Wake Word Modeling

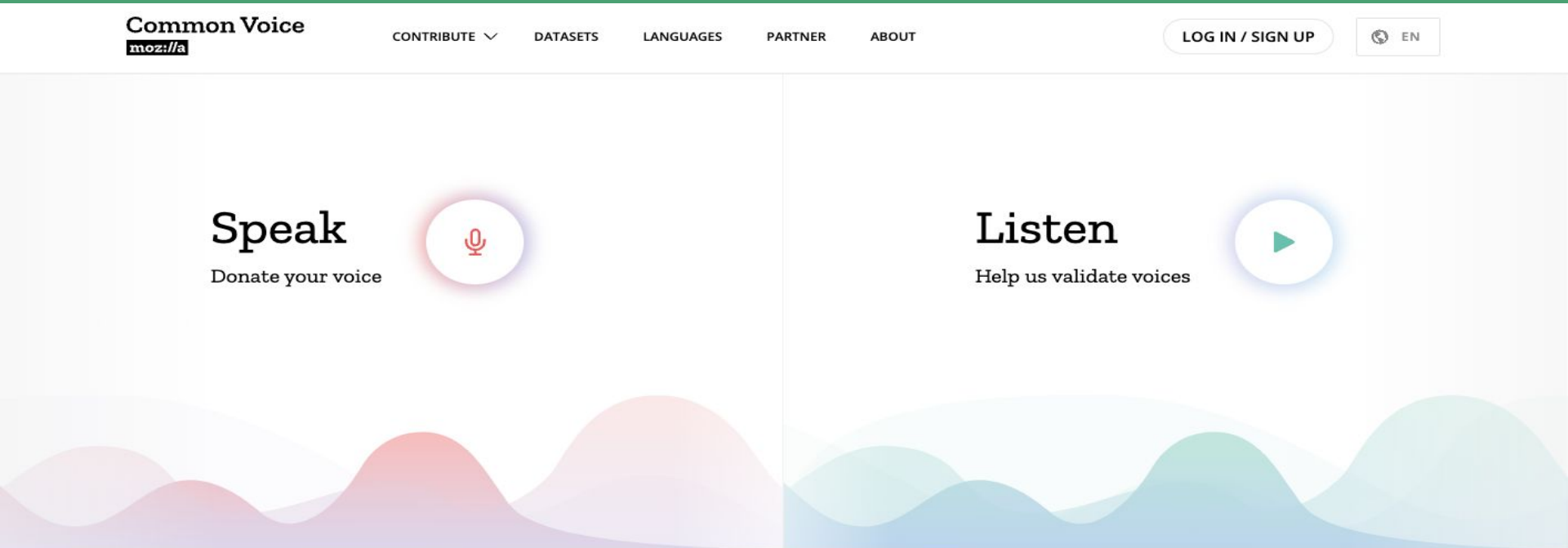
By Ricky Everest

Overview

- Context
- Data
- EDA
- Features
- Models

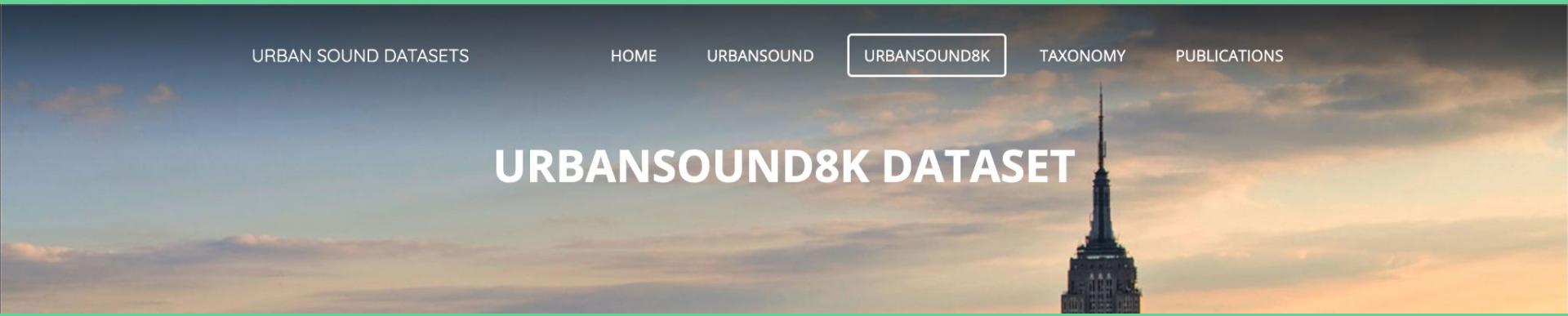
DATA 2000+ Common Voice Samples

- Random Comments
- Wide Range Of Mic Quality



DATA 800+ Urban Sound Samples

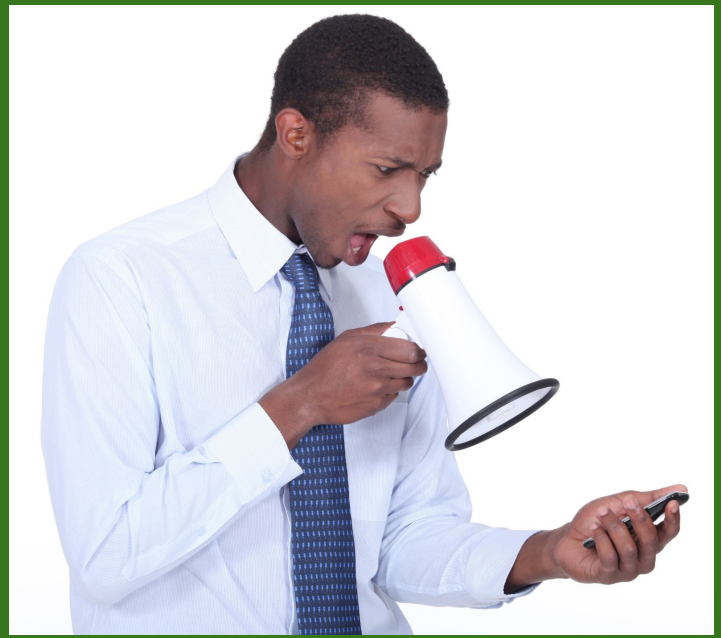
- Environmental noise
- City noise
- City animal noise



