Context classification of using Speech recognition

Team: Subhajit Chatterjee (SXC190070), Sai Sreekar Reddy (SXS190008)

Introduction:

Presently, a lot of work in predictive analysis domain is focused on analyzing natural language, i.e. text, audio and video data. Amongst these, speech recognition is a vastly explored field with numerous applications, ranging from voice-based control of smart appliances, voice-based querying, etc.

Besides, analyzing the speech audio content and the quality can help classify the context of the same, e.g. what topic does the speech pertains to, what is tonality of the speaker, etc.

Classifications like

- Hate Speech? – Y/N
- Suffering from anxiety/depression? – Y/N
- A fraudulent caller? – Y/N

And many others can be achieved which can be used in solving many practical problems

Target: Classify the context of a speech based on the content and quality of the sound

Intended Methodology:

1. Gather speech audio content based on the classification intended to be achieved (Need to discuss on this). Source: YouTube (Interviews, podcasts, etc.), self-recorded content

2 a). Analyzing content:

- Using the speech recognition APIs (google, Bing, sphinx, etc.) provided through python speech_recognition library to convert speech to text.
- Convert text to a corpus -> remove filler words-> tokenize in n-grams-> create document-term-matrix-> get term frequency-inverse document frequency for keywords-> vectorized text
- Create labels (e.g. Hate Speech/ Not) against each of this vectorized examples

b) Analyzing audio quality:

- Extract audio features (python librosa project provides for getting waveform output from audio input)
- Retrieve features pertaining to audio frequency, loudness, rhythm, etc. (studying more on this)
- Use these features in the data along with vectorized text to create final training data

3-Train classifiers on this data

- Try Decision Tree, Ensemble, SVM, ANN algorithms (Need suggestion on this)

4-Predict the context of new audio input