# Trabalhos Práticos: Análise de Voz Através de Machine Learning

Utilizar técnicas de Machine Learning (Neural Networks e Deep Learning p.e.) para automatizar análise de voz. Possíveis abordagens:

- Speech Recognition
- Speaker Recognition/Identification
- Speaker Demographic Identification (Age, gender, accent,...)
- Speech Synthesis
- Computer Aided Diagnosis of Speech Disorders (e.g. Parkinson)

## Ferramentas para Speech Recognition:

### Modelos:

- DeepSpeech[1] by Mozilla[2] (Recursive Neural Networks + Language Model w/ TensorFlow): **existe modelo pré-treinado**,
- Very Deep CNN[3] for spectrograms
- Deep Bidirectional Long Short-Term Memory Network[4]

### Datasets:

**Common Voice[5]**

500h de inglês anotado, com demografia

**ASVspoof 2015[6] e 2017** (Automatic Speaker Verification Spoofing and Countermeasures Challenge)

Contains speech collected from 106 speakers. The dataset comprises three distinct, non-overlapping subsets for training, development and evaluation. The training and development subsets contain both genuine speech and spoofed speech. All examples of the latter are generated artificially using one of five different, well-known speech synthesis or voice conversion spoofing algorithms and one of two different vocoders.

1 https://arxiv.org/pdf/1412.5567.pdf
2 https://github.com/mozilla/DeepSpeech
3 https://arxiv.org/pdf/1610.03022.pdf
4 http://www.cs.toronto.edu/~graves/asru_2013.pdf
5 https://voice.mozilla.org/
6 https://datashare.is.ed.ac.uk/handle/10283/853

**TIMIT**[7] Acoustic-Phonetic Continuous Speech Corpus

TIMIT contains broadband recordings of 630 speakers of eight major dialects of American English, each reading ten phonetically rich sentences. The TIMIT corpus includes time-aligned orthographic, phonetic and word transcriptions as well as a 16-bit, 16kHz speech waveform file for each utterance.

## Packages para Python para Deep Learning:

**TensorFlow,** SKLearn, PyTorch, Caffe

7 https://catalog.ldc.upenn.edu/LDC93S1