DATA

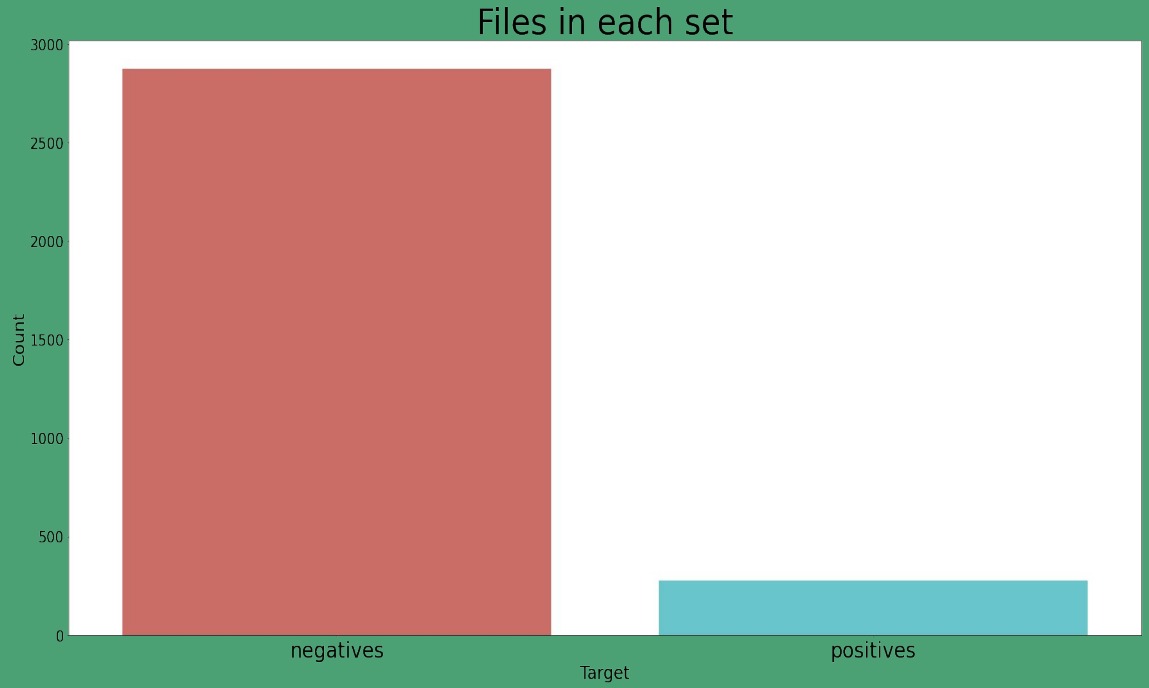
200 + Wake Word Samples(initial)

PHONE



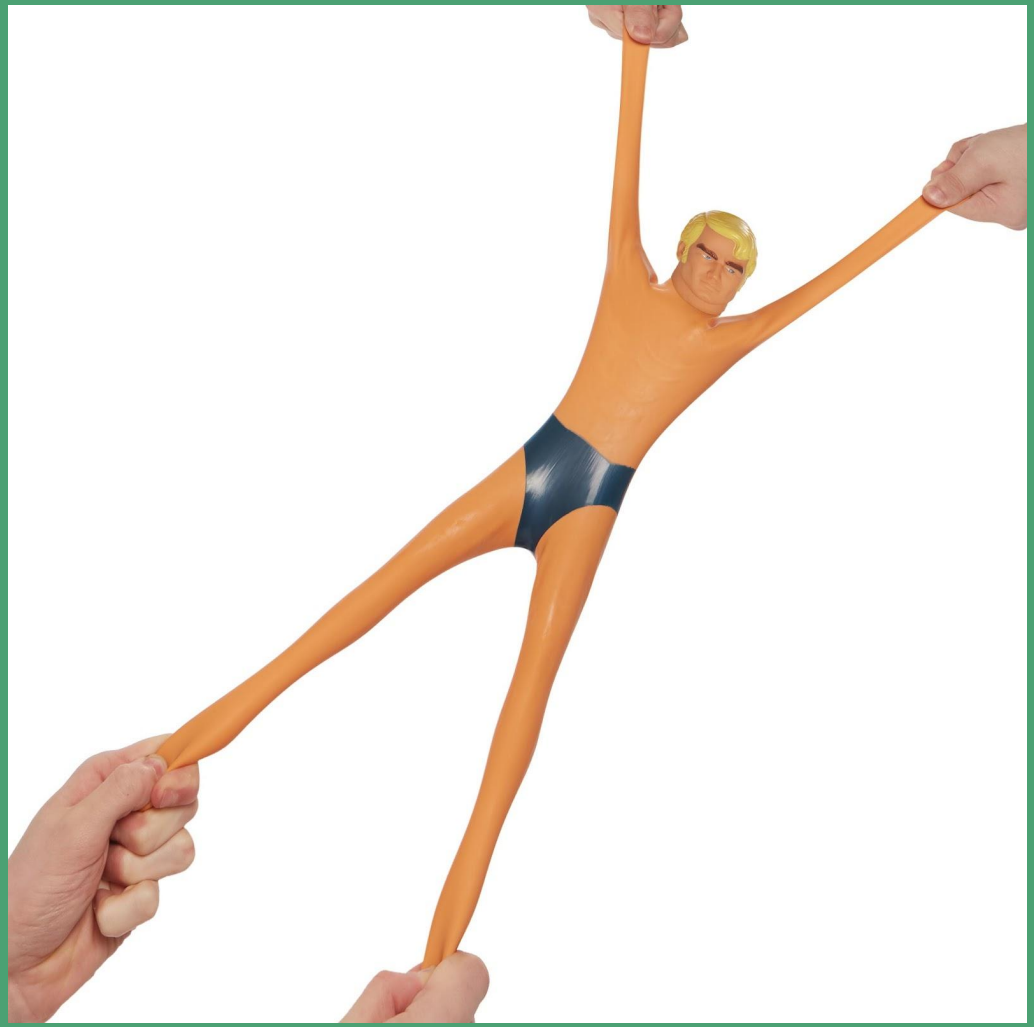
PYAUDIO(high quality) AND
SOUND DEVICE LIBRARY(sounded terrible)

