Group 4

Class: ECS 171

Professor Rafatirad

Group Members:

Nicole Lazovsky, Michael Rimboim, Raymond Luciano Iacobacci, Sanskriti Jain, Eric Liu, Jiu Choe

Note:

Most of the files in this repository are Google Co-Lab files.

To run them on your machine:

1. Open Google Colab editing environment in browser

2. Open new file, you will be prompted to either: open from computer, google drive, or github -> choose github

3. Enter the following: mrimboim/ECS171GRP2

4. Click open

5. Run all cells

6. In the 3rd codeblock in each file, you will be prompted to upload a file. Upload "annotations.xlsx" (also found in the directory).

7. Allow the program to finish running

To run the FastText models:

In order to run the FastText models locally, you will need to install the FastText vectorizer tool onto your machine. To do so, enter this link https://fasttext.cc/docs/en/crawl-vectors.html and follow the installation instructions. Unfortunately, we could not upload the files into our repository as they are too heavy. Remember to adjust the pathways to match the ones on your machine in the following line:

ftmodel = fasttext.load\_model("fastText/cc.en.300.bin")

Instructions for navigating this repo:

1) The "Experimental Models" directory:

Contains all the models we have developed in the process of finding the optimal model for our task.

There, you will find many models that use GLOVE vectorization with neural nets of varying complexity.

2) The "Final Models" directory:

Contains:

a) tfidf-NaiveBayes Prototype

This is our TF-IDF final model version.

Contains the actual model as well as the tests of its ability to capture linguistic nuance.

b) FastText\_Pre\_Improvement

This is our FastText model prior to the architectural and hyperparameter optimizations and prior to the regularization.

This version is discussed in more detail in the project report.

c) FastText\_Pre\_Regularization

This is our FastText model after the architectural and hyperparameter optimizations and prior to the regularization.

This version is discussed in more detail in the project report.

d) FastText\_Final\_Version

This is our final FastText model, after architectural and hyperparameter optimizations and after the regularization.

This version is the final model for our project, for all intents and purposes.

This version contains the paragraph generalization.

e) Other text files containing the articles used to test the praragraph generalization.

3) The "Frontend" directory:

Contains the "Frontend" file, this file contains the Tf-IDF model (as this is the model we choose to apply to the front end)

This file also contains the commands needed to start the front end server and features.

Instructions to use:

1) Run the file using the instructions provided above.

2) Open the front end using the following link:

https://byh4ibmy52x4hglp.anvil.app/5DS5KSTNOAM4O2C3L322DXC7

3) Enter your test sentence or input and click the "Detect Bias" button

Here are the test sentences used in the demo for convenience:

The transgender effort to suppress any recognition that men and women are different and complementary would not matter except for the movement’s political alliance with wealthy progressives and radical feminists who wish to destroy the political power of the male-and-female family.

Radical Virginia Citizens Defense League has organized a rally in Richmond for January 20, and the event is already attracting out-of-state pro-gun folks, including a potential horde of white nationalists.

4) The “Experimental Visualization” File

Contains our dataset exploration and the graphs discussed in the dataset analysis section of the report.