussed. We conclude the thesis in Chapter 5.

2 Methods

This chapter presents the methods of our study. Section 2.1 contains information about the Reddit Corpus data. We describe the logistic regression model and the features in Section 2.2. In Section 2.3, we describe how we evaluate the model and features.

2.1 Reddit Corpus

We use the Reddit Corpus from the Cornell Conversational Analysis Toolkit (ConvoKit). ConvoKit is an open-source toolkit that can be used in Python to make conversation analyses [14]. Further, ConvoKit contains conversational datasets that can be used with the toolkit, like the Reddit Corpus. The Reddit Corpus from ConvoKit is a collection of corpora from Reddit data, and is built from the Reddit Corpus of Pushift.io [51]. The data structure of the conversational datasets consists of "Corpus", "Conversation", "Utterance" and "Speaker" classes. These classes in ConvoKit's conversational datasets have a hierarchy, in which each Corpus consists of a collection of Conversations, each Conversation consists of their corresponding Utterances and each Utterance has one Speaker. However, each Speaker can have multiple Utterances.

We use two different corpora from the Reddit Corpus, namely the subreddit r/climate and r/climateskeptics Corpus. It contains data starting from its inception until October 2018 [19]. We are interested in the conversational part of the data, so the Conversation classes. In context of Reddit, the Conversation classes are the discussions that took place on Reddit, where every discussion consists of Utterances. Figure 1 shows an example of a Conversation which consists of three Utterances.

```
Looks this guy also hated climate change deniers.

--------------

The horror of our times is not so much that so many people are
completely without the ability to think for themselves, and grab
whatever point of view is trendy and stylish; no, the horror is that
these complete nincompoops have managed to take hold of so much
power.

Their retarded points of view dominate the media. They control the
actions of most politicians. They even manage to get those who
disagree with them in public fired from their jobs and ostracized
from their careers. It's sickening, it's horrifying, and it's
saddening. The fascists of the left have taken control, and we here
are among the very few trying to stand up to their insanity.

--------------

      > The fascists of the left have taken control

       Did Hillary win? Please someone wake me up! :O
```

Figure 1: Example of a Conversation from the r/climate subreddit. It consists of three Utterances. The dashed line separates the Utterances.

Table 1: Statistics of the subreddits r/climate and r/climateskeptics Corpus.

	r/climate	**r/climateskeptics**
Conversations	16821	26247
Utterances	88334	259580
Speakers	8552	6600

As mentioned, we use two subreddits, i.e. communities: r/climate and r/climateskeptics. We took care to select subreddits that have different opinions on the scientific consensus on climate change, by evaluating the description and rules of their community. The subreddit r/climateskeptics is a community where discussions take place about the critique on climate science. The rules given by the moderators of r/climateskeptics is that users cannot disparage the subreddit as a whole [53]. The subreddit r/climate is a community where discussions take place about climate science. The rules given by the moderators of this subreddit are that science denial and conspiracy theories are not allowed [52].

The r/climate Corpus consists of 44887 Conversations and the r/climateskeptics consists of 38780 Conversations. As mentioned, the Conversations here are the discussions on the subreddit, which consists of Utterances (see Figure 1). However, not every Conversation was successful, in the sense that there are Conversations that did not have any Utterances. These unsuccessful Conversations are removed from the dataset, since they did not lead to a discussion and thus would act like noise for the classifier. Therefore, we removed in total 28066 r/climate Conversations and 12533 r/climateskeptics Conversations, which left us with 16821 r/climate Conversations and 26247 r/climateskeptics Conversations. Table 1 shows the statistics of the two different corpora. The r/climateskeptics community has less Speakers than the r/climate community. However, r/climateskeptics has more Conversations and Utterances which indicates a higher engagement of the r/climateskeptics community.

2.2 Logistic Regression

To answer the research question, *How do two opposing online communities on Reddit differ in the way they talk about climate change?*, and the hypotheses, we develop a logistic regression model to predict which community a conversation belongs to.

Logistic regression is a supervised learning classification algorithm. It is an extend of linear regression, but instead of fitting a linear equation, logistic regression fits the logistic, or sigmoid, function:

$$f(x) = \frac{1}{1+e^{-\beta_0+\beta_1 x+\ldots+\beta_n x}}.$$

A logistic regression model predicts the probability that an observation belongs to the positive class (in our case, the r/climateskeptics community). The corresponding probability is a value that varies between 0 and 1. We use the logistic regression model as a classifier, which means that an observation either belongs to class 0 or 1. The threshold is set at 0.5, which means that:

if the prediction probability ≥ 0.5, then the model classifies $y = 1$, and
if the prediction probability < 0.5, then the model classifies $y = 0$.

The logistic regression model calculates for each feature the logistic regression coefficient $\beta_1, ..., \beta_n$, where n is the number of features. These coefficients tell us if the feature has a negative or positive effect on the probability of the target variable. By taking the exponent of the coefficient, we get the odds ratio. For example, if a feature has a coefficient of 0.06, then the odds ratio is $e^{0.06} = 1.06$. This can be interpreted as: an increase of the feature by one-unit changes the odds of being part of class 1 by a factor of 1.06. For negative coefficients, we can recalculate their odds ratio by: $\frac{1}{\text{odds ratio}}$. We interpret this as: an one-unit increase of the feature changes the odds of being part of class 0 by a factor of $\frac{1}{\text{odds ratio}}$. The p-value of the feature and their corresponding coefficient tell us if the feature has a significant effect on the target variable. In our case the target variable is a binary variable, where 0 means that a conversation is part of r/climate and 1 means that a conversation is part of r/climateskeptics. We develop our logistic regression model using the Python package *Statsmodels*, which calculates, besides the coefficients, the p-values of the features.

Our goal is to find textual features that are important in distinguishing conversations between the subreddit r/climate and r/climateskeptics. Thus, we develop a logistic regression model based on textual features. We divided the features into eight different feature groups, which we use to create different classification models. All features except the Lexical features consist of a dictionary of words, i.e. cues. Table 2 shows the number of cues and examples of those cues for every feature.

Features 1-15 in Table 2 are the Politeness Strategies. For these features, the *politenessStrategies* framework from ConvoKit is used, which uses the politeness strategies from the study of Danescu-Niculescu-Mizil et al. [24]. For every conversation the number of a certain politeness strategy is counted and divided by the number of words. Features 16-47 in Table 2 uses the Linguistic Inquiry and Word Count (LIWC) dictionary [48]. The LIWC dictionary consists of 6400 words and word stems that belong to different categories. We use the *liwc* Python package to generate these features and normalise the features by dividing the cues per conversation by the number of words. Features 48-51 in Table 2 are the simple lexical features, like the number of characters, words, digits and the average word length in a conversation. The features 48-51 are the only features that are not in the same range as the other features, which has an effect on the size of the coefficients.

The feature groups Politeness Strategies (PS), Psychological Processes (PP) and Pronouns (Pron) are based on the hypotheses. To investigate the writing style of the two communities, we develop the feature groups Style (S), Lexical (Lex) and Other Style (OS). The features in the Style feature group uses the style dimensions from the study of Danescu-Niculescu-mizil et al. [23]. We create the Lexical feature group, since some studies also use lexical features to capture style [41, 59]. Additionally, we add the feature group Other Style (OS), which captures more stylistic dimensions. Moreover, the two communities can differ in the time orientation of their writing or in the use of informal language, respectively these are captured in the feature group Time Orientation (TO) and Informal Language (IL).

Table 2: Features for the Logistic Regression model.

Features		**Example**	**# Cues**
Politeness Strategies (PS)	1. Deference	good, great, interesting	8
	2. Gratitude	thanks, thank, appreciate	3
	3. Greeting	hi, hello, hey	3
	4. Positive	good, inspire, joy	2006
	5. Negative	absurd, accusing, concerned	4783
	6. Please	... please ...	1
	7. Please (start)	Please...	1
	8. Hedges	suggest, assume, usually	130
	9. By the way	by the way	3
	10. Factuality	in fact, truth, reality	15
	11. Apologising	sorry, apologize	10
	12. Direct question	where, what, why	4
	13. Direct start	So..., Then..., But...	5
	14. Subjunctive	could, would	2
	15. Indicative	can, will	2

Table 2: (continued).

