



QUEEN



Facultad de Estudios Superiores

Acatlán

Proyecto Final – Módulo IV

Predicción de Autores en Canciones de Queen

Cardoso Olvera Edgar David

Diplomado en Ciencia de Datos

Junio 2020



GREATEST HITS II

Índice general

Índice general	2
1. Introducción	3
1.1. Objetivo	3
1.2. Metodología del Análisis	3
1.3. Fuente de Datos	3
2. Análisis Exploratorio de Datos	4
2.1. Introducción a los Datos de Audio para Deep Learning	4
2.2. Análisis de Datos	5
2.3. Visualización de Datos	6
3. Conjunto de Entrenamiento y Prueba	7
4. Aumento de Datos Data Augmentation	8
4.1. Time Shifting	8
4.2. Speed Tune	8
4.3. White Noise	8
4.4. Procesamiento	8
5. Modelación Supervisada Deep Learning	10
5.1. Objetivo	10
5.2. Procesamiento Antes de Modelación	10
5.3. Red Neuronal y Arquitectura	10
5.4. Resultados de Modelación	10
6. Conclusiones	12

Capítulo 1

Introducción

En 1970, cuatro grandes músicos formaron una de las mejores bandas de la historia y mi banda favorita, Queen. Los nombres de los cuatro integrantes son Brian May, Freddie Mercury, John Deacon y Roger Taylor; cada uno de ellos escribió alguna de las 155 canciones que se encuentra en los 14 álbumes de estudio de la banda. Como gran fanático de Queen he podido reconocer quien ha escrito la canción solo con escuchar los instrumentos de la misma, pero he querido probar si la inteligencia artificial me puede ayudar en esta tarea.

1.1. Objetivo

El objetivo principal de este proyecto es conocer y practicar como es el funcionamiento del procesamiento y predicción de datos de audio. Con lo anterior, se desarrollará un algoritmo que pueda predecir el autor de una canción de Queen solo por la información obtenida de dicha canción.

Al no tener un punto de comparación dentro de este tipo de tareas y también, al no saber si un experto en Queen podría predecir de manera correcta quien ha escrito la canción solo con escucharla, la métrica esperada deberá ser superior al 25 % de exactitud, ya que esto permitirá decir que el algoritmo no está prediciendo de manera aleatoria los compositores.

1.2. Metodología del Análisis

Al obtener todas las canciones de Queen se podrá analizar la longitud, autor y frecuencia de cada canción. Esto permitirá decidir si es necesario, aumentar la cantidad de los datos para mejorar el ajuste del modelo.

1.3. Fuente de Datos

Las canciones se obtuvieron de manera legal al comprar la discografía completa de Queen. Estas vienen en formato mp3 y tienen una calidad de CD, es decir, 16 bits y 44.1 kHz. Se puede observar, que no se cuenta con la mejor calidad de audio que se puede obtener hoy en día, pero funciona correctamente para el objetivo de este proyecto.

