

Detección acústica de especies salvajes

Dante Bermúdez Marbán
Aprendizaje Profundo

Contexto

- El cambio climático es uno de los problemas más grandes que enfrentamos todos
- Selvas importantes debido a que absorben gran parte del dióxido de carbono



Contexto

— — —



Descripción del problema

— — —



kaggle

Para cada audio, indicar
qué **especies** están
presentes

- Problema de
clasificación **multi
etiqueta**

- train_tp.csv
- train_fp.csv
- train/*.flac
- test/*.flac

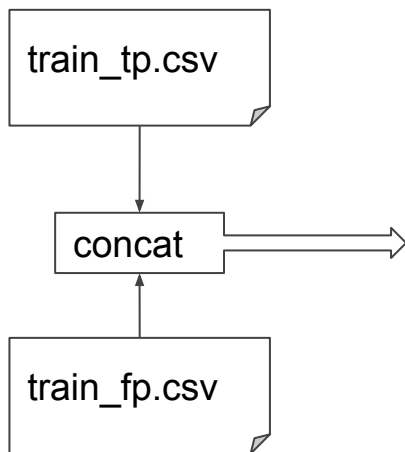
Hay datos con
falsos positivos



CSVs

	recording_id	species_id	songtype_id	t_min	f_min	t_max	f_max
0	003bec244	14	1	44.5440	2531.250	45.1307	5531.25
1	006ab765f	23	1	39.9615	7235.160	46.0452	11283.40
2	007f87ba2	12	1	39.1360	562.500	42.2720	3281.25
3	0099c367b	17	4	51.4206	1464.260	55.1996	4565.04
4	009b760e6	10	1	50.0854	947.461	52.5293	10852.70
...
1211	fe8d9ac40	13	1	53.4720	93.750	54.0960	843.75
1212	fea6b438a	4	1	43.5787	2531.250	45.7653	4031.25
1213	ff2eb9ce5	0	1	15.2267	5906.250	16.0213	8250.00
1214	ffb8d8391	5	1	14.3467	4781.250	16.6987	10406.20
1215	ffb9a7b9a	18	1	40.3200	3187.500	41.0133	5062.50

CSVs



	recording_id	species_id	songtype_id	t_min	f_min	t_max	f_max	is_tp	duration	bandwidth
0	003bec244	14	1	44.5440	2531.250	45.1307	5531.25	True	0.5867	3000.000
1	006ab765f	23	1	39.9615	7235.160	46.0452	11283.40	True	6.0837	4048.240
2	007f87ba2	12	1	39.1360	562.500	42.2720	3281.25	True	3.1360	2718.750
3	0099c367b	17	4	51.4206	1464.260	55.1996	4565.04	True	3.7790	3100.780
4	009b760e6	10	1	50.0854	947.461	52.5293	10852.70	True	2.4439	9905.239
...
7776	ffd88cd84	14	1	3.2000	2531.250	3.7867	5531.25	False	0.5867	3000.000
7777	ffebe7313	1	1	35.2000	3843.750	36.0960	5625.00	False	0.8960	1781.250
7778	fff163132	17	4	22.5547	1312.500	25.0880	7406.25	False	2.5333	6093.750
7779	fff163132	14	1	37.0827	2531.250	37.6693	5531.25	False	0.5866	3000.000
7780	fffb79246	6	1	8.3573	562.500	10.5013	4406.25	False	2.1440	3843.750

8997 rows × 10 columns

s0 s1 s2 s3 s4 s5 s6 s7 s8 s9 ... s14 s15 s16 s17 s18 s19 s20 s21 s22 s23

recording_id

00204008d -1 -1 -1 -1 0 -1 -1 -1 0 -1 ... -1 -1 -1 -1 -1 -1 -1 0 -1 -1

003b04435 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ... -1 -1 -1 -1 -1 -1 -1 -1 0 0

003bec244 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ... 1 -1 -1 -1 -1 -1 -1 -1 -1 -1

005f1f9a5 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ... -1 -1 -1 -1 -1 -1 -1 -1 -1 -1

006ab765f -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ... -1 -1 -1 -1 -1 -1 -1 -1 -1 1

...

ffc6031f8 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ... -1 -1 -1 0 -1 -1 -1 -1 -1 0

ffd88cd84 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ... 0 -1 -1 0 -1 -1 -1 -1 -1 -1

ffebe7313 -1 0 -1 -1 -1 -1 -1 -1 -1 -1 -1 ... -1 -1 -1 -1 -1 -1 -1 -1 -1 -1

fff163132 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ... 0 -1 -1 0 -1 -1 -1 -1 -1 -1

