01

Eduard Alfonso Caballero
Iván Rodrigo Castillo
María Camila Aparicio

ARTIFICIAL INTELLIGENCE II - 2020-1

SPEECH TO COMMAND

Objetivo General:
Reconocer comandos de voz simples en ingles para ejecutar acciones en dispositivos con micrófono.

¿Por qué hacerlo?
Desarrollar una solución con deep learning que permita mejorar la accesibilidad indirecta de dispositivos inteligentes.

¿Qué modelos se utilizaron?
DNN,RNN,LSTM y GRU

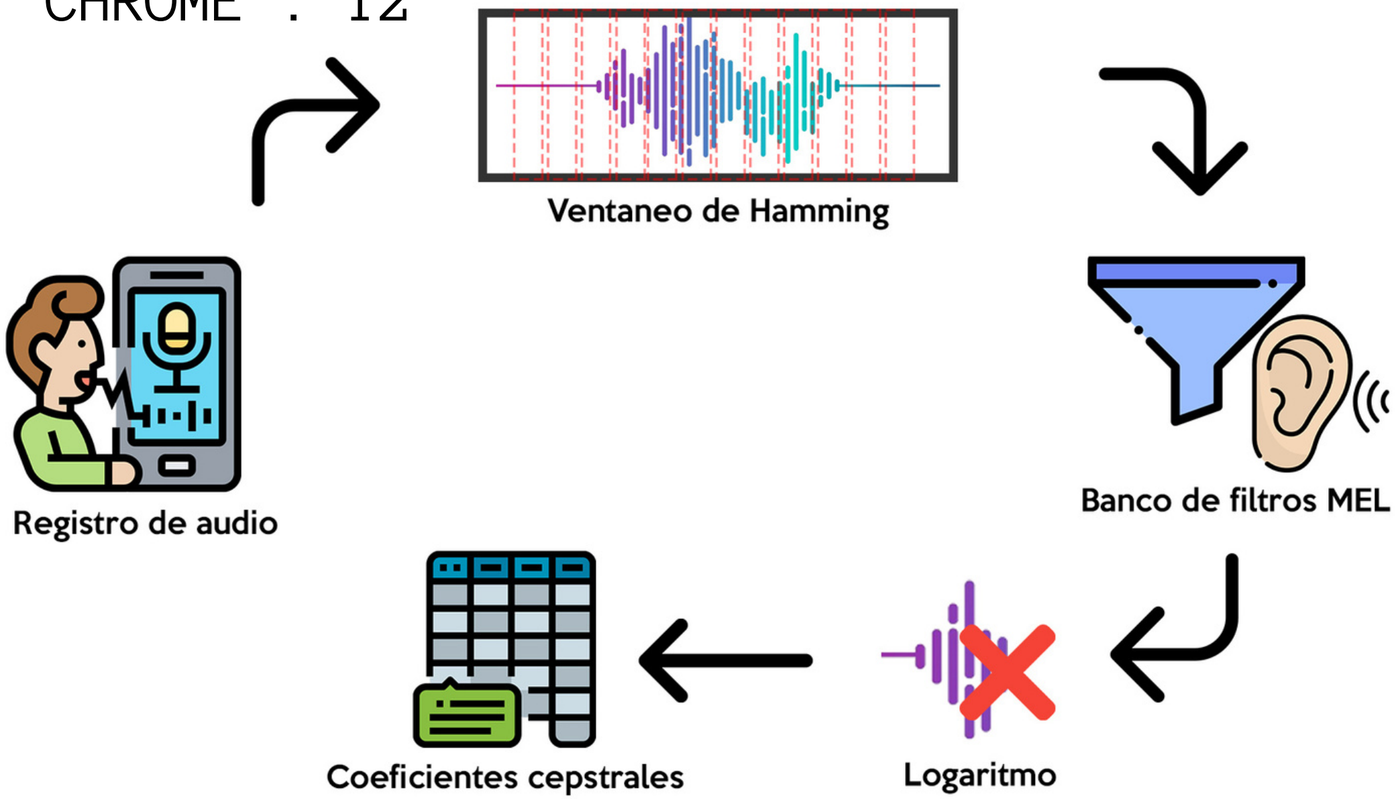
DATASET:
65.000 Audios de 1 seg
30 categorías