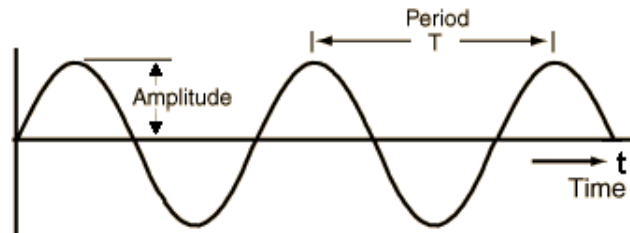
Capítulo 2

Análisis Exploratorio de Datos

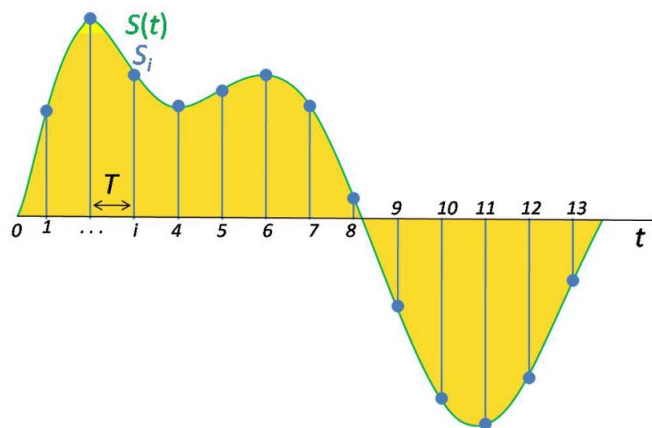
El análisis exploratorio de la información tiene como objetivo reconocer patrones significativos en nuestros datos, encontrar irregularidades en la información y comprende de manera rápida y eficaz el conocimiento que estos transmiten. También brindará información relevante para crear visualizaciones para un mejor entendimiento.

2.1. Introducción a los Datos de Audio para Deep Learning

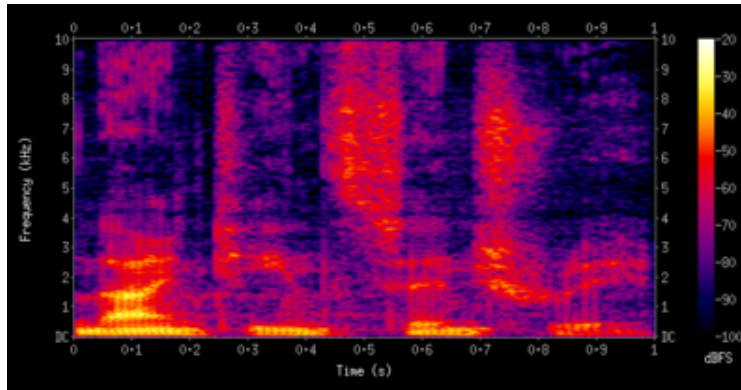
El sonido es generado por la vibración en la presión del aire, esto genera onda de las cuales se puede medir su intensidad, amplitud, frecuencia y longitud.



Al no tener una forma de representar las ondas de manera digital, se recrea la medida de amplitud de la onda en un intervalo de tiempo, por lo que la onda no será perfectamente suave. Cada una de estas medidas se les conoce como muestras y juntando un segundo de estas muestras se les conoce como la tasa de muestras o sample rate. La calidad de audio del CD tienen un sample rate de 44.1 kHz es decir que por cada 10 segundos de canción se tendrán 441,000 muestras.



En la actualidad la mejor forma de trabajar con audio es por medio de las Redes Neuronales Convolucionales (CNN), es decir trabajar con una imagen del audio. Lo anterior, se logra por medio del uso de espectrogramas de cada una de las canciones. El espectrograma es una imagen de la señal a través del tiempo, la intensidad de los colores en esta imagen representa la amplitud de cada frecuencia



Los espectrogramas se generan por una transformación de Fourier para descomponer la señal en frecuencias. La transformación se aplica a una ventana de tiempo y se aplica un escalamiento logarítmico para normalizar las frecuencias. Dentro de este proyecto se utiliza una transformación no lineal para obtener el espectrograma llamado Mel Scale.

2.2. Análisis de Datos

En la siguiente tabla se muestra la distribución de canciones que se tiene por cada uno de los artistas.

