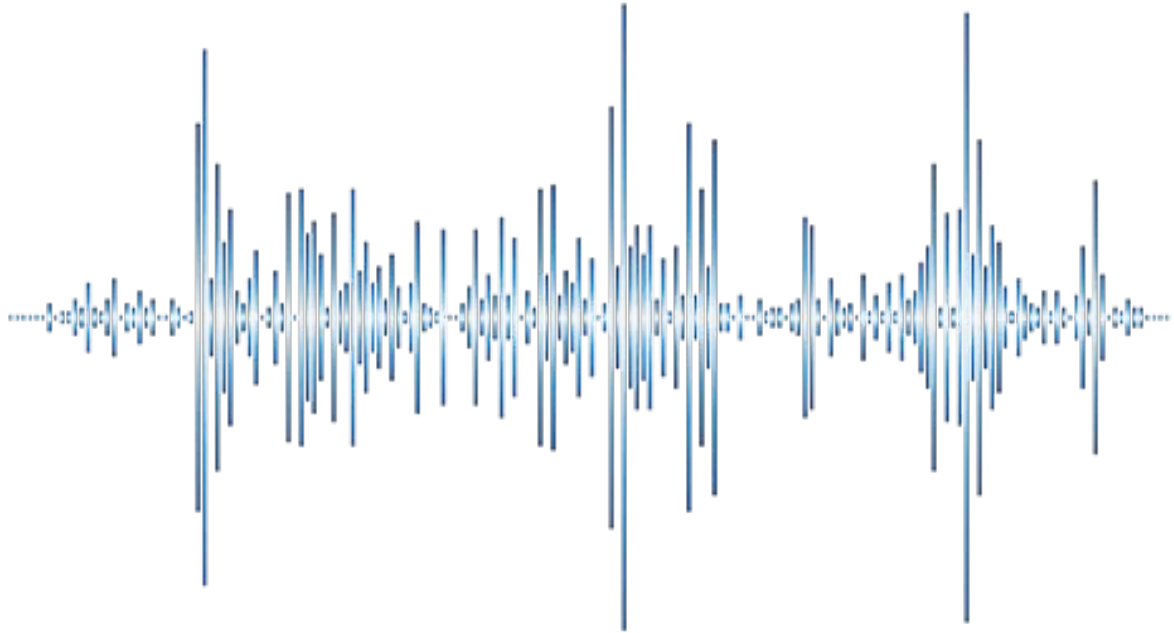


Speech Emotion Recognition

by Cezar Peixeiro

Recognizing emotions - Why ?

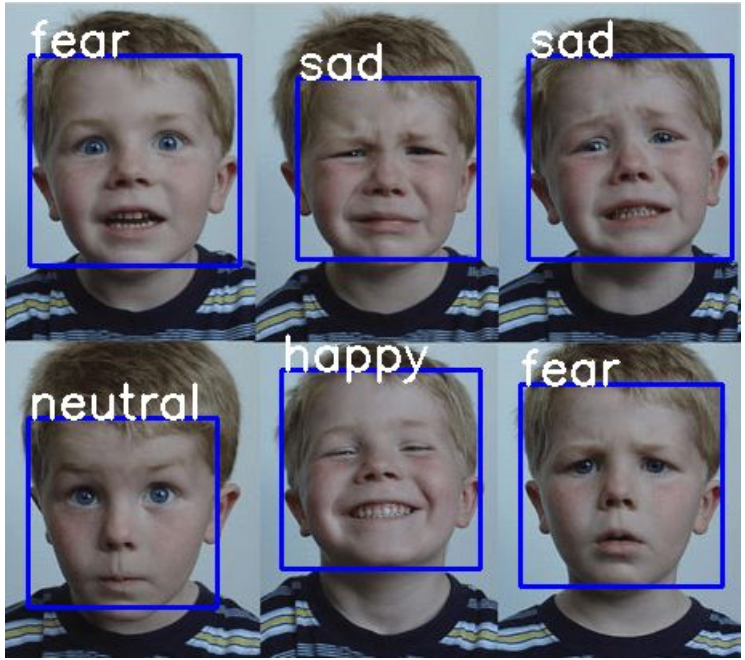
Behavior

Risks

Satisfaction

Most popular methods:

- Facial



Pros:

Facial expressions of emotion are innate rather than a product of cultural learning*

