

CLASSIFICADOR DE SIRENES

Antonio Apolinario, lara Yasmin, Paulo Sérgio

DATASET

Animals	Natural soundscapes & water sounds	Human, non-speech sounds	Interior/domestic sounds	Exterior/urban noises
Dog	Rain	Crying baby	Door knock	Helicopter
Rooster	Sea waves	Sneezing	Mouse click	Chainsaw
Pig	Crackling fire	Clapping	Keyboard typing	Siren
Cow	Crickets	Breathing	Door, wood creaks	Car horn
Frog	Chirping birds	Coughing	Can opening	Engine
Cat	Water drops	Footsteps	Washing machine	Train
Hen	Wind	Laughing	Vacuum cleaner	Church bells
Insects (flying)	Pouring water	Brushing teeth	Clock alarm	Airplane
Sheep	Toilet flush	Snoring	Clock tick	Fireworks
Crow	Thunderstorm	Drinking, sipping	Glass breaking	Hand saw

- Composto por 2000 arquivos de audio no formato .wav
- 50 classes de sons ambiente
- 40 sons para cada classe, entre elas cachorro, gato, sirene e etc...
- 5 segundos, 44.1 kHz;
- Sem necessidade de pré-processamento;

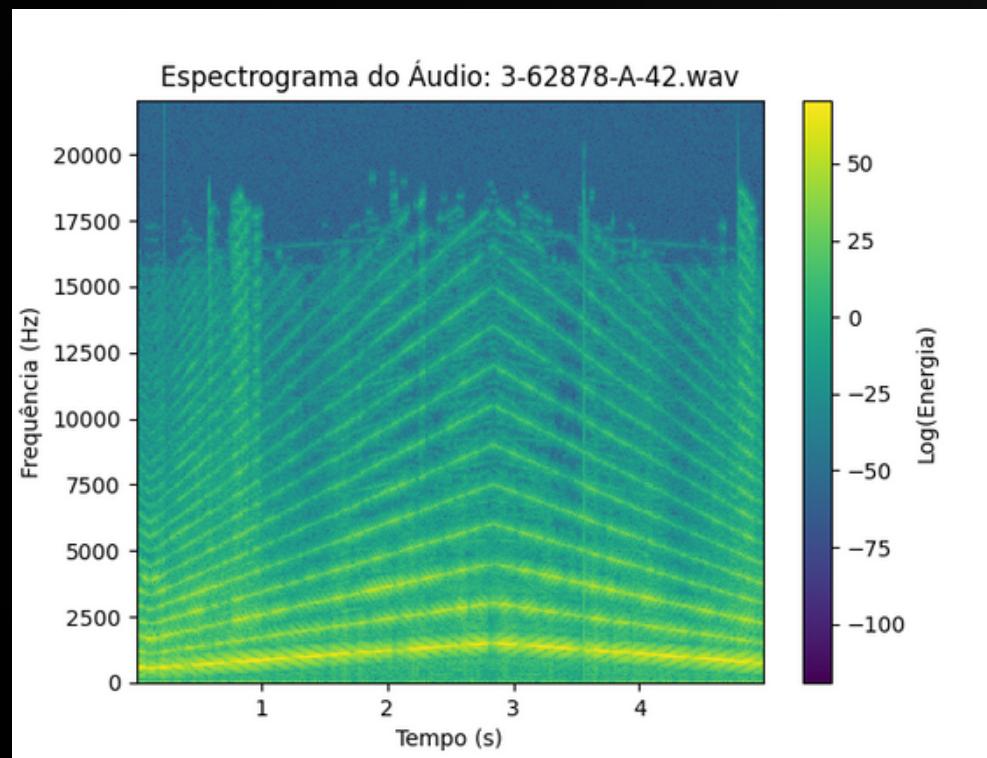
filename	fold	target	category	esc10	src_file	take

EXTRATORES DE CARACTERÍSTICAS

- Transformada rápida de Fourier (FFT);
- Loudness;
- Spectroíde Central;