Artista	# Canciones
Brian	50
Freddie	57
John	23
Roger	25
Total	155

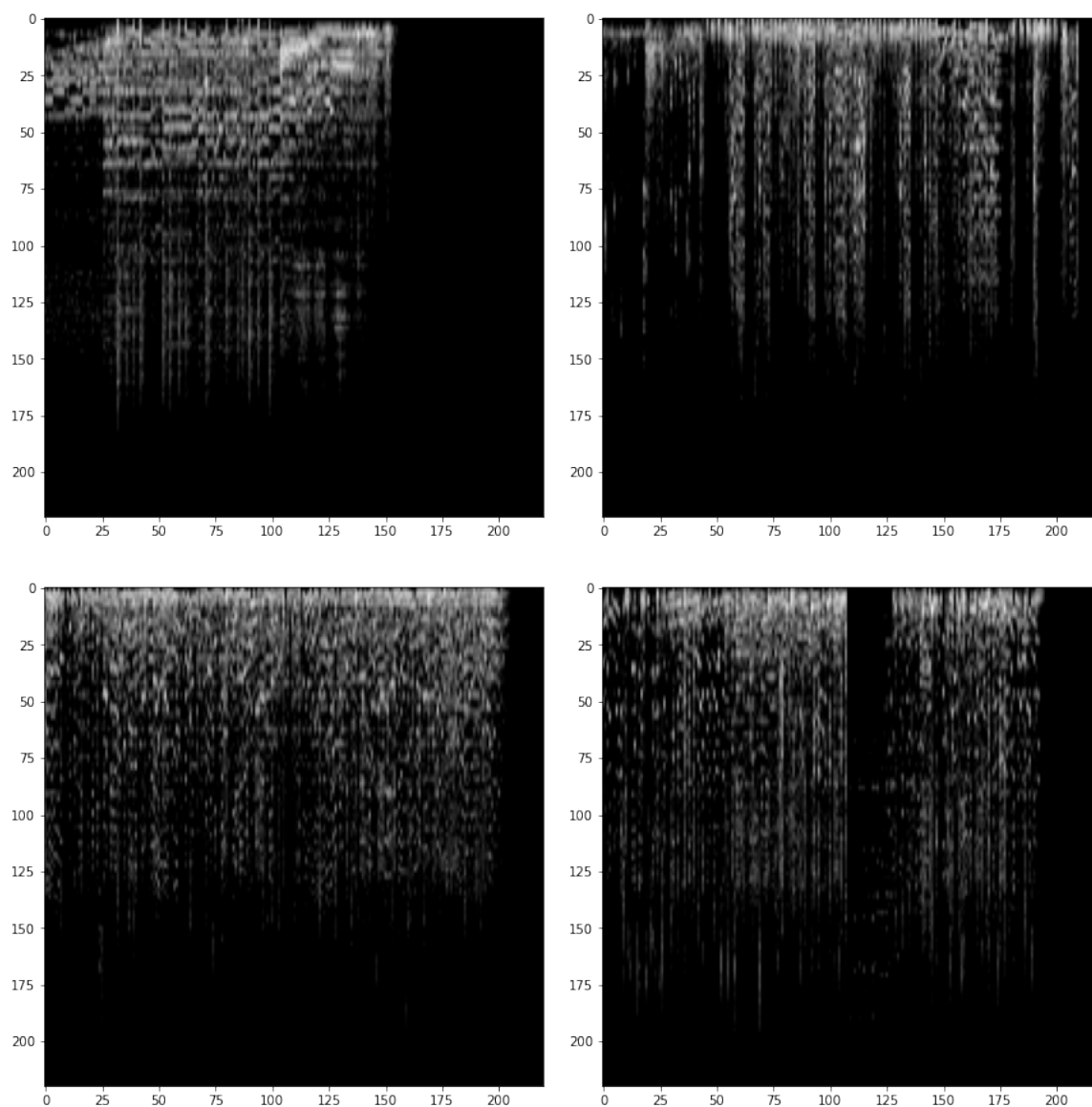
Igualmente se puede notar las distribución de longitud de las canciones.

	Descripción
Media	3.629314
Desv. Estándar	1.315526
Min	0.934313
25 %	2.887401
50 %	3.590095
75 %	4.389007
Max	8.347429

En promedio las canciones duran más de tres minutos y se pueden notar canciones menores a un minuto de longitud y otras con más de ocho minutos. Esta diferencia notable en la longitud de canciones se da por los distintos álbumes de la banda. Las composiciones cortas se pueden encontrar en el álbum Flash, ya que este funcionó como banda sonora para la película homónima, las composiciones más largas se pueden encontrar en el álbum A Night at the Opera, el cual fue un parte aguas para la banda, en él se encuentran canciones extensas como The Prophet's Song y Bohemian Rhapsody.

2.3. Visualización de Datos

En esta sección se observarán los resultados de los espectrogramas obtenidos con las transformaciones explicadas anteriormente



Es muy probable que cualquier ser humano no encuentre mucha diferencia entre estas cuatro imágenes, pero gracias a la Inteligencia Artificial y Deep Learning ahora es posible comprender las características y diferencias que hay entre cada uno de los espectrogramas.

