Final Project Proposal for PPOL670

Shiying Wang

4/17/2020

Project Option

Option 2: Individual Project

Question of Interest

Using supervised machine learning model to solve the NLP classification problem:
Which Sustainable Development Goals(SDG) are linked in the country's National Determined Contributions(NDC) document?

Keywords

Sustainable Development Goals, National Determined Contributions, Text Analysis, NLP, Classification

Data Source

1. Nationally Determined Contributions submissions: NDC Registry, UNFCCC.
2. Sustainable Development Goals: Sustainable Development Goals Knowledge Platform, United Nations
3. NDC-SDG linkages analysis data: Climate Watch, World Resources Institute.

Summary

The idea of this project comes from my work at the World Resources Institute. The NDC-SDG linkages module on our Climate Watch data platform aims to identify potential alignment between the targets, actions, policy measures and plans in countries' Nationally Determined Contributions (NDCs) and the targets of the Sustainable Development Goals (SDGs). Currently we are doing the analysis manually by going through the full NDC document, finding specific targets, actions or policies, then tagging that text to a specific sustainable development goals and targets. (on our website the goals and targets are tagged to the html file)

However, sometimes there can be an intensive submission of NDC which we may get more than 5 new submissions per day (for example around earth day), and some NDC can be around 50 or even 100 pages. So it could be hard for us to do the rapid analysis and review for all of them. A classification model which could help us to pre-identify the text to possible sustainable development goals and targets will be helpful.

In our current NDC-SDG linkages database, there are already 9696 observations coming from the analysis on over 190 countries NDC submissions. For each row we have the data for the specific NDC text, corresponding sustainable development goals (from Goal 1 - 17), corresponding targets under that SDG (such as Target 1.1, 1.2. . . 17.1, 17.2. . .), sectors, country ISO and other informations. I would like to use some supervised machine learning classification model to generate the possible goals and targets for a specific text. Wordcloud

graphs can also be produced to show the trending word for each SDG. What's more, more in-depth analysis which group countries by political regions/income levels could also be interesting to show the different trends of climate actinos in different countries.