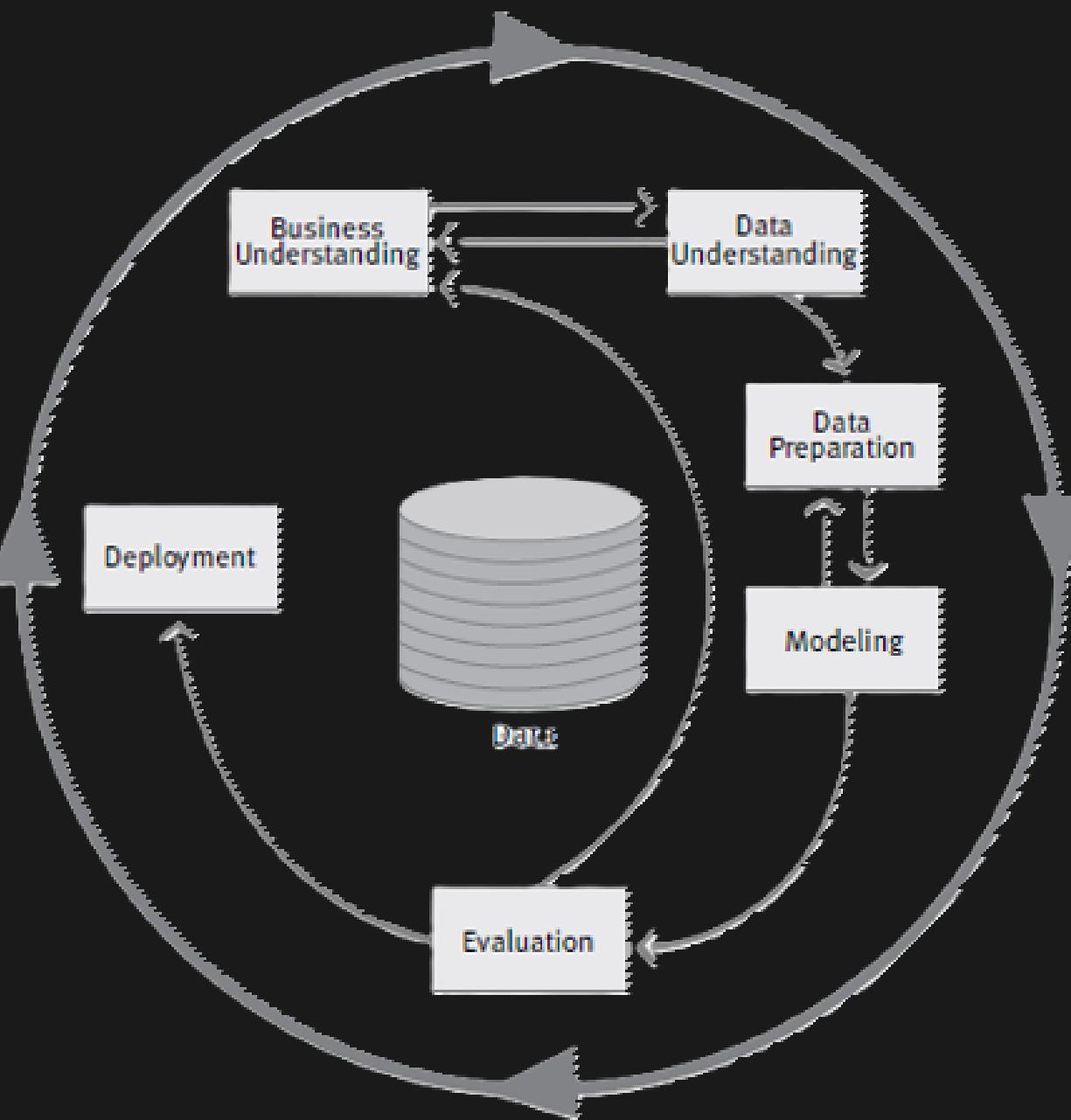




[HTTPS://BERGASPAR.GITHUB.IO/PORTFOLIO/](https://bergaspar.github.io/portfolio/)

