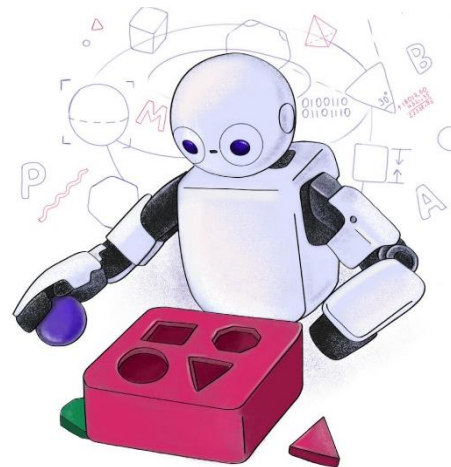


TP557 - Tópicos avançados em IoT e Machine Learning: *Identificação de palavras chave*



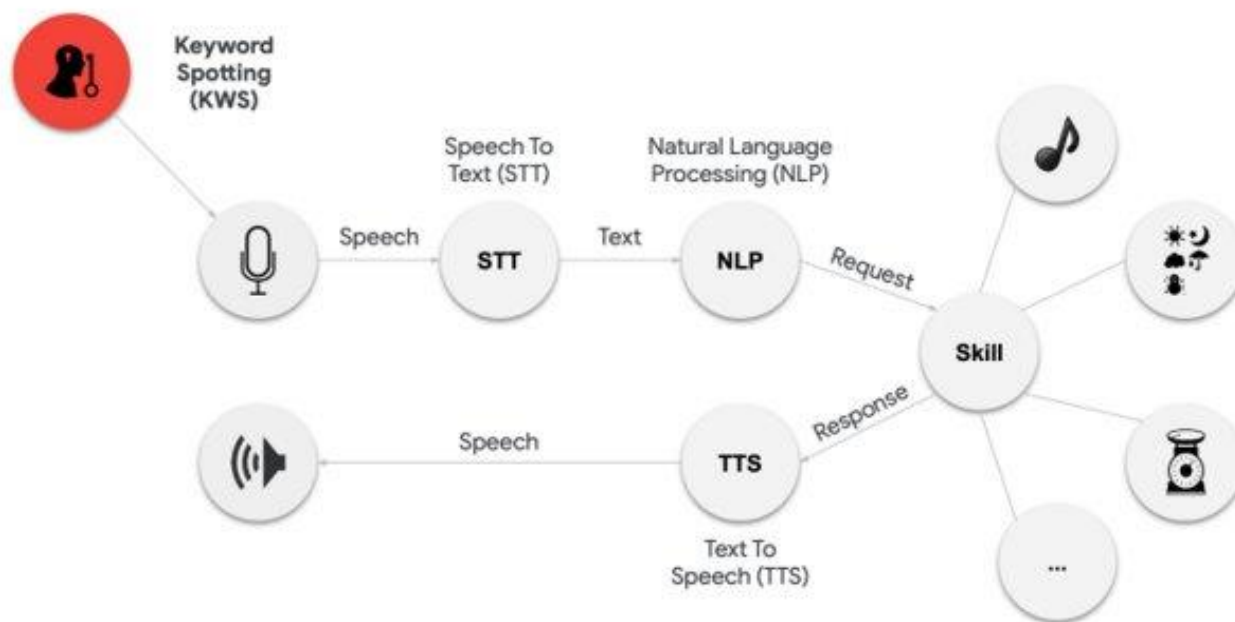
Inatel

Crédito: Marcelo Rovai-Unifei

Samuel Baraldi Mafra
samuelbmafra@inatel.br

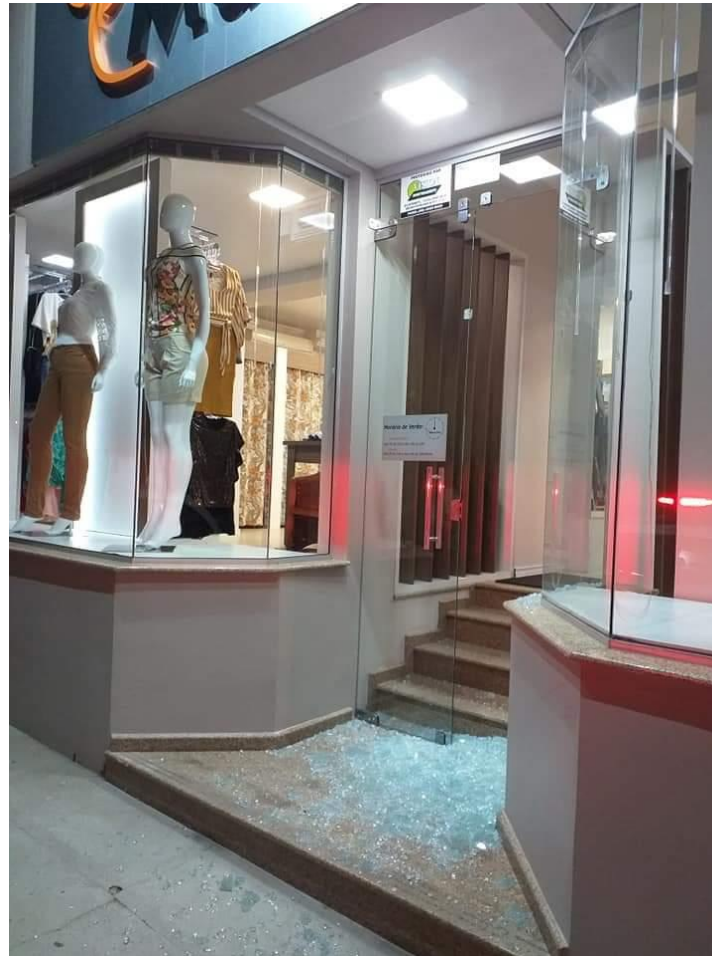
Identificação de palavras chave

- A localização de palavras-chave (KWS) é uma técnica importante para aplicações de fala, que permite aos usuários ativar dispositivos falando uma frase de palavra-chave.



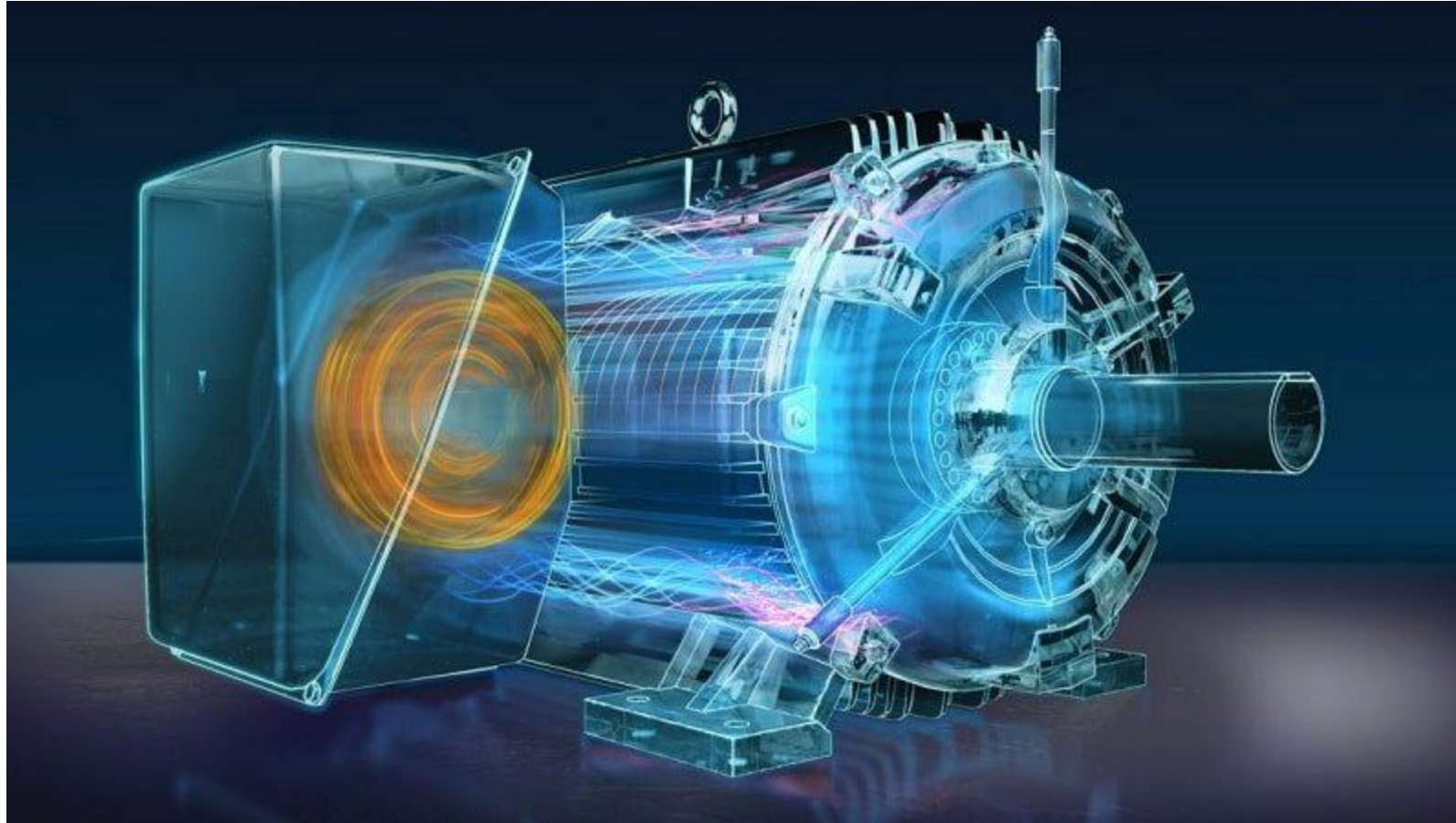
Identificação de palavras chave

- Aplicações: Identificação de vidro quebrado (segurança)



