

# Sentiment Analysis of Twitter Discourse in the Context of UN Climate Conferences

James Hughes, Ameer A Saleem, Vignesh Sridhar, Holly Taylor.  
Supervised by Janique Krasnowska and Fernando A Zepeda

17 March 2023

## **Abstract**

The following document entails a student-led end-to-end data science project, the goal of which being to investigate the changes in perceptions towards climate change over time. The project involved a review of the current literature and available data, sentiment analysis, neural network model training and evaluation of the developed model, as well as projections for the future.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Literature review and strategy</b>	<b>3</b>
2.1	Key literature and implications . . . . .	3
2.2	Scraping Tweets . . . . .	4
2.3	COP Conferences . . . . .	4
2.4	Word Embeddings . . . . .	4
<b>3</b>	<b>Data Analysis</b>	<b>4</b>
3.1	Approach . . . . .	4
3.2	COP24 . . . . .	5
3.3	COP25 . . . . .	5
3.4	COP26 . . . . .	5
3.5	Results . . . . .	5
<b>4</b>	<b>The Model</b>	<b>5</b>
4.1	Training . . . . .	5
4.2	Evaluation . . . . .	5
<b>5</b>	<b>Conclusion</b>	<b>5</b>
5.1	Key takeaways . . . . .	5
5.2	What next? . . . . .	5
<b>6</b>	<b>References</b>	<b>5</b>

# 1 Introduction

We discussed many different avenues for research and project focus initially. Below are detailed a few of these ideas and their resolutions:

- Analyse pledges made by political leaders and conduct some data-backed investigation into how much they keep to their word.  
**Pros:** Plenty of data available, can be investigated in many different scopes (e.g. national vs global, a few select politics vs advocates for a particular party, etc.).  
**Cons:** The language used by such politicians is constrained, which could make it difficult to train a model. Also, it would be difficult to associate a metric with keeping promises, as this could be subjective and it is unlikely that a database encapsulating this with objective labels exists. The scope of individuals to investigate is also somewhat constrained, regardless of the aforementioned scope.
- Investigate the changes of attitudes of the public towards climate change following different climate summits.  
**Pros:** A genuinely interesting question, the implications of which could be used to forecast how attention on climate change could evolve through the years.  
**Cons:** Very ambitious scope for a student-led project- the title needs refined in order for the problem to be well-defined.

We decided to amalgamate the above two project briefs to brainstorm a few more specific and achievable outcomes:

1. Investigate how the perceptions between groups of ‘notable figures’<sup>1</sup> regarding climate change have evolved over the past decade.
2. Investigating the short-term shifts in opinions following individual climate change summits.
3. Investigating the long-term consequences of these summits across the years.

From here, we conducted a literature review to ascertain which of these avenues would be the most fruitful to pursue.

## 2 Literature review and strategy

### 2.1 Key literature and implications

[Insert information about the literature on the NLP Teams Channel]

---

<sup>1</sup>We will later elaborate on what we mean by this.

## 2.2 Scraping Tweets

The literature suggested that online tweets would prove useful as an abundant data source for the project. While Twitter has an API for scraping tweets, we settled on the use of `snsrape`, which appeared easier to implement.

Twitter is a prevalent platform with immense usage, meaning that scraping within an arbitrary timeframe will yield a lot of irrelevant data. It makes reasonable sense to focus on time intervals before, during and after COP summits; this is the time when most ‘notable figures’ would be discussing climate change and voicing their respective opinions on it the most. This also extends to the public, providing us a greater density of climate-related data. Since Twitter does not allow making the text of tweets public, any Twitter-related dataset must undergo a process called hydration.

## 2.3 COP Conferences

COP, standing for ‘Conference Of the Parties’, is the main decision-making body regarding any major decisions regarding climate change. COP meetings are organised by the United Nations and involve almost every country in the world. The goal of such summits is to discuss what major governmental authorities can achieve in the fight against climate change. COP meetings usually take place every year and aim to review the global progress made in mitigating the impact of climate change. The first COP summit took place in Berlin in March 1995. At the time of writing, the most recent COP summit, COP 27, took place in Egypt in November 2022.

## 2.4 Word Embeddings

# 3 Data Analysis

## 3.1 Approach

We began by focussing our attention on dates close to/during COP summits. Due to how much usage Twitter gets nowadays, we were forced to scrape tweets containing the string ‘#COP-’ within a time frame of a single day at a time, yielding tens of thousands of tweets in one go. Our goal at this stage was to break down the composition of these tweets, with the aim of gaining insights into the most commonly-used words pertaining to climate change and trying to spot any general patterns within the tweets of each scrape.

We experimented with a few different analysis methods to help give us a better understanding of the underlying patterns behind the data. We began with K-means clustering, allowing the computer to try identifying categories on its own.

### **3.2 COP24**

### **3.3 COP25**

### **3.4 COP26**

### **3.5 Results**

## **4 The Model**

We implemented a neural network model in Tensorflow to learn the patterns of label associations with the tweets.

### **4.1 Training**

In order to train the model, we required a dataset with labels attributed to each tweet. This could be done by a computer with clustering, but defeats the purpose of model training, since we would have no feasible way of validating so many rows of labelled data. There exists a Climate Change Twitter Dataset that, among many other things, attributes a value in the interval  $[-1, 1]$  with 1 denoting positive sentiment,  $-1$  denoting negative sentiment and 0 indicating a neutral stance on the topic. We decided on using this dataset [cite properly later] consisting of 1.6 million tweets consisting across 15 different languages. We started with just the tweets written in English, with the project potentially extending to tweets in other languages if time allowed. Note that the tweets in these datasets also require hydration to use.

### **4.2 Evaluation**

## **5 Conclusion**

### **5.1 Key takeaways**

### **5.2 What next?**

## **6 References**