INFORMACIÓN

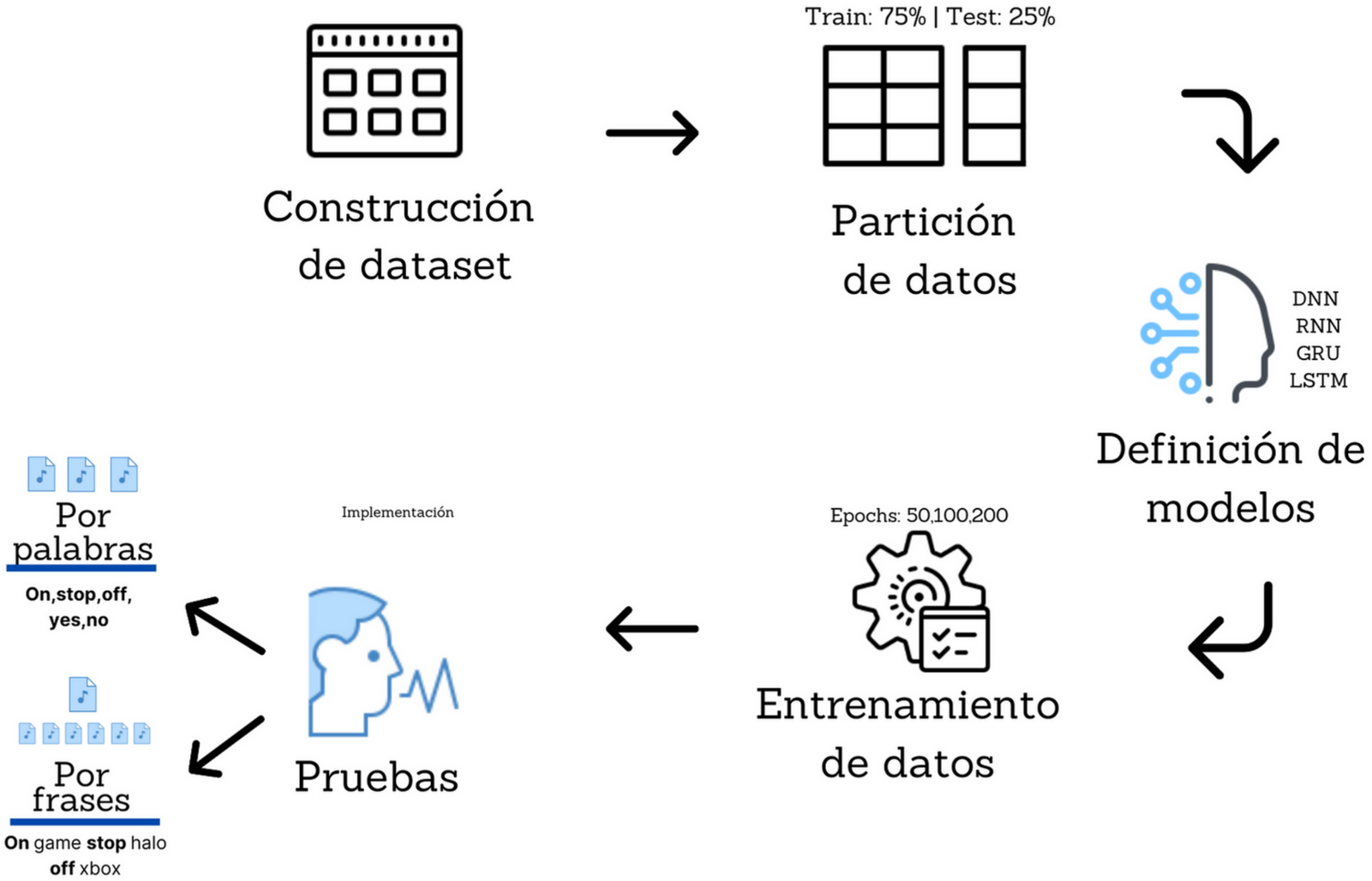
Universidad Industrial de Santander
TensorFlow
librosa
scikit learn
NumPy