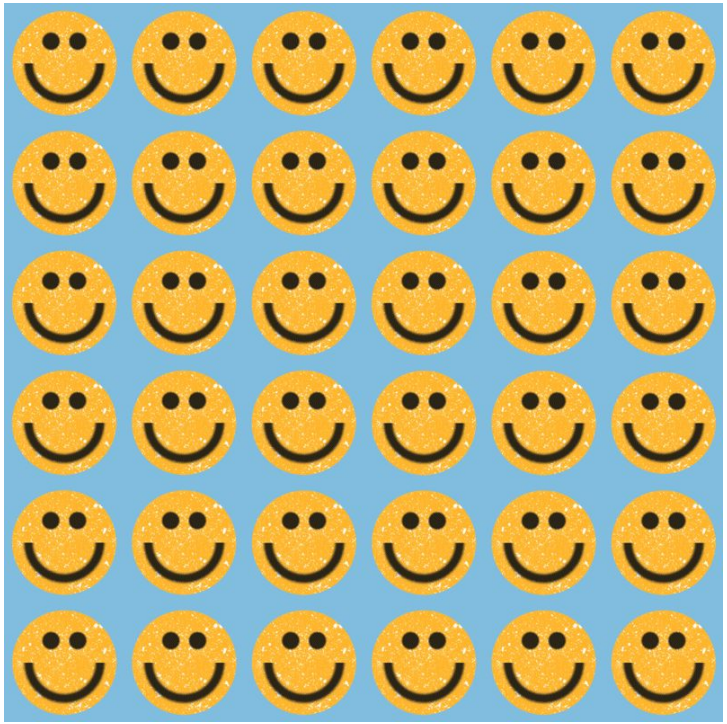
Cons:

Emotional analysis technology assigns more negative emotions to people of certain ethnicities than to others*

Facial expressions don’t reflect our feelings. Instead of reliable readouts of our emotional states, they show our intentions and social goals.*

Most popular methods:

- Text and Speech-to-Text

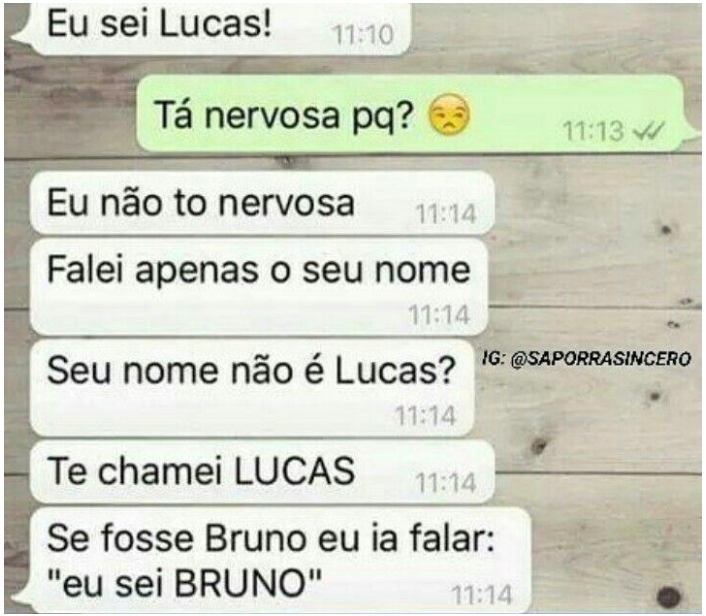


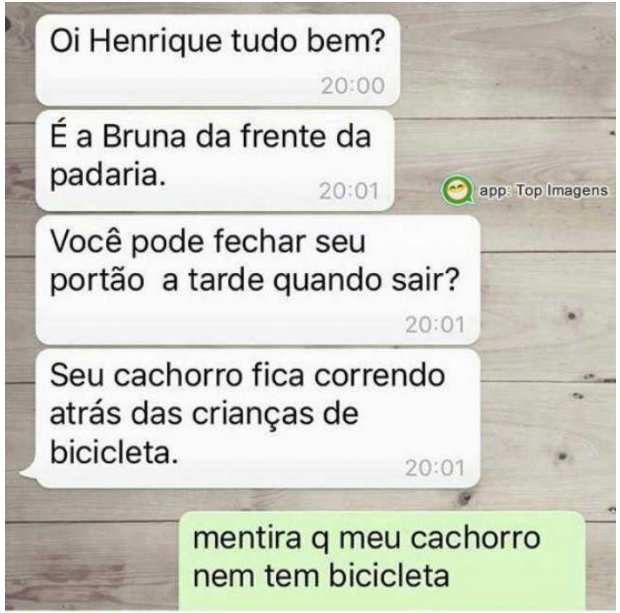
Sentiment
≠
Emotion

Words don’t show intention

Most popular methods:

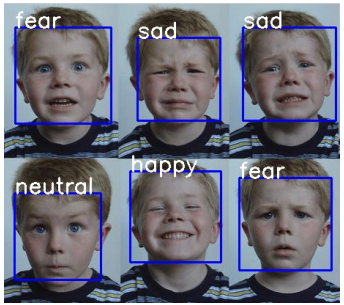
- Text and Speech-to-Text



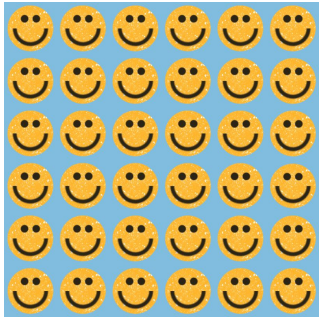


Not even humans interpret emotions ~~(or meaning)~~ properly from text

Together is better!

