Daily Log

Monday October 28

Continued to fetch transcripts from both the code and the demo site, updating the spreadsheet dataset.

Wednesday October 30

Continued to fetch transcripts from both the code and the demo site, updating the spreadsheet dataset. Also continued exploring and getting acquainted with the various documentations of the following APIs for text summarization, the next major task in our project: https://deepai.org/machine-learning-model/summarization, https://www.meaningcloud.com/products/automatic-summarization, https://aylien.com/text-api/summarization/, https://aylien.com/text-api/summarization/.

Friday November 1

Continued to fetch transcripts from both the code and the demo site, updating the spreadsheet dataset. In addition, continued looking into the Mellin and discrete cosine transform algorithms for audio processing for the future to compare preprocessing effects on accruacy and hopefully have even more accurate models working if our proof of concepts first works with text summarization as well.

Friday November 8

Continued to fetch transcripts from both the code and the demo site, updating the spreadsheet dataset. Started trying to implement some basic functions from the deepAI and meaningcloud text summarizations above in Python, resolving some bugs with importants and model parameterization and usage. Also continued looking for other APIs on IBM Watson's list and elsewhere online.

Timeline

Date	Goal	Met
October 25th	Start researching for initial parts of the video transcript summarization and in tandem, continue to build the dataset over time	Yes, found free API's for text summarization and explored their documentations and looked more into audio processing algorithms for pre-processing/normalization for results comparision in the future as we try to optimize the algorithm
November 1st	Decide on a text summarization API and start to implement code/API requests on a test TED transcript I have in the dataset and continue to scale dataset in parallel	Yes, decided to initially start with the deepAI and meaningcloud APIs and still continuing to look for others in tandem while also continuing to scale the dataset with transcripts from IBM Watson and the demo site
November 8th	Discuss with Mr. White whether need to go about implementing Mellin and Discrete Cosine Transform or any other audio processing algorithms; checkpoint: have a successful summarization working and determine what next steps are if any issues arise; continue to scale dataset in parallel	No, haven't discussed with Mr. White yet, will actually do so later on when we get to comparing results of our models' preprocessing once our initial proof-of-concept is working for text summarization as well.
November 15th	Continue to research text summarization APIs and algorithms, continue to scale transcripts dataset in parallel	
November 22th	Start implementation of algorithm utilizing decided upon APIs's various methods and comparing results, continue to scale transcripts dataset in parallel	

Reflection

Starting this text summarization component of our project has been fun as it's fresh and new challenges to resolve and APIs to explore. Hopefully I can rather quickly find a robust API as we eventually did with IBM Watson for speech-to-text and start making queries using our transcripts dataset as input. I'll continue to scale the dataset in tandem as well, of course. Looking forward to another great, productive quarter ahead!