MEETING NOTES JAN 26 2025

Ryan: The current alg only gives us Partisan/or not results, we need to make a classifier to somehow turn it into left right or neutral.

0-shot classification: takes in a LLM (Bart iterates every row and tries to reclassify the dataset into the 3 classes we want to observe, but requires large processing power)

+ We also have 3-4 rows in our current dataset that have nothing for article text. Alg will not run properly because of this.
+ Confidence scores : [0,1].

We go over the code with Ryan:
+ Putting placeholder for US and replacing it back again after lowercasing.
+ Otherwise, the normal cleaning process is just like last time.
+ Spellcheck package (fasttext) (commented out bc it wasn't working)
+ Uses SVM to classify data (source code's algorithm)
+ Pre-ML EDA:
    - Word popularity
    - Unique words
+ ML:
    - PCA for classification, then Ryan tries to use the 0 shot thing to reclassify, his computer couldn't run it for too long

Jonathan gives an example of Roberta:
+ Just raw Roberta text.

So far, we haven't made adjustments to the ML model.