Identificação de palavras chave

- Aplicações: Detecção de anomalias (indústria)



Identificação de palavras chave

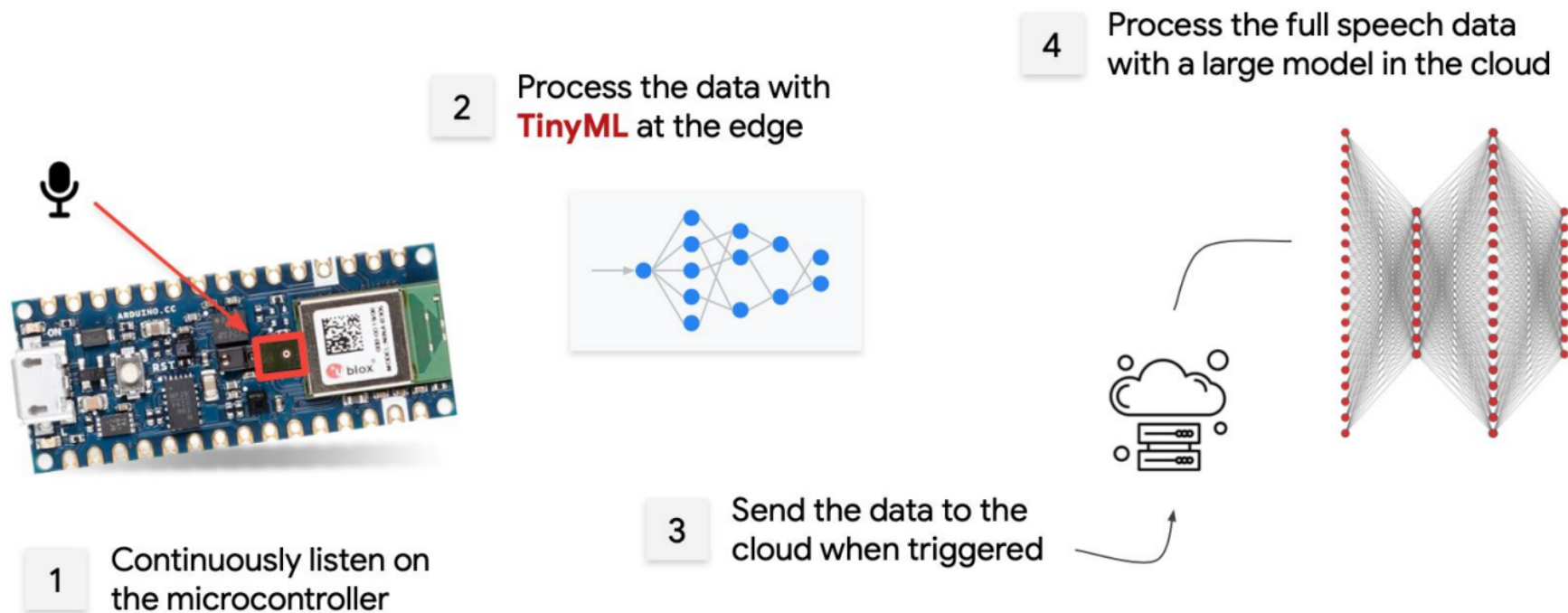
- Aplicações: detecção de roncos e choros



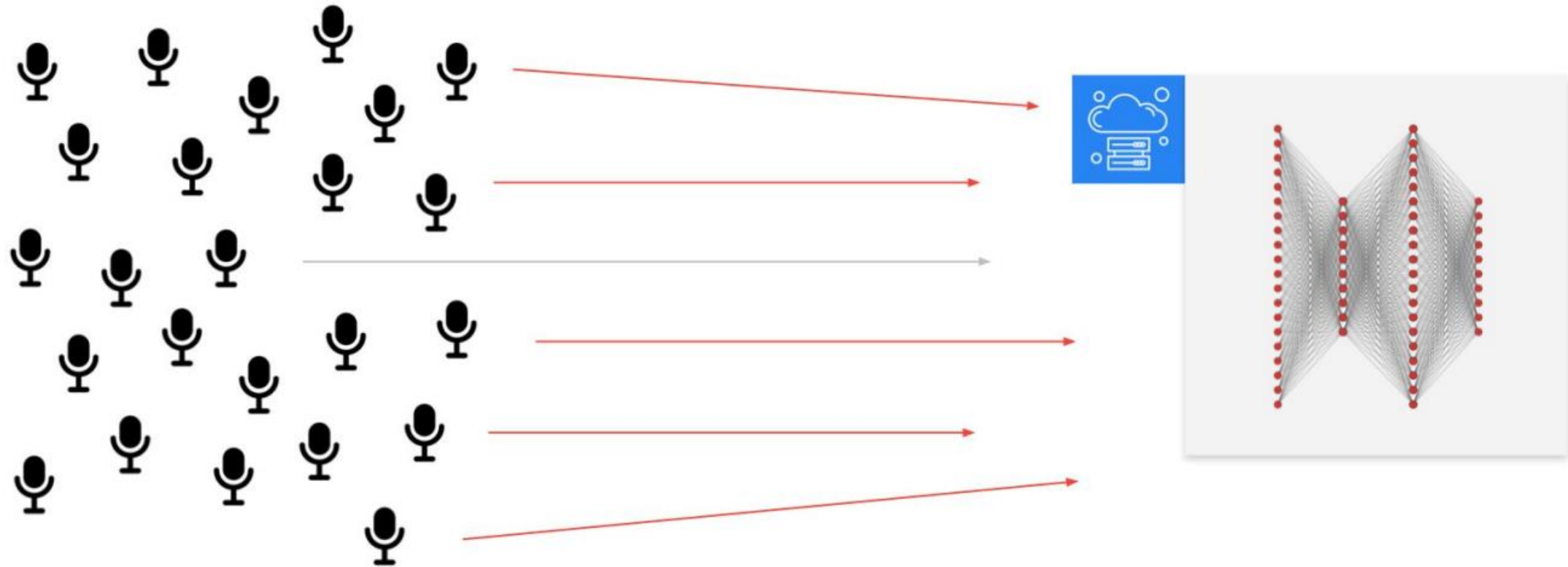
Identificação de palavras chave

- Desafios:
- Banda e latência;
- Precisão e personalização;
- Segurança e privacidade;
- Bateria e memória.