# PROJECT TIMELINE



# OBJETIVOS Y CRITERIOS

## Objective n° 1

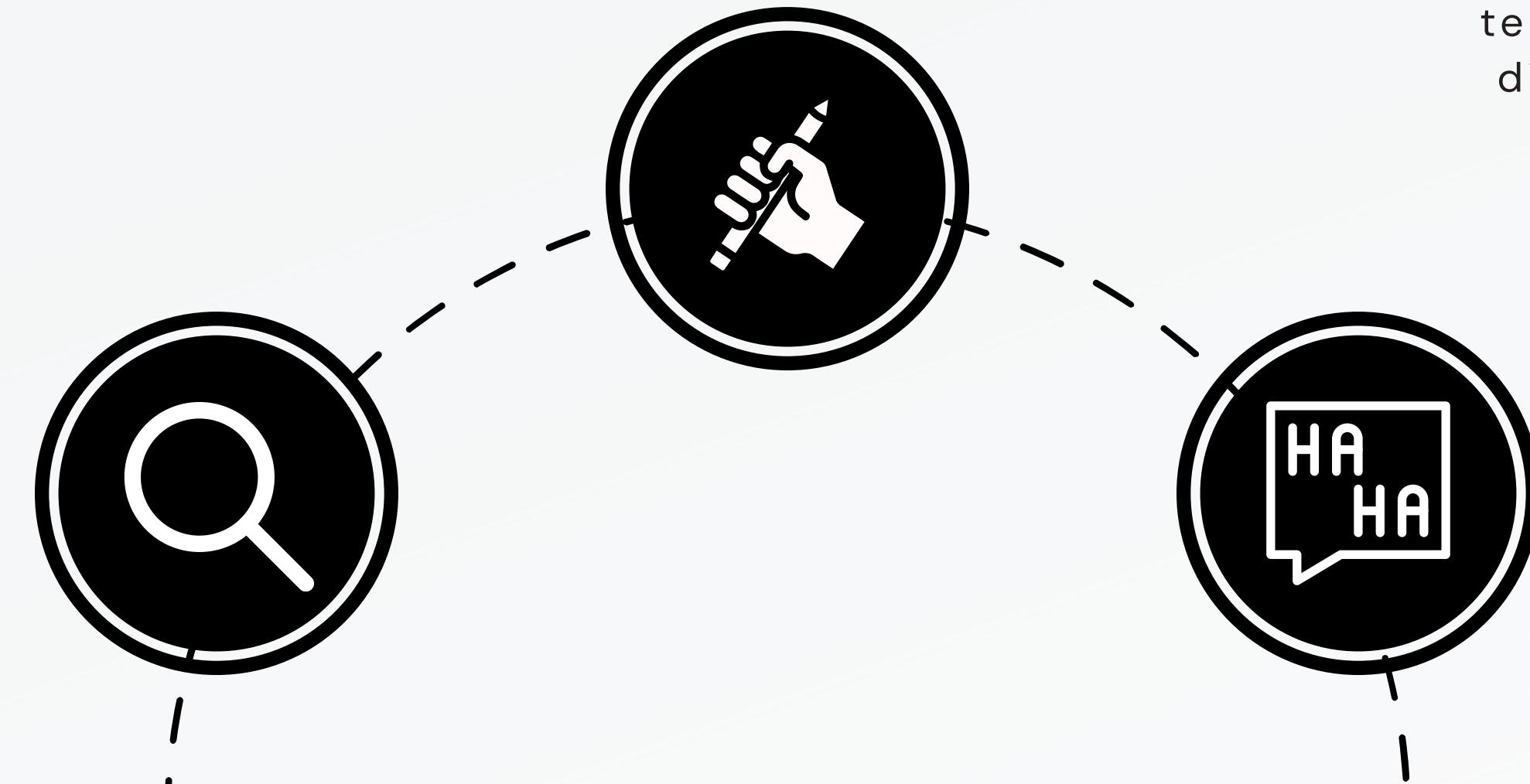
Identificar frases sarcásticas y distinguir entre noticias falsas y noticias legítimas.

