



# PROCESSAMENTO DE SINAIS DE ÁUDIO PARA IDENTIFICAÇÃO DE IDIOMA

Machine Learning / DSP  
Márcio Luiz Bezerra Lopes Júnior

# INTRODUÇÃO

## Problema

- Como identificar o idioma de um sinal de áudio sem nenhum conhecimento prévio?

## Objetivo

- Explorar o uso de algoritmos de aprendizagem de máquina para resolver este problema



# CONJUNTO DE DADOS

Áudio de formato .mp3

Mozilla Common Voice

# MODELOS

## Random Forests

- Entrada: dataframe contendo diversas características extraídas dos sinais

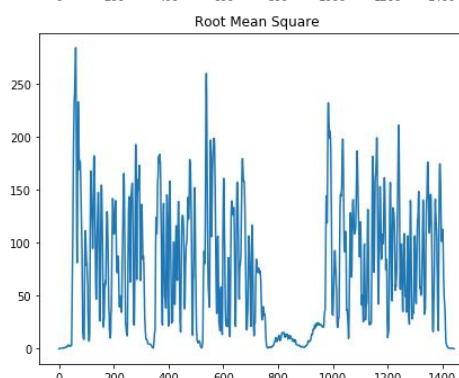
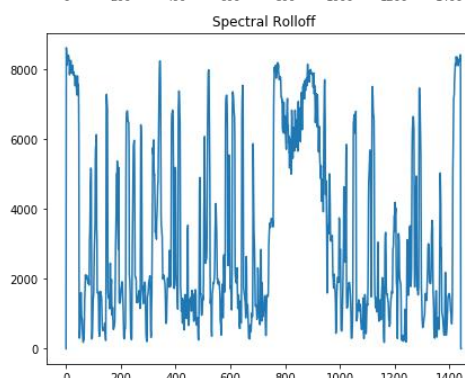
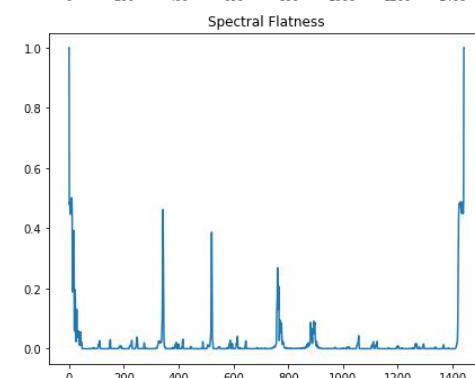
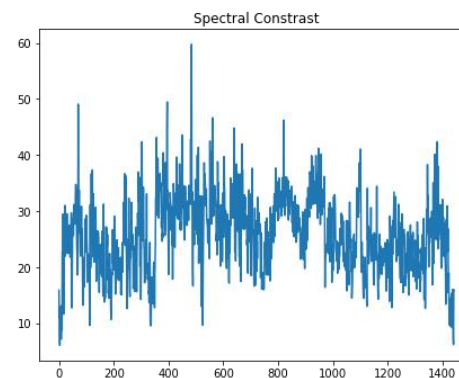
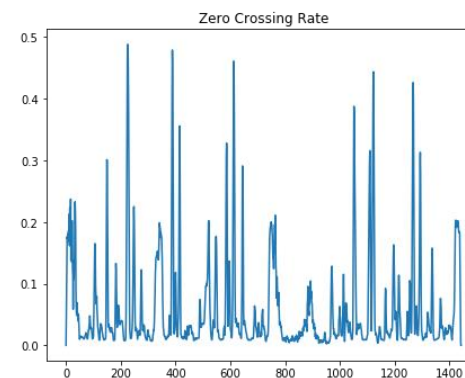
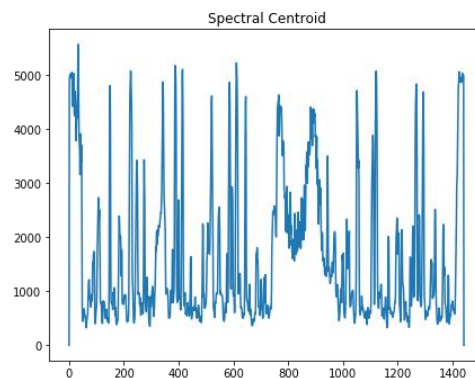
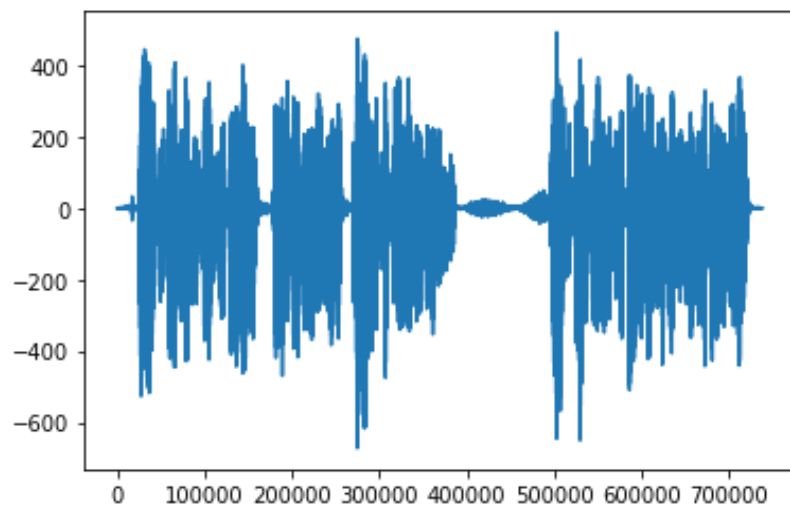
## Rede Neural Convolucional