Capítulo 3

Conjunto de Entrenamiento y Prueba

Al separar el conjunto de entrenamiento y prueba para el modelo de redes neuronales, se tomo el 80 % de las canciones escritas por cada uno de los autores de manera aleatorio. La división es necesaria para medir el performance del modelo y hacer mejoras dentro del mismo.

Al tener solo 155 canciones en total se necesita más información para poder obtener resultados relevantes en el modelo. En el siguiente capitulo se revisarán 3 formas de aumentar la cantidad de datos al trabajar con señales de audio.

Capítulo 4

Aumento de Datos | Data Augmentation

Por lo visto en el capítulo anterior, al tener tan poca cantidad de canciones es muy difícil general un modelo que generalice correctamente las canciones y su autores, es por esto que en esta sección se explicarán las técnicas utilizadas para generar más datos con las misma canciones originales del grupo.

4.1. Time Shifting

Una canción tendrá la misma instrumentación y características aunque esta inicie 15 segundos después o termine 20 segundos antes, este es el objetivo del Time Shifting. Al tener una canción se agregaran momento en blanco al inicio o al final de la canción y esta tendrá la misma duración que la original. Esto generará nueva data con las mismas características.

4.2. Speed Tune

Al alargar o acortar la longitud de la canción se tendrá una diferencia en la velocidad, si esta longitud se modifica de tal manera que casi no sea perceptible para el oído humano pero si para la computadora se obtendrá una nueva canción distinta para la computadora pero con las misma características que la original.

4.3. White Noise

Finalmente, el último método para genera nuevos datos es agregar un ruido blanco de fondo en la canción. El ruido blanco es aquel sonido de interferencia que es comúnmente conocido, este se agrega de fondo a la canción y con esto se tendrá nueva información para el modelo.

4.4. Procesamiento

Los tres métodos se pueden aplicar las veces que se quieran y combinar los distintos métodos para generar más información. Por lo anterior, para cada una de las canciones del conjunto de entrenamiento se le aplicaron 15 filtros de Time Shifting, a cada uno de los 15 filtro más el original se le aplicó el Speed Tuning y a cada uno de ellos se le aplicó el filtro de White Noise, por lo que por cada una de las canciones de entrenamiento se obtuvieron 64 nuevas canciones. Esto se limito a las 64 nuevas canciones ya que por la capacidad del equipo de computo solo se podía manejar esta cantidad. En la siguiente tabla se puede observar el aumento de dichos datos.

	Original	Augmented
Brian	40	2,560
Freddie	46	2,944
John	19	1,216
Roger	20	1,280
Total	125	8,000

Con la tabla anterior se puede observar que se paso de 125 canciones originales que se tenían para entrenamiento a 8,000, con esto se tendrá un mejor performance en el modelo.

Capítulo 5

Modelación Supervisada | Deep Learning

5.1. Objetivo

El objetivo principal de esta sección es encontrar un algoritmo o modelo de Deep Learning que puede predecir, con el mínimo error y máxima generalización, el compositor de las canciones del grupo Queen. Esto se logrará con los espectrogramas de 125 canciones y sus respectivo aumentación de datos y se medirá el performance del modelo con un conjunto de 30 canciones de prueba.

5.2. Procesamiento Antes de Modelación

Con la intención de mejorar los resultados en la modelación, se estandarizó el valor de cada pixel dividiéndolo entre 255 lo que permite un mejor entrenamiento en las redes neuronales.

5.3. Red Neuronal y Arquitectura

El modelo utilizado para la predicción de los autores es una red neuronal con dos capas convolucionales y dos capas densas más la capa de salida, en la cual se regresa un vector de cuatro entradas con la probabilidad de cada uno de lo autores en escribir dicha canción. Para evitar el conocido overfitting se utilizan métodos de regularización dentro de las capas densas y el método de early stopping para no sobre ajustar dicho modelo.

5.4. Resultados de Modelación

En la siguientes tablas se muestran los resultados del entrenamiento y prueba, en ellos no se puede notar un bajo overfitting gracias al filtro de early stopping utilizado en la red neuronal.