fffb79246 -1 -1 -1 -1 -1 -1 0 -1 -1 -1 -1 ... -1 -1 -1 -1 -1 -1 -1 -1 -1 -1

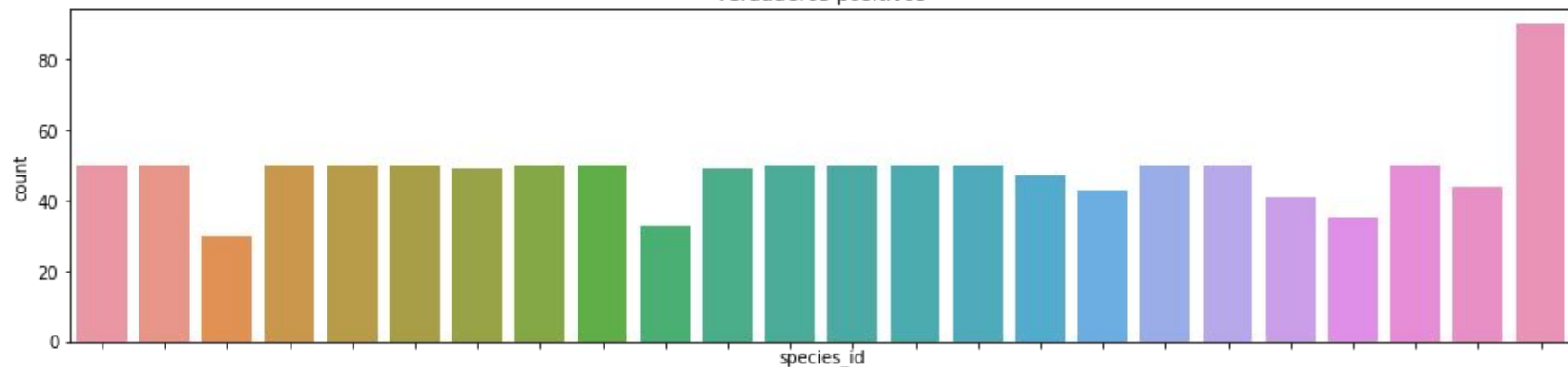
Duplicados

	recording_id	species_id	songtype_id	t_min	f_min	t_max	f_max	is_tp
19	03b96f209	16	4	30.9333	3093.75	32.7680	4593.75	True
20	03b96f209	16	4	49.6320	3093.75	51.4667	4593.75	True

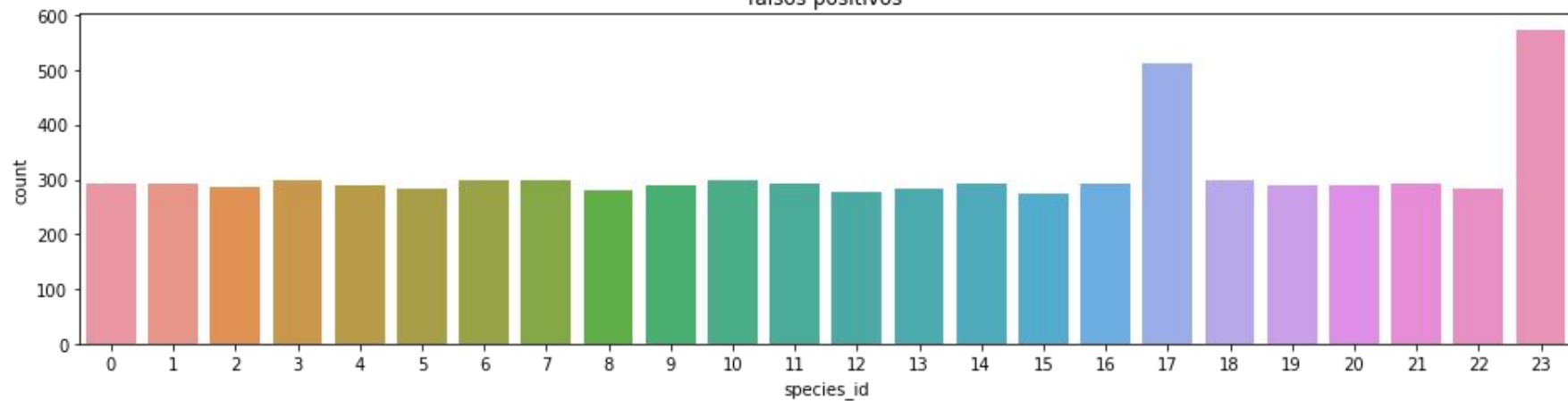
	recording_id	species_id	songtype_id	t_min	f_min	t_max	f_max	is_tp
56	0c48ed342	17	1	50.0320	1312.50	56.4853	3937.50	True
1617	0c48ed342	2	1	53.6640	468.75	55.4400	3000.00	False
1618	0c48ed342	19	1	47.4933	281.25	49.0453	2812.50	False
1619	0c48ed342	17	4	43.5413	1312.50	46.0747	7406.25	False

