



Report 2nd Phase

Aditya Arvind

203230006

Mtech System and Control

Phase 2 report

In Phase 2 I tried various speech recognition API like Google Speech Recognition, Vosk API, Deep Speech, python libraries like Speech Recognition Library. We used moviepy library to extract audio from video then we used the speech recognition library to get text from the audio. Libraries like pyaudio were also used.

The problem we faced was that the API's and the libraries were having problems understanding Indian accent and technical jargon terms like XGboost were not properly identified. There was also a problem of noise. Noise in the video severely affected the text quality I got after audio to text conversion. Thus it was decided that we will further train the model from Vosk API using transfer learning. Thus currently I am working on transfer learning.