RETORNO DOS EXTRATORES

FFT

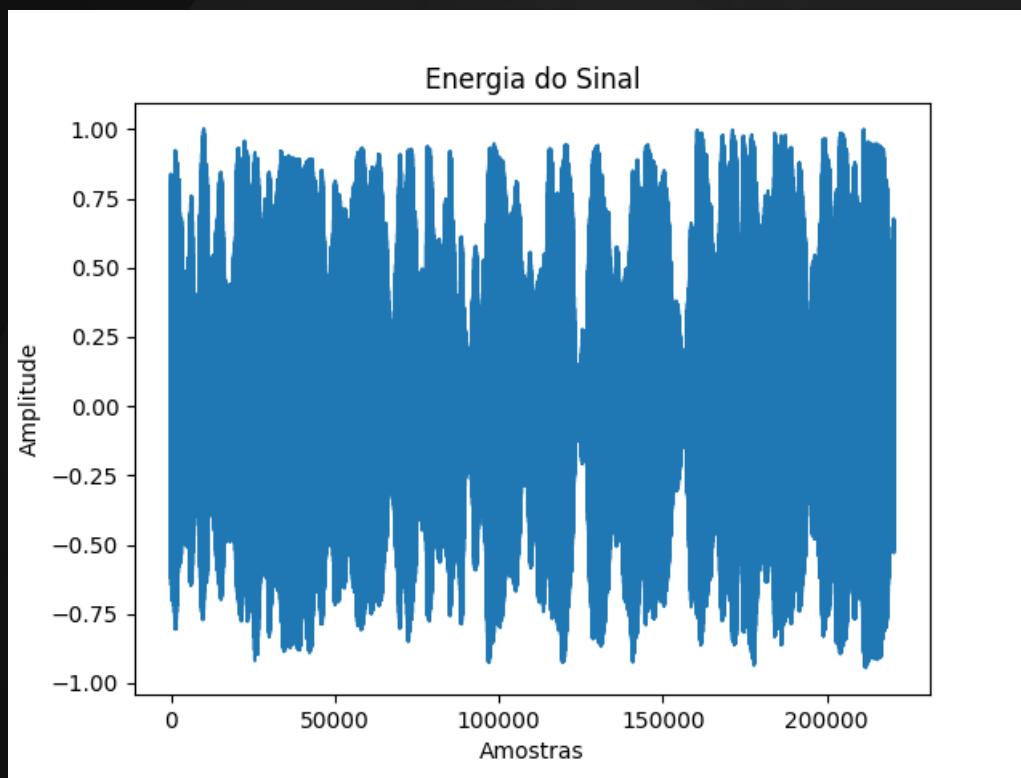


Espectro do sinal: Frequência ao longo do tempo (s).



Identificação de padrões de frequências presentes.

LOUDNESS



Energia ao longo da quantidade de amostras.



Diferentes tipos de sirenes = níveis de intensidade variados.

SPECTRO CENTROÍDE

Standard	Standard
1-31482-A-42.wav	1267.351699440817
1-31482-B-42.wav	1324.936476493196
1-54084-A-42.wav	923.5752554684269
1-76831-A-42.wav	2388.810697915736
1-76831-B-42.wav	1933.226747274848
1-76831-C-42.wav	1943.049223428071
1-76831-D-42.wav	2065.735127546538
1-76831-E-42.wav	2149.051290323320