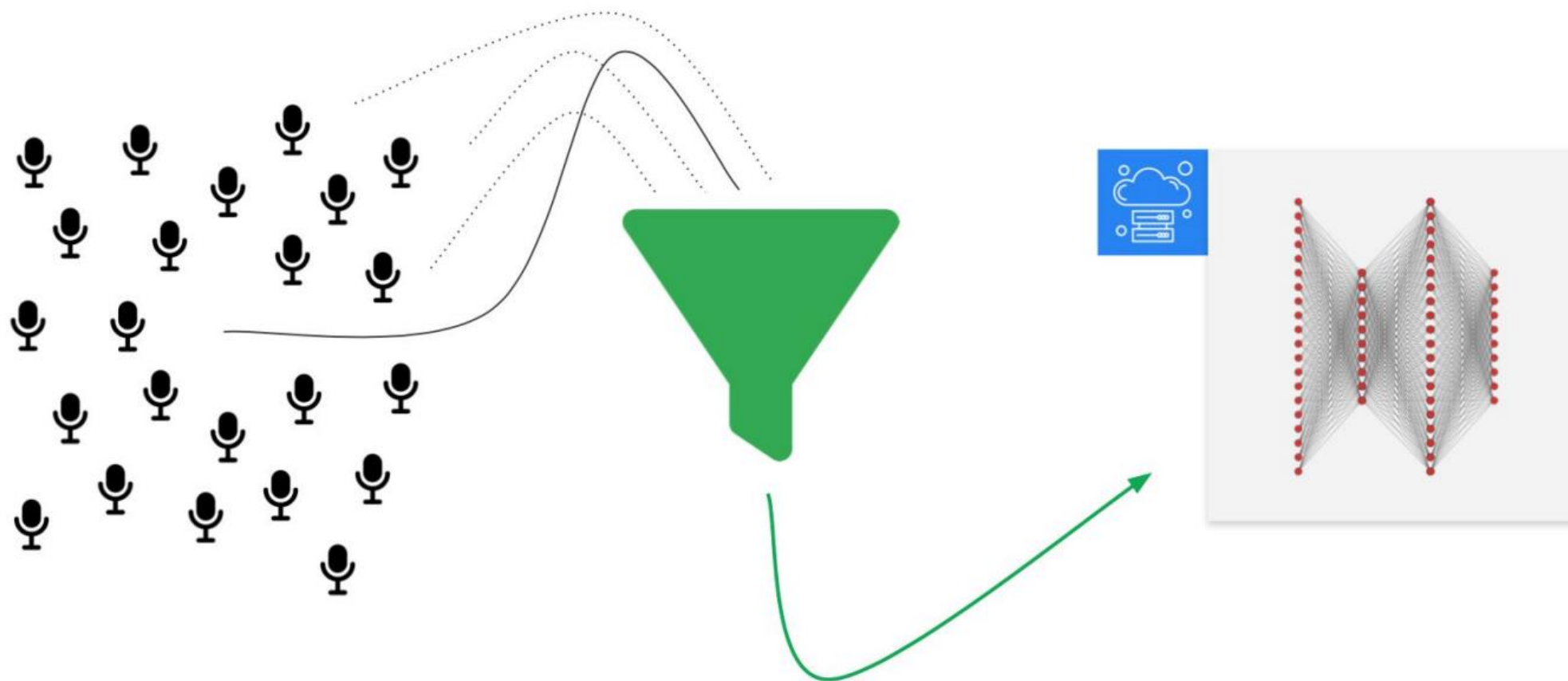
Identificação de palavras chave



Identificação de palavras chave



Identificação de palavras chave



Identificação de palavras chave

Como construímos um bom conjunto de dados?

- Quem são os usuários?
- O que eles precisam?
- Que tarefa eles estão tentando resolver?
- Como eles interagem com o sistema?
- Como o mundo real torna isso difícil?

Identificação de palavras chave

Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition

Pete Warden
Google Brain
Mountain View, California
petewarden@google.com

<https://arxiv.org/pdf/1804.03209.pdf>

Identificação de palavras chave

- Primeira versão 10 palavras
- Yes
- No
- Left
- Right
- Go
- Stop
- One
- Two
- Four
- Six

Identificação de palavras chave

- Segunda versão: 35 palavras
- 2618 voluntários
- Mais de mil gravações de cada palavra.

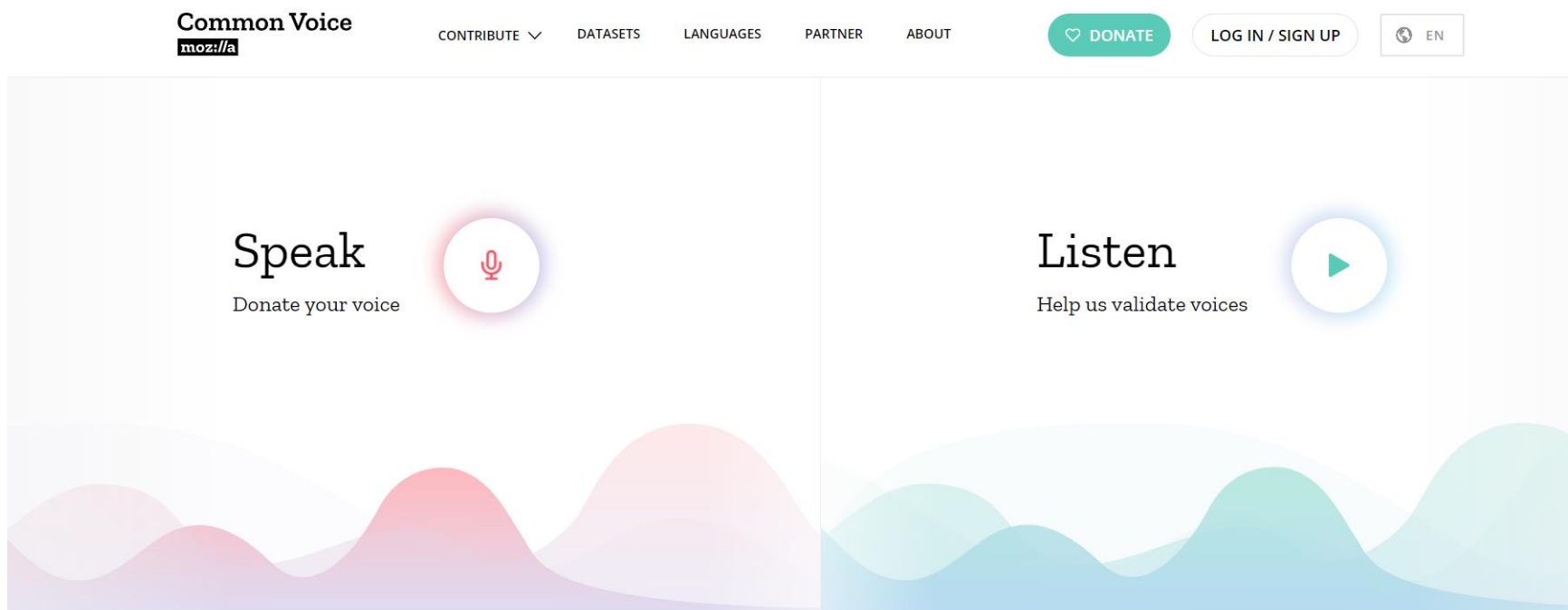
Word	Number of Utterances
Backward	1,664
Bed	2,014
Bird	2,064
Cat	2,031
Dog	2,128
Down	3,917
Eight	3,787
Five	4,052
Follow	1,579
Forward	1,557
Four	3,728
Go	3,880
Happy	2,054
House	2,113
Learn	1,575
Left	3,801
Marvin	2,100
Nine	3,934
No	3,941
Off	3,745
On	3,845
One	3,890
Right	3,778
Seven	3,998
Sheila	2,022
Six	3,860
Stop	3,872
Three	3,727
Tree	1,759
Two	3,880
Up	3,723
Visual	1,592
Wow	2,123
Yes	4,044
Zero	4,052

Identificação de palavras chave

- Validação do dataset
- Verificação de palavras ditas erradas
- Retirar gravações baixas
- Retirar momentos de mudo.

Identificação de palavras chave

- Mozilla common voice
- <https://commonvoice.mozilla.org/en>



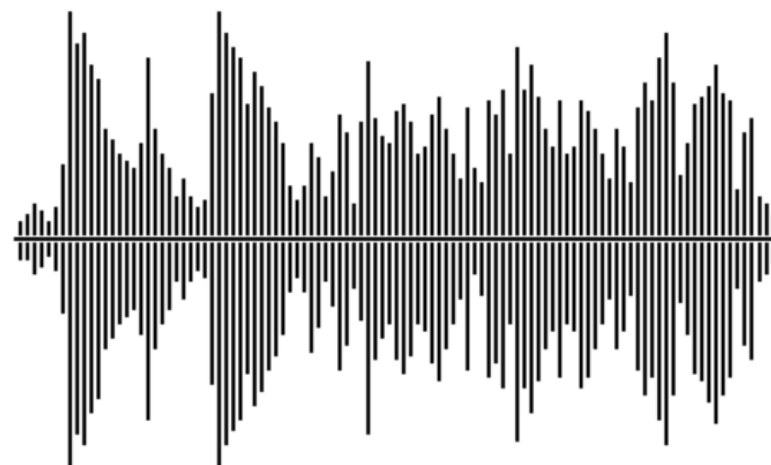
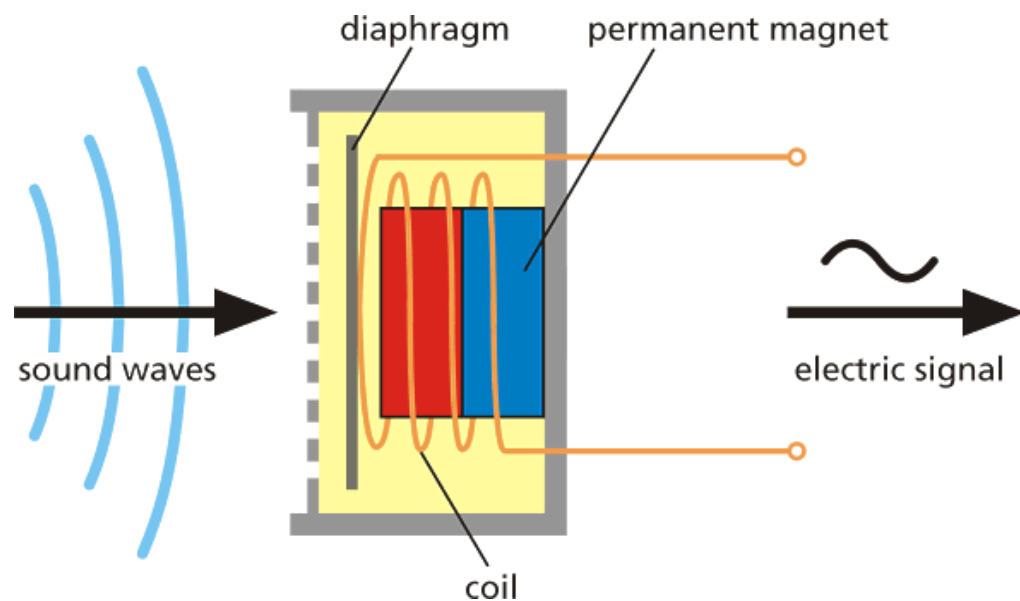
Identificação de palavras chave