Features		**Example**	**# Cues**
Psychological Processes (PP)	16. Anger	annoyed, aggressive, hate	230
	17. Anxiety	scares, anxious, fear	116
	18. Sadness	sad, worthless, sobbing	136
Pronouns (Pron)	19. First person singular	I, mine, myself	24
	20. First person plural	we, us, our	12
	21. Second person	you, yourself, your	30
	22. Third person singular	he, she, her, his	17
	23. Third person plural	they, their, them	11
	24. Impersonal pronouns	it, anybody, nobody	59
Style (S)	25. Articles	a, the, an	3
	26. Certainty	never, undeniable, obviously	113
	27. Conjunctions	but, also, and	43
	28. Discrepancy	hope, unusual, problem	83
	29. Negations	no, not, don't	62
	30. Prepositions	between, about, for	74
	31. Quantifiers	A few, whole, tons	77
	32. Tentative	perhaps, vaguely, undecided	178
	33. Insights	know, understand, solution	259
Time Orientation (TO)	34. Past focus	ago, did, had	341
	35. Present focus	is, today, am	424
	36. Future focus	will, soon, tommorow	97
Informal Language (IL)	37. Swear words	damn, pussy, fuck	131
	38. Netspeak	bruh, yup, jk	209
	39. Assent	k, ok, ah	36
	40. Nonfluencies	er, hm, huh	19
	41. Fillers	idontknow, imean, ohwell	14
Other Style (OS)	42. Auxiliary verbs	are, be, ain't	141
	43. Common Adverbs	very, generally, especially	140
	44. Comparisons	better, worst, scariest	317
	45. Interrogatives	why, who, how	48
	46. Numbers	thousands, dozen, zero	36
	47. Causation	because, depend, effect	135
Lexical (Lex)	48. No. characters	-	-
	49. No. words	-	-
	50. No. digits	-	-
	51. Average word length	-	-

2.3 Evaluation

We have two Reddit corpora, namely the r/climate and r/climateskeptics corpus. We label the r/climate corpus as class 0 and the r/climateskeptics corpus as class 1. The data is split into a training (60%), development (20%) and test set (20%). Our data is imbalanced since we have more r/climateskeptics conversations than r/climate conversations. We make sure that the ratio's between conversations of r/climate and r/climateskeptics are the same in every set. The ratio for this is 39% of r/climate conversations and 61% of r/climateskeptics conversations. In Table 3 the exact number of conversations in each set can be seen.

To handle the imbalanced data, we use random oversampling on the train dataset. Random oversampling balances the classes by replicating random observations of the minority class [10]. Batista et al. [10] compared different methods that handle imbalanced data. Their results show that random oversampling is competitive in comparison to more complex oversampling methods [10]. In our case the r/climate community is the minority class. So, from our train dataset, we randomly duplicate a conversation that is part of the r/climate community. We do this until the

Table 3: Number of conversations of each subreddit in train and test dataset.

	r/climate	r/climateskeptics	Total
Train	15747 (10093*)	15747	25840
Development	3364	5250	8614
Test	3364	5250	8614
Total	16821	26247	43068

* amount of conversations before oversampling.

minority class, i.e. r/climate, in the train dataset is balanced with r/climateskeptics.

We create models with different feature group combinations. Our baseline model is a random classifier, which we develop by using the *DummyClassifier* from *sci-kit learn*. This classifier predicts conversations as being part of r/climate or r/climateskeptics uniform at random. To evaluate the different models, we use a confusion matrix. Figure 2 shows the confusion matrix in terms of our classes:

		Predicted class	
		r/climateskeptics	**r/climate**
Actual class	**r/climateskeptics**	True Positive (TP)	False Negative (FN)
	r/climate	False Positive (FP)	True Negative (TN)

Figure 2: Confusion matrix for binary classifications.

We use the confusion matrix to calculate the recall and precision. Recall is the proportion positives that are correctly classified. In our case, this means the proportion that a conversation of the subreddit r/climateskeptics is also classified as part of the subreddit r/climateskeptics. Recall is calculated as follows:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

Precision is the proportion true positives of all positives. In our case, this means the proportion correctly classified r/climateskeptics conversations of all classified r/climateskeptics conversations. Precision is calculated as follows:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

We apply random oversampling only on the train dataset. Since our development and test dataset are still imbalanced, we use the macro F_1 score. The macro F_1 score is a suitable metric for imbalanced data since it equally accounts both classes [28]. This metric is calculated by taking the average of the F_1 score of each class. The F_1 score is the harmonic mean of the precision and recall of a class, which is calculated as follows:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}}$$

And the macro F_1 score is calculated by the following formula:

$$\text{Macro } F_1 = \frac{1}{2} \sum_{i=0}^{2} F_1 \text{ score}_i$$

The range of the macro F_1 score is in the range [0,1], in which a high macro F_1 score means a good performance of the model on both classes and a low macro F_1 indicates to a low performance of the model on both classes [28].

3 Analysis

This chapter presents the results and analysis of our model. We investigate how the two different subreddits differ in the way they talk on Reddit. To investigate this, we develop logistic regression models with different feature combinations. We start this chapter by analysing the different feature combinations (Section 3.1). In Section 3.2, we show the performance of our final model. Section 3.3 analyses the features used in our final model and Section 3.4 presents two conversations in which the model made large prediction errors.

3.1 Feature Combinations

The results of the different models on the development dataset can be seen in Table 4. The random baseline model has a macro F_1 score of 0.496. Every model improves this random baseline. Most models in Table 4 are more successful in classifying the r/climateskeptics conversations, i.e. lower F_1 score for r/climate than r/climateskeptics. This might be the result of the imbalanced development dataset, since we only used random oversampling on the train dataset.

We highlighted an interesting pattern in Table 4. As we compare the models that uses the feature groups individually, our results show that Pronouns (Pron) has the highest macro F_1 score (0.581). Adding Style (S), Informal Language (IL), Time Orientation (TO), Psychological Processes (PP), Politeness Strategies (PS) and Other Style (OS) sequentially to this model improves the macro F_1 score further and eventually leads to the best combination (model in bold in Table 4) in terms of the macro F_1 score. The sequence in which the features are added reflects on the importance of these features in this model, in which Pronouns has the highest influence and Other Style the lowest.

Table 4: Evaluation of the feature combinations on the development dataset. The highlighted models show the pattern that leads to the best model in terms of the macro F_1 score. The best model is in bold.

Features	**F_1 Score r/climate**	**F_1 Score r/climateskeptics**	**Macro F_1 Score**
Baseline	0.444	0.549	0.496
PS	0.449	0.637	0.543
PP	0.561	0.450	0.505
Pron	0.519	0.643	0.581
S	0.418	0.654	0.536
TO	0.418	0.654	0.536
Lex	0.582	0.476	0.529
IL	0.351	0.722	0.537
OS	0.504	0.602	0.553
PS + P	0.560	0.658	0.609
Pron + S	0.563	0.680	0.622
Lex + P	0.566	0.629	0.598
IL + P	0.554	0.676	0.615

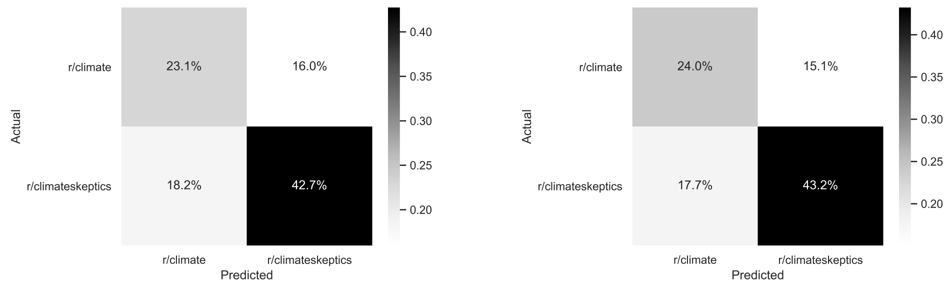
Table 4: (continued).

Features	**F_1 Score r/climate**	**F_1 Score r/climateskeptics**	**Macro F_1 Score**
PS + S + Pron	0.567	0.688	0.628
S + Pron + PP	0.567	0.684	0.625
S + Pron + OG	0.566	0.686	0.626
Pron + S + IL	0.564	0.701	0.633
PS + S + Pron + IL	0.566	0.706	0.636
Pron + S + IL + TO	0.568	0.706	0.637
S + Pron + PP + IL	0.533	0.705	0.619
S + Pron + OS + IL	0.565	0.704	0.634
PS + S + Pron + PP + IL	0.569	0.707	0.638
PS + S + Pron + PP + TO	0.575	0.696	0.636
S + Pron + PP + OS + TO	0.577	0.695	0.636
Pron + S + IL + TO + PP	0.573	0.708	0.641
PS + S + Pron + PP + OS + TO	0.578	0.702	0.640
Pron + S + IL + TO + PP + PS	0.575	0.712	0.643
PS + S + Pron + PP + TO + Lex	0.586	0.686	0.636
S + Pron + PP + IL + TO + Lex	0.586	0.691	0.639
PP + Pron + S + TO + Lex + IL + OS	0.584	0.693	0.638
PS + Pron + S + TO + Lex + IL + OS	0.581	0.694	0.637
PS + PP + S + TO + Lex + IL + OS	0.572	0.678	0.625
PS + PP + Pron + TO + Lex + IL + OS	0.581	0.682	0.632
PS + PP + Pron + S + Lex + IL + OS	0.578	0.690	0.634
Pron + S + IL + TO + PP + PS + OS	**0.575**	**0.714**	**0.644**
PS + PP + Pron + S + TO + Lex + OS	0.585	0.689	0.637
PS + PP + Pron + S + TO + Lex + IL	0.585	0.694	0.640
All	0.584	0.698	0.641

Note: PS = Politeness Strategies, PP = Psychological Processes, Pron = Pronouns, S = Style, TO = Time Orientation, L = Lexical, IL = Informal Language, OS = Other Style.