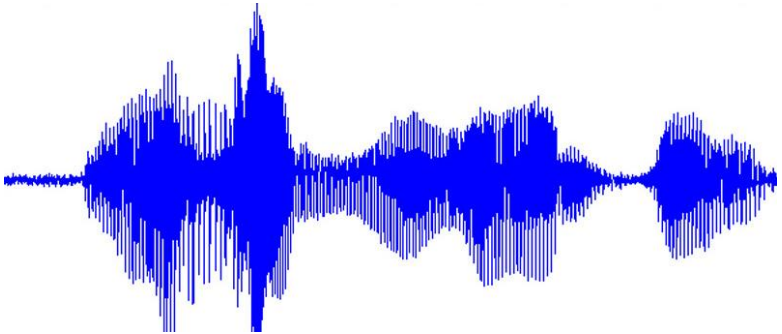


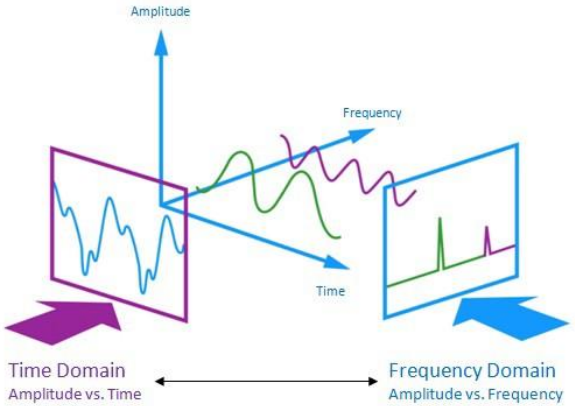
+



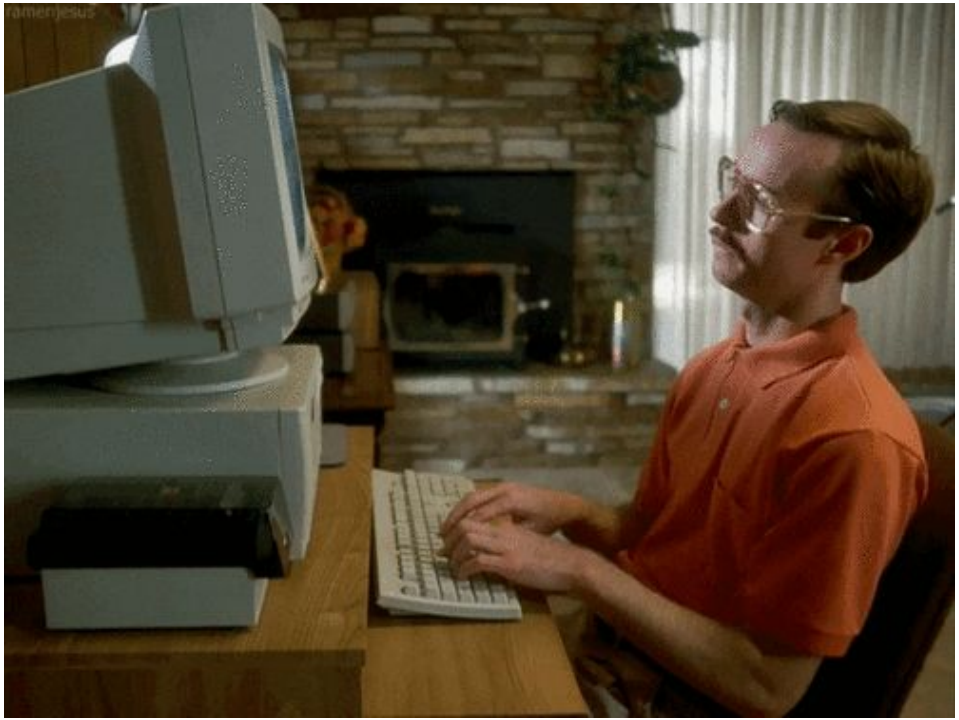
+ ?

Audio Analysis!





Let’s talk about some analysis!



The Dataset

The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)

24 actor x 60 records
(1440 audio files)

The Dataset

* Modality (01 = full-AV, 02 = video-only, 03 = audio-only)

* Vocal channel (01 = speech, 02 = song)

* Emotion (01 = neutral, 02 = calm, 03 = happy, 04 = sad, 05 = angry, 06 = fearful, 07 = disgust, 08 = surprised)

* Emotional intensity (01 = normal, 02 = strong). NOTE: There is no strong intensity for the 'neutral' emotion

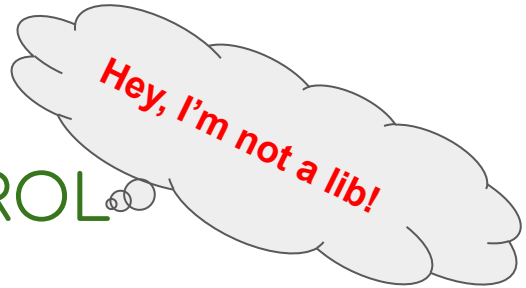
* Statement (01 = "Kids are talking by the door", 02 = "Dogs are sitting by the door")

* Repetition (01 = 1st repetition, 02 = 2nd repetition).