Distribución de especies

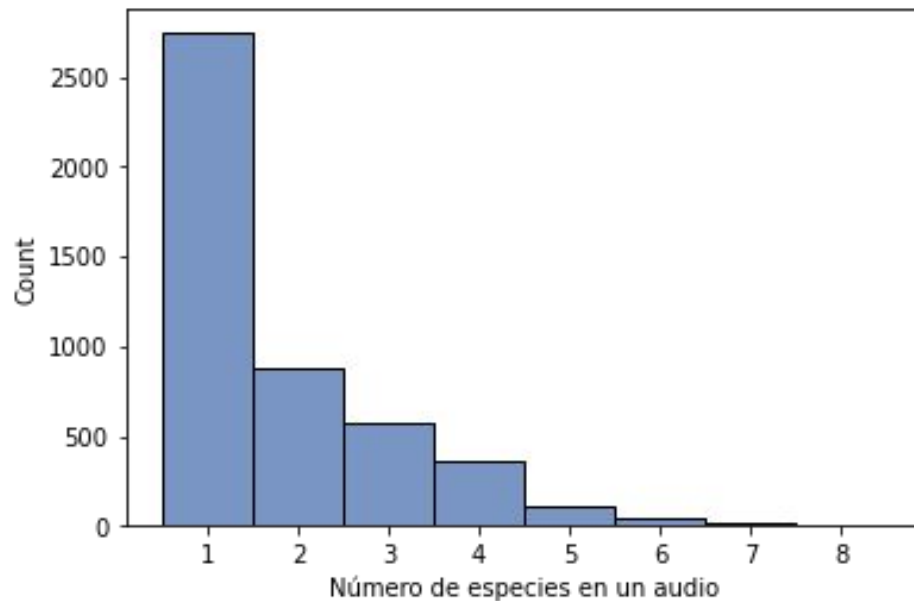
verdaderos positivos



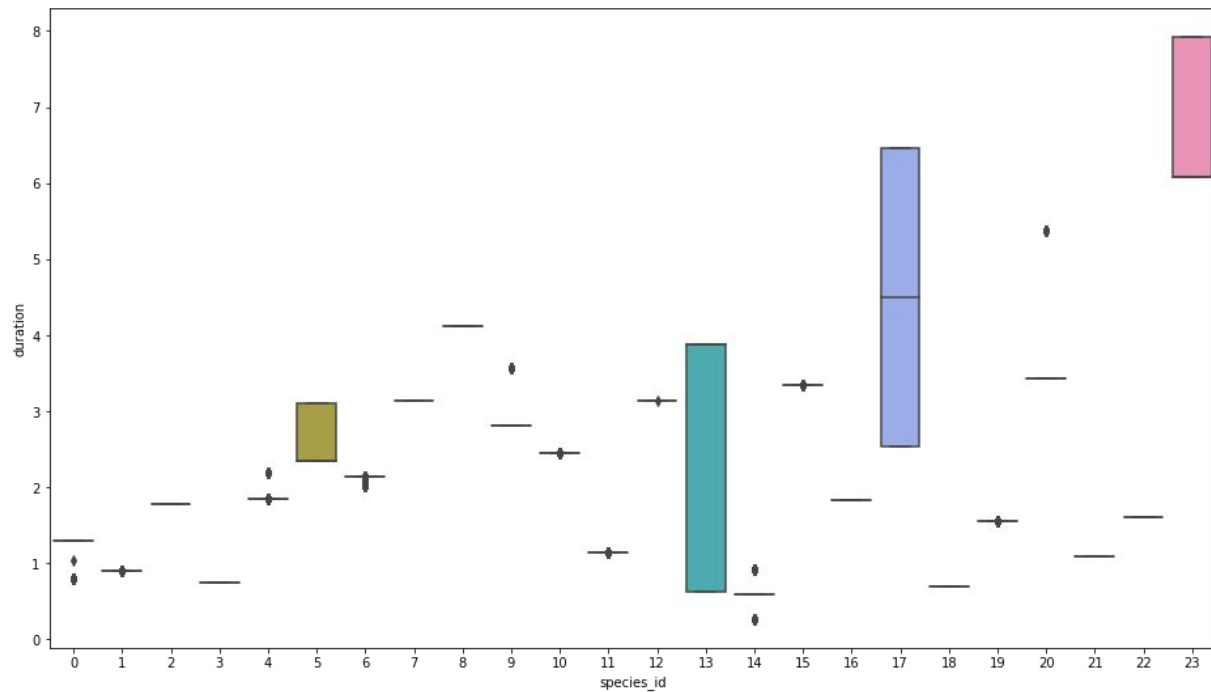
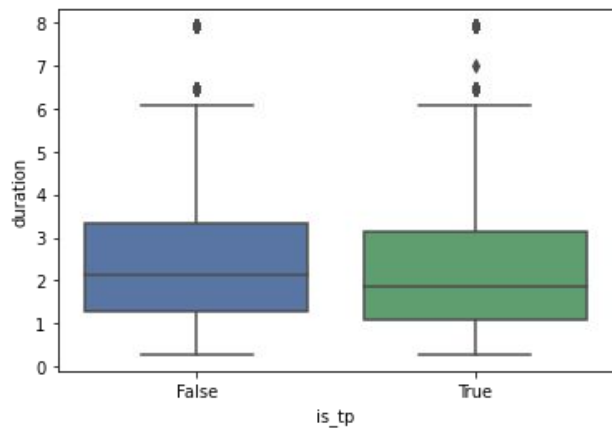
falsos positivos