EDA + Preprocessing



EDA + Preprocessing

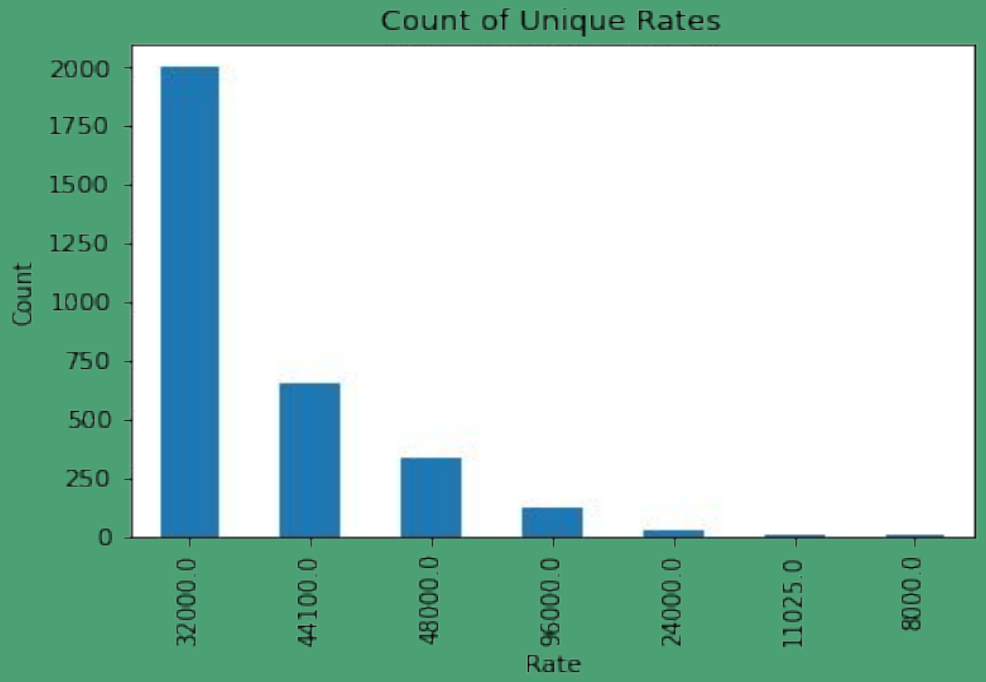
- Pyrubberband
- Librosa
- Changed Pitch
- Kept Audio Length



EDA + Preprocessing

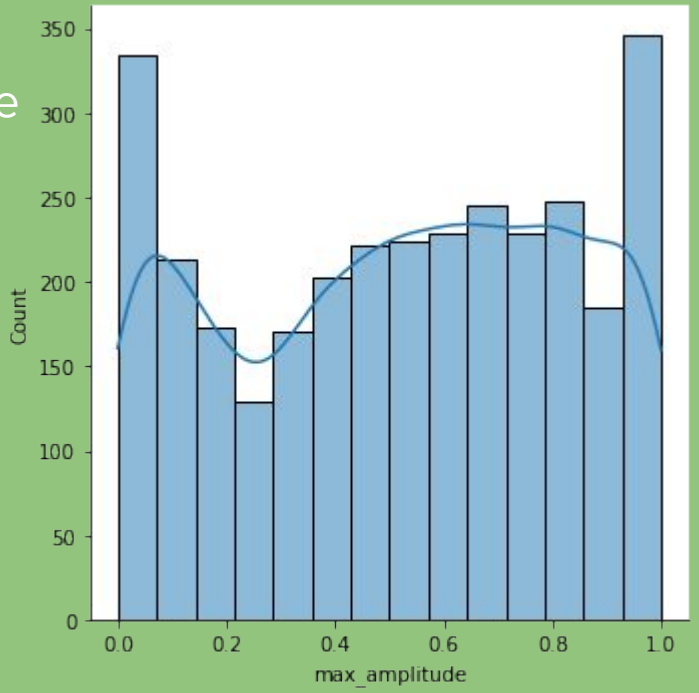
EDA + Preprocessing

- Sample Rate
- Like Hi or Low Def.

