Clasificación de géneros musicales

Integrantes
Alberto Olvera Trejo
Ricardo Bernabé Nicolás
Miguel Ángel Romero González

CONTENIDO

Motivación

A lo largo de este curso aprendimos a usar varias herramientas para clasificar datos y hacer predicciones.

Nosotros decidimos aplicar estos conocimientos para ver si se puede lograr clasificar géneros musicales

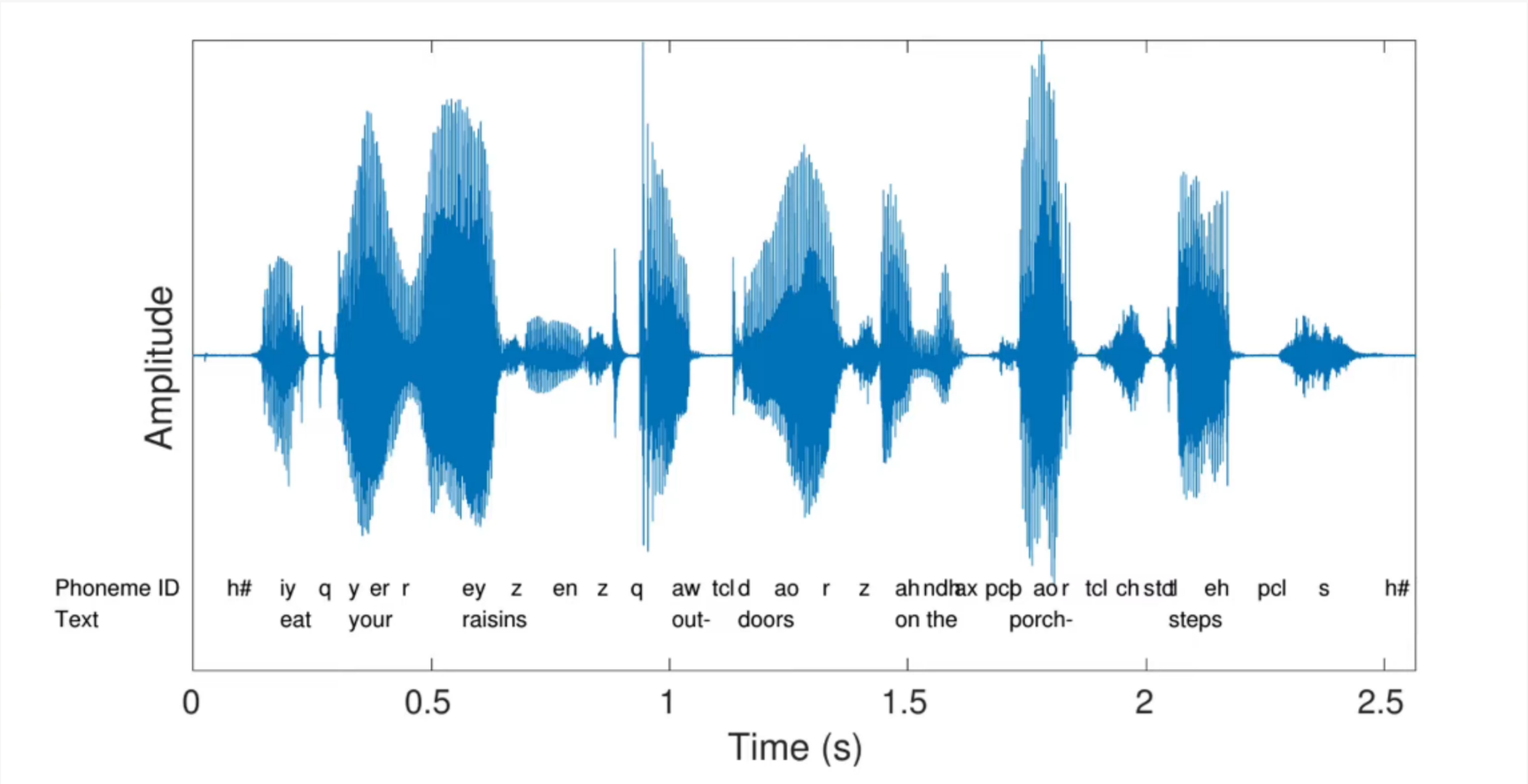
Introducción

Si nosotros oímos una pieza musical, dependiendo de nuestra experiencia y que tan abiertos nuestros horizontes musicales, quizá logremos identificar a qué género pertenece.

Para que una computadora pueda analizarla, debemos de analizarla de manera especial.