CARACTERÍSTICAS

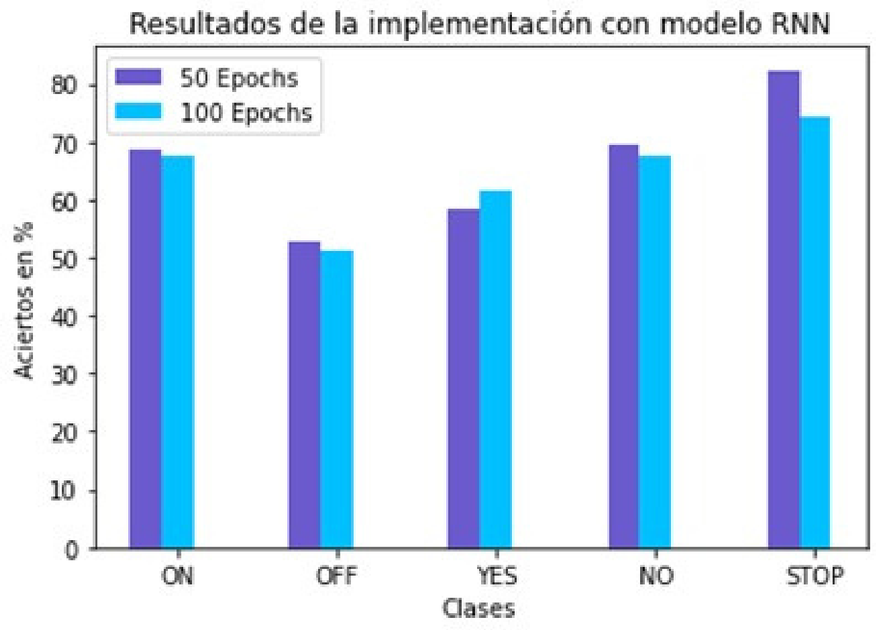
MFCC : 80 | CHROME : 12

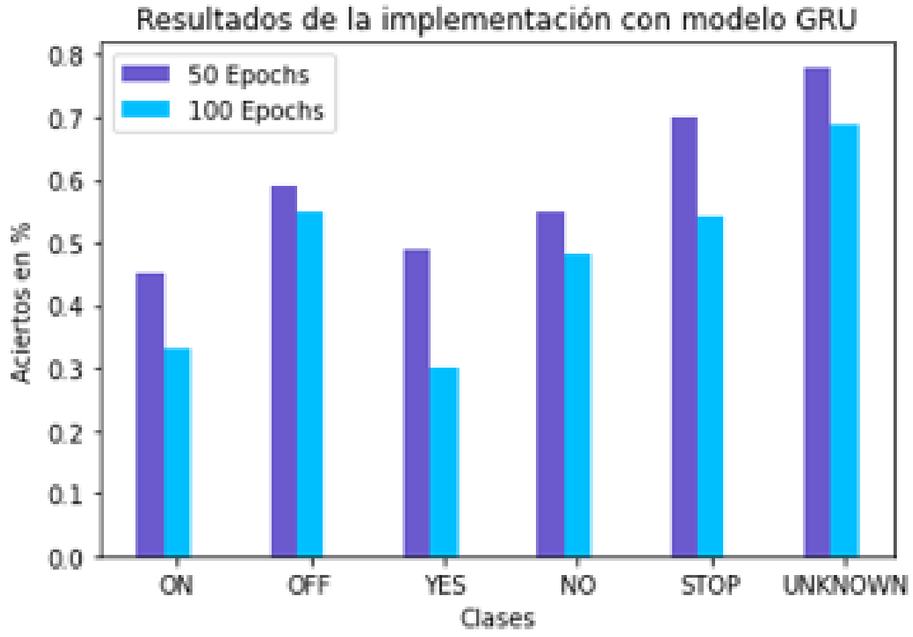


METODOLOGÍA



RESULTADOS DEL DATASET





RESULTADOS DE LA IMPLEMENTACIÓN

Análisis de Resultados RNN	Comandos		
	# comandos detectados	# comandos reales	% Acierto
	3	5	60%
	4	5	80%
	5	5	100%
	4	5	80%
	5	5	100%
Totales	21	25	84%

Análisis de Resultados DNN	Comandos		
	# comandos detectados	# comandos reales	% Acierto
	3	2	67%
	2	2	100%
	3	3	100%
	3	2	67%
	1	1	100%
Totales	12	10	83%