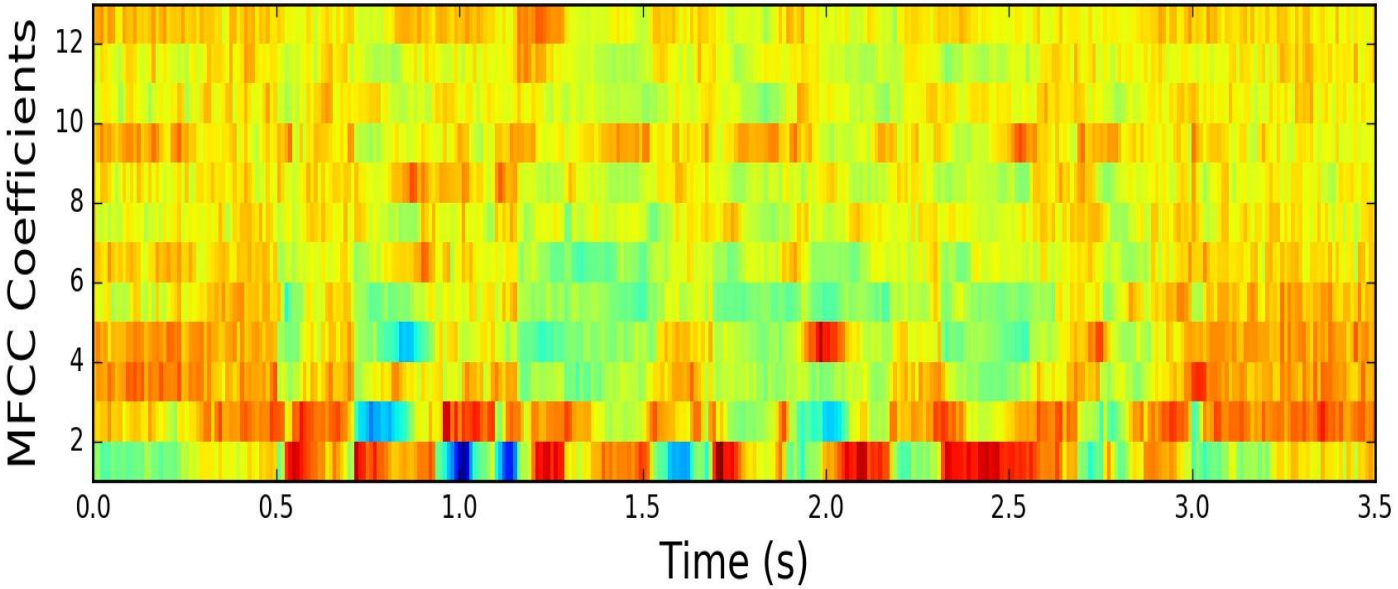


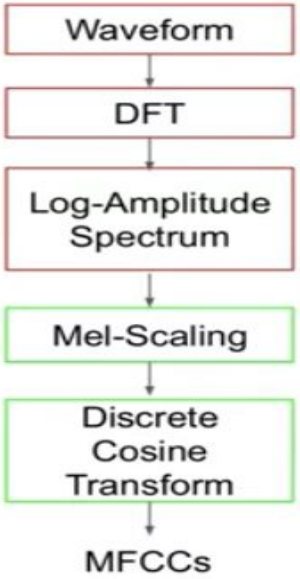
EDA + Preprocessing

- Decibels
- Some clips not audible



MFCCs

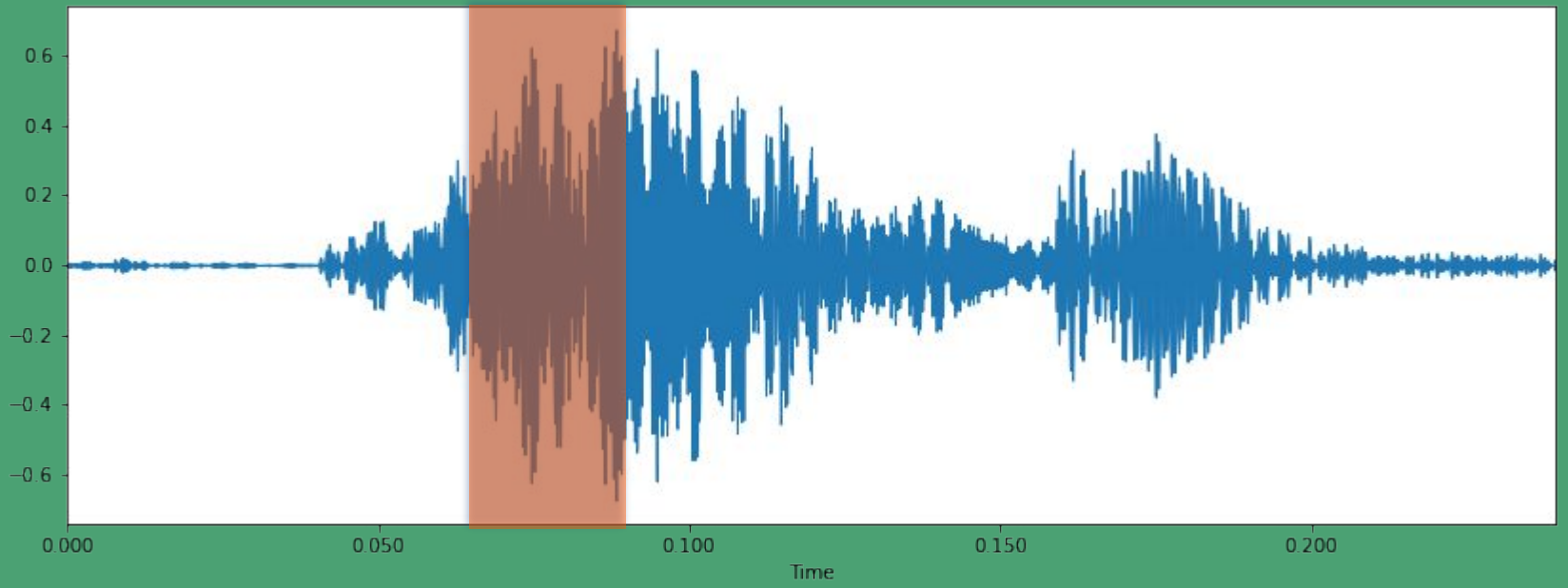




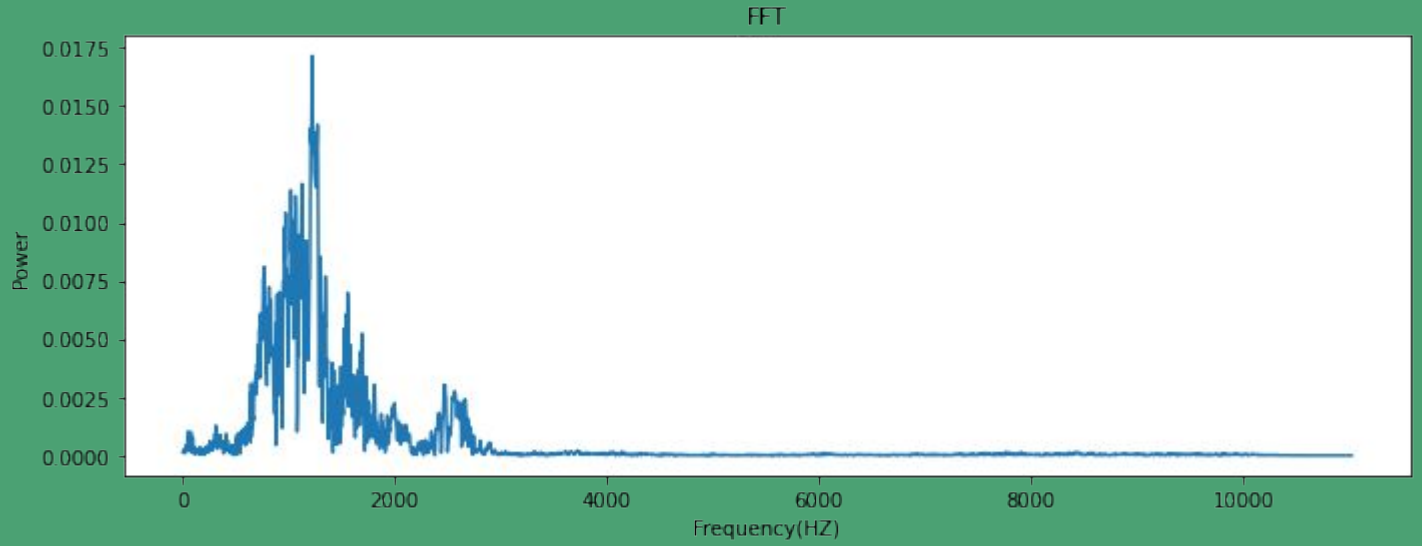