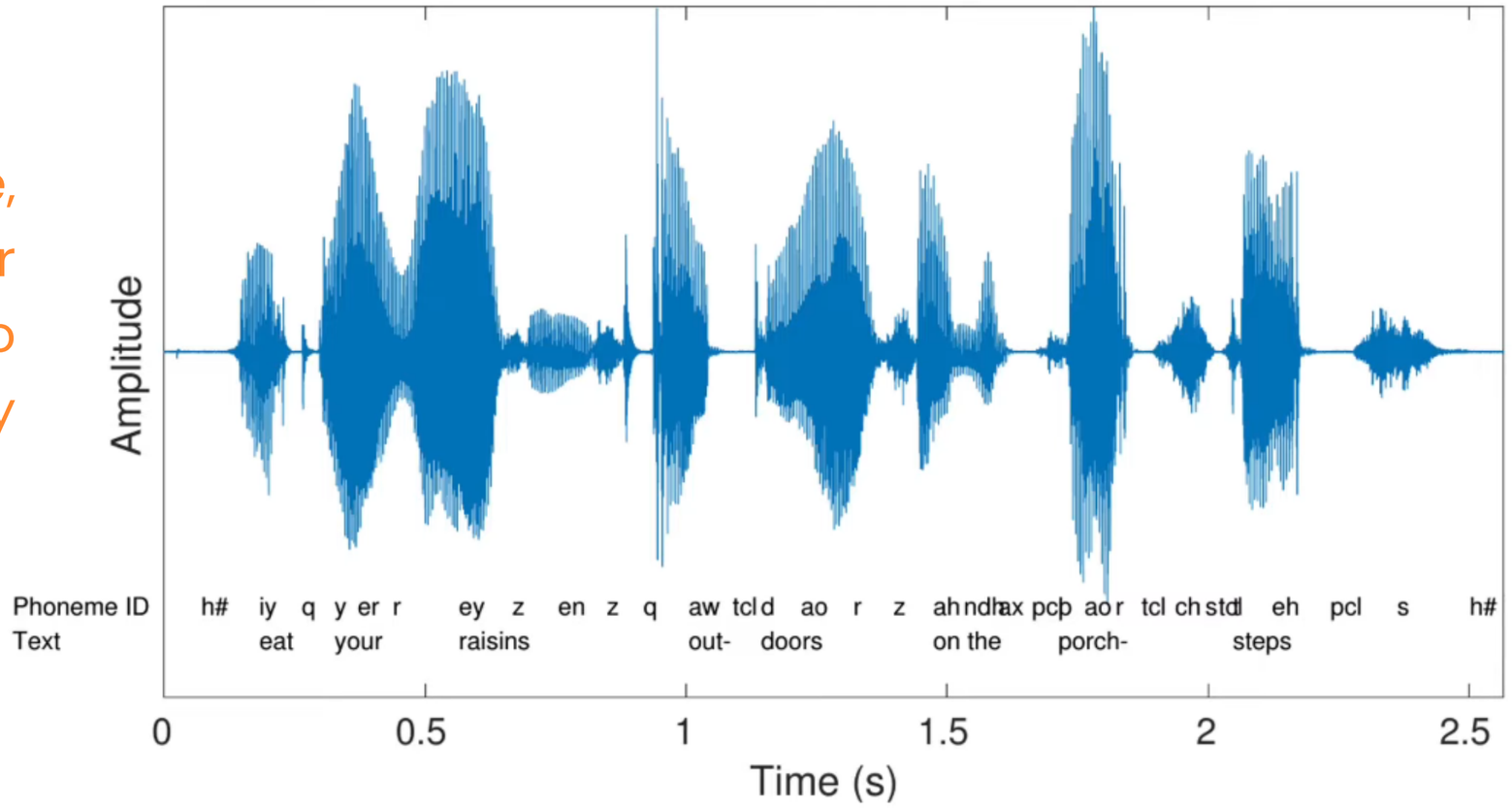
Audio Sample

Para procesar una canción debemos de analizar la muetra de audio que genera



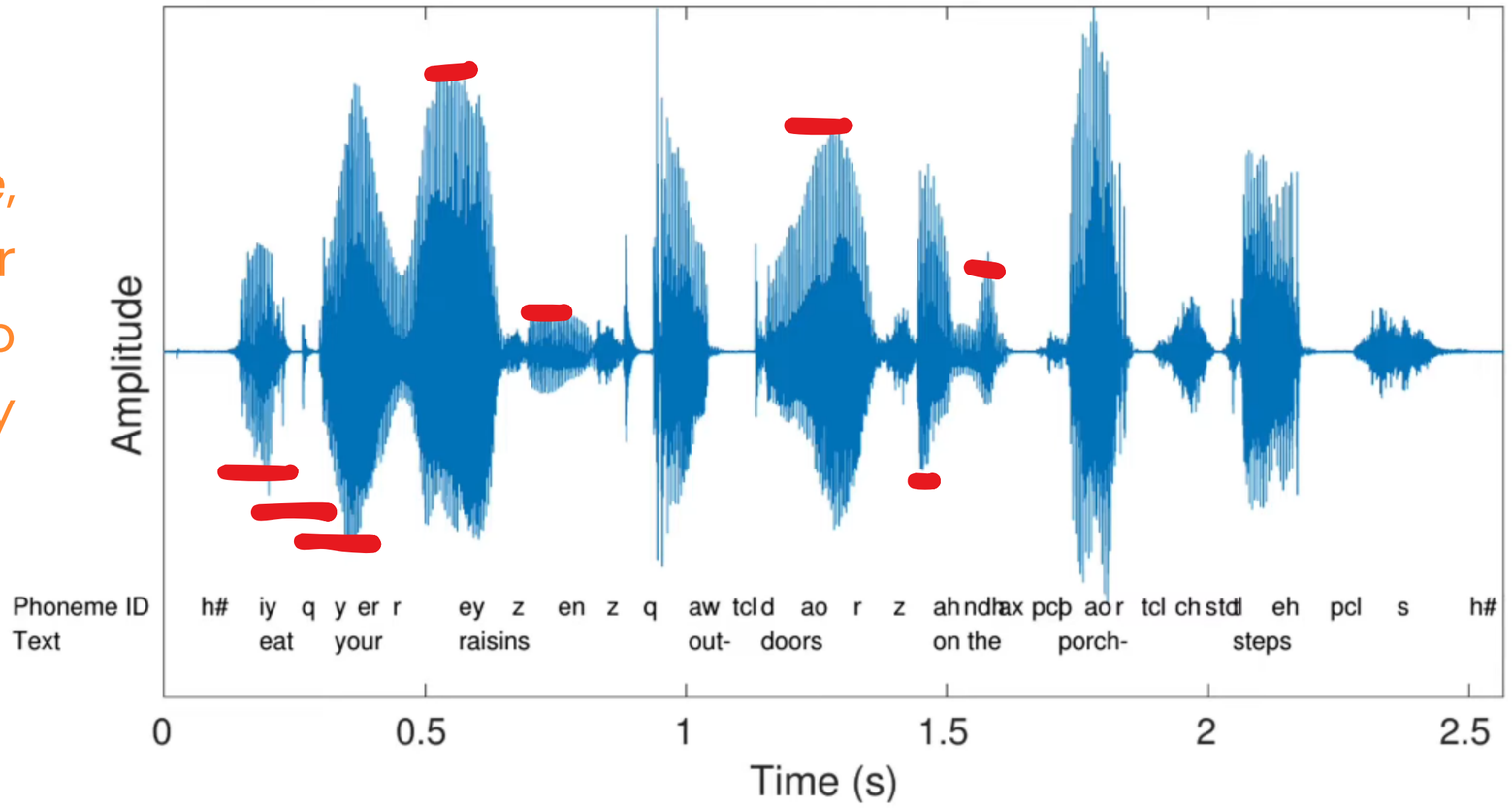
Windowing

La muestra es muy grande, por lo que debemos de tomar partes pequeñas (por lo general de 25ms) y procesarlas