3.2 Final Model

The final model consists of seven feature groups: Pronouns, Style, Informal Language, Time Orientation, Psychological Processes, Politeness Strategies and Other Style. We use the final model to evaluate the model on the test dataset. Figure 3 shows the confusion matrix for the development dataset and test dataset.



(a) Development dataset **(b)** Test dataset

Figure 3: Confusion matrices of final model.

Table 5: Performance of final model.

	F_1 Score r/climate	F_1 Score r/climateskeptics	Macro F_1 Score
Development	0.575	0.714	0.644
Test	0.594	0.725	0.660

The performance of the model is almost the same on both development and test dataset, which can also be seen in the comparable macro F_1 score on the development and test dataset (Table 5). Table 5 also shows that the F_1 score of the r/climateskeptics class is higher than the F_1 score of the r/climate class. This clearly shows that the model is more successful in classifying the r/climateskeptics conversations, which might be the result of the imbalanced development and test dataset, since we only used random oversampling on the train dataset.

3.3 Features

3.3.1 Politeness Strategies

Table 6 shows the results of the Politeness Strategies features [24]. We have ten positive politeness strategies that are correlated to being polite and five negative politeness strategies that are correlated to being impolite [24]. It is worth noting that most Politeness Strategies features have $p > 0.05$. Those features did not have a significant effect on classifying conversations.

Our results show that conversations including the positive politeness features gratitude (*thanks, appreciate*), greeting (*hi, hello, hey*) or positive lexicon (*good, inspire, joy*) and the negative politeness feature direct question (*what, why, how*) are more likely to be part of the r/climate community. Thus, three out of ten positive politeness features and one out of five negative politeness features are more likely to be part of the r/climate community. Especially the positive politeness features gratitude and greeting have a large effect in classifying the r/climate conversations. An example of a conversation that uses gratitude and greeting can be seen in Figure 4. In Figure 4, the first speaker uses a greeting when sharing a project, while another speaker uses gratitude to thank the initial speaker.

Table 6: Feature coefficients and p-values of Politeness Strategies. Highlighted features have $p \leq 0.05$.

Features		**Coefficient**	**Odds ratio**	**p-value**
Politeness Strategies (PS)	1. Deference - *polite*	3.58	35.90	0.00
	2. Gratitude - *polite*	-11.13	68511.92*	0.00
	3. Greeting - *polite*	-19.90	500000000*	0.02
	4. Positive lexicon - *polite*	-3.18	24.11*	0.00
	5. Negative lexicon - *impolite*	0.08	1.08	0.81
	6. Please - *polite*	-7.16	1287.62*	0.26
	7. Please (start) - *impolite*	2.46	11.65	0.66
	8. Hedges - *polite*	1.25	3.48	0.37
	9. By the way - *polite*	-0.80	2.23*	0.99
	10. Factuality - *impolite*	4.59	98.32	0.00
	11. Apologising - *polite*	1.40	4.04	0.47
	12. Direct question - *impolite*	-6.17	477.08*	0.00
	13. Direct start - *impolite*	-1.15	3.14*	0.27
	14. Subjunctive - *polite*	-7.15	1277.37*	0.41
	15. Indicative - *polite*	22.44	5554926507.31	0.14

* The odds ratios of the negative coefficients are recalculated with $\frac{1}{oddsratio}$.

```
Hello fine redditors of /r/climate

I wanted to share a small project of mine that I have been developing in
the evenings and weekends.

Our goal is to simply raise awareness about the current melt conditions of
the polar ice caps.  The app can be a great tool to use if ever engaging in
discussion with anyone who is passionate about climate science.

It is simple, but we hope you all like it and find it useful.
We would love to do a lot more with it, and perhaps in the future if we can
get more traffic we will be able to.
Until then, we hope to provide at least a little bit of help in drawing
attention to the issue.

Thanks!
--------------
    Great idea!
    Any plans for an Android version?
    --------------
        Thanks! and yes, an Android version is in progress already.

Indirect greeting   Gratitude
```

Figure 4: Conversation from the r/climate subreddit. Green shows the indirect greeting and blue shows the gratitude cues.

Further, our results show that conversations including the positive politeness feature deference (*great, good, interesting*) or negative politeness feature factuality (*in fact, reality, truth*) are more likely to be part of r/climateskeptics. Thus, one out of ten politeness features and one out of five impoliteness features are more likely to be part of the r/climateskeptics community.

Our first hypothesis suggests that the r/climateskeptics uses less politeness cues in their conversations than the r/climate community. As mentioned, we find three *politeness* features significant for r/climate and only one *politeness* feature significant for r/climateskeptics. In terms of the *impoliteness features*, we find a significant feature for both communities each. We find mixed results, i.e. we find significant *politeness* and *impoliteness* features for each community. However, we find more evidence that the r/climate community is more polite. Three significant *politeness* features were significant for classifying r/climate conversations and two of those features (gratitude and greeting) show a large odds ratio. Hence, we find no evidence to reject our first hypothesis that the r/climateskeptics uses less politeness cues in their conversations than the r/climate community.

We suggested our first hypothesis since Kaiser [35] found that climate skeptics not only attack individuals with opposing views, but also attack each other in their own community. Kaiser [35] argued that this could be since there is not a clear consensus in climate skepticism. An example, in which there are two speakers who respond defensively to each other while using factuality (impolite) cues, can be seen in Figure 5.

```
perhaps if you had actually read all 500 questions, you would understand
why there are 500 questions. but you won't because you enjoy being right
even when you're wrong. (i just listened to BOTH sides of the argument)
      --------------
      Actually, since then I *have* looked into both sources.  I haven't
      through all of both (yet)

Factuality
```

Figure 5: Part of a conversation from the r/climateskeptics subreddit. Red shows the factuality cues.

Table 7: Feature coefficients and p-values of Style. Highlighted features have $p \leq 0.05$.

Features		Coefficient	Odds ratio	p-value
Style (S)	25. Articles	3.08	21.79	0.00
	26. Certainty	9.77	17418.09	0.00
	27. Conjunctions	0.91	2.49	0.03
	28. Discrepancy	-4.31	74.20*	0.00
	29. Negations	0.21	1.24	0.66
	30. Prepositions	0.63	1.88	0.01
	31. Quantifiers	-0.46	1.59*	0.32
	32. Tentative	2.37	10.67	0.00
	33. Insights	2.50	12.23	0.00

* The odds ratios of the negative coefficients are recalculated with $\frac{1}{oddsratio}$.

3.3.2 Style

In Table 7 we show the results of the Style features [23]. Most Style features have $p \leq 0.05$, and thus have a significant effect in classifying conversations. However, most of these features do not show a high odds ratio.

Our second hypothesis suggests that the r/climate community uses more negation cues (*no, not, never*) in their conversations than the r/climateskeptics community. The use of negations has been brought together with inhibition [61], in which inhibition can be a response to the impacts of climate change [25]. An example of this is when supporters of the consensus of climate change cannot properly express the impacts of climate change [25]. Our results show that the use of negations in conversation are more likely to be part of the r/climateskeptics community. However, the odds ratio is rather small and the feature has $p > 0.05$. Thus, we do not find evidence of a correlation between the use of negations and the classification of conversations. Hence, we cannot reject the null hypothesis that there is no difference in the use of negations between the r/climate and r/climateskeptics community.