Questões a serem levantadas

- Quais sons devem ser gravados?
- Qualidade do microfone;
- Ruído ambiente

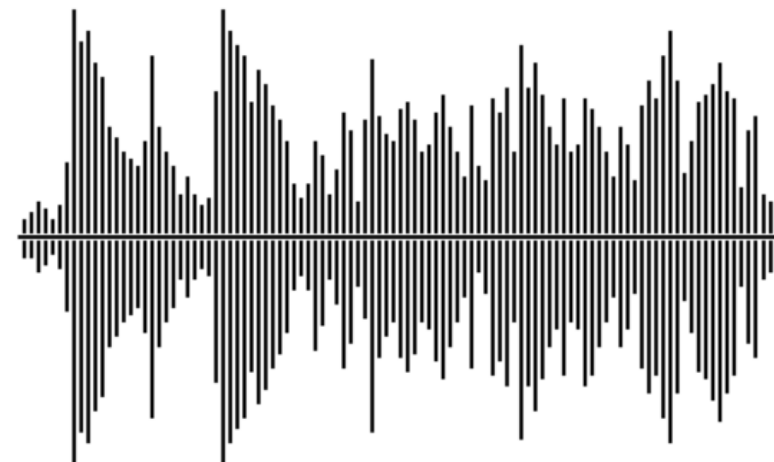
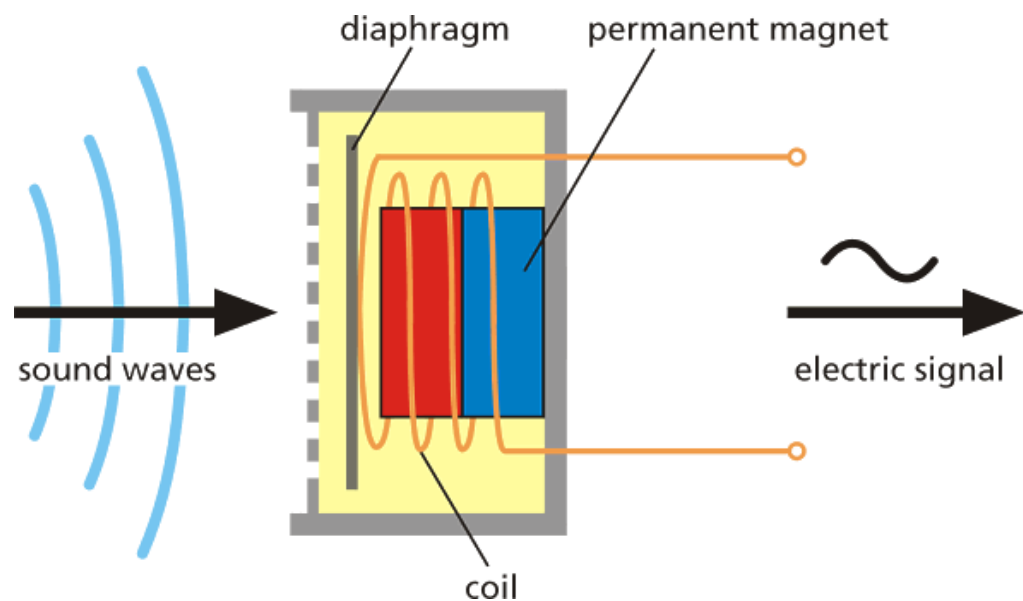
Identificação de palavras chave

Sinal de áudio

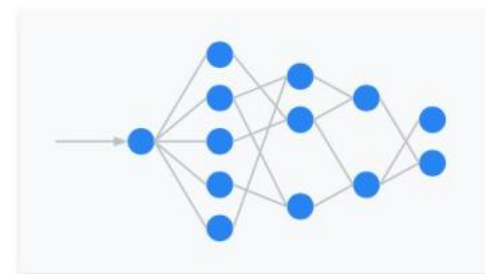


Identificação de palavras chave

Sinal de áudio



16kHz signal



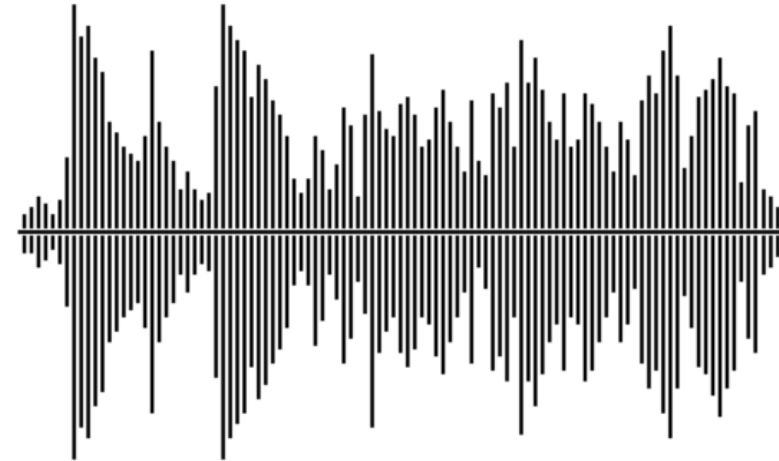
Inatel

CAMINHOS
QUE CONECTAM
COM O FUTURO

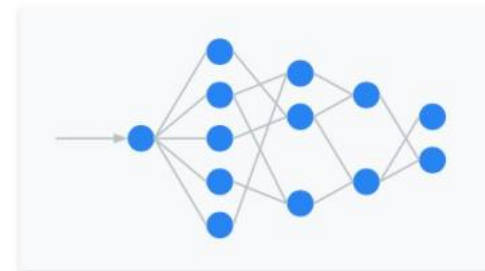
Identificação de palavras chave

Sinal de áudio

- 16 mil amostras por segundo
- Não dá para jogar direto como entrada de rede neural