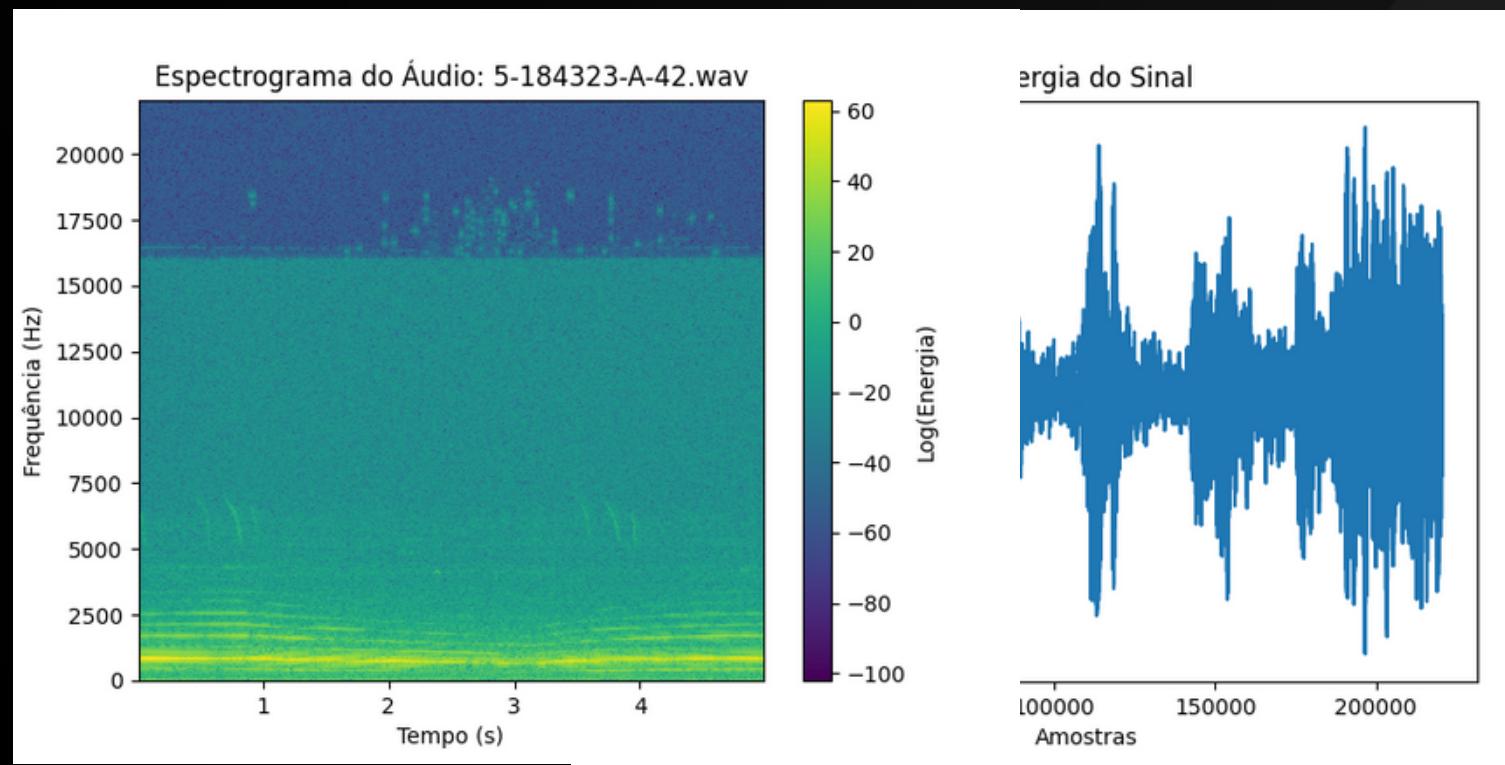
Centro de massa da frequência - Onde estão concentradas as frequências predominantes.



Predominância de frequência mais alta ou baixa, dependendo da classe.

COMO RELACIONAR OS EXTRATORES?

FFT + LOUDNESS + SPECTRO CENTROÍDE



Standard
1267 . 351699440817
1324 . 936476493196
923 . 5752554684269
2388 . 810697915736
1933 . 226747274848
1943 . 049223428071
2065 . 735127546538
2149 . 051290323320

- Normaliza as imagens (escala cinza);
- Transforma todas as imagens em um vetor 1D;
- Combina as 3 características em um vetor:

Shape of combined features array: (40, 614401)
(audios, qntTotal de carac)

COMO SEPARAR E ROTULAR?