In terms of the Style features, we see that certainty (*never, undeniable, obviously*) has the largest effect in classifying conversations as part of the r/climateskeptics community. The use of certainty cues has been correlated to individuals who seek for information that is in line with their points of view [13]. Another study found that individuals who use more certainty cues are unlikely to change their points of view when confronted with other information [49]. However, further research could done on why the r/climateskeptics community uses more certainty cues. In Figure 6, we show an example of a conversation, in which certainty cues are used by r/climateskeptics users.

```
I can tell you this with absolute certainty:  Ben Stein doesn't know a damn
thing about science.  He always sucked at science questions on his little
quiz show, and he still sucks at science now.
       --------------
        You think that this is what it is about?  It is about the
        policing of thoughts on global warming.
            --------------
            Exactly. The warmists are controling the apparatus of
            social censorship, and using it liberally to censor anyone
            of any public stature who speaks out against them.
----------------------------------------------------------------------------
Certainty
```

Figure 6: Part of a conversation from the r/climateskeptics subreddit. Pink shows the certainty cues.

Table 8: Feature coefficients and p-values of Pronouns. Highlighted features have $p \leq 0.05$.

Features		Coefficient	Odds ratio	p-value
Pronouns	19. First person singular	-0.54	1.72	0.30
(Pron)	20. First person plural	-11.57	105496.36*	0.00
	21. Second person	11.36	86074.22	0.00
	22. Third person singular	8.15	3469.57	0.00
	23. Third person plural	10.19	26747.51	0.00
	24. Impersonal pronouns	1.3467	3.8448	0.000

* The odds ratios of the negative coefficients are recalculated with $\frac{1}{oddsratio}$.

3.3.3 Pronouns

In Table 8, we show the results of the Pronouns features. Noteworthy is that almost all Pronouns features have a significant effect ($p \leq 0.05$) in classifying conversations. We suggested in our third hypothesis that the r/climate community uses more first person singular pronouns cues (*I, me, mine*) than the r/climateskeptics community. We hypothesised this since the use of first person singular pronouns can correlate to emotional pain [55], which can be a result of climate change [21]. Our results show that the first person singular pronouns have $p > 0.05$, which means that this feature does not have a significant effect in classifying conversations. Hence, we cannot reject the null hypothesis that there is no difference in the use of first person singular pronouns between the r/climate and r/climateskeptics community.

Furthermore, the results in Table 8 show that conversations including first person plural pronouns (*we, us, our*) are more likely to be part of the r/climate community. In Figure 7a, it can be seen that a r/climate speaker talks about ”our planet” and that both speakers uses ”we” when talking about who is affected by climate change. A previous study found that the use of first person plural pronouns by individuals may lead to an increased group cohesion [61]. Group cohesion refers to the degree that individuals of a group feel connected, and also indicates to which degree they work together to achieve their goals [5]. It might be that the r/climate community uses more first person plural pronouns to strengthen their bond to fight climate change together.

Table 8 also shows that conversations including second person (*you, yourself, your*), third person singular (*he, she, his*) and plural (*they, their, them*), and impersonal (*it, anybody*) pronouns are more likely to be part of the r/climateskeptics community. An example of a conversation including these features can be seen in Figure 7b. Table 8 shows that second person, third person singular and plural have a large odds ratio. It is not clear to us why the r/climateskeptics community uses more second person pronouns. Further research could be done to see why this is the case.

A possible explanation that conversations including third person singular and plural are more likely to be part of r/climateskeptics involves the term *out-group.* An out-group is a group in which the individual does not identify with [4]. For example, a climate skeptic who does not identify itself with the supporters of climate change. A previous study suggested that the use of third person pronouns might reflect on experiencing the out-group, in which more attention is given to that out-group [38]. This may indicate that the r/climateskeptics community is more conscious about their opposing group (supporters of the consensus of climate change).

```
Not our planet - we are just renting. Always a
temporary arrangement. We're about to get
evicted for not looking after the place.
        --------------
        No, we are after burning the apartment
        down. At the moment, we chocking on
        the fumes as the entire structure
        falls apart. When we're finished,
        there will be no apartment, because
        earth will be a dead planet.

First person plural   Second person
Third person singular   Third person plural
Impersonal
```

(a) r/climate

```
Yes, it must be a conspiracy of the men in
black with their black-box methods. And if you
can link it with the lizard-men you might
double the number of your active supporters :)
        --------------
        It sounds like they are using Michael
        Mann's hockey stick model.
        --------------
        Jonova is the woman who managed to
        find the scandal. Except there is no
        scandal because she completely
        misinterpreted everything to come up
        in a way that made deniers happy. And
        then her crappy statistics spreads in
        the denier blogosphere and becomes
        part of the growing arsenal of
        misinformation. Depressing.

First person plural   Second person
Third person singular   Third person plural
Impersonal
```

(b) r/climateskeptics

Figure 7: Part of conversations using the significant pronouns features. (a) shows a r/climate conversation and (b) shows a r/climateskeptics conversation. Purple shows the first person plural pronouns cues, red shows second person pronouns, yellow shows the third person singular pronouns, green shows the third person plural pronouns and blue shows the impersonal pronouns cues.

3.3.4 Psychological Processes

Table 9 shows the results of the Psychological Processes features, in which we only found a non-significant ($p > 0.05$) effect for the feature anger. In our fourth hypothesis, we suggest that the r/climate community uses more sadness cues (*sad, worthless*) in their conversations than the r/climateskeptics community. Sadness can arise as a response to climate change related losses [21]. Our result shows indeed that conversations including sadness cues are more likely to be part of the r/climate community. Hence, we find no evidence to reject the fourth hypothesis that the r/climate community uses more sadness cues in their conversations than the r/climateskeptics community. In Figure 8a, it can be seen that a r/climate speaker uses sadness cues as a response to climate change related losses. However, further research could be done to investigate the correlation between sadness and the r/climate community.

The fifth hypothesis suggests that the r/climate community uses more anger cues (*hate, annoyed*) in their conversations than the r/climateskeptics community. We hypothesised this since anger could also be a response to climate change related losses [21]. Our results indicate that conversations consisting anger are more likely to be part of the r/climate community. However, we find no significant effect for the feature anger ($p > 0.05$). Hence, we cannot reject the null hypothesis that there is no difference in the use of anger cues between the r/climate and r/climateskeptics community.

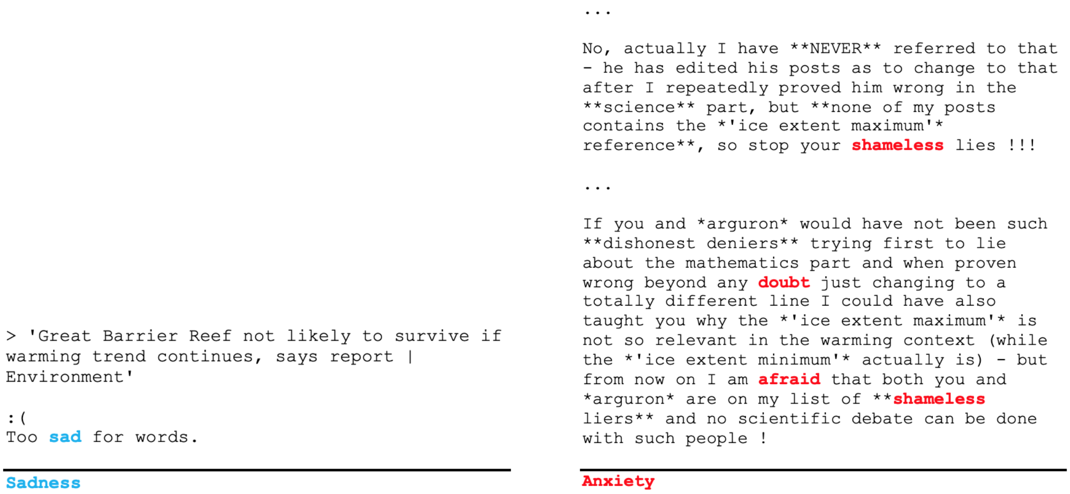
The sixth hypothesis suggests the subreddit r/climateskeptics uses more anxiety cues (*scares, anxious*) in their conversations than the opposing subreddit r/climate. Anxiety has been brought

Table 9: Feature coefficients and p-values of Psychological Processes. Highlighted features have $p \leq 0.05$.

Features		**Coefficient**	**Odds ratio**	**p-value**
Psychological	16. Anger	-0.34	1.40*	0.67
Processes	17. Anxiety	9.55	14071.68	0.00
(PP)	18. Sadness	-2.13	8.41*	0.03

* The odds ratios of the negative coefficients are recalculated with $\frac{1}{oddsratio}$.

together with responding defensively [47]. We find strong evidence in our model that conversations including the feature anxiety are more likely to be part of the r/climateskeptics community (see Table 9). Hence, we find no evidence to reject our sixth hypothesis that the r/climateskeptics community uses more anxiety cues in their conversations than the r/climate community. While it might be that the r/climateskeptics community uses more anxiety cues, since they respond more defensively to each other [47], see Figure 8b for an example, further research could be done to explain why anxiety cues are more used in the r/climateskeptics community.