Windowing

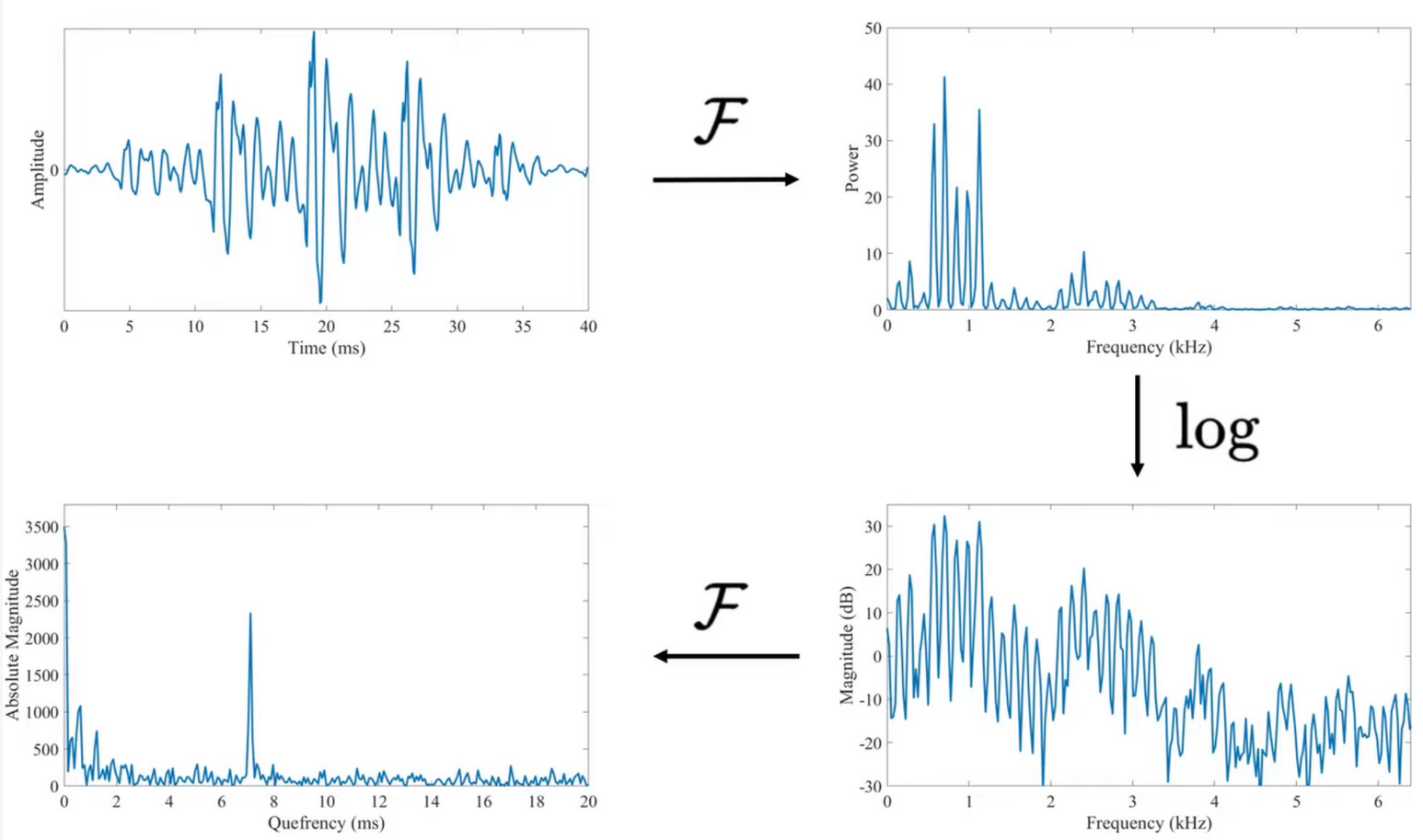
La muestra es muy grande, por lo que debemos de tomar partes pequeñas (por lo general de 25ms) y procesarlas



Matematicas complicadas

Cepstrum

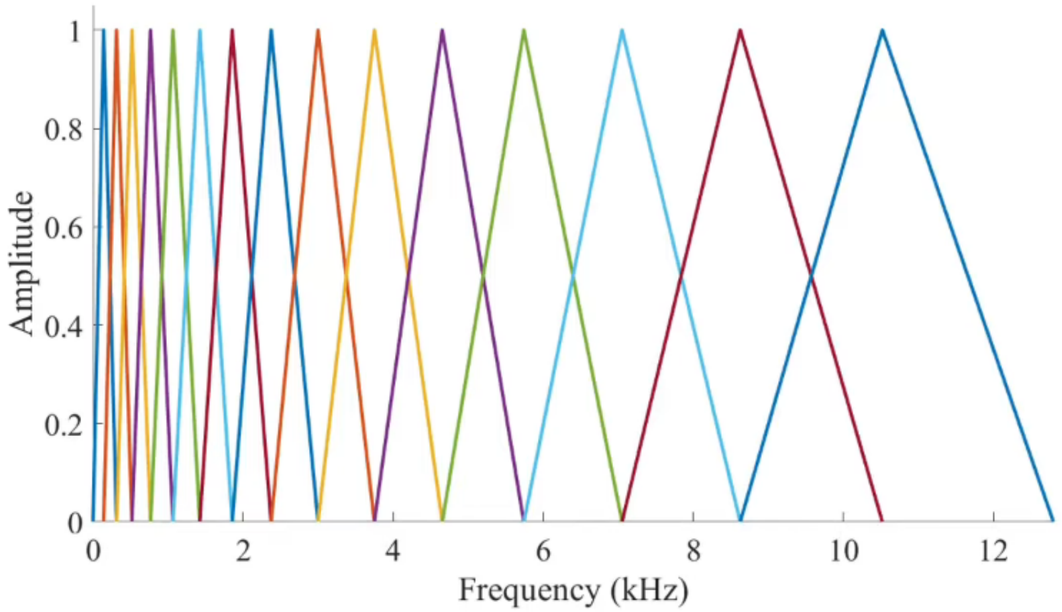
$$C_p = \left| \mathcal{F}\left\{ \log\left(\left| \mathcal{F}\{f(t)\} \right|^2 \right) \right\} \right|^2$$



Triangular Filterbank

Un paso extra para poder normalizar mejor nuestras ventadas y hacer un espectrograma más simple