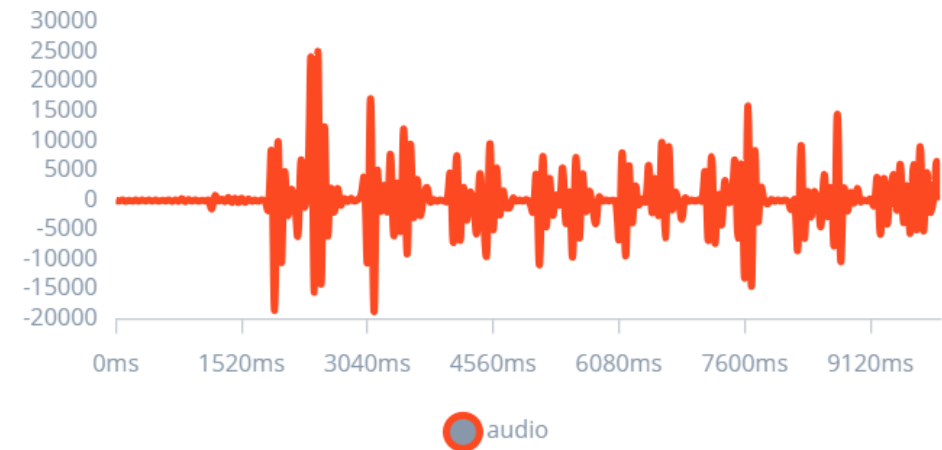
16kHz signal



Identificação de palavras chave

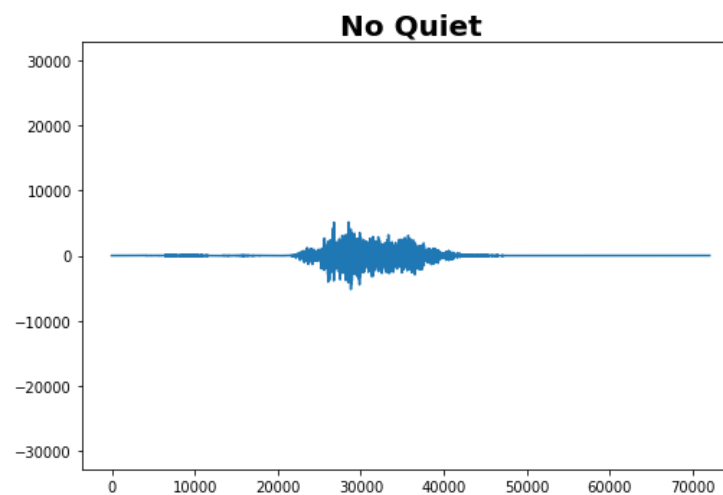
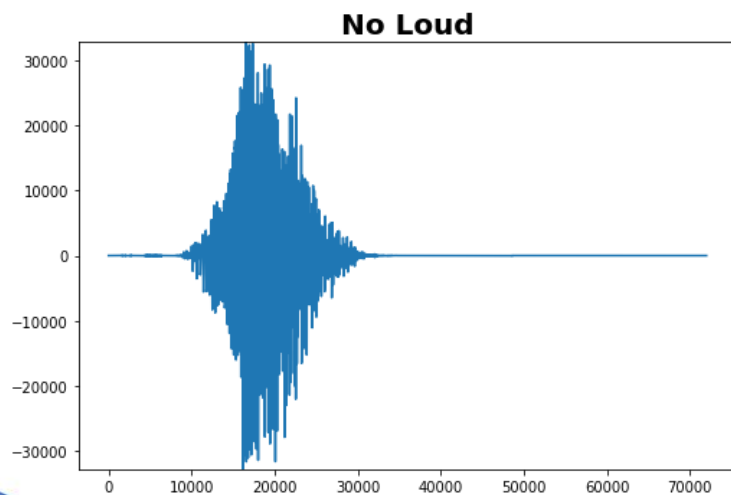
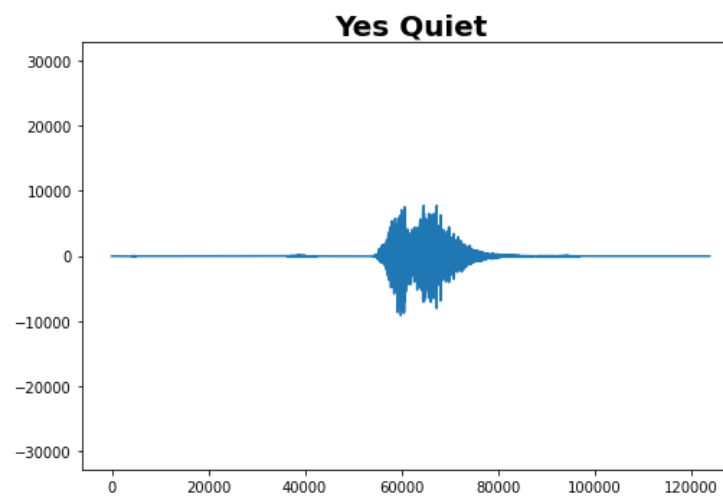
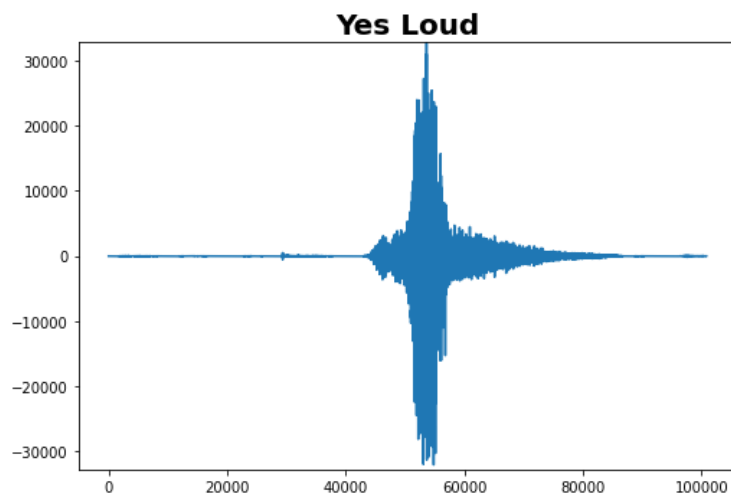
Sinal de áudio ligado/desligado

- Onde a palavra começa?
- Como alinhar?
- Como retirar partes importantes do sinal



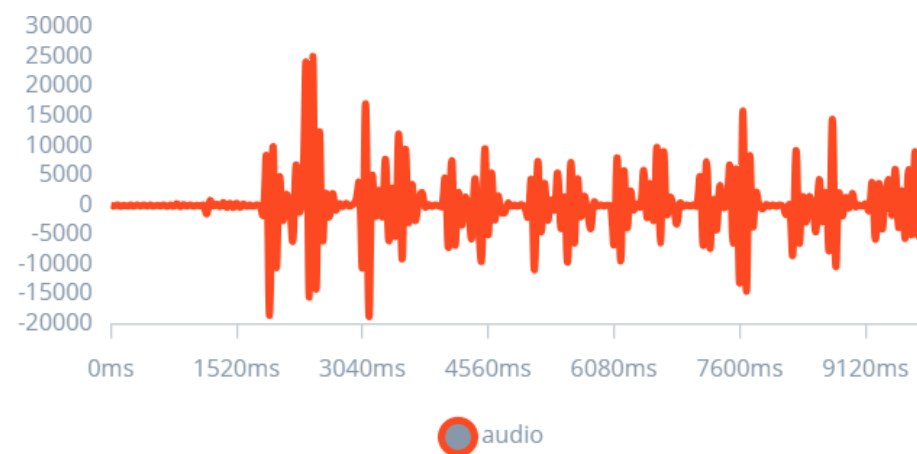
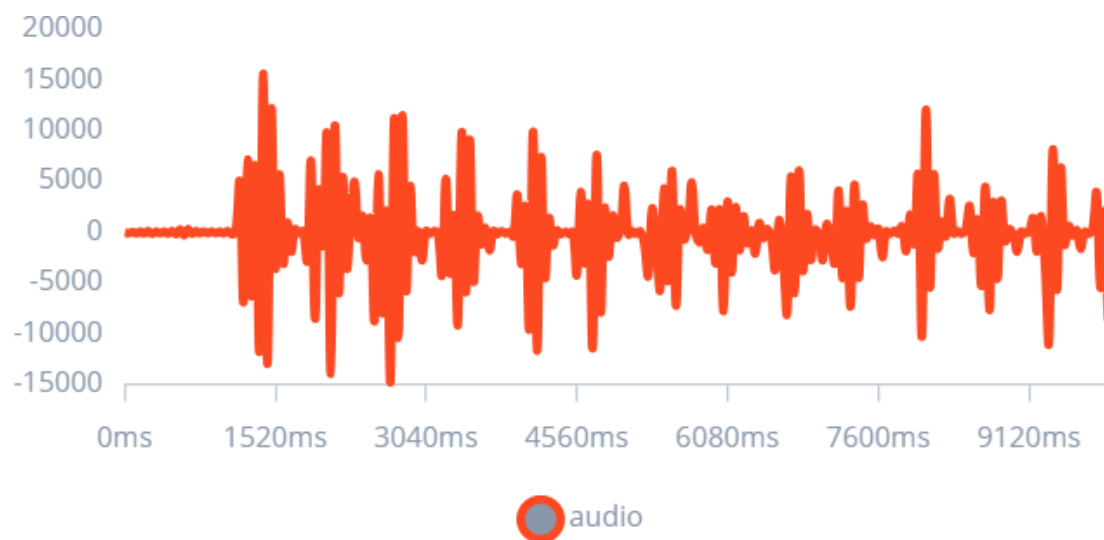
Identificação de palavras chave