(a) r/climate using sadness

(b) r/climateskeptics using anxiety

Figure 8: Part of conversations. (a) shows a r/climate conversation using sadness and (b) shows a r/climateskeptics conversation using anxiety. Blue shows the sadness cues and red the anxiety cues.

3.3.5 Time Orientation

Table 10 shows the results of the Time Orientation features. We find significant effects of present and future focus ($p \leq 0.05$). Conversations including present focus (*is, today, am*) or future focus (*will, soon*) cues are more likely to be part of the r/climate community. In Figure 9a, a r/climate speaker uses the present focus when talking about the current climate issues and future focus when talking about climate issues that will arise in the future. In Figure 9b, it can be seen that a r/climateskeptics speaker uses the present focus since the user is focusing on the validity of the current climate science and not on the consequences that may happen in the future. It is interesting that the odds ratio for future focus is higher than the present focus, which might be connected with the r/climate community being more focused on future climate policies and consequences.

Table 10: Feature coefficients and p-values of Time Orientation. Highlighted features have $p \leq 0.05$.

Features		**Coefficient**	**Odds ratio**	***p*-value**
Time	34. Past focus	0.00	1.00	0.99
Orientation	35. Present focus	-1.38	3.98*	0.00
(TO)	36. Future focus	-7.19	1329.84*	0.00

* The odds ratios of the negative coefficients are recalculated with $\frac{1}{oddsratio}$.

Who **expects** coral bleaching to end? **It's** caused by increased temperatures so it **makes** sense to assume that only lower temperatures **will** stop it. But those **aren't** happening. **It's** been **predicted** for years that all tropical coral reefs **are** toast.

Present Focus **Future Focus**

(a) r/climate

I agree that there **has** been a 1.5mm average rise per year in sea level as per this data yes. Previous to this we **have** no data. And no one **is** arguing that CO2 levels **have** increased.

Where however it stops **is** the correlation between this rise and CO2 and the acceleration caused by CO2.

...

I **get** the hypothesis: rising C02 = sea level rise. I just **don't see** the data backing it up, or better still data from a long enough trend to **support** said hypothesis. There **can be** no conclusion in that respect.

You however **seem** to **have** found one. Which, in itself, **is** a complete mystery.

Present Focus **Future Focus**

(b) r/climateskeptics

Figure 9: Part of conversations using present and future focus. (a) shows a r/climate conversation and (b) shows a r/climateskeptics conversation. Blue shows the present focus cues, green shows the future focus cues.

3.3.6 Informal Language

Table 11 shows the results of the Informal Language features. We show that the r/climate communitys uses more swear words (*fuck, bullshit*), netspeak (*lol, btw)* or assent (*yup, cool*).

It is not clear to us why the r/climate community uses more swear words. For example, Figure 10a shows how a r/climate speaker uses swear words to express how bad climate issues are, while in Figure 10b another r/climate speaker uses swear words to express anger towards climate change deniers. A previous study found that using more swear words is related to a lower agreeableness [40], which may be a possible explanation. Another study found that echo chambers are correlated to using less swear words [26]. This might explain why the r/climateskeptics community uses less swear words, since it may act like an echo chamber [43]. However, further research could investigate why the r/climate community uses more swear words and why it might be that the r/climate community is less of an echo chamber.

Furthermore, we find a significant effect of nonfluencies (*ah, er, hm*). Conversations including nonfluencies are more likely to be part of r/climateskeptics. A previous study argued that the use of nonfluencies could correlate to an underlying uncertainty about the topic [42]. This is interesting, since it contradicts the finding that the r/climateskeptics community uses more certainty cues than the r/climate community (see Subsection 3.3.2). So, it is still unclear to us why the r/climateskeptics uses more nonfluencies. In Figure 11, a r/climateskeptics speaker uses nonfluencies to express sarcasm towards climate policies. Further research could be done on why the r/climateskeptics community uses more nonfluencies.

Table 11: Feature coefficients and p-values of Informal Language. Highlighted features have $p \leq 0.05$.

Features		Coefficient	Odds ratio	p-value
Informal Language (IL)	37. Swear words	-5.96	389.32*	0.00
	38. Netspeak	-6.90	989.07*	0.00
	39. Assent	-2.36	10.58*	0.00
	40. Nonfluencies	6.98	1070.36	0.00
	41. Fillers	4.53	92.78	0.41

* The odds ratios of the negative coefficients are recalculated with $\frac{1}{oddsratio}$.

Well then we're waaaaay ahead of the game,
because just last week, we hit over
pre-industrial temperatures.

Well the average probably isn't going
to be that high. If it is we're **fucked**

And just because it isn't the
that high yet doesn't mean that
were not still **fucked**, because
we are.

Yeah, we'll just get
fucked a bit later.

Climate's going
to **fuck** us
discretely then
it's going **fuck**
us hard.

Swear words

(a) r/climate

Fuck off denier tard. Deniers - talk about a s
pecies on the verge of extinction. Hey **retard**,
so you haven't noticed anything unusual
lately? like all the AGW jacked weather disast
er smashing the **shit** out of the infrastructure
and tearing up people's lives every single
day? No? I guess there is no wifi signal when
your head's up your **ass**.

Swear words

(b) r/climate

Figure 10: Part of conversations using swear words. (a) and (b) shows a r/climate conversation. Yellow shows the swear words cues.

That's three billion of your tax money, folks. Hope you all enjoy paying
for .... **hmmm**, what are we paying for, anyway? Al Gore needs another
mansion on the seashore? Michael Mann is looking to hire a really good
lawyer? **Well**, all those nice people who make metal windmills that kill
birds and rust solid after ten years need to put in backyard swimming
pools, I guess.

Nonfluencies

Figure 11: Utterance of a conversation from the r/climateskeptics subreddit. Green shows the nonfluencies cues.

3.3.7 Other Style

Table 12 shows the results of the Other Style features. While most Other Style features did have a significant effect ($p < 0.05$), most of those did not have a large effect. Interrogatives show the largest odds ratio, in which conversations including interrogatives are more likely to be part of the r/climateskeptics community. Interrogatives are words that are associated to questions, e.g. *how, when, what*. This feature differs from the feature direct question (part of the Politeness Strategies). The feature direct question only accounts the cues where words like *how* or *what* are at the beginning of a sentence, whereas the feature interrogatives also accounts these words when they are used in the middle of sentences. An example of this can be found in Figure 12.

Table 12: Feature coefficients and p-values of Other Style. Highlighted features have $p \leq 0.05$.

Features		Coefficient	Odds ratio	p-value
Other Style (OS)	42. Auxiliary verbs	2.6803	14.5902	0.000
	43. Common Adverbs	-1.5087	4.5210*	0.000
	44. Comparisons	-0.2238	1.2509*	0.627
	45. Interrogatives	6.1136	451.9702	0.000
	46. Numbers	2.3876	10.8872	0.001
	47. Causation	-0.8746	2.3979*	0.029

* The odds ratios of the negative coefficients are recalculated with $\frac{1}{oddsratio}$.

A previous study found that the use of more interrogatives was related to a higher perception of truthfulness [31]. The r/climateskeptics community might try to deceive readers that what they are saying is the truth. This may be a possible explanation on why conversations using interrogatives are more likely to be part of the r/climateskeptics community. However, further research could be done to investigate the relation between interrogatives and the r/climateskeptics community.

You are fucking pathetic. Smug bastard telling everyone **what** to believe **when** you don't understand the basics yourself. Typical clueless liberal.

Interrogatives

Figure 12: Utterance of a conversation from the r/climateskeptics subreddit. Blue shows the interrogatives cues.

3.4 Error Analysis

To gain an understanding of the behaviour of our model, we analysed cases in which the model made large prediction errors. We analysed ten conversations that were classified with a high predicted probability as r/climateskeptics, while they were actually r/climate conversations. Furthermore, we analysed ten conversations that were classified with a high predicted probability as r/climate, while they were actually r/climateskeptics conversations. From those, we manually chose one wrongly classified r/climate conversation and one wrongly classified r/climateskeptics conversation to illustrate why the model classified these wrongly.

First, the wrongly classified r/climate conversation i.e. predicted: r/climateskeptics, actual: r/climate. In Figure 13, we show an utterance of this wrongly classified conversation. It might not be unexpected that the model classified this conversation as r/climateskeptics instead of r/climate. Figure 13 shows that it includes a lot of certainty cues, third person plural pronouns and impersonal pronouns. These features were more likely to be part of the r/climateskeptics community. Furthermore, the speaker of this utterance (Figure 13) tends to be more of a r/climateskeptics user as this speaker has critique on the consensus on climate change. Further research could focus on the different views in climate skepticism.