Palabras
on
off
yes
no
stop

Frases
On game **Stop** halo **off** xbox
Off digital clock **not** widget
not audio **stop** volume **not** dvd
music **on** juanes **yes** la camisa negra **yes**
not computer display

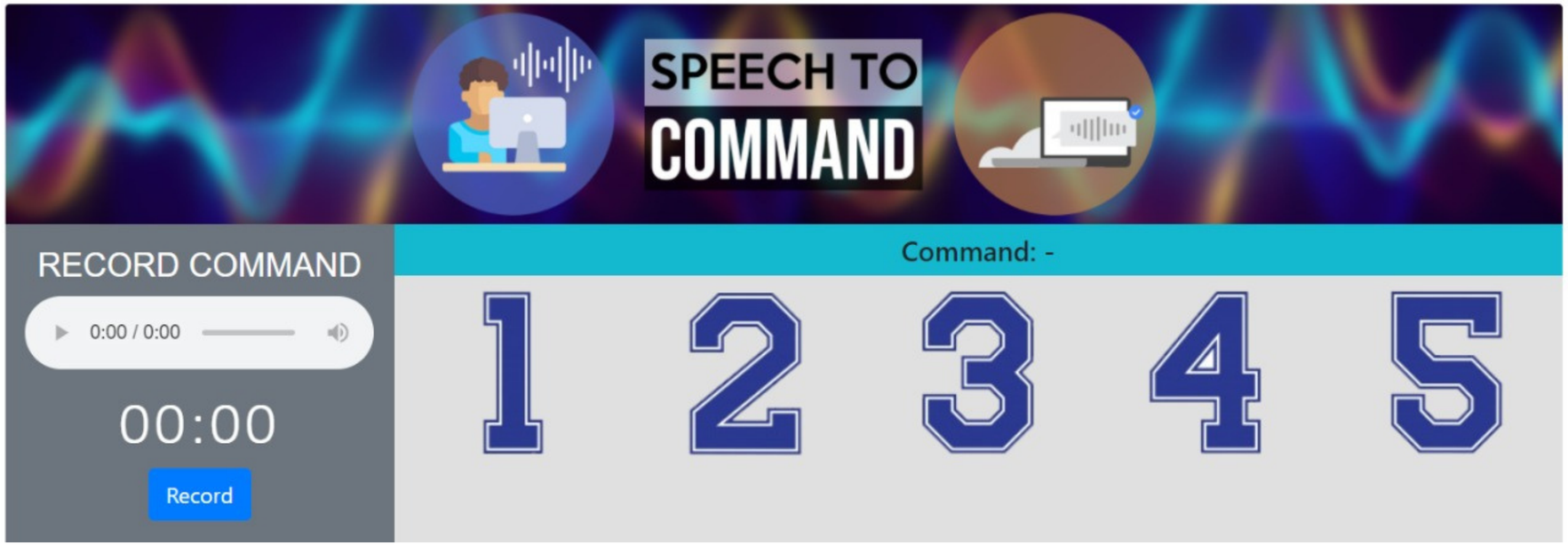
CONCLUSIONES

- Las características frecuenciales MFCC y Chroma, en conjunto, son una representación que garantiza una predicción aceptable de los comandos de voz. Se evidenció que el ruido influye significativamente en los resultados debido a la alta sensibilidad de los micrófonos convencionales.

- La detección de comandos en una frase es un reto de mayor complejidad en comparación a la detención de comandos por palabras, ya que el tratamiento de la frase captada depende de su duración y nivel de potencia para poder identificar cada palabra en esta.

- Una arquitectura sencilla como la DNN puede lograr un rendimiento igual o mayor de óptimo con respecto a las redes neuronales recurrentes. En las pruebas reales, el modelo con más aciertos es el DNN (Adamax). Por otro lado, para las pruebas con la partición test del dataset, el modelo GRU y RNN simple poseen una mayor precisión.

TRABAJO A FUTURO



REFERENCIAS

- https://www.kaggle.com/c/tensorflow-speech-recognition-challenge
- https://towardsdatascience.com/how-i-understood-what-features-to-consider-while-training-audio-files-eedfb6e9002b
- http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/