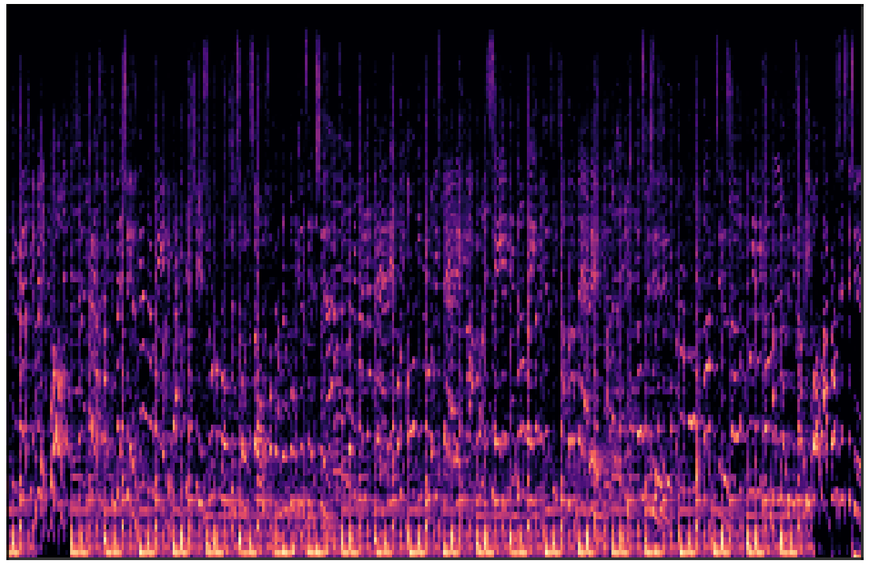
Triangular Filterbank



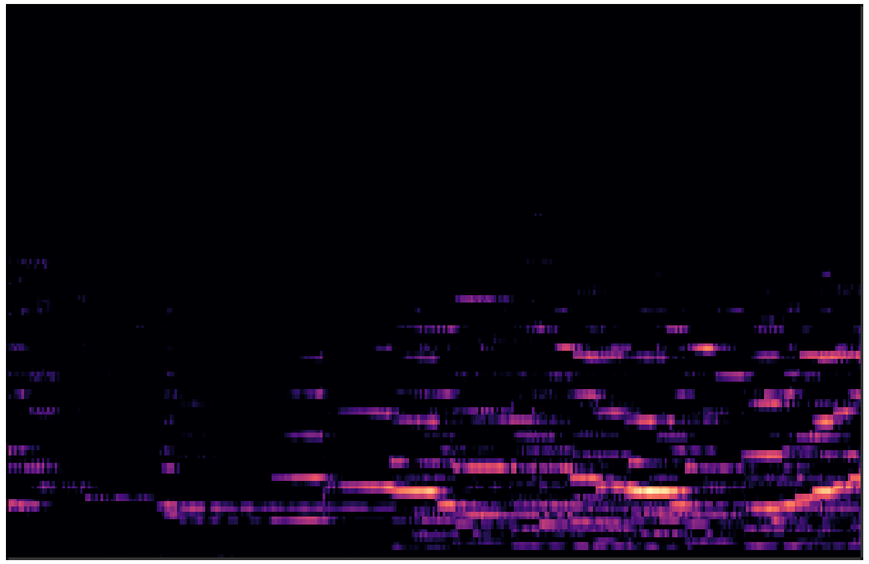
$$u_k = \sum_{h=f_{k-1}+1}^{f_{k+1}-1} w_{k,h} |x_h|^2$$

Espectograma

La unión de cada una de nuestras ventanas da como resultado un espectograma



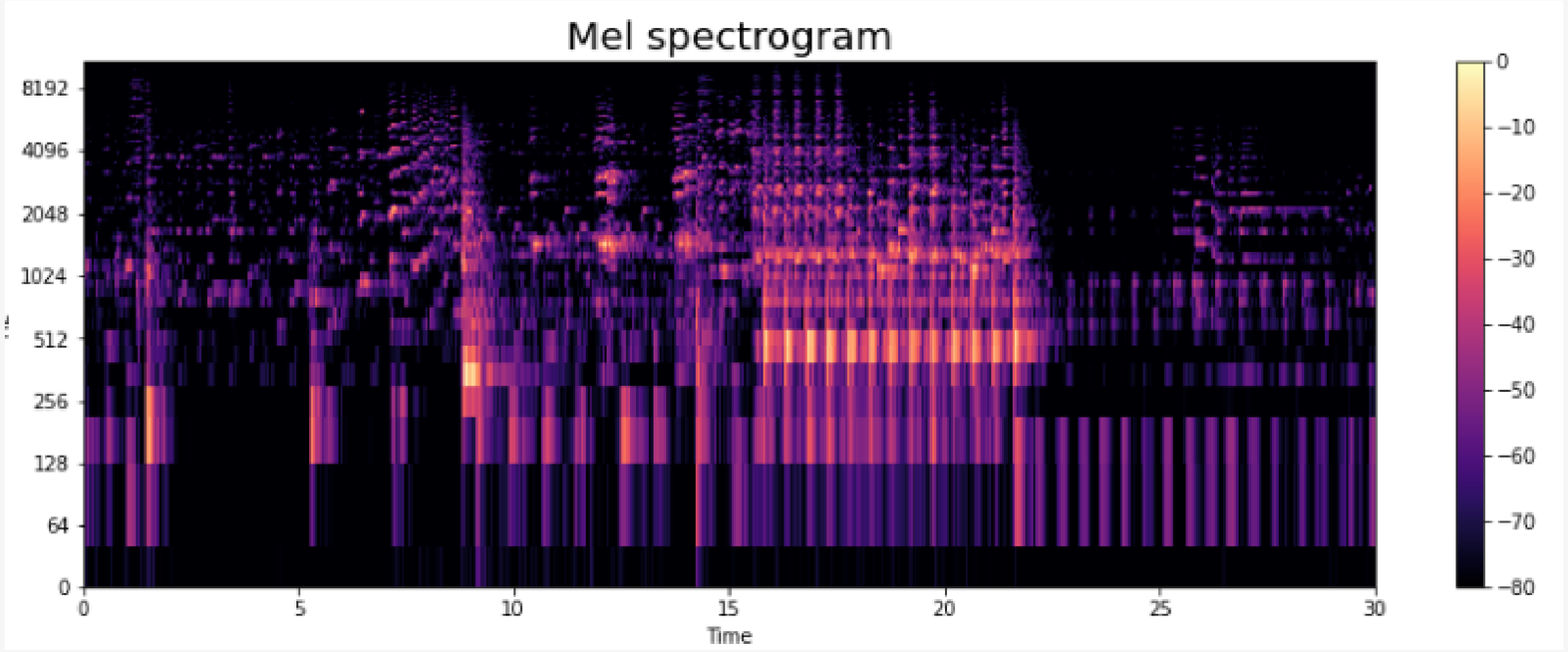
Espectrogama de una canción de hiphop



Espectrogama de una canción M clásica

Espectograma

Aplicando el triangular bank en cada una de nuestras ventana, conseguimos un espectrograma más simple