It's **truly** like trying to deprogram a cult member. My parents are Mormons, and **they** come out with **exactly** the same behavior as most global warmists. **They** refuse to even see the evidence **they** don't like, as if **it** doesn't exist. **They've** invested so much into apologist debate and believe **they** are righteous in **their** hatred of evil CO2. **Nothing** is going to sway **them**. Even if you show **them** evidence from the same source **they** use, such as NASA, **that** rising CO2 is in **fact** greening the planet at a faster rate now than we cut down trees, **they** ignore **it**. **Everything** **they** believe is **true**, and **everything** **must** point to **that**. There is **nothing** good **that** comes from evil, therefore evil CO2 **must** only cause bad **things** to happen. The narrative is **that** the Earth is dying very soon, and **it's** our punishment for being evil. **It's** **all** cult thinking.

Certainty **Third Person Plural** **Impersonal Pronouns**

Figure 13: Utterance of a conversation that the model wrongly classified. Actual: r/climate, Predicted: r/climateskeptics. Red shows the certainty cues, blue the third person plural cues and green the impersonal pronouns cues.

Second, the wrongly classified r/climateskeptics conversations, i.e. predicted: r/climate, actual: r/climateskeptics. An utterance of this wrongly classified conversation can be found in Figure 14. The coloured features in Figure 14 show the features that were more likely to be part of the r/climate community, and especially first person plural had a large odds ratio. The speaker in

Figure 14 seems to be against the climate policies, such as the CO_2 taxes. However, it also seems that the speaker is not necessarily against climate action. This might indicate to the not so clear common ground in climate skepticism [35].

Our bills in the UK **are** already ridiculous, **we** **cannot** afford them to **go** up any more over **bullshit**. I **mean** I fully **support** throwing money at technologies like the Algae oil project in Spain and the viable commercial extraction of Uranium from sea water because **we** **will** **eventually** **need** to deal with the oil issues.

But please, CO2 taxes and windfarm subsides, **leave** it out. Enough.

Swear **First Person Plural** **Present Focus** **Future Focus**

Figure 14: Utterance of a conversation that the model wrongly classified. Actual: r/climateskeptics, Predicted: r/climate. Red shows the swear cues, blue the first person plural cues, yellow the present focus cues and green the future focus cues.

4 Discussion

The goal of this thesis was to investigate how two opposing communities, r/climate and r/climateskeptics, differ in the way they talk about climate change. We investigated this by developing a logistic regression model using textual features.

We found that the r/climate community used more first person plural pronouns and words related to gratitude and greeting in their conversations than the r/climateskeptics community. Since the politeness strategies gratitude and greeting had a large odds ratio, it may indicate that the r/climate community is more polite in their conversations than the r/climateskeptics community. However, it is not clear to us why they use more words related to gratitude and greeting than the r/climateskeptics community. Further research could be done to explain this finding. The use of first person plural pronouns may lead to an increased group cohesion (i.e. the degree that individuals feel connected and to which degree they want to achieve their group goals [5]) [61]. It could be that the r/climate users use more first person plural pronouns to increase their bond to fight climate change together.

Further, we found that the r/climateskeptics community used more second person pronouns, third person pronouns and words related to certainty and anxiety in their conversations than the r/climate community. It is not clear to us why the r/climateskeptics community uses more second person pronouns. Further research could be done to investigate why the r/climateskeptics community uses more second person pronouns. It might be that individuals who use third person pronouns are more conscious of the out-group (i.e. the group that the individual does not identify with [4]) [38]. This may indicate that the r/climateskeptics community is more conscious of their opposing group (supporters of the consensus of climate change). More research could be done to see why the r/climateskeptics community uses more third person pronouns. Previous studies found that individuals who use certainty cues seek for information that is in line with their points of view [13] and that those individuals are unlikely to change their points of view when confronted with other information [49]. However, further research could be done on why the r/climateskeptics community uses more certainty cues.

Our first limitation affects the use of more anxiety cues by the r/climateskeptics community. We found that the anxiety category of LIWC also consists of the word stem *alarm∗*. Thus, words like *alarmism*, *alarmist* and *alarmists* were counted as anxiety cues. Climate skeptics label climate scientist and their opposing group as climate alarmists [36]. Climate alarmists are (supposedly) individuals that blindly support the consensus of climate change [36]. We found that 19002 r/climateskeptics conversations contained *alarm∗* words, whereas only 1100 r/climate conversations contained *alarm∗* words. However, these words most likely did not capture anxiety, but contributed a lot to the large odds ratio of the feature anxiety.

The second limitation involves the different views within the two communities. We expected the r/climateskeptics community to have different views in their community, as it may be that there is no clear common ground in climate skepticism [35]. However, it might also be that there is not always a clear consensus in the r/climate community. An example can be seen in Figure 15. Therefore, further research could consider the different points of views of both communities which may help with understanding both communities better.

Our third limitations is that the r/climateskeptics users also commented in the r/climate community, and vice versa. We had 8552 r/climate speakers and 6600 r/climateskeptics speakers.

```
For the climate I gave up meat on 1-2-94, best thing I ever did.
--------------
...

--------------
it is low, /r/climate has been invaded by vegans who only post anti-meat
articles. it's been like this for months.
    --------------
    how about fuck you and your anti-meat stance?
    --------------
    How about you try to reduce the population rather than limit the
    amounts of specific foods they can eat?

    You'll have much more success in telling people they can't make more
    people than telling them they can't have more meat.
```

Direct Question

Figure 15: Part of a conversation from the r/climate subreddit. Orange shows the direct question cues.

In total there were 1034 speakers who were active in both communities. Since we classified conversations, it might be that some conversations consisted of utterances of both opposing sides of the consensus of climate change, which may have affected our model. Therefore, further research could focus on classifying utterances rather than classifying conversations.

A fourth limitation involves the sociodemographic characteristics of Reddit users [3]. The Reddit population is not comparable to the general population, which makes it harder to generalise to the offline world. Reddit users tend to be younger than the general population and are mostly male [3]. However, our findings are still relevant for further research on Reddit.

Another limitation is that we indirectly assumed that using more, for instance, politeness cues reflects on a more polite conversation. Speakers might have quoted other speakers or articles, which may included politeness cues. However, such quotes do not reflect on a more polite conversation. The same can be said about the psychological processes features (anger, sadness and anxiety). Therefore, further research could focus on the context in which those cues are used instead of counting the number of cues per conversation.

5 Conclusion

This study investigated how the r/climate and r/climateskeptics community differ in the way they talk about climate change. We developed a logistic regression model to classify which community a discussion belongs to. We found various features that had a large effect in classifying conversations. Our most important findings are that conversations including words related to certainty, second person pronouns and third person pronouns were more likely to be part of the r/climateskeptics community. Whereas conversations including first person plural pronouns and words related to gratitude and greeting were more likely to be part of the r/climate community. Our findings could improve chatbots in context of climate change. Such chatbots can be improved with knowledge about the preferences and values of the user to stimulate conversation flow and the engagement of the user [15]. The features that had a large effect in distinguishing r/climate and r/climateskeptics conversations could be implemented in these chatbots. By doing so, the chatbot may better distinguish between individuals that are climate skeptics and those that are not. Thus, our finding may help the chatbot to give personalised answers depending on if it is talking to a climate skeptic or not. Future work could consider focusing on the classification of utterances of the two opposing communities. Distinguishing between utterances can help gaining a deeper understanding of the online climate change discourse to counteract the polarisation.