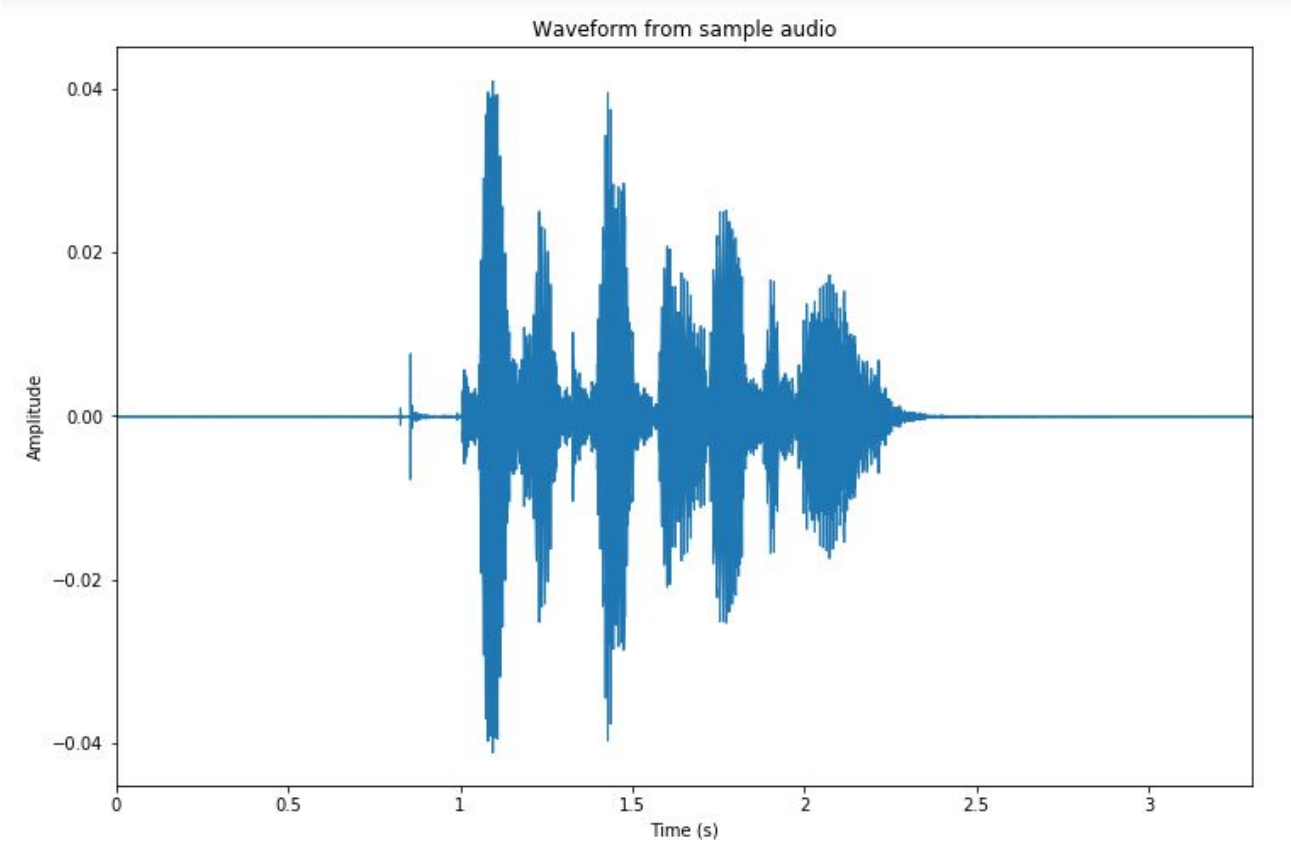
* Actor (01 to 24. Odd numbered actors are male, even numbered actors are female)

Getting Started: New libs to learn!

