POP77142 Assignment 2: Text Analysis

Before Submission

- Make sure that you can run all cells without errors
- You can do it by clicking `Kernel`, `Restart & Run All` in the menu above
- Make sure that you save the output by pressing Command+S / CTRL+S
- Rename the file from `02_assignment.ipynb` to `02_lastname_firstname_studentnumber.ipynb`
- Use Firefox browser for submitting your Jupyter notebook on Blackboard.

Overview

In this assignment you will need to analyse the debates of the 33rd session of the Dáil Éireann (Irish Parliament) that was in sitting between 2020 and 2024. The complete debate records for that session are available on Blackboard as a compressed CSV file. Do note that the dataset is quite large , it contains ~600K individual speeches and takes about 0.5GB of disk space when uncompressed.

The dataset is structured as follows:

dail	vol	no	date	speaker_name	speaker_role	constituency	party	text

where:

`dail` - is the number of the Dáil (e.g. 33rd Dáil)

`vol` - is the volume number of the debates (e.g. 1000)

`no` - is the number of the debate in the volume (e.g. 1)

`date` - is the date of the debate (in YYYY-MM-DD form, e.g. 2020-01-01)

`speaker_name` - is the name of the speaker

`speaker_role` - is the role of the speaker (e.g. TD, Minister, etc.)

`constituency` - is the constituency of the speaker

`party` - is the party of the speaker

`text` - is the text of the speech

Note that some of the texts belong to the outside speakers, such as, e.g. external experts, witnesses, etc. Another aspect of this data to keep in mind is that some of the recorded speeches are in Irish. You can choose to use those in your analysis or exclude them.

Part 1: Modelling Topics

In this part of the assignment you will need to model the topics of the speeches in the Dáil. You can use any method that you think is most appropriate for this task. You can adopt any of the other number of avenues: dictionary methods, topic modelling, supervised learning, LLMs. You can also choose to use the metadata in the dataset to inform your analysis.

Part 2: Modelling Ideology

In this part of the assignment you will need to model the ideology of the speakers in the Dáil. There could be a number of ways to tackle this problem, from more traditional methods, such as, e.g. dictionary-based approaches, to more advanced methods, such as, e.g. supervised learning and LLMs. You are free to choose the method that you think is most appropriate for this task.