Bibliography

[1] M. Aklin and J. Urpelainen. Perceptions of scientific dissent undermine public support for environmental policy. *Environmental Science and Policy*, 38:173–177, 2014. ISSN 14629011. doi: 10.1016/j.envsci.2013.10.006. 2

[2] A. Amaya, R. Bach, F. Keusch, and F. Kreuter. New Data Sources in Social Science Research: Things to Know Before Working With Reddit Data. *Social Science Computer Review*, pages 1–18, 2019. ISSN 15528286. doi: 10.1177/0894439319893305. 1

[3] A. Amaya, R. Bach, F. Keusch, and F. Kreuter. New Data Sources in Social Science Research: Things to Know Before Working With Reddit Data. *Social Science Computer Review*, pages 1–18, 2019. ISSN 15528286. doi: 10.1177/0894439319893305. 23

[4] American Psychological Association. outgroup – APA Dictionary of Psychology, . URL `https://dictionary.apa.org/outgroup`. 15, 22

[5] American Psychological Association. group cohesion – APA Dictionary of Psychology, . URL `https://dictionary.apa.org/group-cohesion`. 15, 22

[6] American Psychological Association. inhibition – APA Dictionary of Psychology, . URL `https://dictionary.apa.org/inhibition`. 3

[7] W. R. Anderegg, J. W. Prall, and J. Harold. Reply to O'Neill and Boykoff: Objective classification of climate experts. *Proceedings of the National Academy of Sciences of the United States of America*, 107(39):2010, 2010. ISSN 00278424. doi: 10.1073/pnas.1010824107. 2

[8] E. Aumayr and J. Chan. Reconstruction of Threaded Conversations in Online Discussion Forums. *Artificial Intelligence*, pages 26–33, 2011. URL `http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewPDFInterstitial/2840/3279`. 4

[9] J. Bao, J. Wu, Y. Zhang, E. Chandrasekharan, and D. Jurgens. Conversations Gone Alright: Quantifying and Predicting Prosocial Outcomes in Online Conversations. 2021. doi: 10.1145/3442381.3450122. URL `http://arxiv.org/abs/2102.08368{%}0Ahttp://dx.doi.org/10.1145/3442381.3450122`. 4

[10] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, 6(1):20–29, 2004. ISSN 1931-0145. doi: 10.1145/1007730.1007735. 8

[11] C. Beer. *The rise of online communities.* 2020. ISBN 9789150627787. URL `https://blog.globalwebindex.com/chart-of-the-week/online-communities/`. 2

[12] E. F. Bloomfield and D. Tillery. The Circulation of Climate Change Denial Online: Rhetorical and Networking Strategies on Facebook. *Environmental Communication*, 13(1):23–34, 2019. ISSN 17524040. doi: 10.1080/17524032.2018.1527378. URL `https://doi.org/10.1080/17524032.2018.1527378`. 1

[13] L. A. Brannon, M. J. Tagler, and A. H. Eagly. The moderating role of attitude strength in selective exposure to information. *Journal of Experimental Social Psychology*, 43(4):611–617, 2007. ISSN 00221031. doi: 10.1016/j.jesp.2006.05.001. 14, 22

[14] J. P. Chang, C. Chiam, L. Fu, A. Z. Wang, J. Zhang, and C. Danescu-Niculescu-Mizil. ConvoKit: A Toolkit for the Analysis of Conversations. 2020. URL `http://arxiv.org/abs/2005.04246`. 5

[15] A. P. Chaves and M. A. Gerosa. How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human-Computer Interaction*, 37(8):729–758, 2021. ISSN 15327590. doi: 10.1080/10447318.2020.1841438. 2, 24

[16] D. Choi, J. Han, T. Chung, Y. Y. Ahn, B. G. Chun, and T. Kwon. Characterizing conversation patterns in reddit: From the perspectives of content properties and user participation behaviors. *COSN 2015 - Proceedings of the 2015 ACM Conference on Online Social Networks*, pages 233–243, 2015. doi: 10.1145/2817946.2817959. 4

[17] M. D. Choudhury, H. Sundaram, A. John, and D. D. Seligmann. What makes conversations interesting? Themes, participants and consequences of conversations in online social media. *WWW'09 - Proceedings of the 18th International World Wide Web Conference*, pages 331–340, 2009. doi: 10.1145/1526709.1526754. 4

[18] M. Cinelli, G. de Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences of the United States of America*, 118(9), 2021. ISSN 10916490. doi: 10.1073/pnas.2023301118. 1

[19] ConvoKit. Reddit Corpus (by subreddit) — convokit 2.4.5 documentation. URL `https://convokit.cornell.edu/documentation/subreddit.html`. 5

[20] J. Cook, D. Nuccitelli, S. A. Green, M. Richardson, B. Winkler, R. Painting, R. Way, P. Jacobs, and A. Skuce. Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, 8(2), 2013. ISSN 17489326. doi: 10.1088/1748-9326/8/2/024024. 1

[21] A. Cunsolo and N. R. Ellis. Ecological grief as a mental health response to climate change-related loss. *Nature Climate Change*, 8(4):275–281, 2018. ISSN 17586798. doi: 10.1038/s41558-018-0092-2. 3, 15, 16

[22] M. Dahiya. A Tool of Conversation: Chatbot. *International Journal of Computer Sciences and Engineering*, 5(5):158–161, 2017. ISSN 2347-2693. URL `http://www.ijcseonline.org/pub{_}paper/27-IJCSE-02149.pdf`. 2

[23] C. Danescu-Niculescu-mizil, M. Gamon, and S. Dumais. Mark my words! Linguistic style accommodation in social media. *Proceedings of the 20th International Conference on World Wide Web, WWW 2011*, pages 745–754, 2011. doi: 10.1145/1963405.1963509. 7, 14

[24] C. Danescu-Niculescu-Mizil, M. Sudhof, J. Dan, J. Leskovec, and C. Potts. A computational approach to politeness with application to social factors. *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 1:250–259, 2013. 4, 7, 12

[25] T. J. Doherty and S. Clayton. The Psychological Impacts of Global Climate Change. *American Psychologist*, 66(4):265–276, 2011. ISSN 0003066X. doi: 10.1037/a0023141. 3, 14

[26] N. Duseja and H. Jhamtani. A Sociolinguistic Study of Online Echo Chambers on Twitter. pages 78–83, 2019. doi: 10.18653/v1/w19-2109. 18

[27] J. Farrell. Politics: Echo chambers and false certainty. *Nature Climate Change*, 5(8):719–720, 2015. ISSN 17586798. doi: 10.1038/nclimate2732. 1

[28] M. Grandini, E. Bagli, and G. Visani. Metrics for Multi-Class Classification: an Overview. pages 1–17, 2020. URL `http://arxiv.org/abs/2008.05756`. 9

[29] R. Grundmann and R. Krishnamurthy. The Discourse of Climate Change: a corpus-based approach. *Critical Approaches to Discourse Analysis across Disciplines*, 4(2):113–133, 2010. 4

[30] K. Haltinner and D. Sarathchandra. Considering attitudinal uncertainty in the climate change skepticism continuum. *Global Environmental Change*, 68(February), 2021. ISSN 09593780. doi: 10.1016/j.gloenvcha.2021.102243. 3

[31] T. Holtgraves and E. Jenkins. Texting and the Language of Everyday Deception. *Discourse Processes*, 57(7):535–550, 2020. ISSN 15326950. doi: 10.1080/0163853X.2019.1711347. URL `https://doi.org/10.1080/0163853X.2019.1711347`. 20

[32] C. C. Howarth and A. G. Sharman. Labeling opinions in the climate debate: A critical review. *Wiley Interdisciplinary Reviews: Climate Change*, 6(2):239–254, 2015. ISSN 17577799. doi: 10.1002/wcc.332. 2

[33] IPCC. Global warming of 1.5°C. Summary for Policymakers. Technical report, 2018. 1

[34] S. M. Jang and P. S. Hart. Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data. *Global Environmental Change*, 32:11–17, 2015. ISSN 09593780. doi: 10.1016/j.gloenvcha.2015.02.010. URL `http://dx.doi.org/10.1016/j.gloenvcha.2015.02.010`. 4

[35] J. Kaiser. Public spheres of skepticism: Climate Skeptics' online comments in the german networked public sphere. *International Journal of Communication*, 11:1661–1682, 2017. ISSN 19328036. 3, 13, 21, 22

[36] J. Kaiser and M. Rhomberg. Questioning the Doubt: Climate Skepticism in German Newspaper Reporting on COP17. *Environmental Communication*, 10(5):556–574, 2016. ISSN 17524040. doi: 10.1080/17524032.2015.1050435. 22

[37] A. Leuschner. Is it appropriate to 'target' inappropriate dissent? on the normative consequences of climate skepticism. *Synthese*, 195(3):1255–1271, 2018. ISSN 15730964. doi: 10.1007/s11229-016-1267-x. 2

[38] M. Lyons, N. D. Aksayli, and G. Brewer. Mental distress and language use: Linguistic analysis of discussion forum posts. *Computers in Human Behavior*, 87(December 2017): 207–211, 2018. ISSN 07475632. doi: 10.1016/j.chb.2018.05.035. URL `https://doi.org/10.1016/j.chb.2018.05.035`. 15, 22

[39] P. Matthews. Why Are people skeptical about climate change? Some insights from blog comments. *Environmental Communication*, 9(2):153–168, 2015. ISSN 17524040. doi: 10.1080/17524032.2014.999694. URL `http://dx.doi.org/10.1080/17524032.2014.999694`. 1

[40] M. R. Mehl, S. D. Gosling, and J. W. Pennebaker. Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90(5):862–877, 2006. ISSN 00223514. doi: 10.1037/0022-3514.90.5.862. 18

[41] T. Neal, K. Sundararajan, A. Fatima, Y. Yan, Y. Xiang, and D. Woodard. Surveying stylometry techniques and applications. *ACM Computing Surveys*, 50(6), 2017. ISSN 15577341. doi: 10.1145/3132039. 7