Dataset

Usamos GTZAN Dataset - Music Genre Classification que contiene 10 géneros distintos y 100 canciones por cada uno, y además contiene el desglose de cada una de las ventanas llamadas mfcc

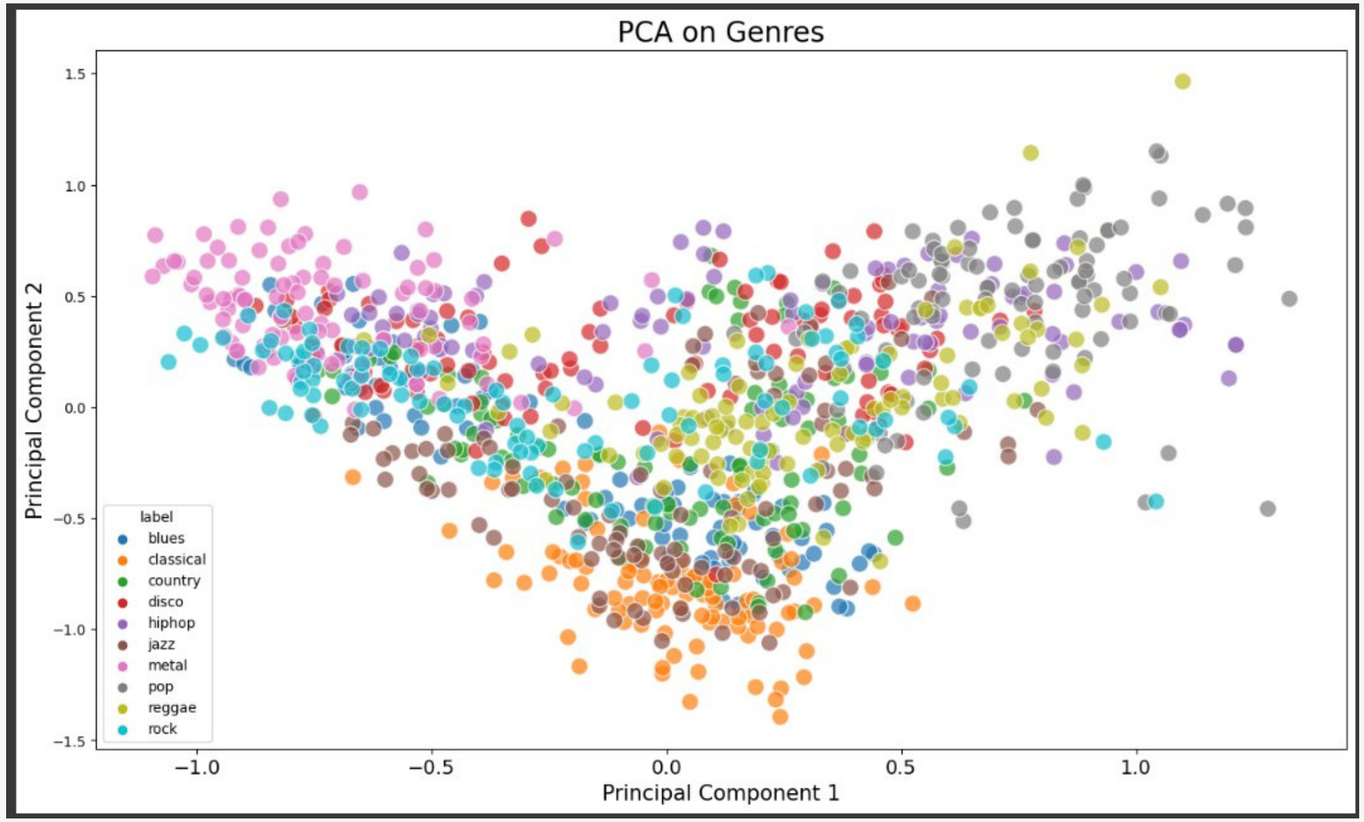
	filename	length	chroma_stft_mean	chroma_stft_var	rms_mean	rms_var	spectral_centroid_mean	spectral_centroid_var	spectral_bandwidth_mean	spectral_bandwidth_var	...	mfcc16_var
0	blues.00000.wav	661794	0.350088	0.088757	0.130228	0.002827	1784.165850	129774.064525	2002.449060	85882.761315	...	52.420910
1	blues.00001.wav	661794	0.340914	0.094980	0.095948	0.002373	1530.176679	375850.073649	2039.036516	213843.755497	...	55.356403
2	blues.00002.wav	661794	0.363637	0.085275	0.175570	0.002746	1552.811865	156467.643368	1747.702312	76254.192257	...	40.598766
3	blues.00003.wav	661794	0.404785	0.093999	0.141093	0.006346	1070.106615	184355.942417	1596.412872	166441.494769	...	44.427753
4	blues.00004.wav	661794	0.308526	0.087841	0.091529	0.002303	1835.004266	343399.939274	1748.172116	88445.209036	...	86.099236

5 rows × 60 columns

mfcc16_var	mfcc17_mean	mfcc17_var	mfcc18_mean	mfcc18_var	mfcc19_mean	mfcc19_var	mfcc20_mean	mfcc20_var	label
52.420910	-1.690215	36.524071	-0.408979	41.597103	-2.303523	55.062923	1.221291	46.936035	blues
55.356403	-0.731125	60.314529	0.295073	48.120598	-0.283518	51.106190	0.531217	45.786282	blues
40.598766	-7.729093	47.639427	-1.816407	52.382141	-3.439720	46.639660	-2.231258	30.573025	blues
44.427753	-3.319597	50.206673	0.636965	37.319130	-0.619121	37.259739	-3.407448	31.949339	blues
86.099236	-5.454034	75.269707	-0.916874	53.613918	-4.404827	62.910812	-11.703234	55.195160	blues