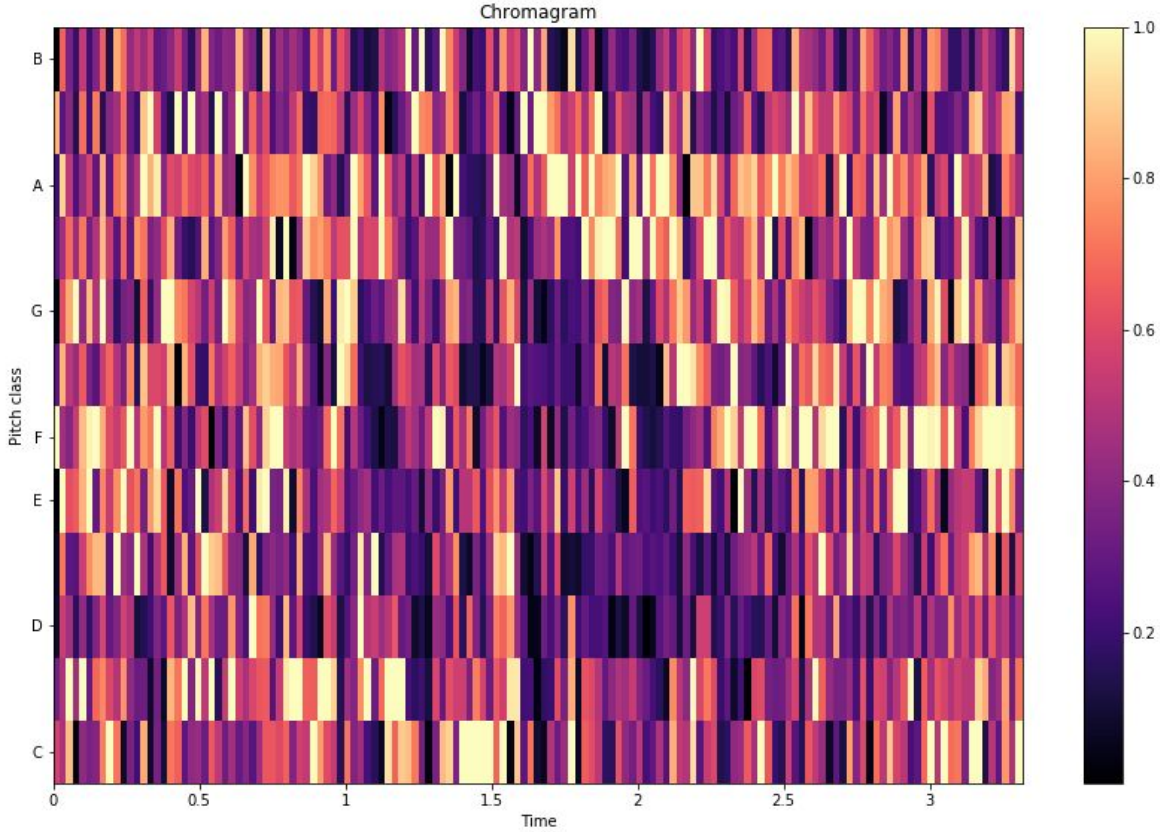
PULSEAUDIO
VOLUME CONTROL
(pauvcontrol)



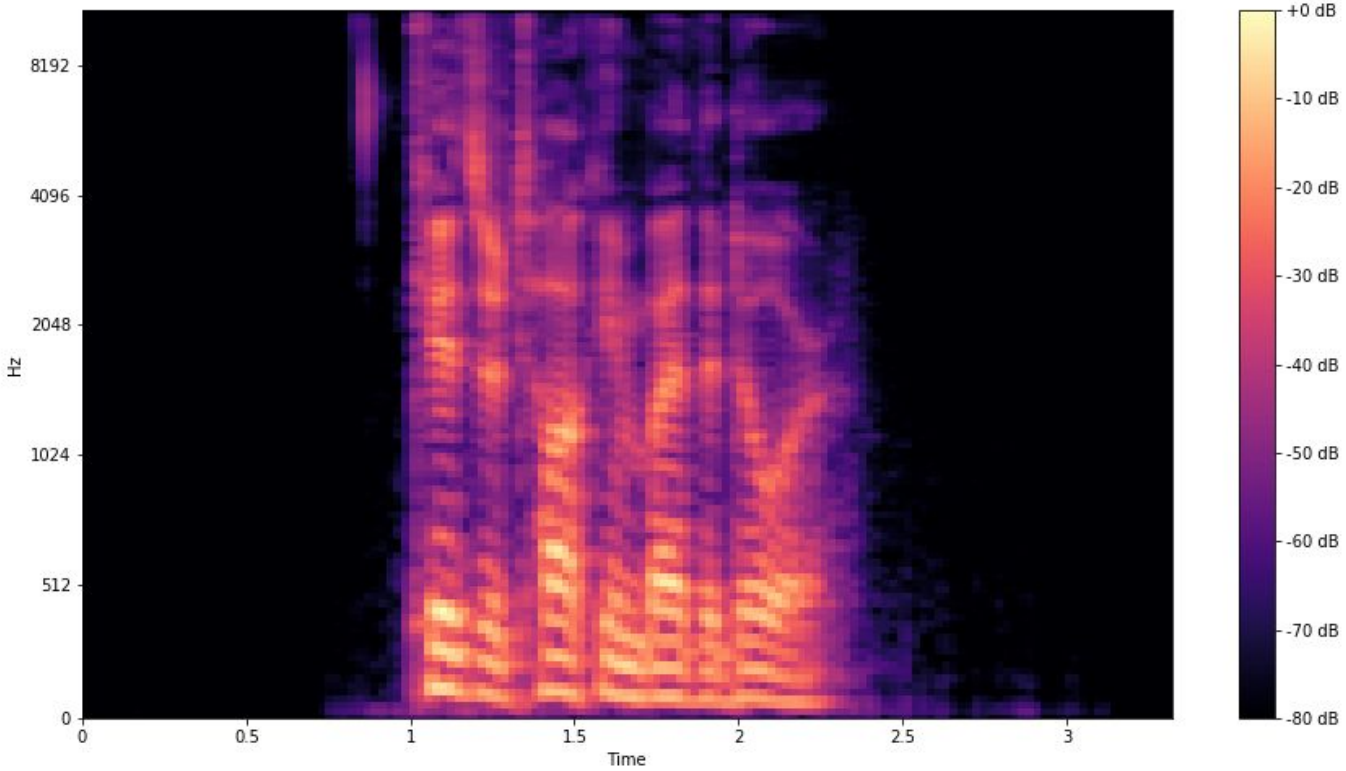
Audio and Features (waveplot)



