

CSE 538 Project Proposal

Team Members:

- Carwyn Collinsworth
- Michael Arkhangelskiy 112119654
- Michael McNeill 112338842

Train a classifier to discriminate on whether the input sentence is from the transcript of a professional speaker (think ted talks or presidential candidate or political figure <https://www.kaggle.com/code/rounakbanik/ted-data-analysis>) or from the average person speaking (we could find a dataset of people asking questions to speakers or just casual conversations). Then we can use a pre-implemented explainable-AI package such as LRP to see which input words (or phrases) seemed to have the biggest impact on perceived 'authority' or professionalism. An example of the output we would get is at the bottom of this colab file https://github.com/yidinghao/interpreting-nlp/blob/master/examples/bert_sst/test_bert_sentiment.ipynb where each word is highlighted by a color representing its importance to the classification.

Notes from meeting:

- Gradient based methods
- Attention is not interpretation
- Look up 2 more methods
- How do you come up with patterns that professional speakers use
- Measure requirements of good public speaking with our datasets
- Compare that with what the model picks up

Previous work & Background:

Previously

Project Detail:

The goal of this project is to

Implementation:

Idk Python probably

Literature Survey:

Papers or something?