- Entrada: Spectrograma – representação visual do sinal

# EXTRAÇÃO DE CARACTERÍSTICAS

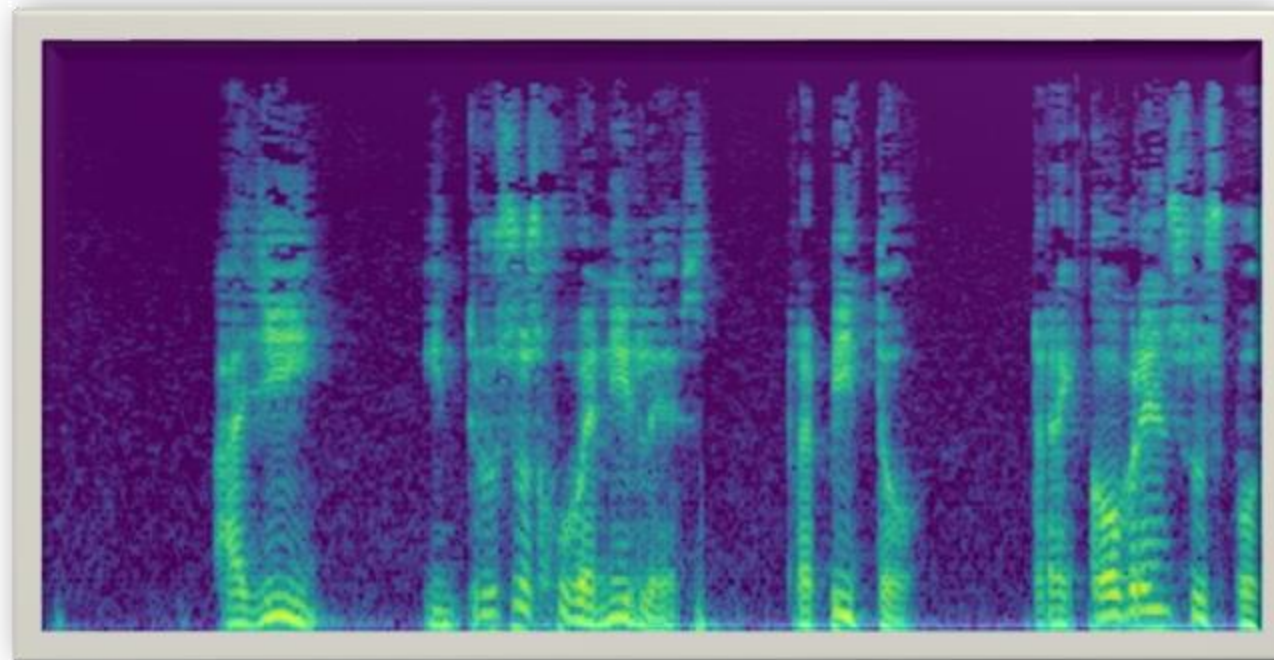
- Centroide espectral
  - Taxa de cruzamento do eixo x
  - Contraste espectral
  - Nivelamento espectral
  - Rolagem espectral
  - Raiz da média quadrática
- 
- Quatro propriedades: *máximo, mínimo, média e desvio padrão*