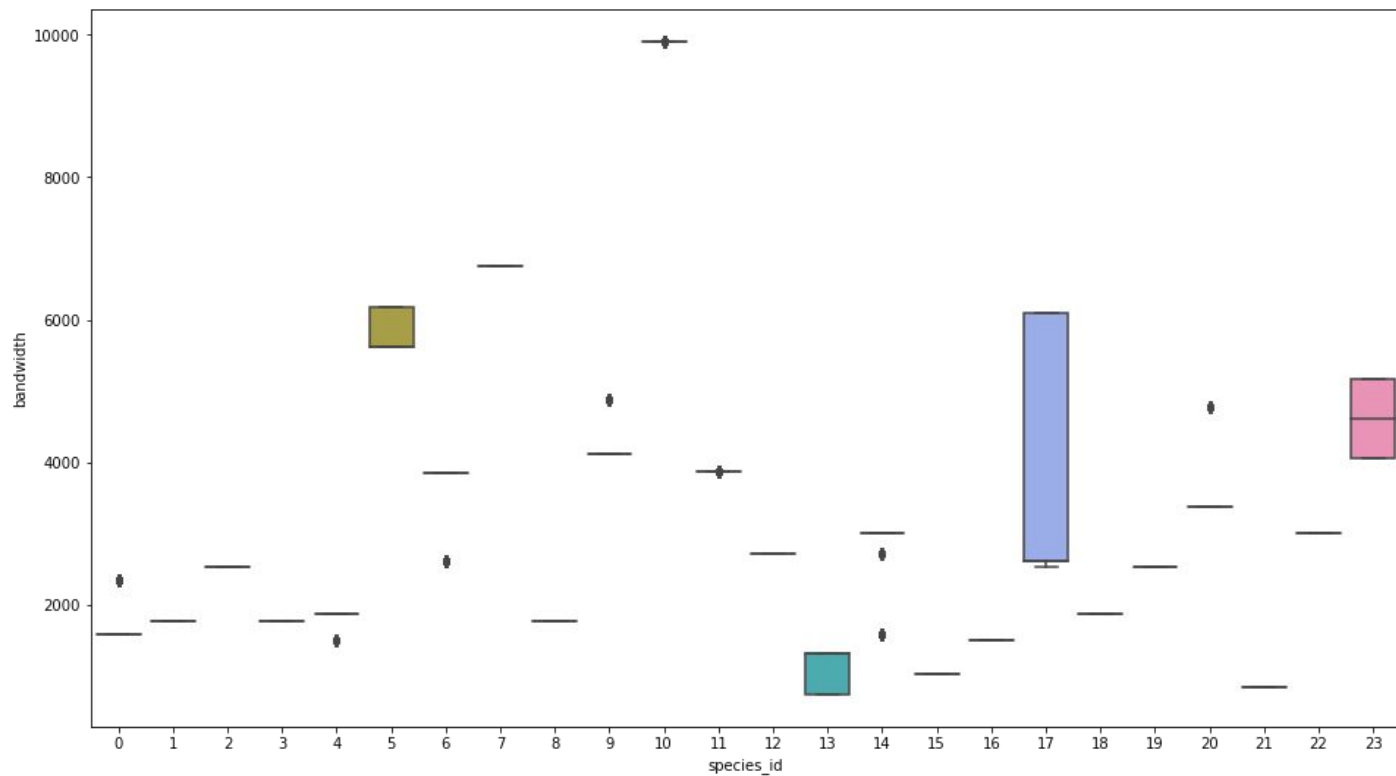
Cantidad de especies por audio



Duración

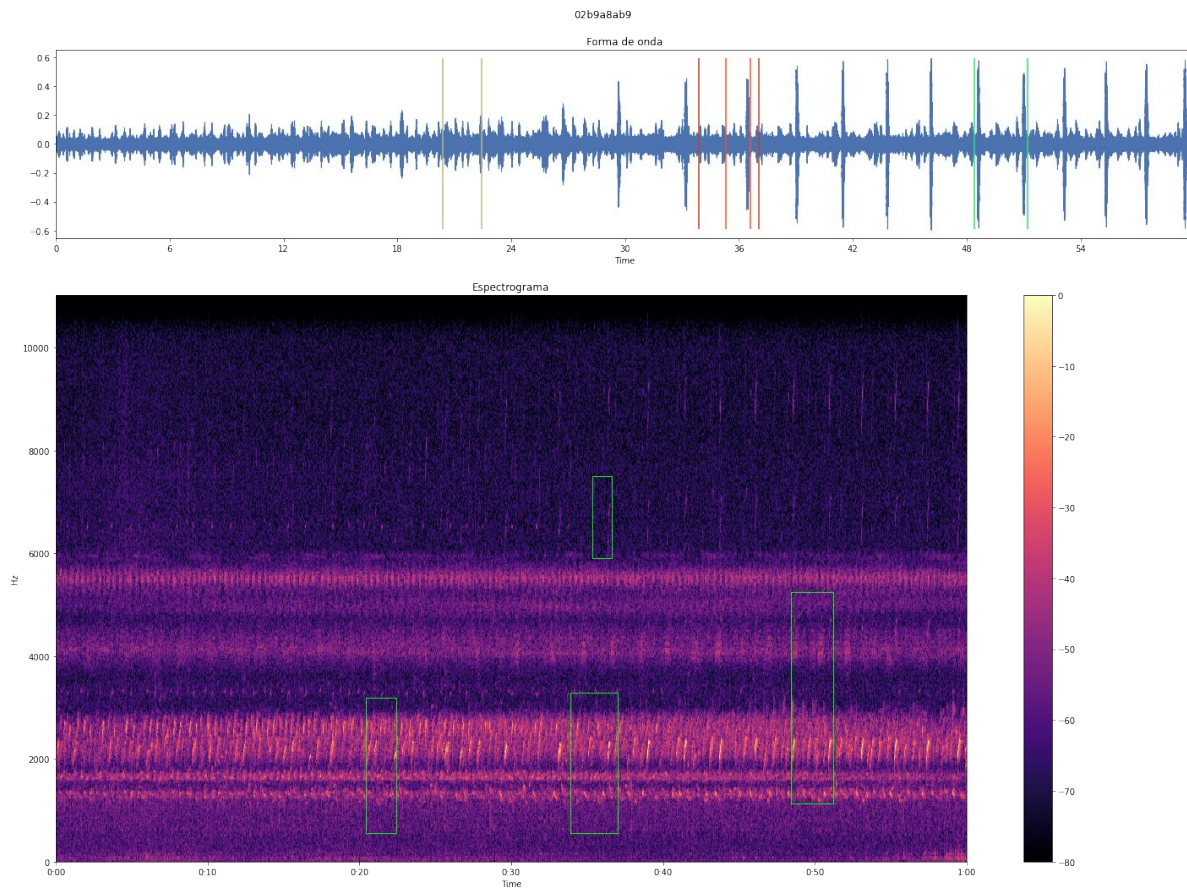


Ancho de banda



Audios

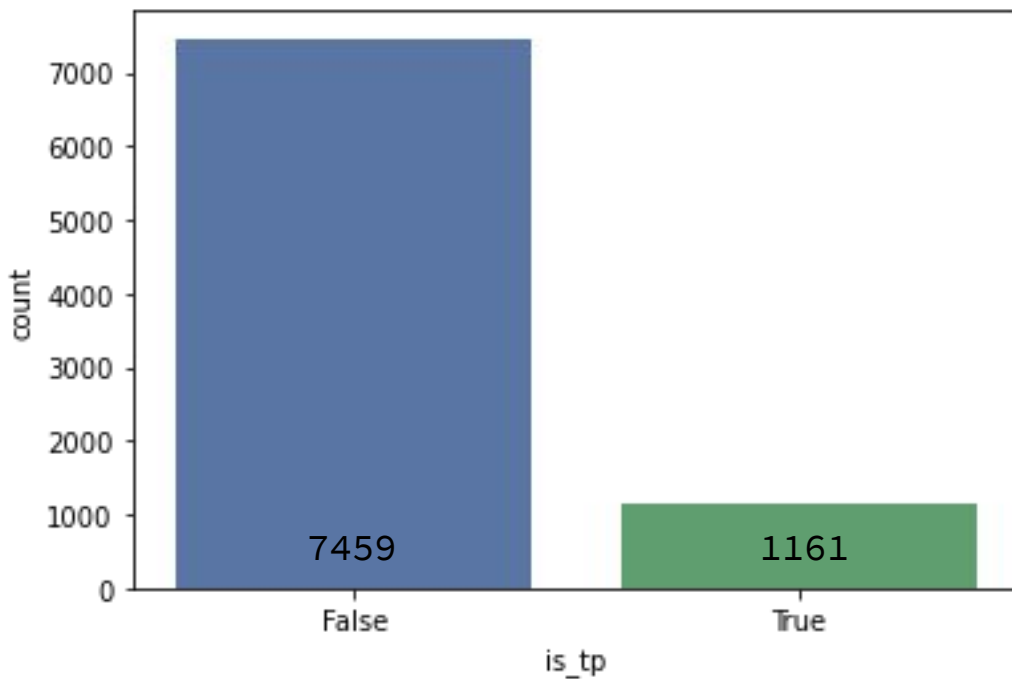
- 1 minuto
- 4727 para entrenamiento
- 1992 para predicción



Aumento de datos

— — —

- Lidar desbalanceo de falsos positivos
- Solamente se aplicó en “audios buenos”.
- Tres técnicas
 - Ruido gaussiano
 - Desplazamiento
 - Cambio de tono