[42] T. Nguyen, D. Phung, and S. Venkatesh. Analysis of psycholinguistic processes and topics in online autism communities. In *Proceedings - IEEE International Conference on Multimedia and Expo*, 2013. ISBN 9781479900152. doi: 10.1109/ICME.2013.6607615. 18

[43] N. M. D. Niezink. Offline Consequences of Echo Chambers. page 10617, 2017. 1, 18

[44] S. J. O'Neill and M. Boykoff. Climate denier, skeptic, or contrarian? *Proceedings of the National Academy of Sciences of the United States of America*, 107(39):2010, 2010. ISSN 10916490. doi: 10.1073/pnas.1010507107. 2

[45] L. Oswald and J. Bright. How do climate change skeptics engage with opposing views? Understanding mechanisms of social identity and cognitive dissonance in an online forum. 2021. URL `http://arxiv.org/abs/2102.06516`. 1, 2

[46] W. Pearce, K. Holmberg, I. Hellsten, and B. Nerlich. Climate change on twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 report. *PLoS ONE*, 9(4):1–11, 2014. ISSN 19326203. doi: 10.1371/journal.pone.0094785. 1

[47] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer. Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 54:547–577, 2003. ISSN 00664308. doi: 10.1146/annurev.psych.54.101601.145041. 3, 17

[48] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn. The development and psychometric properties of LIWC2015. Technical report, University of Texas at Austin, Austin, TX, 2015. 7

[49] J. V. Petrocelli, Z. L. Tormala, and D. D. Rucker. Unpacking attitude certainty: Attitude clarity and attitude correctness. *Journal of Personality and Social Psychology*, 92(1):30–41, 2007. ISSN 00223514. doi: 10.1037/0022-3514.92.1.30. 14, 22

[50] J. Powell. Scientists Unanimous on Anthropogenic Global Warming in 2019. *Bulletin of Science, Technology and Society*, 39(1-2):3, 2019. ISSN 02704676. doi: 10.1177/0270467620922151. 1

[51] Pushshift.io. Reddit Statistics - pushshift.io. URL `https://pushshift.io/`. 5

[52] Reddit Inc. Information about the world's climate, . URL `https://www.reddit.com/r/climate/`. 6

[53] Reddit Inc. Climate Skeptics: Trying to see through the alarmism, . URL `https://www.reddit.com/r/climateskeptics/`. 6

[54] Reddit Inc. Homepage - Reddit, 2021. URL `https://www.redditinc.com/`. 1

[55] S. S. Rude, E. M. Gortner, and J. W. Pennebaker. Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18(8):1121–1133, 2004. ISSN 02699931. doi: 10.1080/02699930441000030. 3, 4, 15

[56] S. Schweizer, S. Davis, and J. L. Thompson. Changing the conversation about climate change: A theoretical framework for place-based climate change engagement. *Environmental Communication*, 7(1):42–62, 2013. ISSN 17524032. doi: 10.1080/17524032.2012.753634. 4

[57] S. Shugars and N. Beauchamp. Why Keep Arguing? Predicting Engagement in Political Conversations Online. *SAGE Open*, 9(1), 2019. ISSN 21582440. doi: 10.1177/2158244019828850. 4

[58] R. Simmons, P. C. Gordon, and D. L. Chambless. Pronouns in marital interaction. *Psychological science*, 16(12):932–936, 2005. ISSN 0956-7976. URL `http://www.ncbi.nlm.nih.gov/pubmed/16313655`. 4

[59] E. Stamatatos. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, 60(March):538–556, 2009. 7

[60] C. R. Sunstein. *#Republic Divided Democracy in the Age of Social Media.* Princeton University Press, 2017. 1

[61] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1): 24–54, 2010. ISSN 0261927X. doi: 10.1177/0261927X09351676. 3, 14, 15, 22

[62] D. Toniuc and A. Groza. Climebot: An argumentative agent for climate change. *Proceedings - 2017 IEEE 13th International Conference on Intelligent Computer Communication and Processing, ICCP 2017*, (September):63–70, 2017. doi: 10.1109/ICCP.2017.8116984. 2

[63] A. Tyagi, J. Uyheng, and K. M. Carley. Affective Polarization in Online Climate Change Discourse on Twitter. *Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2020*, pages 443–447, 2020. doi: 10.1109/ASONAM49781.2020.9381419. 1

[64] C. W. van Eck, B. C. Mulder, and S. van der Linden. Echo Chamber Effects in the Climate Change Blogosphere. *Environmental Communication*, 15(2):145–152, 2021. ISSN 17524040. doi: 10.1080/17524032.2020.1861048. URL `https://doi.org/10.1080/17524032.2020.1861048`. 1

[65] W. Van Rensburg. Climate Change Scepticism: A Conceptual Re-Evaluation. *SAGE Open*, 5(2), 2015. ISSN 21582440. doi: 10.1177/2158244015579723. URL `https://doi.org/10.1177/2158244015579723`. 2

[66] Yale Program on Climate Change Communication. Yale Climate Opinion Maps 2020 - Yale Program on Climate Change Communication, 2020. URL `https://climatecommunication.yale.edu/visualizations-data/ycom-us/`. 1

[67] M. Yeomans, J. Minson, H. Collins, F. Chen, and F. Gino. Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes*, 160(March):131–148, 2020. ISSN 07495978. doi: 10.1016/j.obhdp.2020.03.011. 4

[68] A. X. Zhang, B. Culbertson, and P. Paritosh. Characterizing online discussion using coarse discourse sequences. *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*, pages 357–366, 2017. 4

[69] J. Zhang, J. P. Chang, C. Danescu-Niculescu-Mizil, L. Dixon, N. Thain, Y. Hua, and D. Taraborelli. Conversations gone awry: Detecting early signs of conversational failure. *arXiv*, pages 1350–1361, 2018. ISSN 23318422. 4

[70] D. Zhao and M. B. Rosson. How and why people Twitter: The role that micro-blogging plays in informal communication at work. *GROUP'09 - Proceedings of the 2009 ACM SIGCHI International Conference on Supporting Group Work*, pages 243–252, 2009. doi: 10.1145/1531674.1531710. 4

Appendices

A Coefficients and p-values of Final Model

Table A.1: Feature coefficients and p-values of Final Model

Features		**Coefficient**	**p-value**
Politeness Strategies (PS)	1. Deference	3.5807	0.000
	2. Gratitude	-11.1348	0.000
	3. Indirect (greeting)	-19.8985	0.024
	4. Positive	-3.1827	0.000
	5. Negative	0.0775	0.813
	6. Please	-7.1605	0.260
	7. Please (start)	2.4550	0.661
	8. Hedges	1.2468	0.368
	9. Indirect (btw)	-0.8036	0.987
	10. Factuality	4.5882	0.001
	11. Apologising	1.3951	0.467
	12. Direct question	-6.1677	0.000
	13. Direct start	-1.1456	0.266
	14. Subjunctive	-7.1526	0.406
	15. Indicative	22.4380	0.136
Psychological Processes (PP)	16. Anger	-0.3348	0.665
	17. Anxiety	9.5519	0.000
	18. Sadness	-2.1293	0.027
Pronouns (Pron)	19. First person singular	-0.5430	0.304
	20. First person plural	-11.5664	0.000
	21. Second person	11.3630	0.000
	22. Third person singular	8.1518	0.000
	23. Third person plural	10.1942	0.000
	24. Impersonal pronouns	1.3467	0.000

Table A.1: (continued).

Features		**Coefficient**	***p*-value**
Style (S)	25. Articles	3.0817	0.000
	26. Certainty	9.7653	0.000
	27. Conjunctions	0.9129	0.028
	28. Discrepancy	-4.3068	0.000
	29. Negations	0.2134	0.660
	30. Prepositions	0.6335	0.009
	31. Quantifiers	-0.4625	0.321
	32. Tentative	2.3677	0.000
	33. Insights	2.5035	0.000
Time Orientation (TO)	34. Past focus	0.0025	0.988
	35. Present focus	-1.3823	0.000
	36. Future focus	-7.1928	0.000
Informal Language (IL)	37. Swear words	-5.9644	0.000
	38. Netspeak	-6.8968	0.000
	39. Assent	-2.3588	0.003
	40. Nonfluencies	6.9757	0.000
	41. Fillers	4.5303	0.411
Other Style (OS)	42. Auxiliary verbs	2.6803	0.000
	43. Common Adverbs	-1.5087	0.000
	44. Comparisons	-0.2238	0.627
	45. Interrogatives	6.1136	0.000
	46. Numbers	2.3876	0.001
	47. Causation	-0.8746	0.029