

Predicción del campeón del mundial de fútbol Qatar 2022



Gustavo Alvarado | Statistical Learning II | Septiembre 2022
[Notebook](#)

Idea

La predicción del campeón del próximo campeón del mundial de fútbol en Qatar 2022 utilizando MLP (*Multilayer Perceptron*) el cual es uno de los tipos de redes neuronales más comunes.

Fuentes de datos

Se utilizaron los [resultados](#) de todos los partidos entre selecciones de fútbol entre 1872 y 2022 (alrededor de 40,000 partidos), y el [ranking](#) de selecciones de fútbol de la FIFA a agosto de 2022. También se utilizaron los resultados de las tandas de [penalties](#) en las fases finales de los mundiales de 1982 a 2018.

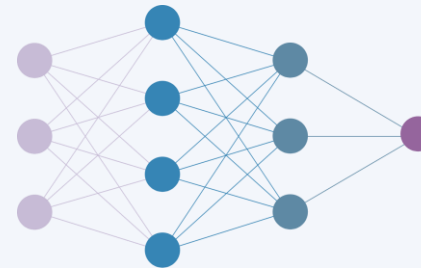


Metodología

3 variables de entrada (ranking equipo local, ranking equipo visitante, cancha neutral) y 2 de salida (goles equipo local, goles equipo visitante).

5 experimentos, con entrenamientos de distintas configuraciones de la red neuronal (número de capas ocultas o número de neuronas por capa) y distintos parámetros de entrenamiento.

Experimento	Exactitud
1	75%
2	78%
3	77%
4	76%
5	76%



Se seleccionaron los modelos utilizados para los experimentos 2 y 3, los cuales presentaron una mejor exactitud en los valores de prueba. También se efectuó un entrenamiento para un modelo adicional para tandas de penalties.

Resultados

Se obtuvieron las predicciones de los resultados de la fase de grupos con los modelos 2 y 3 de forma alterna. Se efectuó el cálculo de la tabla de posiciones de cada grupo para definir la configuración de las fases finales.

Países Bajos 2	Países Bajos 1	Argentina 2 (4)	Inglatera 2	Qatar 1
Gales 1			Inglatera 2	Inglatera 2
Argentina 2	Argentina 2		Francia 1	Francia 2
Dinamarca 1		Inglatera 2 (3)	México 1	México 1
España 2	España 1	Argentina 2 (2)	Bélgica 2	Bélgica 2
Croacia 1			Alemania 1	Alemania 1
Brasil 2	Brasil 2	Brasil 2 (3)	Bélgica 1	Portugal 2
Uruguay 1			Portugal 1	Suiza 1

INGLATERRA

Conclusiones

- Los modelos toman en cuenta distintos factores, no únicamente el ranking.
- Los modelos le dieron un peso adecuado a la localía de Qatar.
- Es posible que la exactitud no haya sido mayor debido a la aleatoriedad que existe el fútbol.

Clasificación de jeroglíficos egipcios



Gustavo Alvarado | Statistical Learning II | Septiembre 2022

[Notebook](#)

Idea

Clasificación de jeroglíficos egipcios utilizando CNN (*Convolutional Neural Network*), el cual es el tipo de red neuronal más efectivo para procesamiento de imágenes.

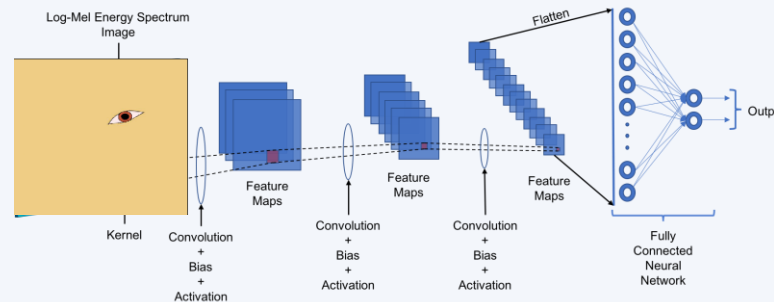
Fuente de datos

Lista de Gardiner de [95 jeroglíficos egipcios](#), cada uno con distintas versiones de los mismos en distintas posiciones (3890 imágenes para entrenamiento y 1039 para prueba).



Metodología

Se reorganizó el data set para poder obtener las imágenes de entrenamiento y de **prueba** con la librería **ImageDataGenerator** de Keras.



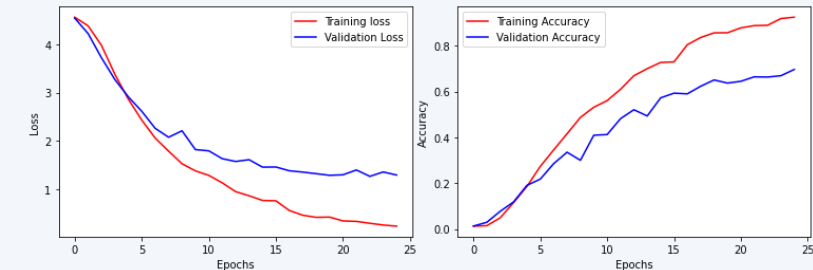
Se redimensionaron las imágenes a un tamaño de 200 x 233. Se definieron 3 operaciones de convolución con **kernels** de 3 x 3, y su respectivo **MaxPooling** de 3 x 3 y activación **relu**. Finalmente se aplicó una operación **Flatten** y una capa oculta de 64 neuronas.