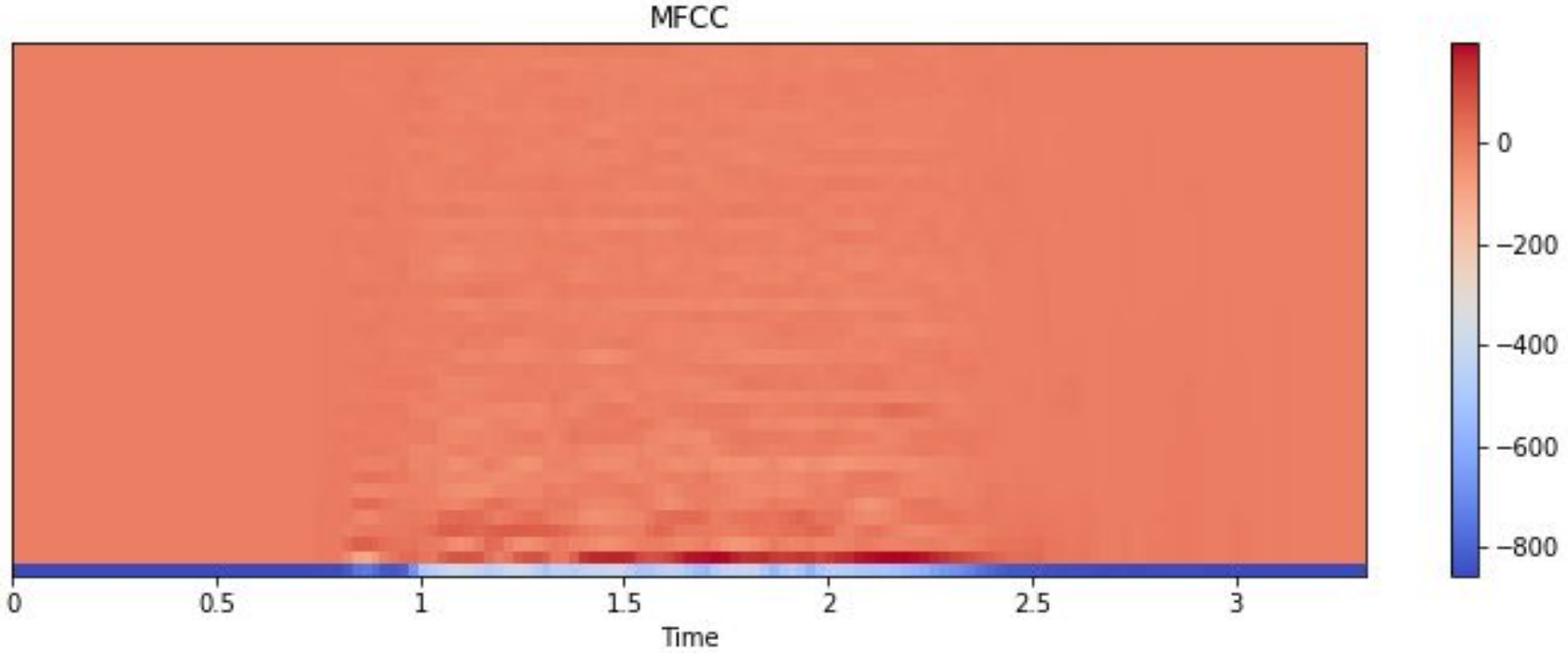
Audio and Features (Chromagram)



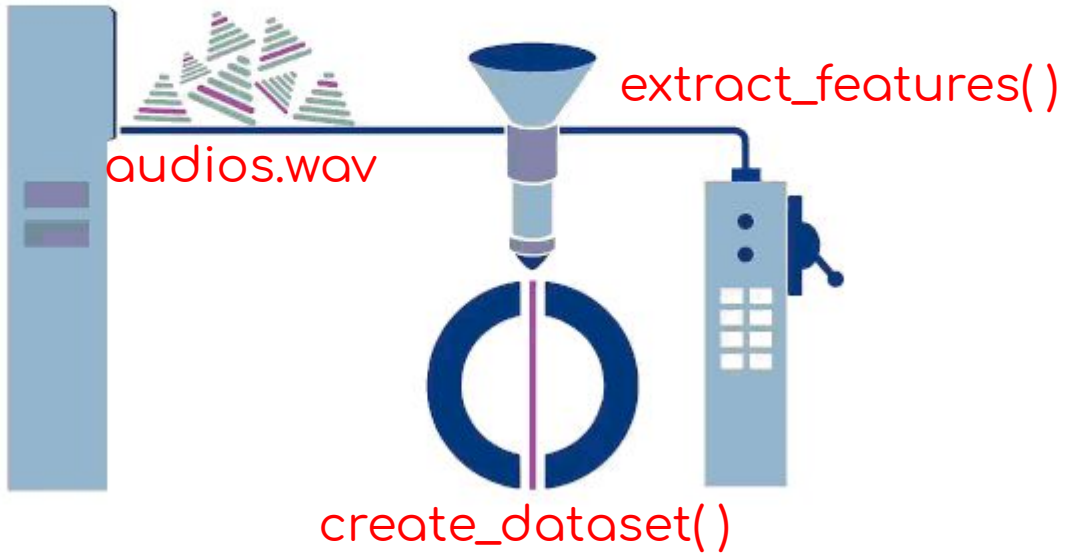
Audio and Features (mel spectrogram)