Borrowed from https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html

Feature Extraction

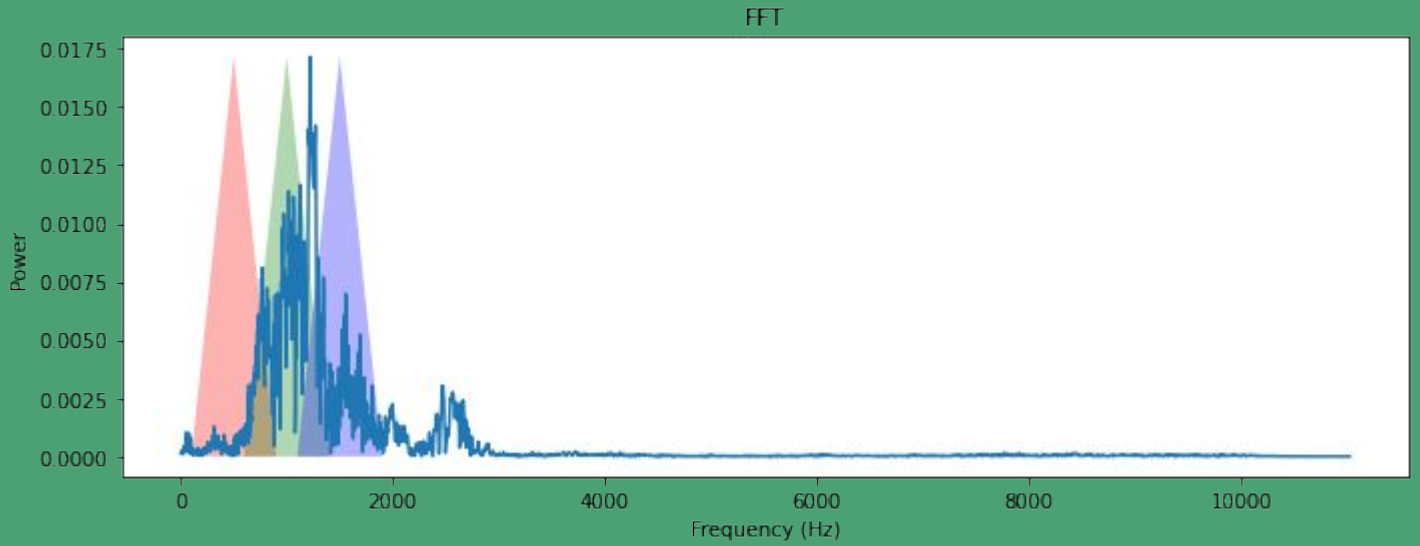
Signal WAVES



FFT

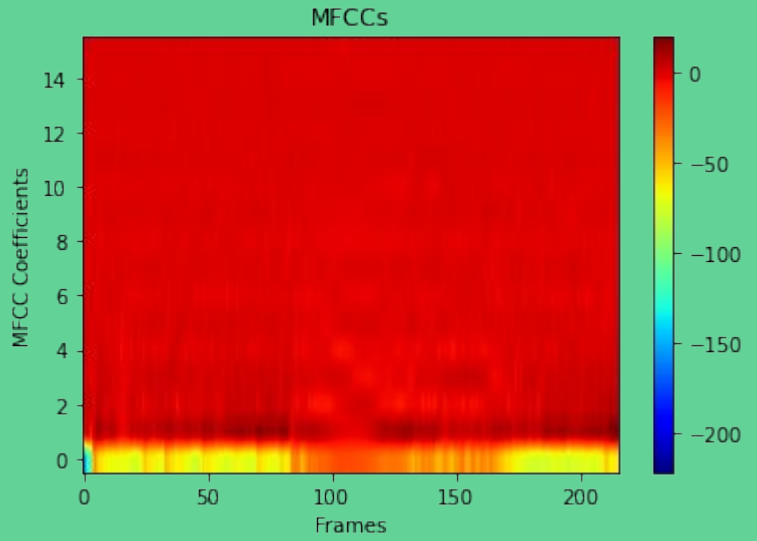

FFT
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
Power
0
2000
4000
6000
8000
10000
Frequency(HZ)