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 198, 231, 32)	896
max_pooling2d (MaxPooling2D)	(None, 66, 77, 32)	0
conv2d_1 (Conv2D)	(None, 64, 75, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 21, 25, 32)	0
conv2d_2 (Conv2D)	(None, 19, 23, 64)	18496
max_pooling2d_2 (MaxPooling2D)	(None, 6, 7, 64)	0
flatten (Flatten)	(None, 2688)	0
dense (Dense)	(None, 64)	172096
dense_1 (Dense)	(None, 95)	6175

Total params: 206,911		
Trainable params: 206,911		
Non-trainable params: 0		

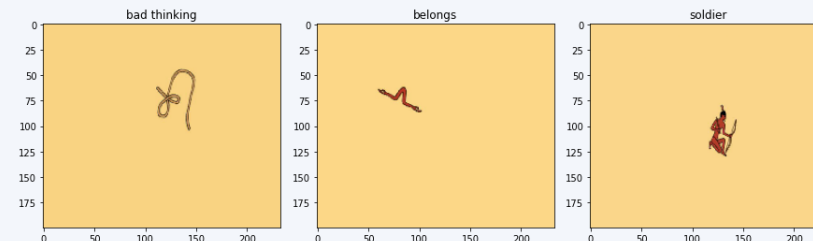
Resultados

En un único experimento se obtuvieron resultados satisfactorios, con una exactitud de alrededor de 92% para los datos de entrenamiento y 70% para los datos de prueba.



Conclusiones

El modelo entrenado dio resultados satisfactorios con los jeroglíficos de prueba.



Predicción del álbum al que pertenece una canción



Gustavo Alvarado | Statistical Learning II | Septiembre 2022
[Notebook](#)

Idea

Predicción del álbum al que pertenece una canción del grupo Metallica, utilizando RNN (*Recursive Neural Networks*) el cual es el tipo de red neuronal más efectivo para procesamiento de audio.

Fuente de datos

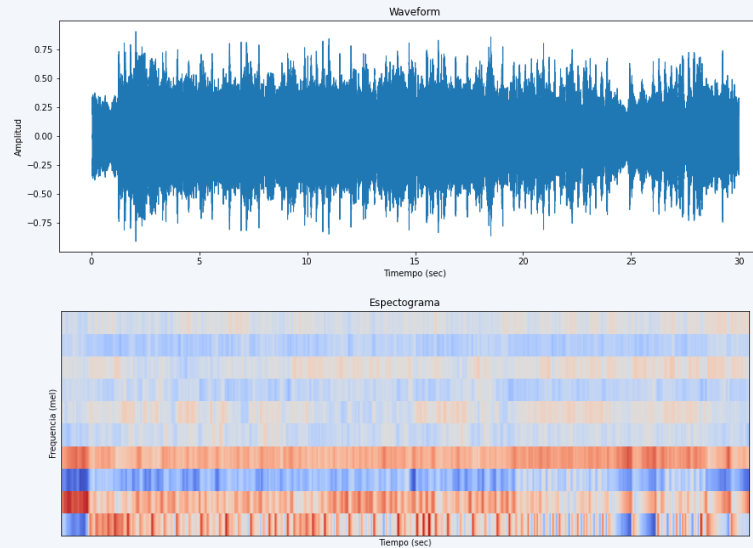
Discografía completa del grupo Metallica obtenido a través del [API de Spotify](#).



Metodología

Por medio API de Spotify se obtuvo un *playlist* que contiene la discografía completa, incluyendo nombre de canción, álbum y URL que incluye una sección de 30 segundos de cada canción, de un total de 100 canciones.

Se separan 90 canciones para entrenamiento y 10 para prueba.



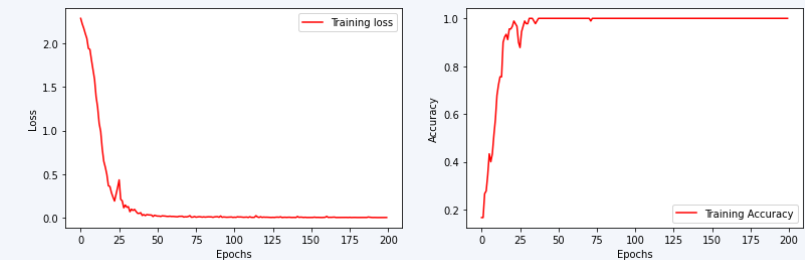
Se descargan las secciones de 30 segundos en formato mp3 en Google Drive. Se convierte a formato wav con **AudioSegment**.

Con **Librosa**, se obtiene el *waveform* y espectrograma, utilizando los **coeficientes de frecuencia Mel** (MFCC) de las 90 canciones de entrenamiento

Se efectuaron dos experimentos de RNN con la variante **LTMS** (*Long Short-Term Memory*), utilizando distintos parámetros y distintas cantidades de capas ocultas

Resultados

El segundo experimento dio resultados con valores cercanos al 100% de exactitud para los datos de entrenamiento. Para los datos de prueba la exactitud **no fue la esperada**.



Conclusiones

- Es posible que la poca cantidad de canciones cause que el modelo no sea efectivo en datos de prueba.
- Se puede observar que las predicciones del modelo que no son acertadas, tienden a tener en promedio 1 disco de diferencia de distancia cronológica.
- Si los discos se agrupan en *clústers* basado en espectros de sonidos similares, es posible que el modelo sea más efectivo.