- Utilizando K-means para agrupar os dados de acordo com suas características

```
#qnt de gp
num_clusters = 4

#inicializacao do modelo kmeans
kmeans = KMeans(n_clusters=num_clusters, random_state=42)
kmeans.fit(X_scaled) #dados normalizados

#obter rotulo p/ cada audio
cluster_labels = kmeans.labels_
```

- Obtém rotulos e mapea os clusters para classes :D

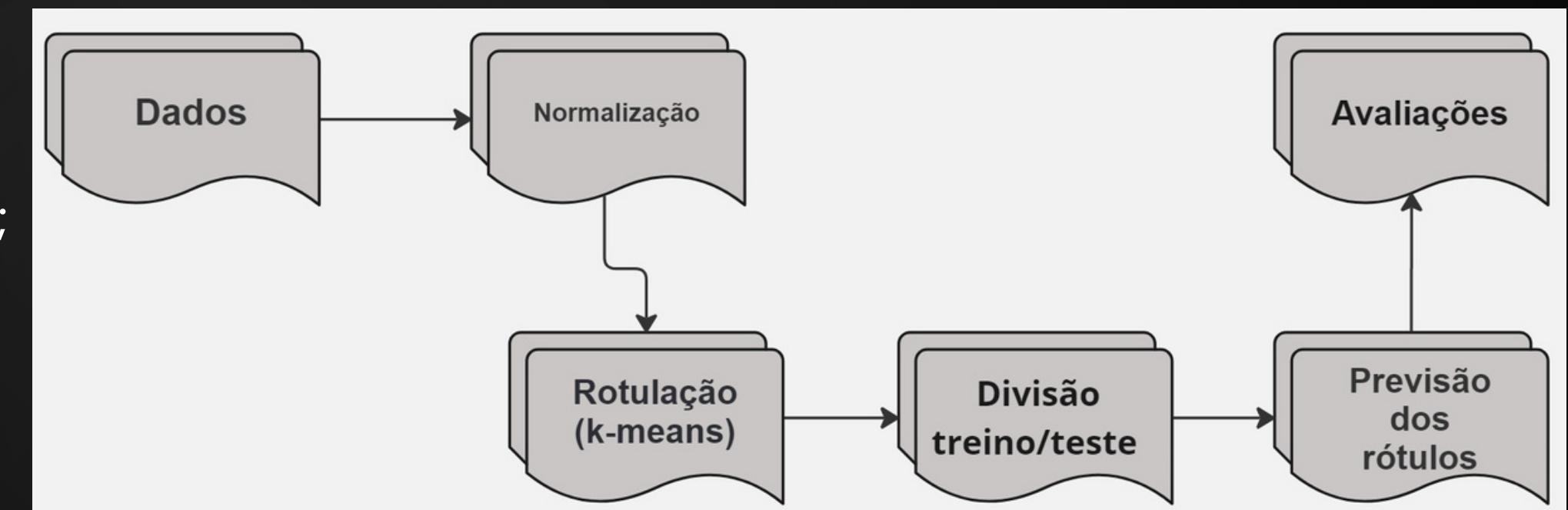
```
def atribuir_rotulos_cluster(cluster_labels):
    #mapeando rotulos de clusters p classes
    mapping = {0: "ambulância", 1: "bombeiro", 2: "polícia", 3: "alarme"}
    rotulos_classe = [mapping[label] for label in cluster_labels]
    return rotulos_classe
```

CLASSIFICADOR SVM

(máquina de vetores de suporte)

Motivações:

- Poucos hiperparâmetros para ajustar;
- Lida bem com peq. conjuntos de dados / ruidosos;
- Menos sensível a outliers;
- Modelo aprende com ruídos além de padrões;



- Hiperparâmetros: kernel, random_state = 40, n_clusters = 4
Variação dos hiperparâmetros → Impacto na classificação
(melhor faixa para teste entre 0.2 e 0.32)
- Classes: Alarme, âmbulância, bombeiro e polícia;