Sinal de áudio YES/NO



Identificação de palavras chave

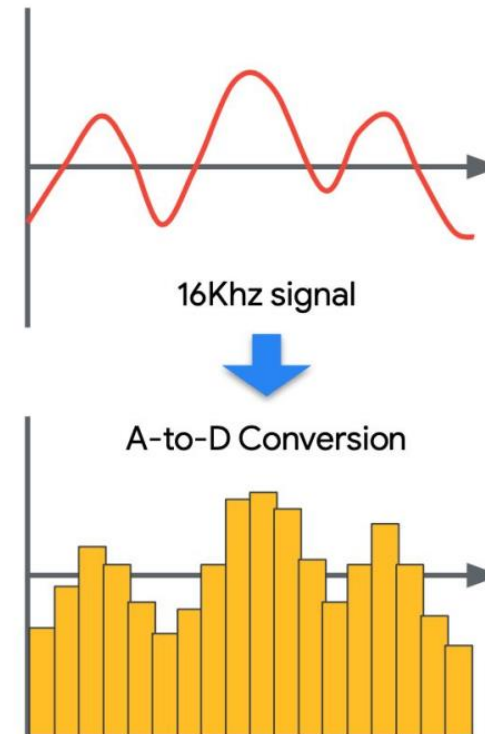
Sinal de áudio ligado/desligado



Identificação de palavras chave

Sinal de áudio

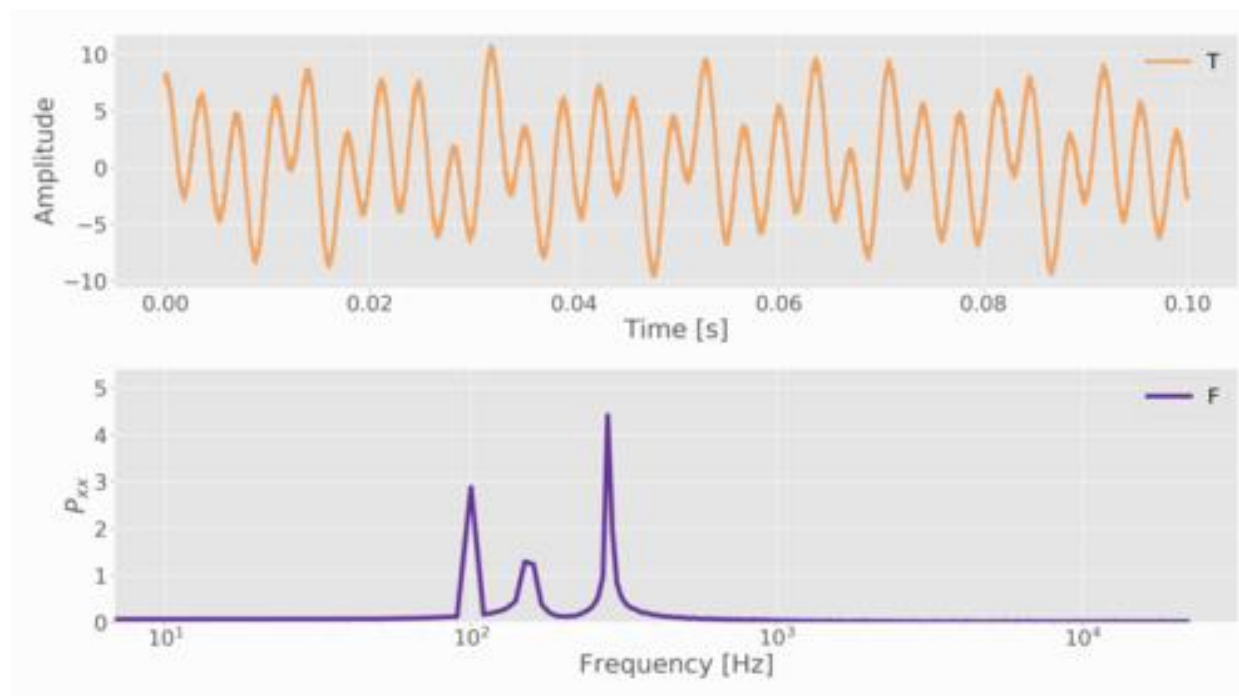
- 16 mil amostras por segundo
- Amostragem A/D



Identificação de palavras chave

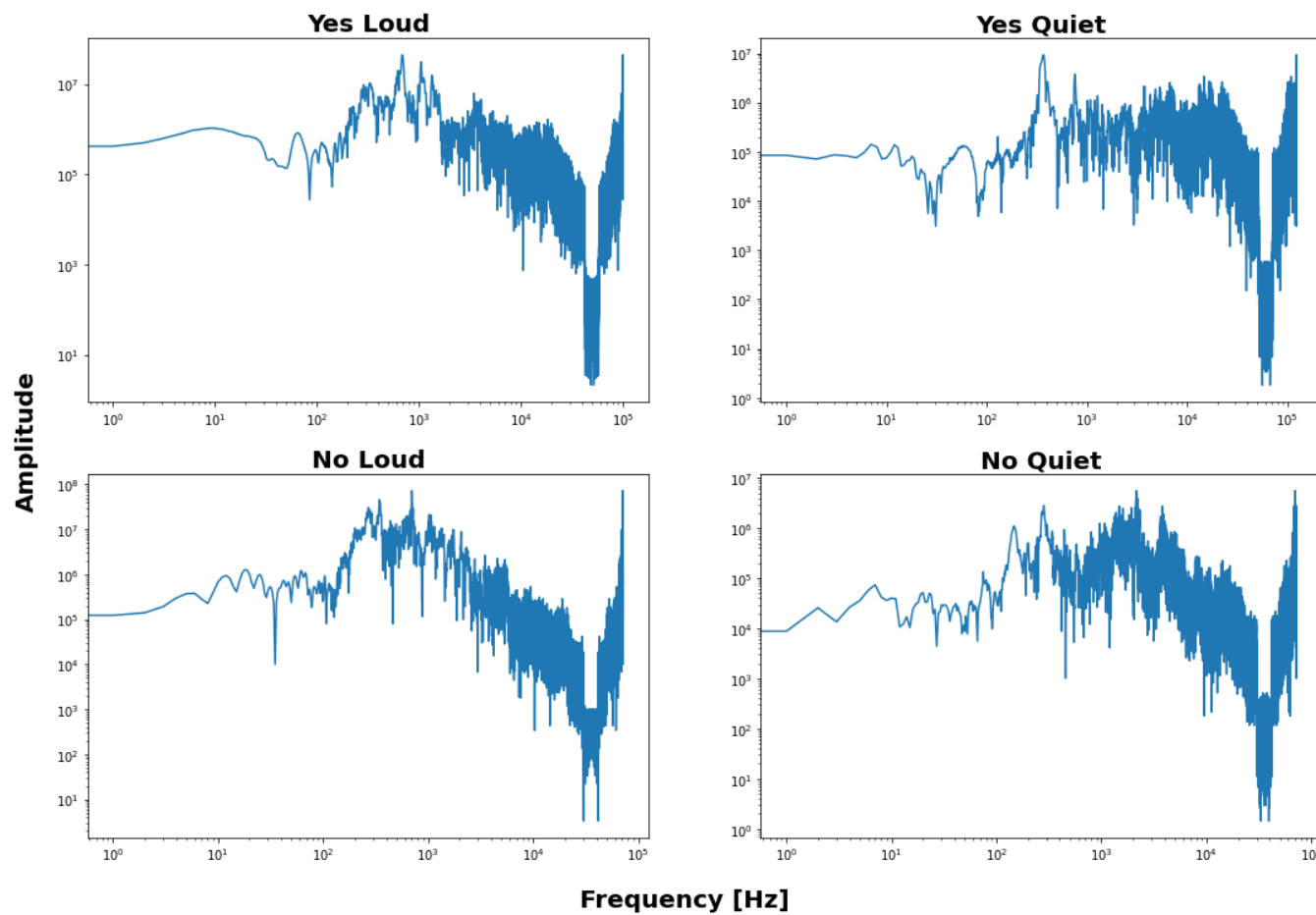
Sinal de áudio

- 16 mil amostras por segundo
- FFT



Identificação de palavras chave

Sinal de áudio YES/NO



Identificação de palavras chave

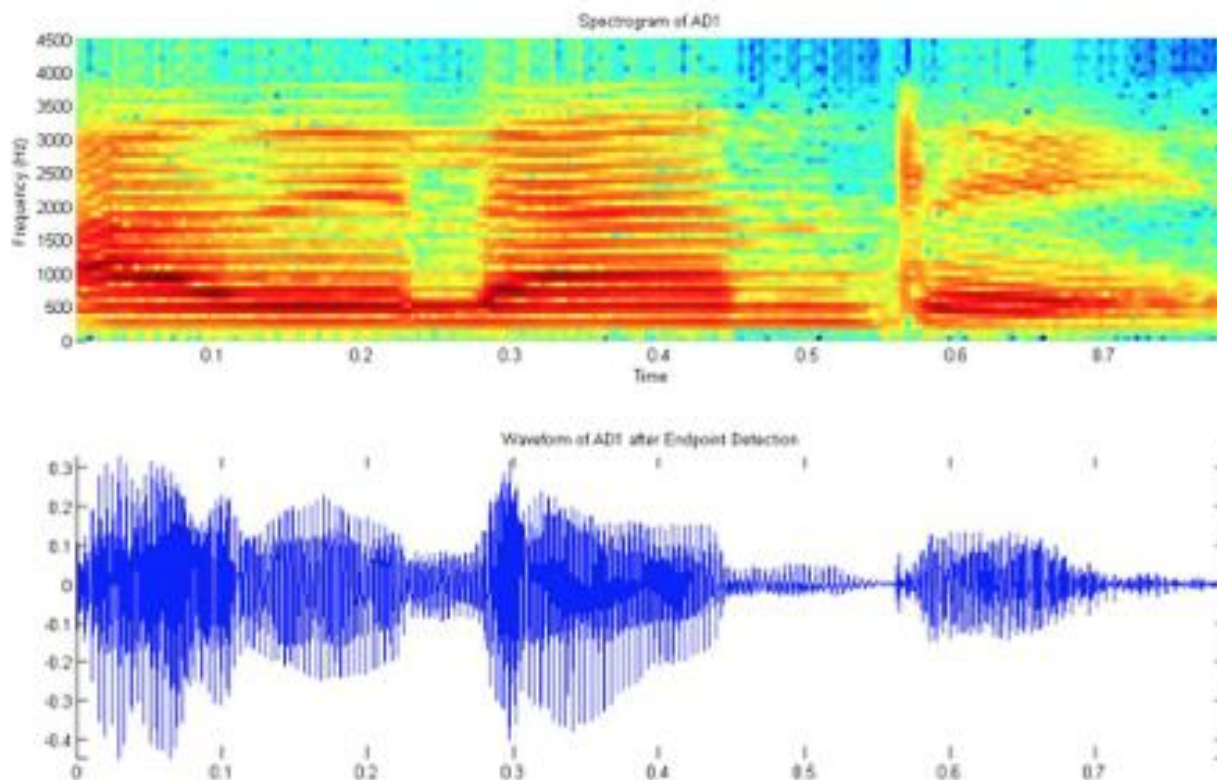
Sinal de áudio

- Espectrograma
- Um espectrograma é uma forma de visualizar a intensidade de um sinal através do tempo e em várias frequências.
- Os espectrogramas são gráficos de duas dimensões (frequência x tempo), com a terceira dimensão, a amplitude, sendo representada pela variação das cores