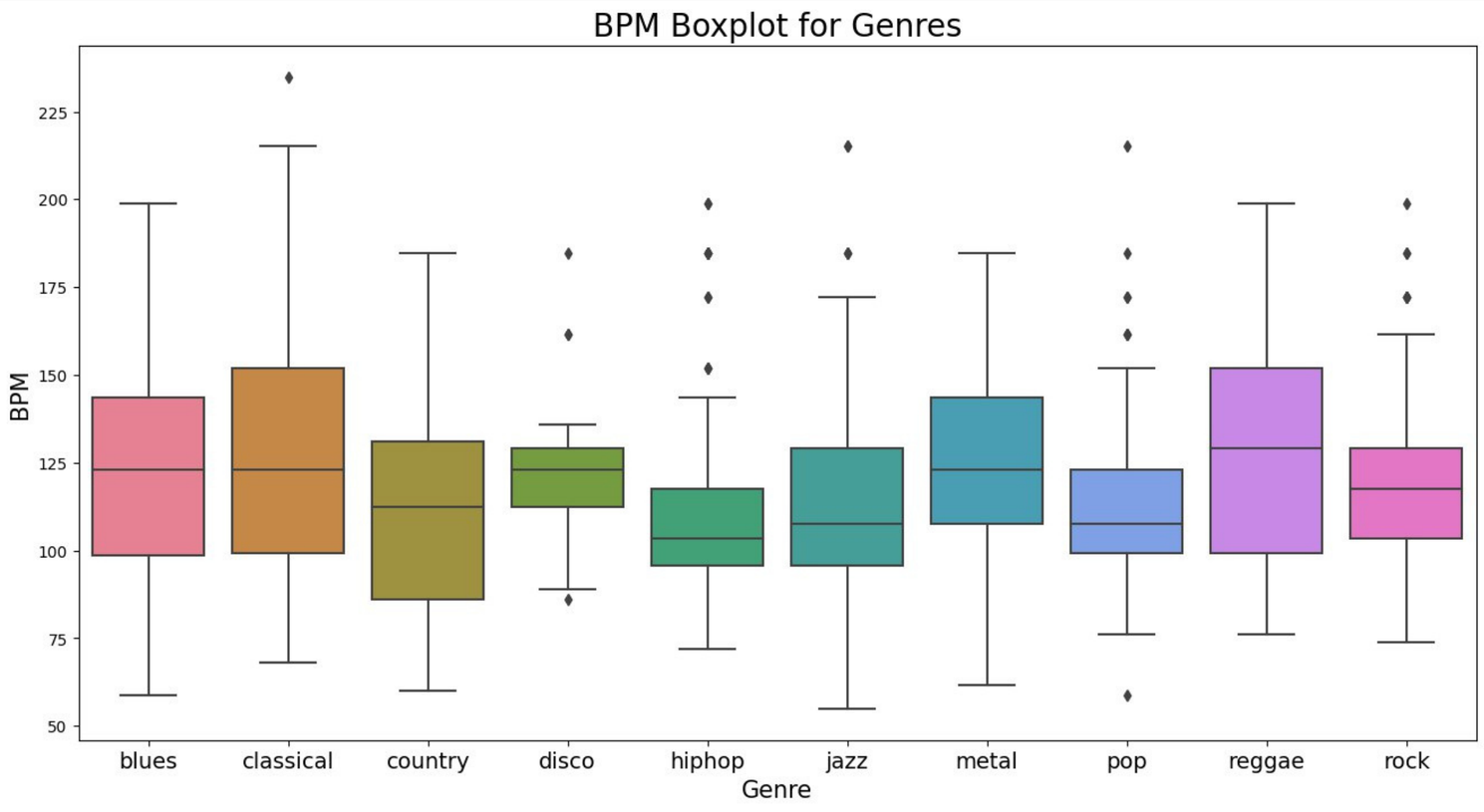
Visualización

Se hizo un PCA para obtener dos componentes principales y poder graficar los diferentes géneros



Visualización

También se hizo un boxplot de los beats por minuto para cada género



Uso de algoritmos para clasificar

Se utilizo Grid Search, el cual es un método para poder encontrar lo mejores hiperparámetros, como por ejemplo para poder encontrar cual es la mejor profundidad en un random forest, o la k óptima en K-Neighbors, etc. Durante el uso de Grid Search se empleo cross validation para mejorar los resultados

Descision Tree C
Logistic Regression

Decision Tree Classifier

	precision	recall	f1-score	support
0	0.44	0.44	0.44	18
1	0.62	0.76	0.68	17
2	0.32	0.41	0.36	17
3	0.27	0.19	0.22	21
4	0.33	0.43	0.38	21
5	0.57	0.72	0.63	18
6	0.57	0.81	0.67	16
7	0.57	0.48	0.52	27
8	0.22	0.22	0.22	18
9	0.30	0.11	0.16	27
accuracy			0.43	200
macro avg	0.42	0.46	0.43	200
weighted avg	0.42	0.43	0.41	200

```
------------------------------
Accuracy of Decision Tree Classifier : 0.435
OrderedDict([('criterion', 'gini'), ('max_depth', 86)])
------------------------------------------------------------
```

Resultados de DTC

	precision	recall	f1-score	support
blues	0.39	0.39	0.39	18
classical	0.94	0.94	0.94	17
country	0.53	0.53	0.53	17
disco	0.25	0.24	0.24	21
hiphop	0.45	0.43	0.44	21
jazz	0.58	0.78	0.67	18
metal	0.52	0.94	0.67	16
pop	0.79	0.70	0.75	27
reggae	0.50	0.39	0.44	18
rock	0.47	0.30	0.36	27
accuracy			0.55	200
macro avg	0.54	0.56	0.54	200
weighted avg	0.54	0.55	0.53	200

```
------------------------------
Accuracy of Logistic Regression : 0.545
OrderedDict([('C', 1000.0), ('penalty', 'l2'), ('solver', 'newton-cg')])
------------------------------------------------------------------------
```

Resultados de Logistic Regression

k-Neighbors
Random Forest

	precision	recall	f1-score	support
blues	0.19	0.28	0.23	18
classical	0.48	0.71	0.57	17
country	0.11	0.12	0.11	17
disco	0.19	0.24	0.21	21
hiphop	0.33	0.29	0.31	21
jazz	0.31	0.28	0.29	18
metal	0.24	0.25	0.24	16
pop	0.50	0.48	0.49	27
reggae	0.09	0.11	0.10	18
rock	0.00	0.00	0.00	27
accuracy			0.27	200
macro avg	0.24	0.27	0.26	200
weighted avg	0.24	0.27	0.25	200

```
------------------------------
Accuracy of K-Neighbors Classifier : 0.27
OrderedDict([('n_neighbors', 6)])
------------------------------------------------------------
```