Aumentado de datos

— — —

Ruido gaussiano

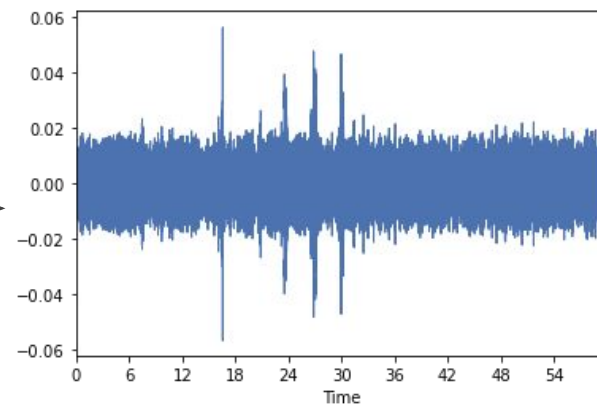
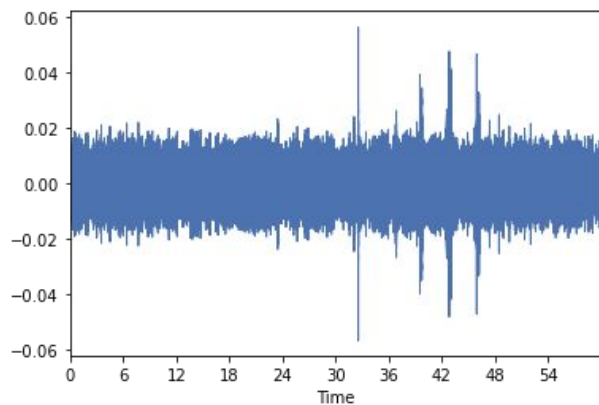
```
audio + factor*ruido_gaussiano = nuevo_audio
```



Aumentado de datos

— — —

Desplazamiento

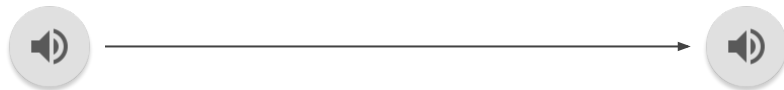


Aumentado de datos

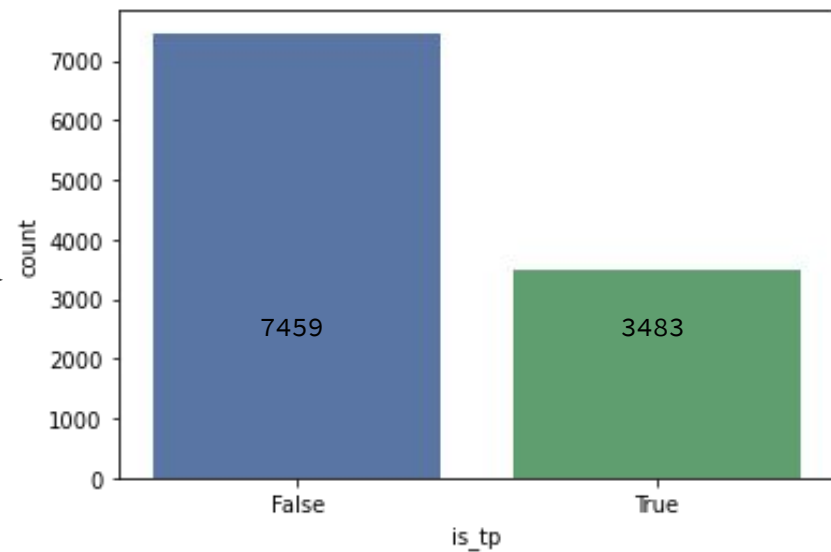
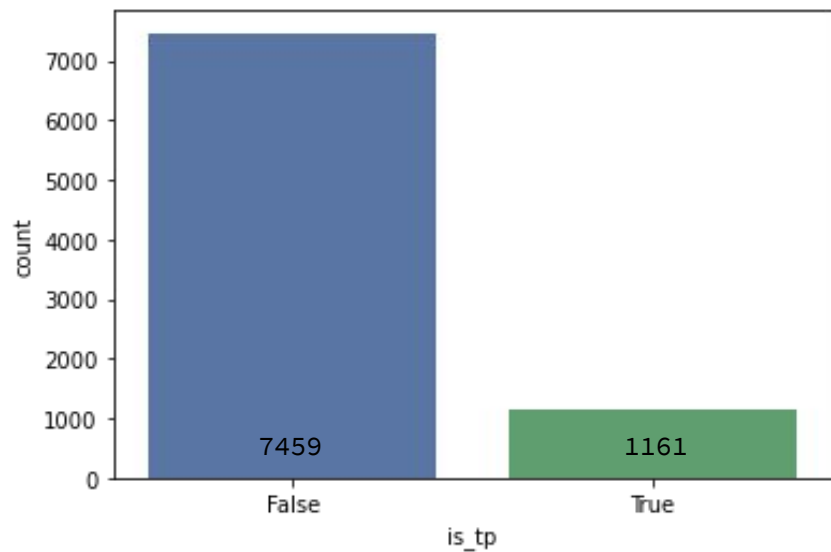
— — —

Cambio de tono

- Se agregaban (o quitaban) uno o dos semitonos
- Se escucha más grave o más agudo