Mel Filters


FFT
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
Power
0
2000
4000
6000
8000
10000
Frequency (Hz)

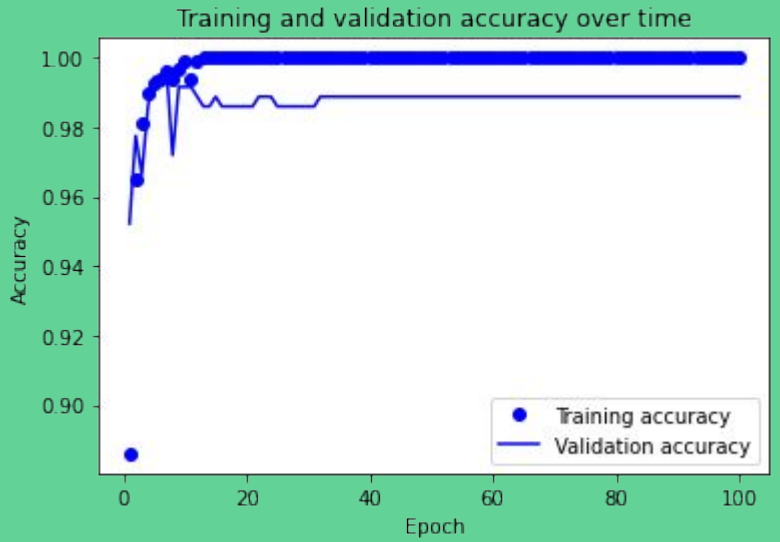
MFCC PARAMS

- Window Length = 25 ms
- Window Step = 1 ms
  - creates small overlapping frames
- NFFT or Number of Fast Fourier Transform points 512
- Coefficients = 16

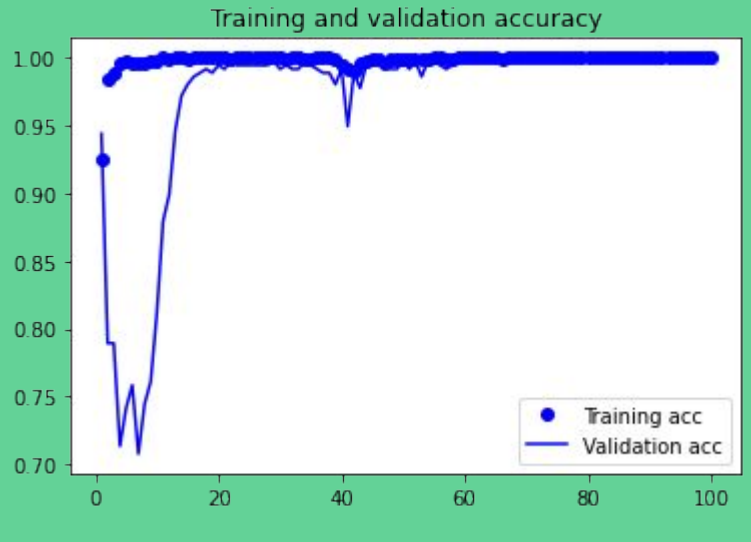


Initial Models - Insufficient Data

RNN Performance

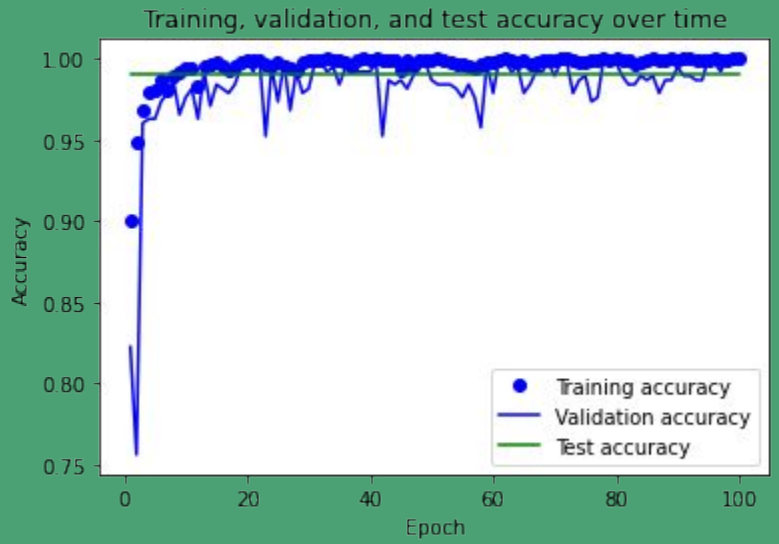


CNN Performance



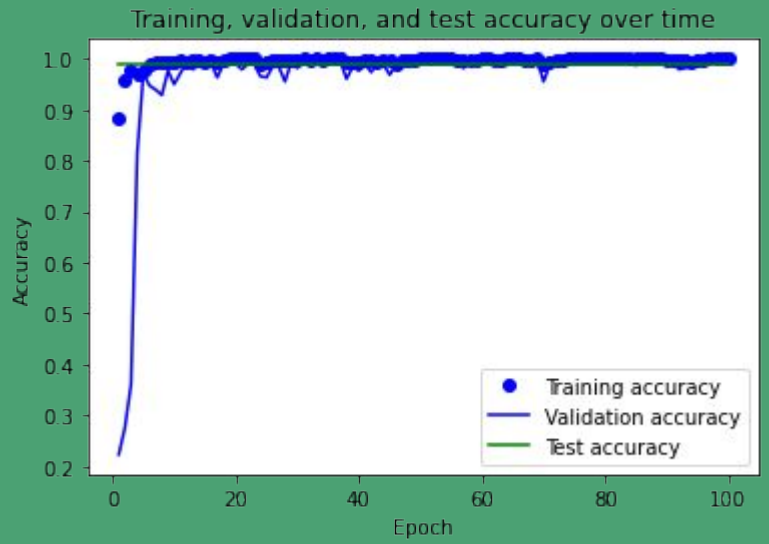
Models with added data - still not enough data