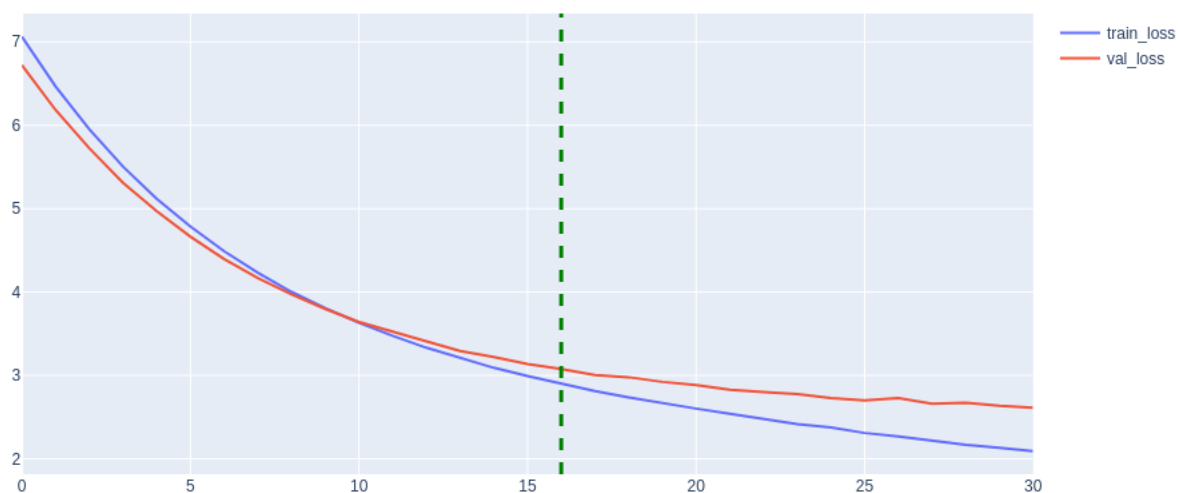
Set de Entrenamiento				
	Precisión	Recall	F1	Soporte
Brian	0.64	0.85	0.73	2398
Freddie	0.63	0.86	0.73	2732
John	0.93	0.18	0.3	1162
Roger	0.95	0.29	0.44	1226
Accuracy			0.66	7518

Set de Prueba				
	Precisión	Recall	F1	Soporte
Brian	1.00	0.80	0.89	10
Freddie	0.52	1.00	0.69	11
John	1.00	0.25	0.40	4
Roger	0.00	0.00	0.00	5
Accuracy			0.67	30

Dentro del set de prueba se puede observar una gran precisión para el guitarrista Brian May y el bajista John Deacon, esto quiere decir que el modelo predice de manera correcta las canciones de estos artistas, pero al observar el recall vemos que solo predijo bien una canción del bajista lo que disminuye el recall y por ende el F1. Para Freddie Mercury es el caso contrario, a él se le asigna la mayoría de las canciones aunque no las haya escrito y con esto puede tener un recall alto en la predicción, lo que da cerca de un 0.70 de F1.

Al ver el accuracy del modelo, es notable el nulo overfitting que existe en el modelo guardado, esto gracias al filtro de early stopping que se tiene dentro del modelo. En la siguiente gráfica se puede notar el momento donde se guarda el mejor modelo (línea color verde) y esto disminuye el overfitting que se puede obtener en esta clase de modelos.

Loss



Capítulo 6

Conclusiones

Al solo conocer el rol principal de cada uno de los integrantes del grupo se puede reconocer cuando un artista ha escrito una canción de Queen, pero esto es difícil y en muchos casos no se tienen una buena exactitud con esta predicción. Al observar los resultados finales del modelo se puede notar que sí existe un patrón específico en las canciones hechas por Brian May y Freddie Mercury, pero es muy difícil diferenciar al autor con las canciones de John Deacon y Roger Taylor.

Finalmente, es notable el accuracy del modelo, ya que este indica que no se hace una predicción aleatoria, sino que se pudo reconocer los patrones principales dentro de las canciones de Queen y su autores. Estas predicciones podrán mejorar con una mayor cantidad de datos y un mayor poder computacional.