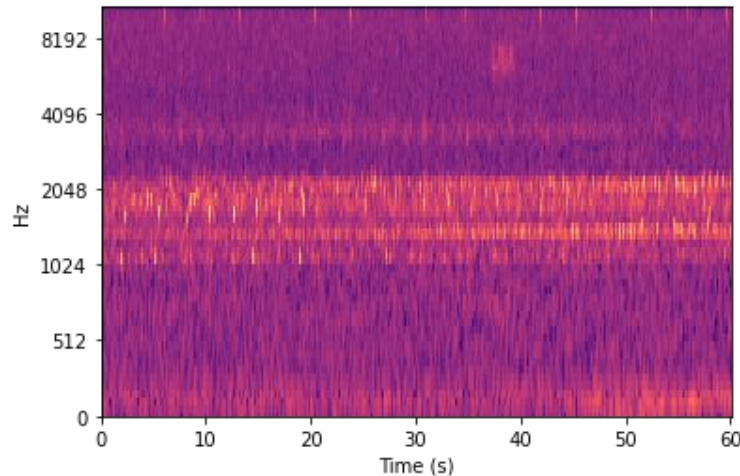
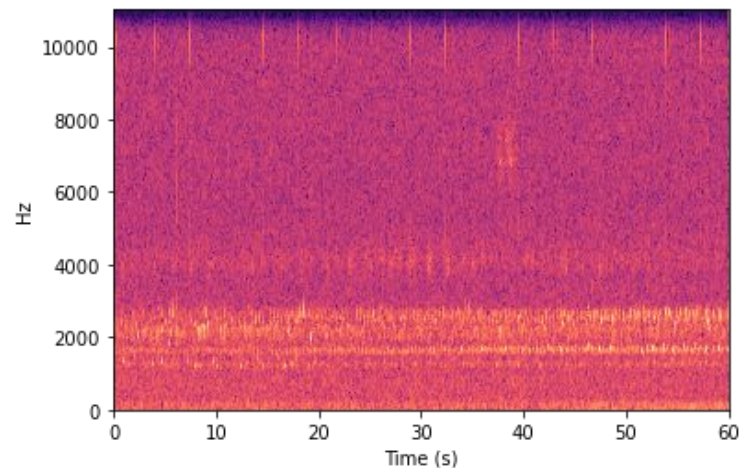
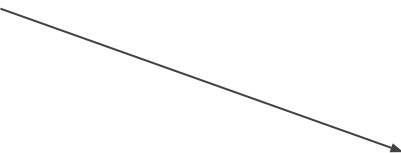
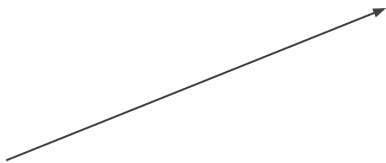
Aumentado de datos



Representaciones tiempo- frecuencia

— — —

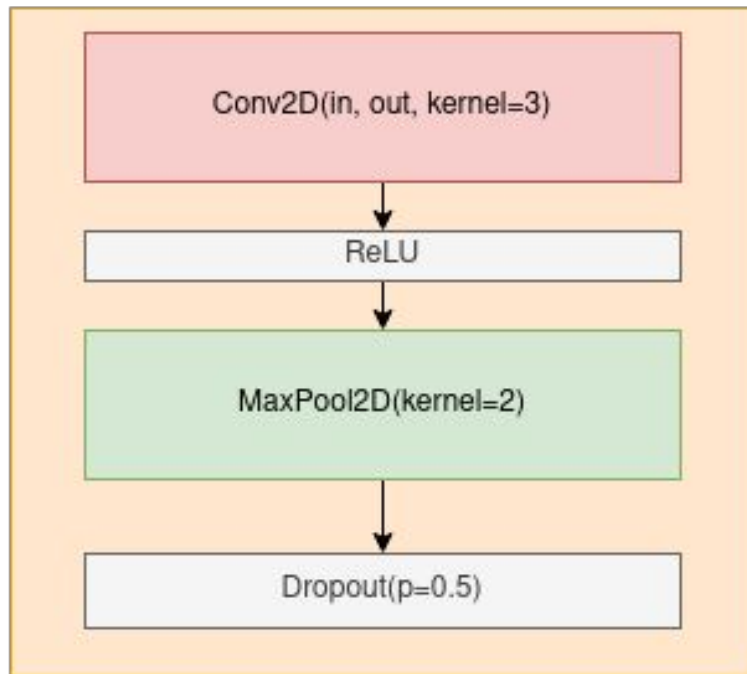
- $N_{\text{FFT}} = 400$
- Espectrograma en escala lineal
- Espectrograma en escala de Mel
 - $n_{\text{mels}} = 64$



Modelo

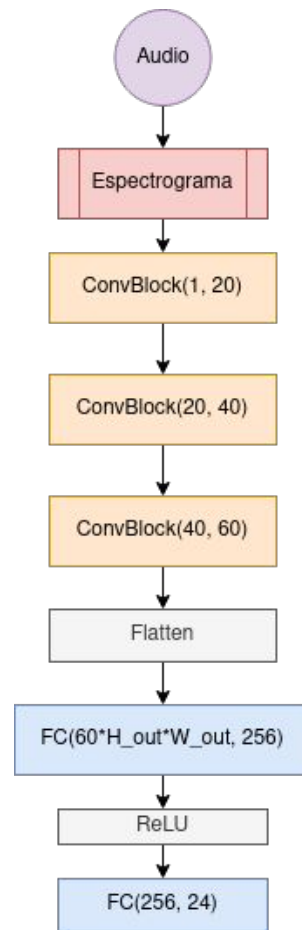
— — —

ConvBlock(in, out)