Resultados de KNN

	precision	recall	f1-score	support
blues	0.87	0.72	0.79	18
classical	0.82	0.82	0.82	17
country	0.57	0.94	0.71	17
disco	0.50	0.62	0.55	21
hiphop	0.84	0.76	0.80	21
jazz	0.79	0.83	0.81	18
metal	0.62	0.94	0.75	16
pop	0.80	0.59	0.68	27
reggae	0.58	0.61	0.59	18
rock	0.69	0.33	0.45	27
accuracy			0.69	200
macro avg	0.71	0.72	0.70	200
weighted avg	0.71	0.69	0.68	200

```
------------------------------
Accuracy of K-Neighbors Classifier : 0.69
OrderedDict([('criterion', 'entropy'), ('max_depth', 8), ('max_features', 'auto'), ('n_estimators', 296)])
------------------------------------------------------------------------------
```

Resultados de Random Forest

GradientBoosting

XGB

```
GBC  Classification
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_gb.py:280:
  warnings.warn(
              precision    recall  f1-score   support

       blues       0.65      0.61      0.63        18
   classical       0.65      0.76      0.70        17
     country       0.41      0.53      0.46        17
       disco       0.36      0.43      0.39        21
      hiphop       0.75      0.71      0.73        21
        jazz       0.61      0.61      0.61        18
       metal       0.61      0.88      0.72        16
         pop       0.62      0.59      0.60        27
      reggae       0.53      0.50      0.51        18
        rock       0.50      0.22      0.31        27

    accuracy                           0.56       200
   macro avg       0.57      0.58      0.57       200
weighted avg       0.57      0.56      0.56       200

------------------------------
Accuracy of GBC Classifier : 0.565
OrderedDict([('criterion', 'friedman_mse'), ('learning_rate', 0.2),
--------------------------------------------------------------------
```

Resultados de GB

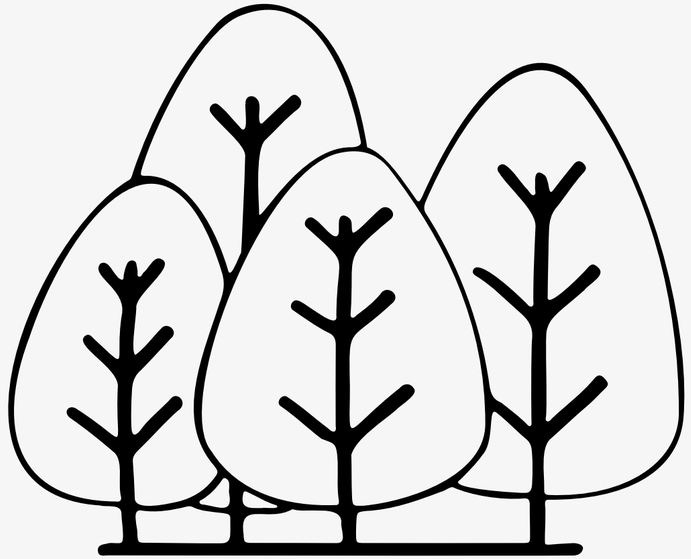
```
Parameters: { "n_stimators" } are not used.

              precision    recall  f1-score   support

           0       0.72      0.72      0.72        18
           1       0.89      0.94      0.91        17
           2       0.41      0.71      0.52        17
           3       0.63      0.57      0.60        21
           4       0.64      0.86      0.73        21
           5       0.81      0.72      0.76        18
           6       0.68      0.81      0.74        16
           7       0.77      0.63      0.69        27
           8       0.58      0.39      0.47        18
           9       0.53      0.37      0.43        27

    accuracy                           0.66       200
   macro avg       0.67      0.67      0.66       200
weighted avg       0.67      0.66      0.65       200

------------------------------
Accuracy of xgb Classifier : 0.655
OrderedDict([('learning_rate', 0.1), ('n_stimators', 200)])
--------------------------------------------------------------------
```

Resultados de XGB

¿Ganador?

¡Random Forest!

Modelo	Accuracy
Decision Tree	0.43
Regresión Logística	0.545
K-Neighbors	0.235
Random Forest	0.69
GradientBoost	0.56
XGB	0.655
Red Multicapa	0.9