Identificação de palavras chave

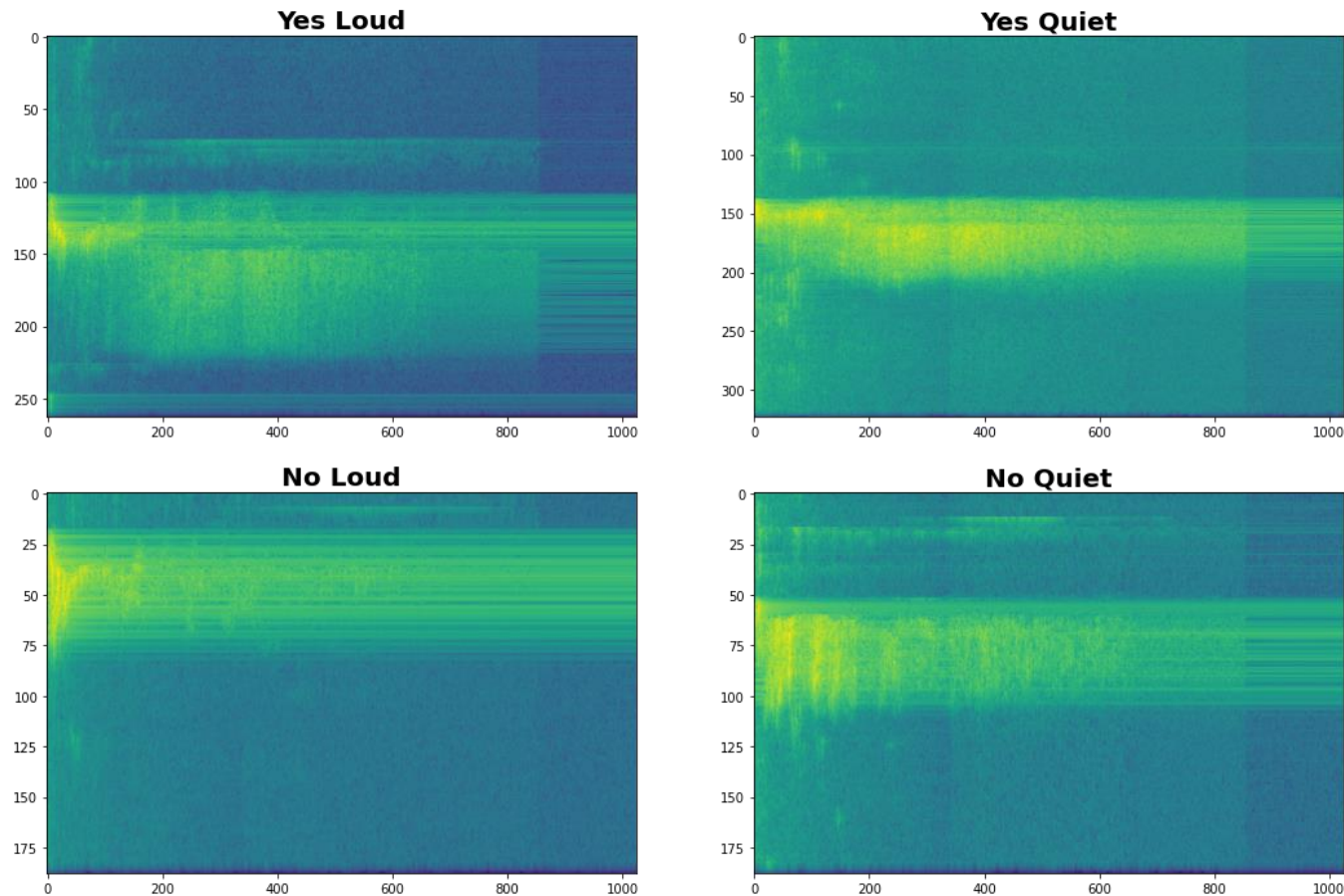
Sinal de áudio

- 16 mil amostras por segundo
- FFT
- Espectrograma
- Janelas



Identificação de palavras chave

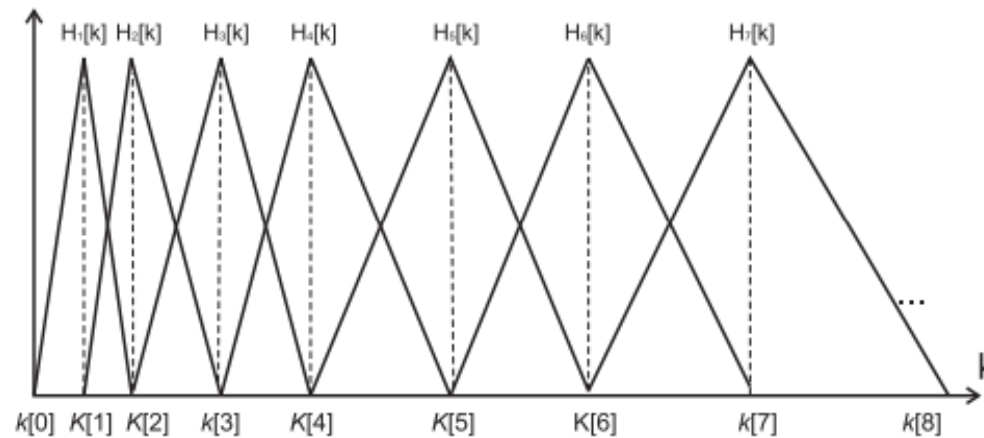
Sinal de áudio



Identificação de palavras chave

Sinal de áudio

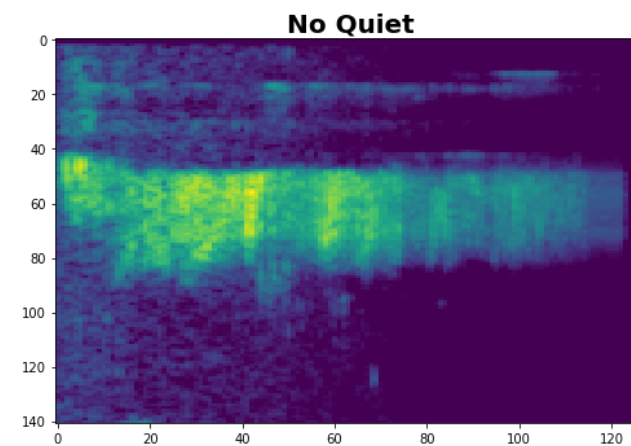
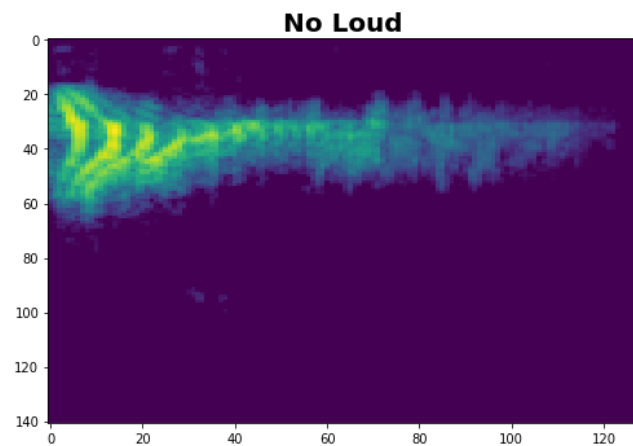
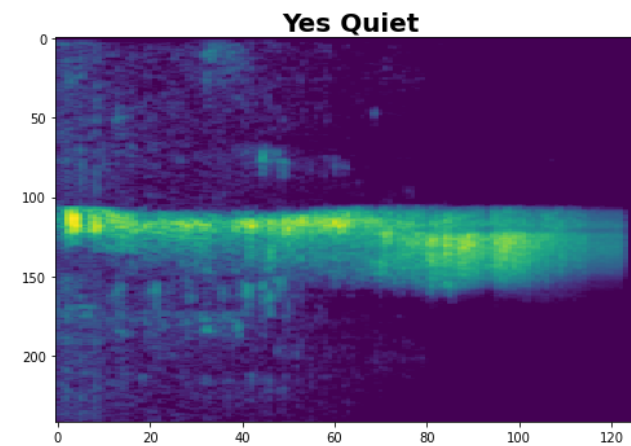
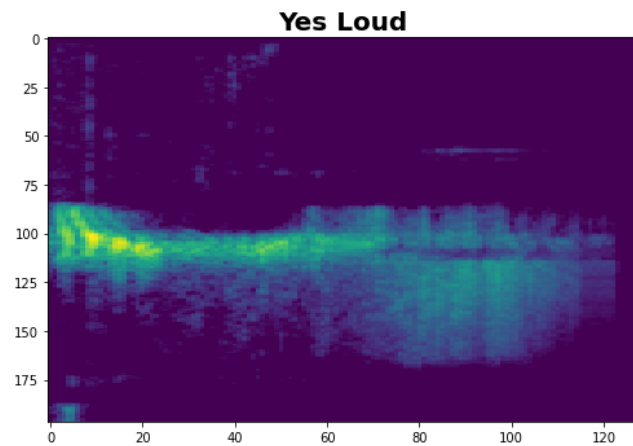
- Filtros MEL
- A escala Mel relaciona a frequência percebida, ou tom, de um tom puro à sua frequência real medida. Os humanos são muito melhores em discernir pequenas mudanças no tom em frequências baixas do que em frequências altas.