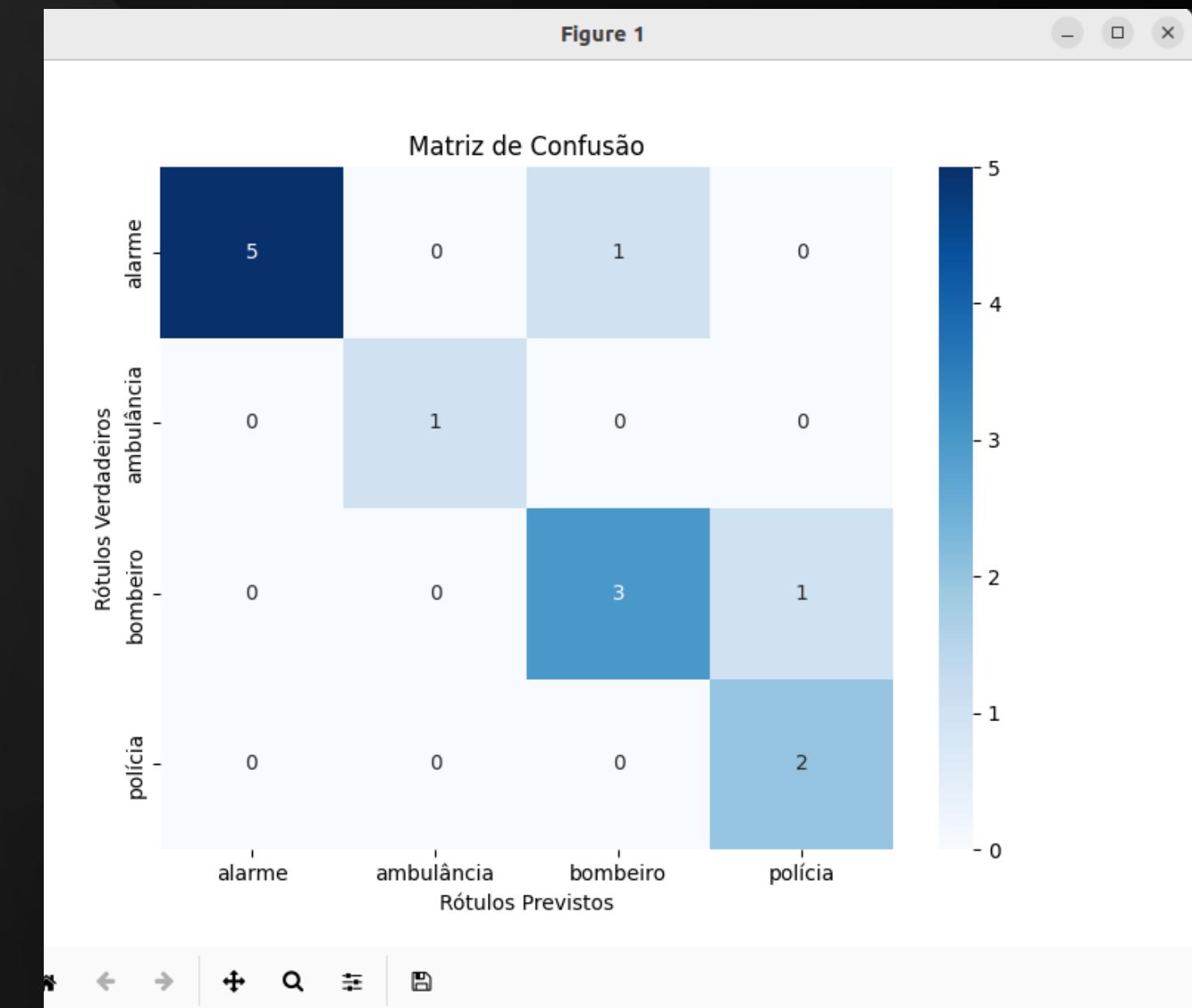
RESULTADOS

84.6%

ACURÁCIA

Relatório de classificação:

	precision	recall	f1-score	support
alarme	1.00	0.83	0.91	6
ambulância	1.00	1.00	1.00	1
bombeiro	0.75	0.75	0.75	4
policia	0.67	1.00	0.80	2
accuracy			0.85	13
macro avg	0.85	0.90	0.86	13
weighted avg	0.87	0.85	0.85	13



DESAFIOS

X

MELHORIAS FUTURAS

- Semelhança entre os áudios;
- Utilização de vetores 1D (classificador);
- Divisão randômica de treino e teste p/ uma base de dados pequena
- Maior base de dados;
- Extração de mais características?!
- Divisão manual treino/teste;
- Mudança de classificador;

OBRIGADO!