Modelo

- Las dimensiones de los mapas de características dependen de la representación tiempo-frecuencia
- Función de pérdida: Entropía cruzada binaria (solamente en especies con anotaciones)
- 5 épocas, en lotes de tamaño 4
- Optimizador: SGD con $lr=0.001$



[-1, -1, -1, -1, 0, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, -1, -1, 1, -1, -1, -1]

Evaluación

— — —

- Label Ranking Average Precision

Para cada etiqueta positiva, ¿qué proporción de las predicciones que tuvieron mejor ranking son verdaderos?

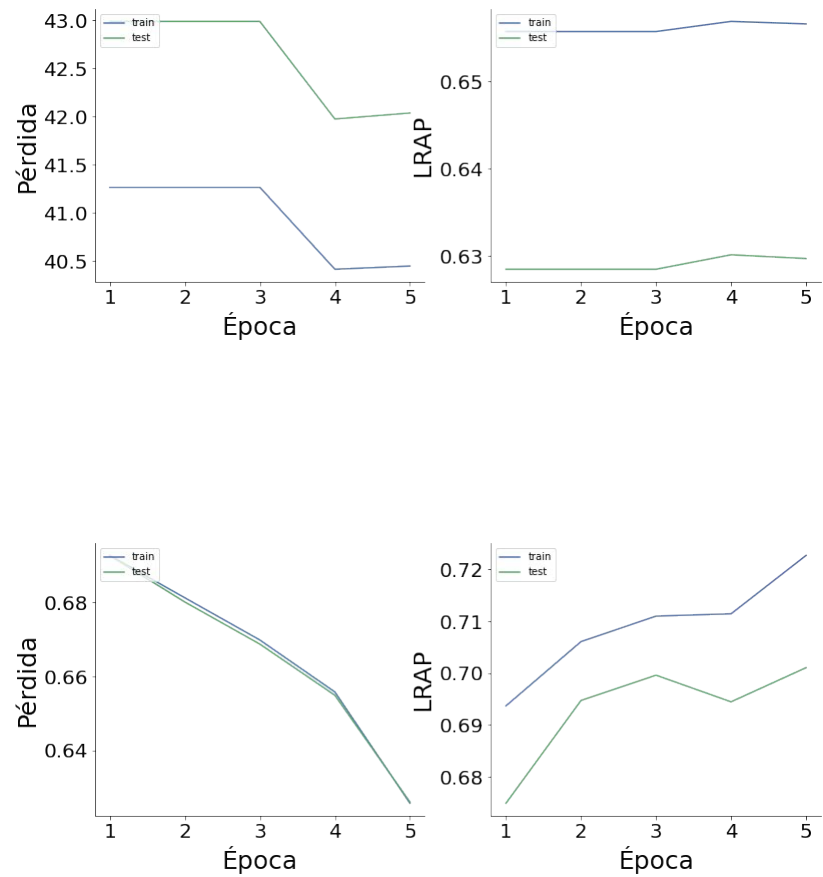
$$LRAP(y, \hat{f}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \frac{1}{||y_i||_0} \sum_{j: y_{ij}=1} \frac{|\mathcal{L}_{ij}|}{\text{rank}_{ij}}$$

$$\mathcal{L}_{ij} = \left\{ k : y_{ik} = 1, \hat{f}_{ik} \geq \hat{f}_{ij} \right\}, \text{rank}_{ij} = \left| \left\{ k : \underset{..}{\hat{f}_{ik}} \geq \underset{..}{\hat{f}_{ij}} \right\} \right|$$

Resultados

— — —

LRAP	Train	Test
Spec	0.657	0.630
Mel	0.723	0.701



Resultados

— — —

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
conv-mel-submission.csv	12 hours ago	1 seconds	1 seconds	0.37756

Complete

[Jump to your position on the leaderboard](#) ▼

Conclusiones

— — —

- Espectrograma en escala de Mel resulta mejor representación
 - Explorar otras representaciones
- Los datos de evaluación podrían estar etiquetados de manera diferente
- Largos de tiempo de preprocesamiento/entrenamiento
 - Si el tiempo no es problema, probar arquitecturas del estado del arte
- Seguir aumentando los datos, por medio de enmascarar los espectrogramas.

Referencias

— — —

- [Pink, J. \(2018\). 3 ways climate change affects tropical rainforests.](#)
- [RFCx. \(2021\). *Our work*](#)
- [Kaggle. \(2020\). Rainforest Connection Species Audio Detection](#)
- [Scikit-learn. \(2020\). 3.3.3.2. Label ranking average precision](#)
- Huzaifah, M. (2017). Comparison of Time-Frequency Representations for Environmental Sound Classification using Convolutional Neural Networks

**Gracias por su
atención**