<https://towardsdatascience.com/getting-to-know-the-mel-spectrogram-31bca3e2d9d0>

Identificação de palavras chave

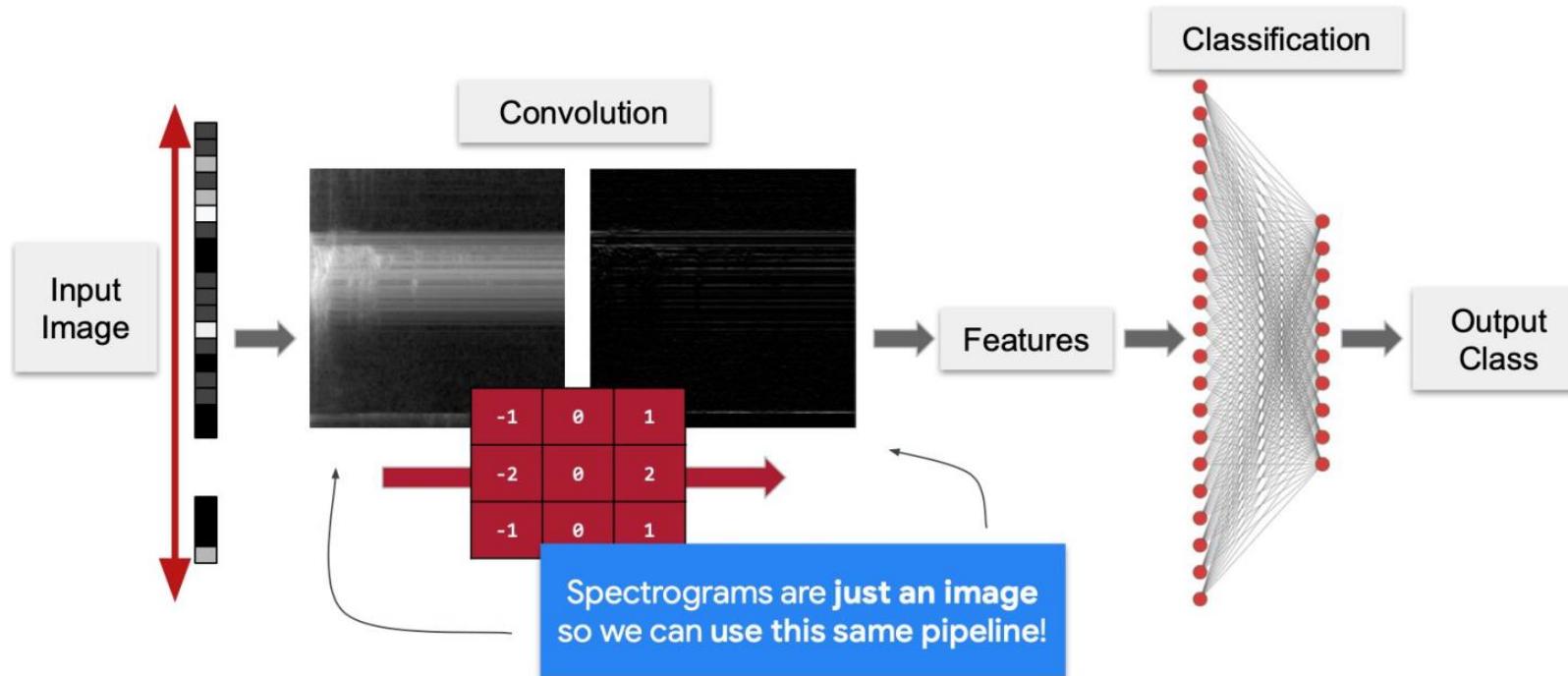
Sinal de áudio



Identificação de palavras chave

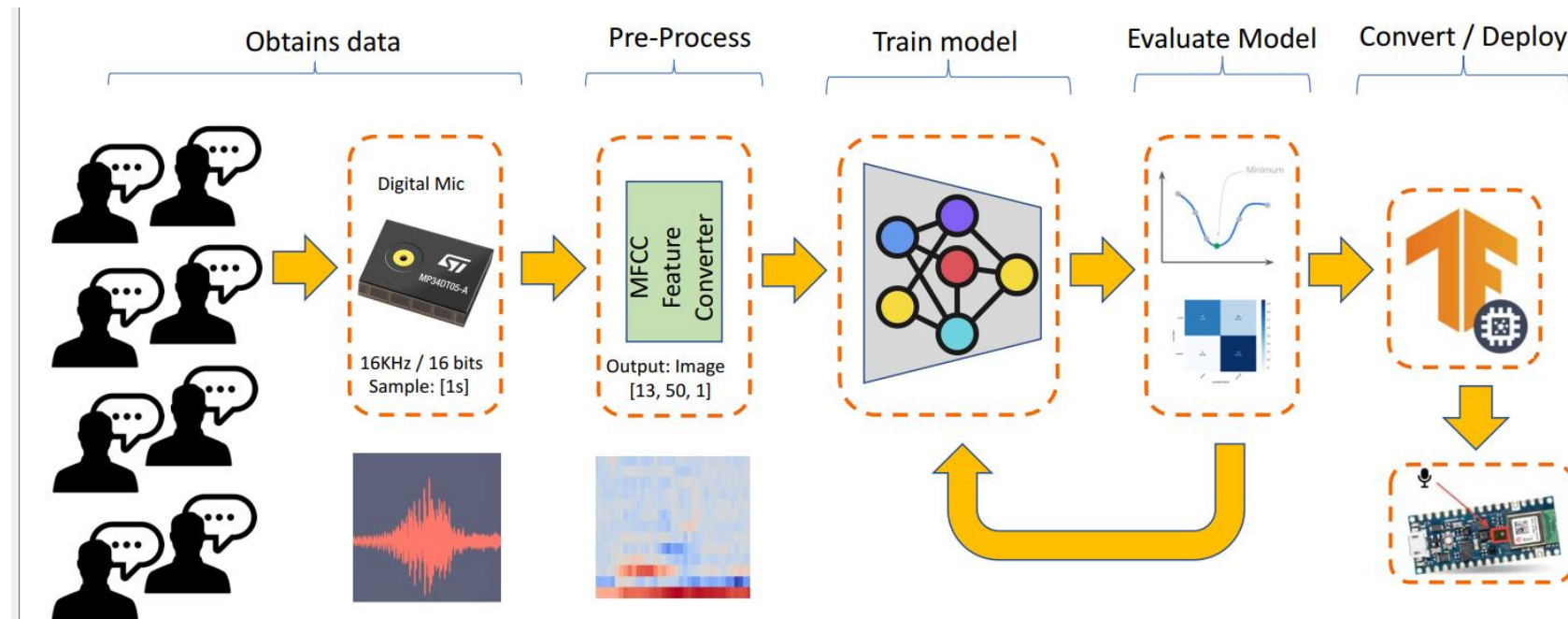
Processo de classificação

A model for **Keyword Spotting**



Identificação de palavras chave

Processo de classificação



Identificação de palavras chave

Trabalho

- Pense em uma aplicação para identificação de palavras chave
- Faça um detalhamento da aplicação
- Quais serão os comandos ou palavras chave?
- Plote os espectrogramas das palavras chave.
- Treine uma rede neural no edge impulse.
- Criar um power point com todas as informações.