CNN Performance



Final validation accuracy: 100.00%
Final test accuracy: 99.07%

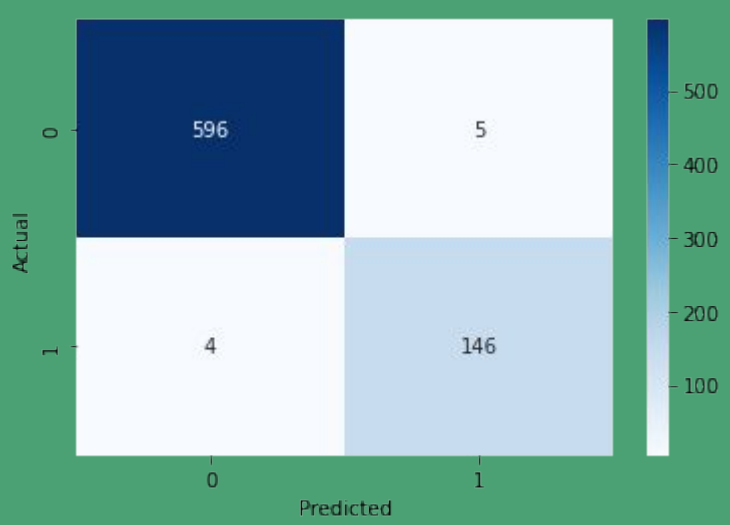
RNN CNN COMBO



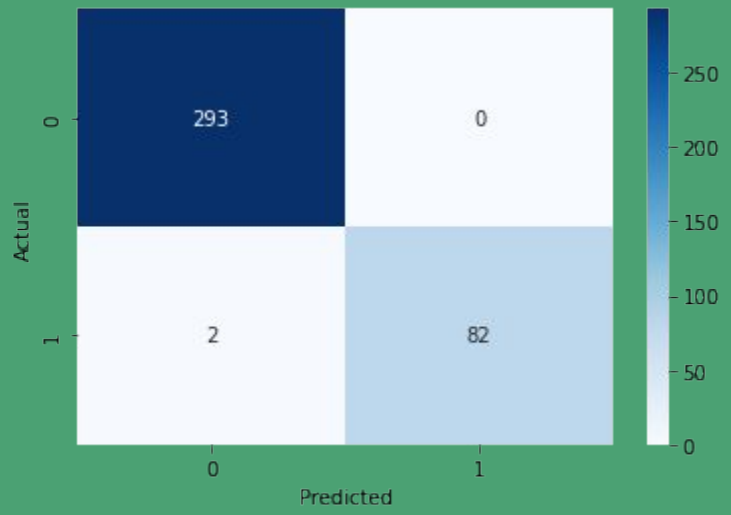
Final training accuracy: 100.00%
Final validation accuracy: 99.47%