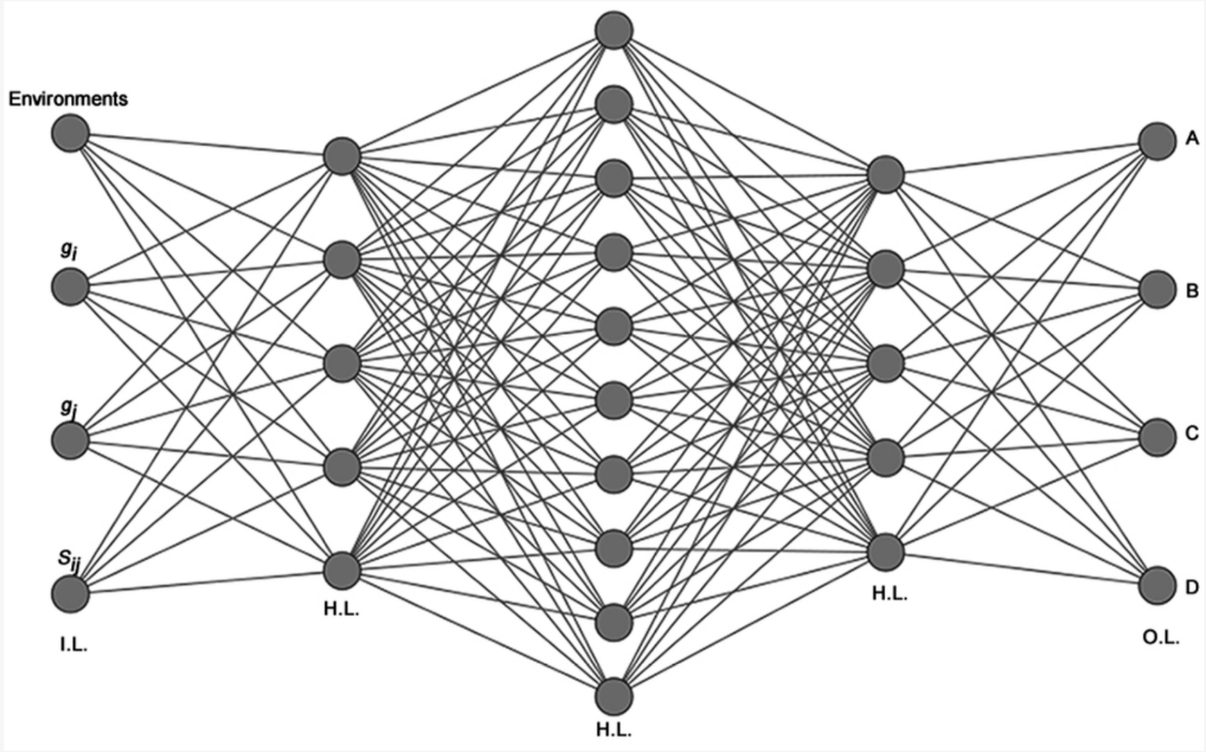


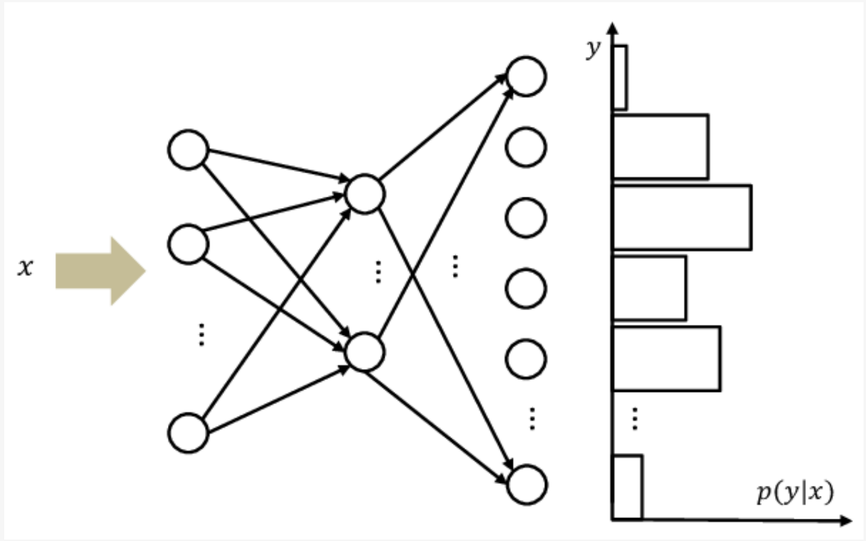
Red neuronal

La arquitectura usada fue la siguiente:

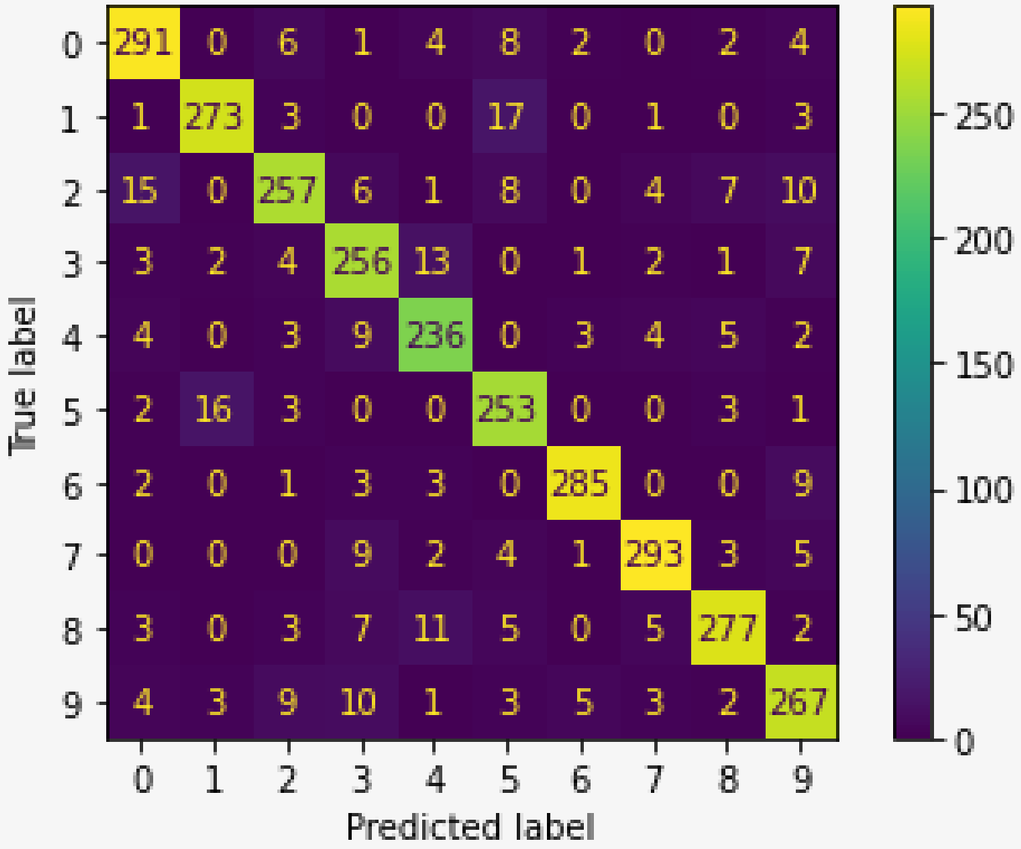
- Capa de 1024 neuronas
- Capa de 512 neuronas
- Capa de 512 neuronas
- Capa de 256 neuronas
- Capa de 128 neuronas
- Capa de 64 neuronas
- Capa softmax para clasificar

Entre cada capa se usó un dropout





Red neuronal



	precision	recall	f1-score	support
0	0.90	0.92	0.91	318
1	0.93	0.92	0.92	298
2	0.89	0.83	0.86	308
3	0.85	0.89	0.87	289
4	0.87	0.89	0.88	266
5	0.85	0.91	0.88	278
6	0.96	0.94	0.95	303
7	0.94	0.92	0.93	317
8	0.92	0.88	0.90	313
9	0.86	0.87	0.87	307
accuracy			0.90	2997
macro avg	0.90	0.90	0.90	2997
weighted avg	0.90	0.90	0.90	2997

Conclusiones

Para algunos géneros, se pueden observar diferencias claras como el metal y el pop, pero para otros como el blues y el Jazz, debido a origenes históricos no tienen diferencias tan claras.

Podemos detectar cómo hay géneros que influencian a otros géneros debido a sus similitudes.

La red neuronal logró clasificar los 10 géneros de manera exitosa