

NLP And Classification of UN Climate Change 2021 Glasgow Summit Tweets

Question/need:

- What is the framing question of your analysis, or the purpose of the model/system you plan to build?
 - **What are the hottest topics online on twitter in regards of the COP26 which is the United Nations Climate Change 2021 in Glasgow.**
- Who benefits from exploring this question or building this model/system?
 - **Understanding the peoples mental understanding and conclusions from the conference and maybe something to study by the governments, realizing the current standing of people on twitter or their country.**

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?
 - **Twitter website by using the APIs from the Twitter developer account.**
- What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?
 - **Using Tweepy to grab tweets surrounding #COP26 #UNCC21 #Climatechange and grabbing rows and columns in regards of tweets. Main column being the text on the tweets.**
- If modeling, what will you predict as your target?
 - **Topic Modeling**
 - **Clustering**
 - **Classification?**

Tools:

- How do you intend to meet the tools requirement of the project?
 - **Pandas, numpy, seaborn ,NLTK, spaCy, sklearn , and Tableau and google sheets**
- Are you planning in advance to need or use additional tools beyond those required?
 - **Maybe, but unsure still.**

MVP Goal:

- What would a minimum viable product (MVP) look like for this project?
 - **MVP: Provide a cluster and classification of the topics of tweets surrounding the UN Climate Change 21 Glasgow Summit .**