**Speech classification**

**Introduction:**

This project deals with the speech classification of human voice into interrogative, exclamatory and declarative types of sentences using deep learning.

**Technologies Used:**

* Python3, ffmpeg, librosa, NumPy, keras

**Input:**

Our input is confined to us accent, where in the speech is extracted from the videos with the help of subtitles(.srt file) using time stamps.

This extracted mp3 files are converted to wav files which inturn got converted into numpy array using ffmpeg library.

This input array is of size 13x48 which is fed as an input array to the neural network made using keras.

**Model:**

Initially we used RNN model for this classification and we got accuracy of 60% which is the network is not getting trained.

Then we modified the network using Pooling and we used a stride of size 3x3 but the accuracy is a bit better.

Then we tries using CNN and we could get the accuracy of about 90%, which means the network is learning something and could give some better results.

**Future plan:**

We expected that data agumentation might be giving better results than previous case and we are working on to stretch, add noise and fasten the speed of audio and serve this as input and try to check the accuracy.

And also use RECORDED VOICE(SPEECH) as input to the network.(this might be the better choice).