Models with added data - still not enough data

RNN+CNN Test

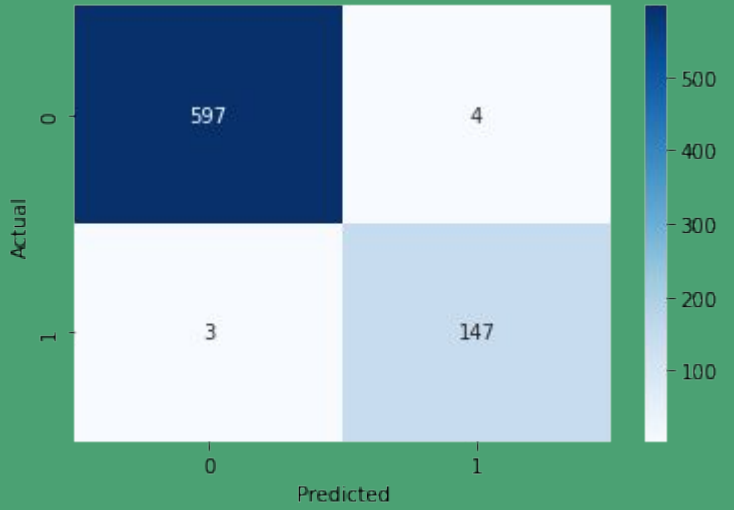


RNN+CNN Validation

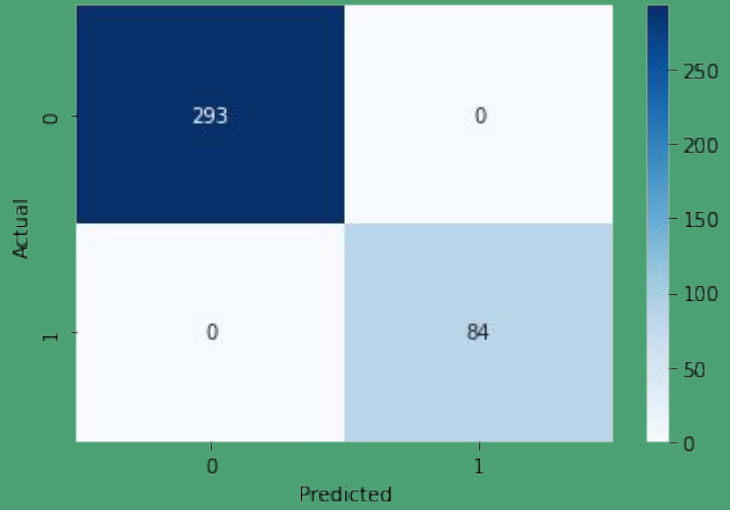


Models with added data - still not enough data

CNN Test

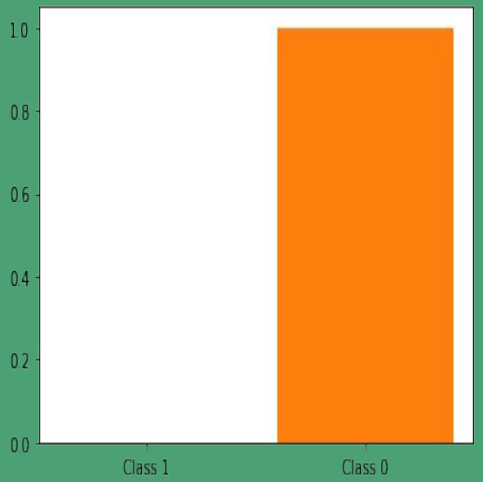


CNN Validation

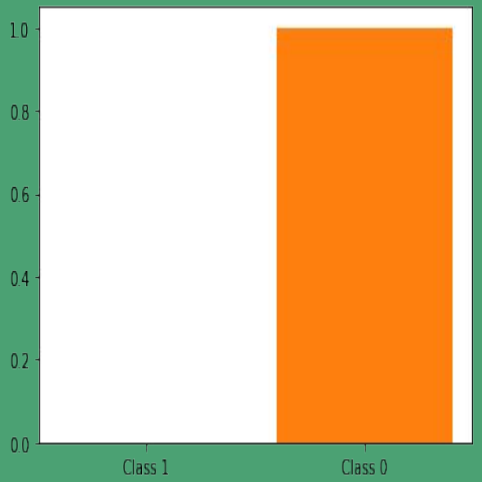


Inference CNN

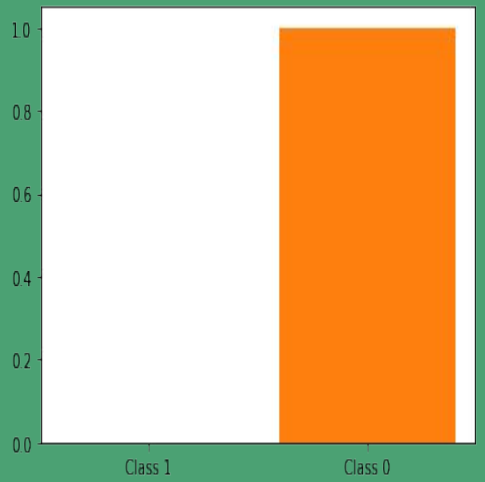
LOW Quality TARGET



COW

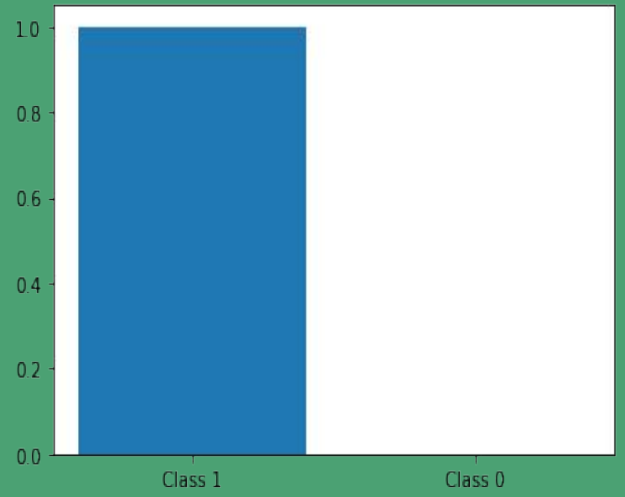


ARTICHOKE

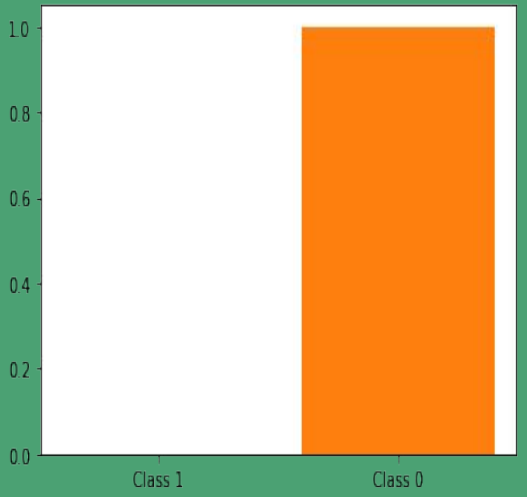


Inference RNN + CNN

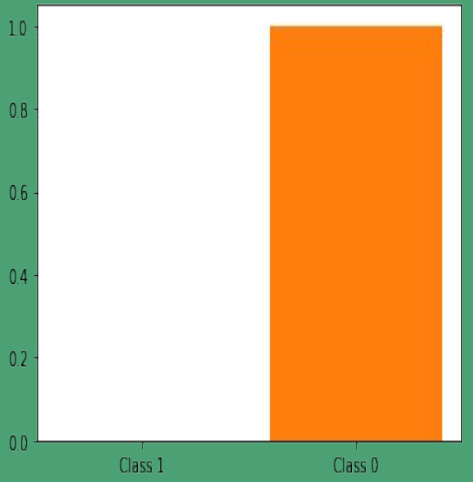
LOW Quality TARGET



COW



ARTICHOKE



Conclusions

- ADD MORE DATA
- CONSIDER AI GENERATED CLIPS
- 11 Labs
- ADD A LANGUAGE MODEL