**Steps after extracting all videos and keeping unique videos**

1. Labelling.R
   1. LOCATION: 4\_Politics/python3\_script/data/Labelling.R
   2. NEED: file with videos
   3. CREATE: 20200504-193926\_joe\_biden\_nlp.csv

All notebooks below can be found at this path: 4\_Politics/python3\_script/data/nlp/notebooks

1. Notebook: Extract\_subtitles.ipynb
   1. USE: Jupyter Notebook
   2. NEED: 20200504-193926\_joe\_biden\_nlp.csv
   3. CREATE: 20200504-193926\_joe\_biden\_nlp\_subs.csv
   4. CREATE: 20200504-193926\_joe\_biden\_nlp\_subs\_clean.csv
   5. CREATE: 20200504-193926\_joe\_biden\_nlp\_subs\_clean\_expanded.csv
2. Notebook: nlp\_EDA\_v3.ipynb
   1. USE: Jupyter Notebook
   2. NEED: 20200504-193926\_joe\_biden\_nlp\_subs\_clean.csv
3. Notebook: nlp\_EDA\_v4.ipynb
   1. USE: Jupyter Notebook
   2. NEED: 20200504-193926\_joe\_biden\_nlp\_subs.csv
4. Notebook: nlp\_model\_training.ipynb
   1. USE: Google Colab
   2. NEED: 20200504-193926\_joe\_biden\_nlp\_subs\_clean.csv
   3. COMMENT: good accuracy, but overfitting.
5. Notebook: nlp\_model\_training\_v2.ipynb
   1. USE: Google Colab
   2. NEED: 20200504-193926\_joe\_biden\_nlp\_subs\_clean\_expanded.csv
   3. COMMENT: Better accuracy, less overfitting.

**How to use Google Colab**

* Files need to be in a google drive. I placed my files in the folder “Colab Notebooks”.
* Colab uses cloud and is a lot faster in training models than your laptop.
* This code is needed to use colab effectively:

from google.colab import drive  
drive.mount('/content/gdrive')

#move current directory to file path where your notebooks and files are placed.

%cd "gdrive/My Drive/Colab Notebooks"