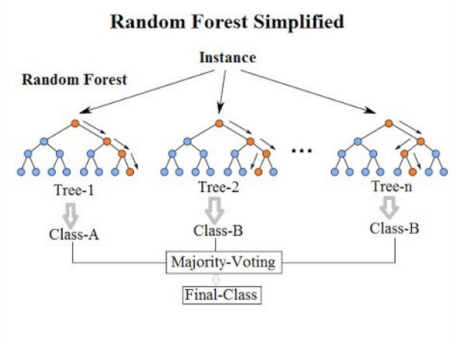
Audio and Features (mfcc)

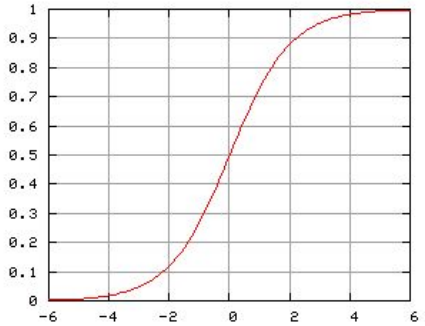


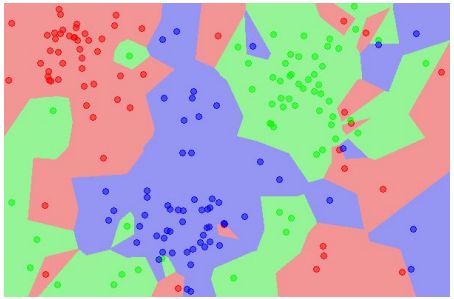
Feature Extraction

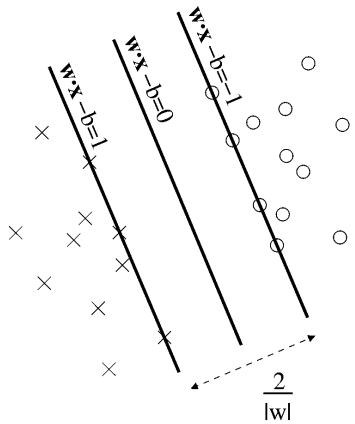


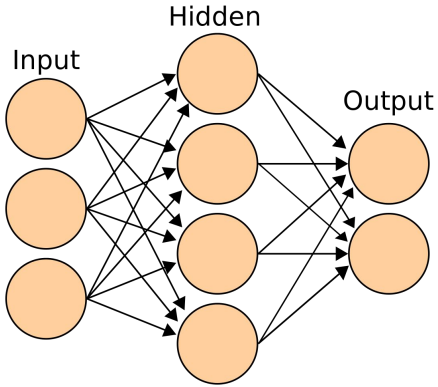
Trying Models



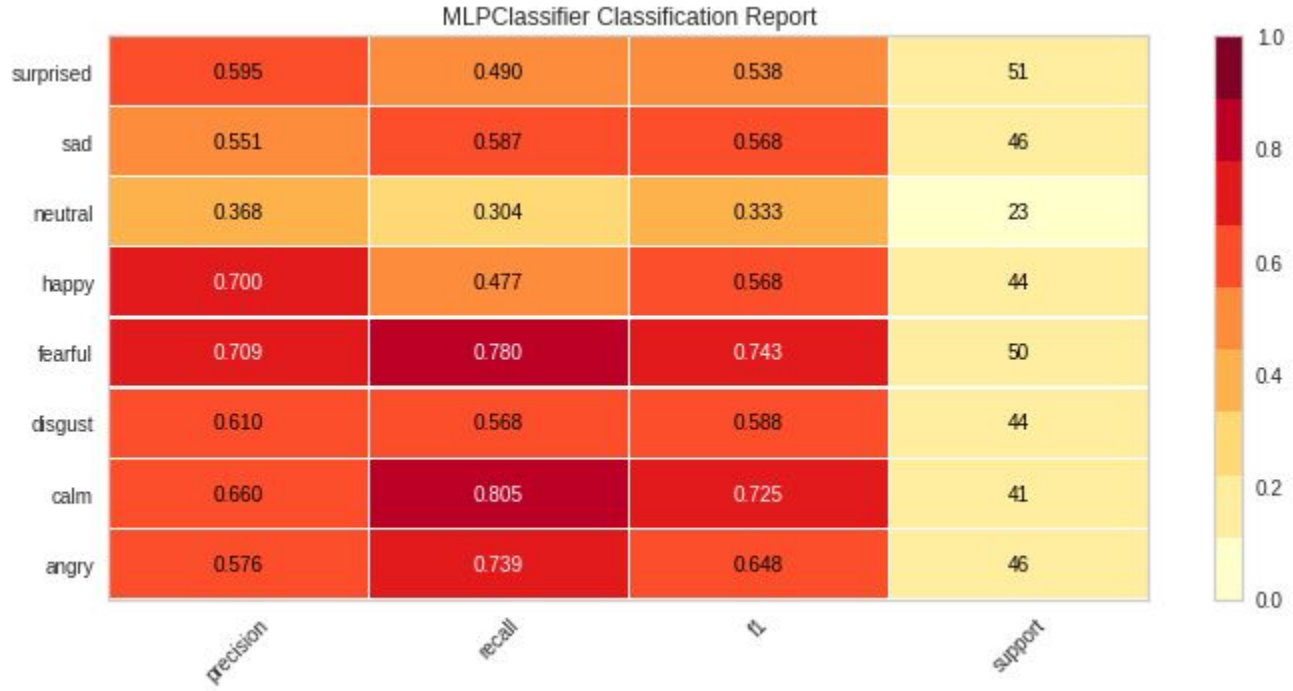