# EXTRAÇÃO DE CARACTERÍSTICAS

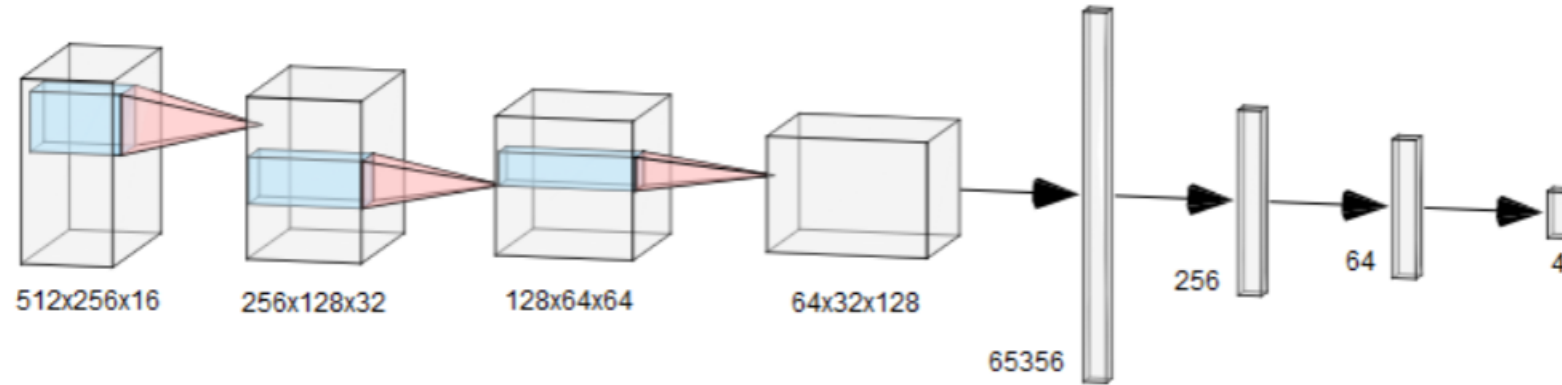


# SPECTROGRAMA

Processamento de Sinal  $\leftrightarrow$  Processamento de Imagem



# REDE CONVOLUCIONAL



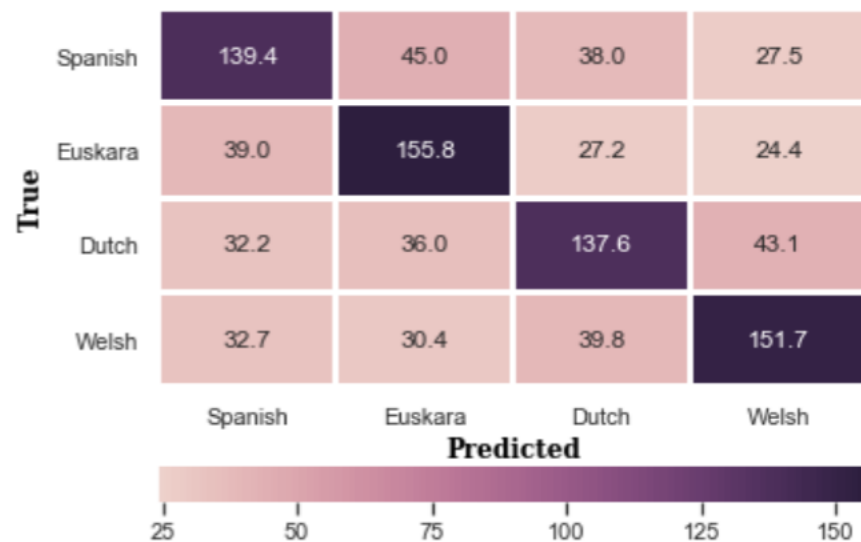


# RESULTADOS

Idiomas seleccionados:

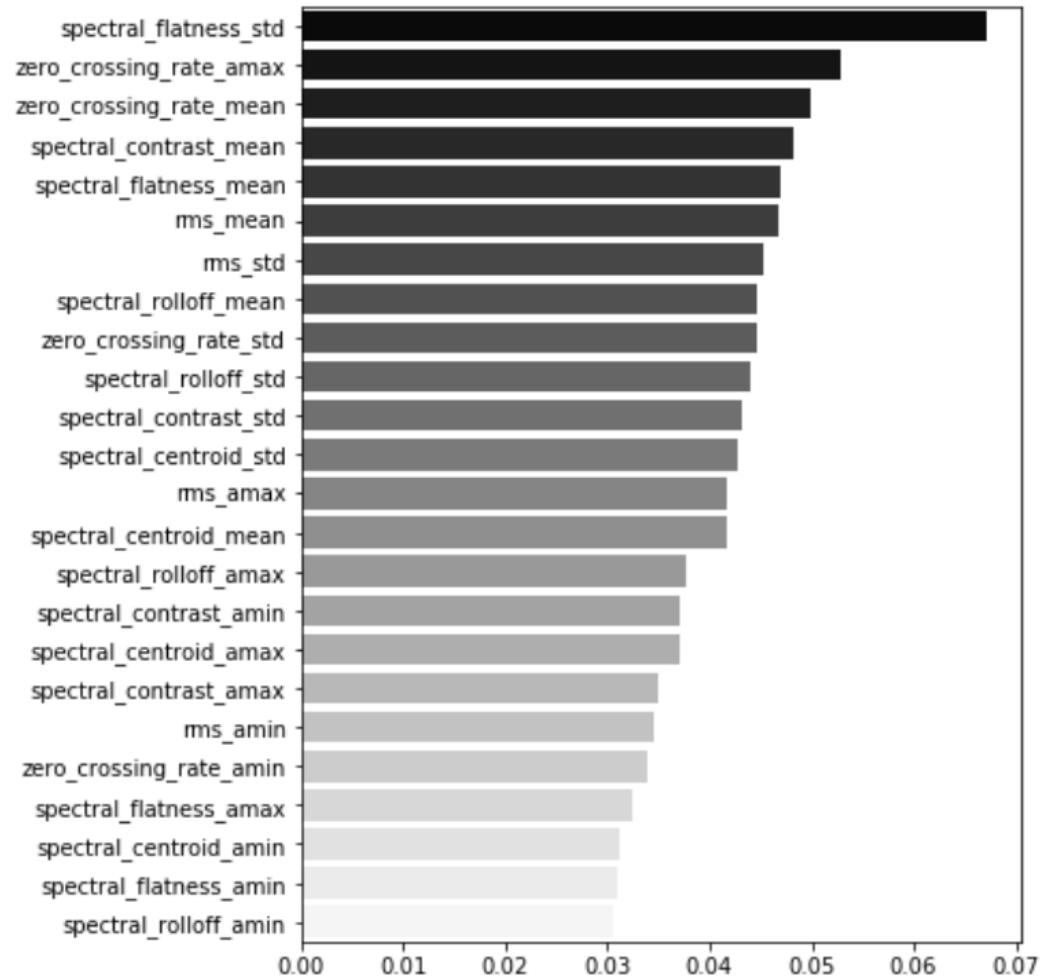
- Espanhol
- Holandês
- Galês
- Basco

# RESULTADOS - RANDOM FORESTS

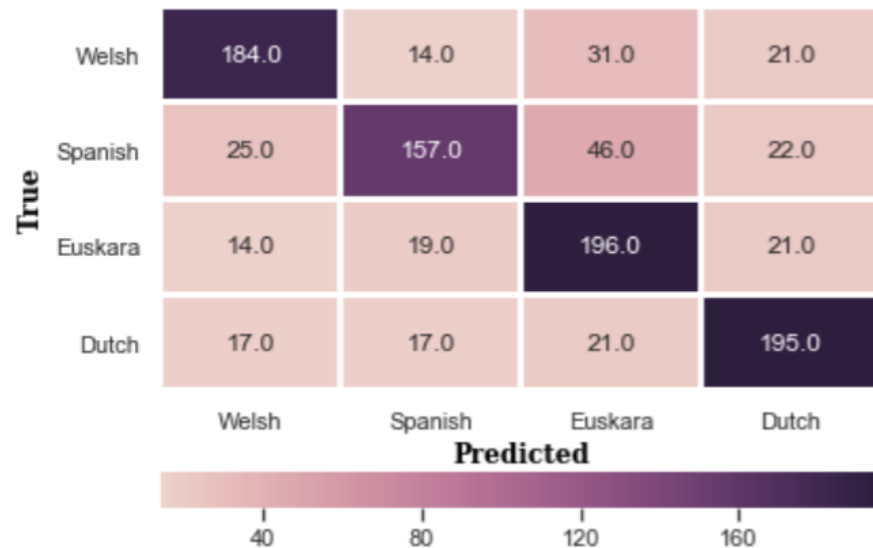


Language	Precision	Recall	F1-score
Welsh	0.59	0.62	0.61
Spanish	0.55	0.51	0.53
Euskara	0.59	0.65	0.62
Dutch	0.57	0.52	0.54
Overall	0.58	0.58	0.58

# FEATURE IMPORTANCE



# RESULTADOS - CNN



Language	Precision	Recall	F1-score
Welsh	0.77	0.74	0.75
Spanish	0.76	0.63	0.69
Euskara	0.67	0.78	0.72
Dutch	0.75	0.78	0.77
Overall	0.74	0.73	0.73



**OBRIGADO!**

PERGUNTAS?