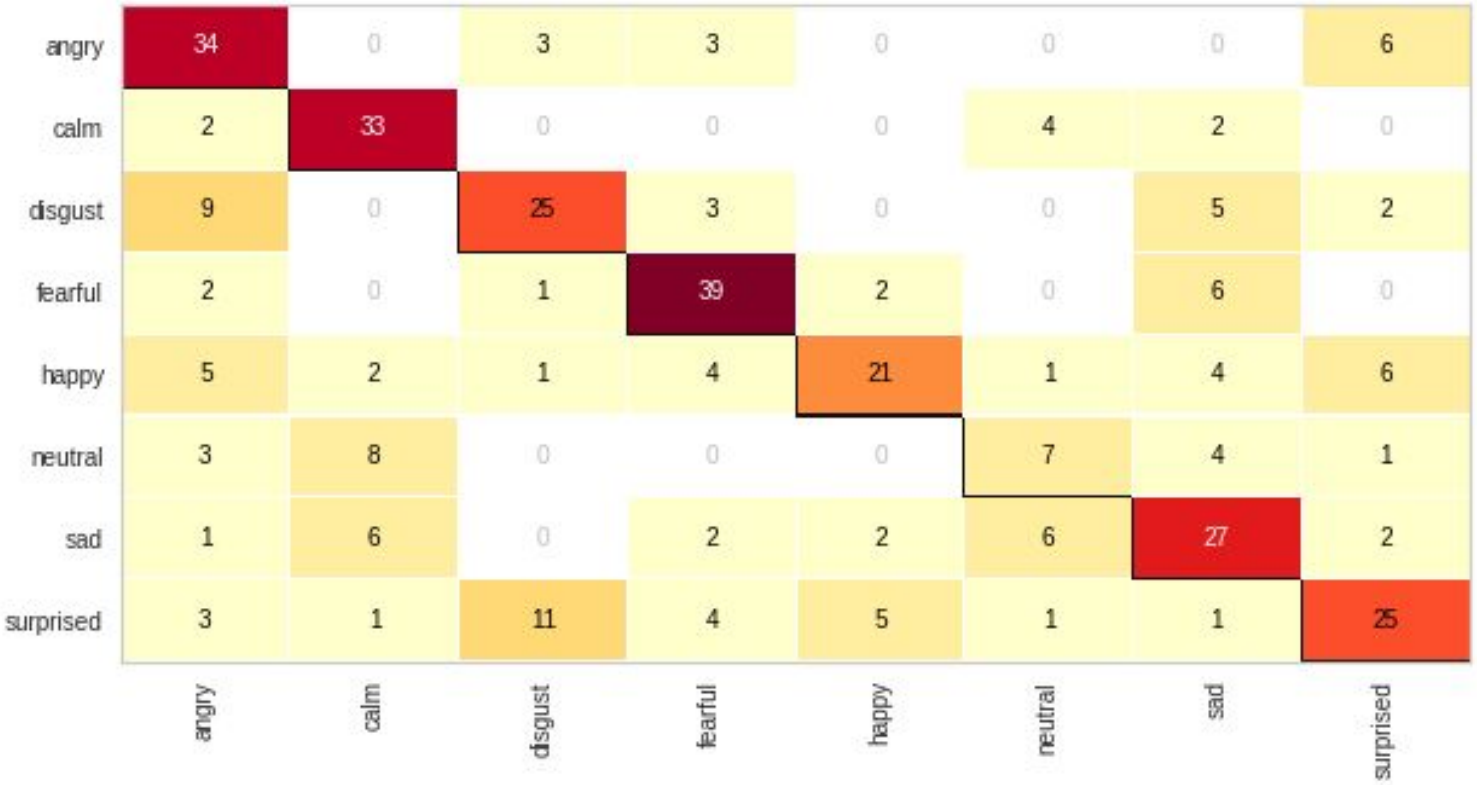




MLPC details (classification report)



MLPC details (confusion matrix)



What about audio in portuguese?

What about live classification?

Jupyter Notebook + YT

Future...

- Create a dataset of portuguese audio;
- Analyze audio data in time
- Train model to be more efficient;