## Objective n° 2

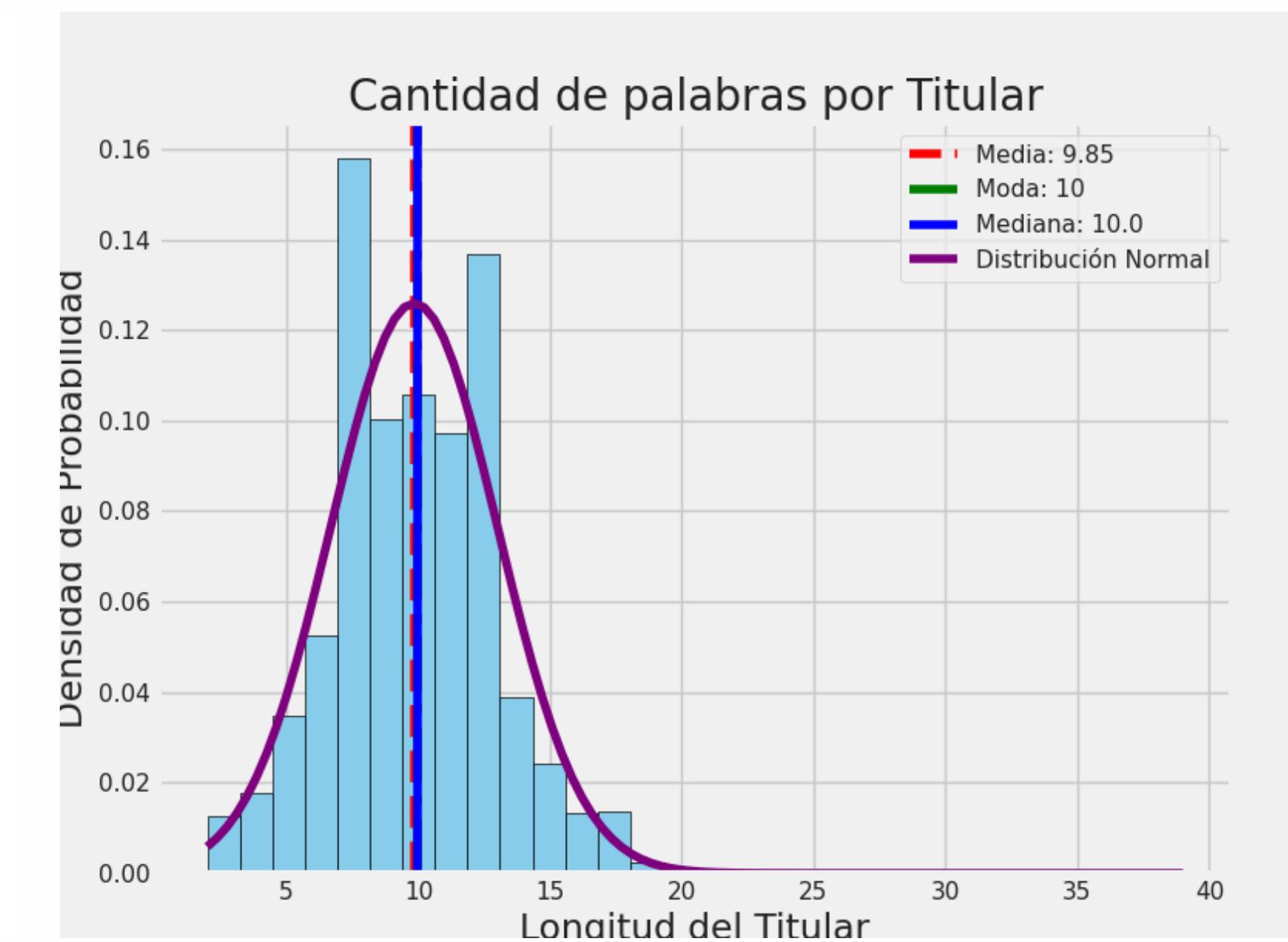
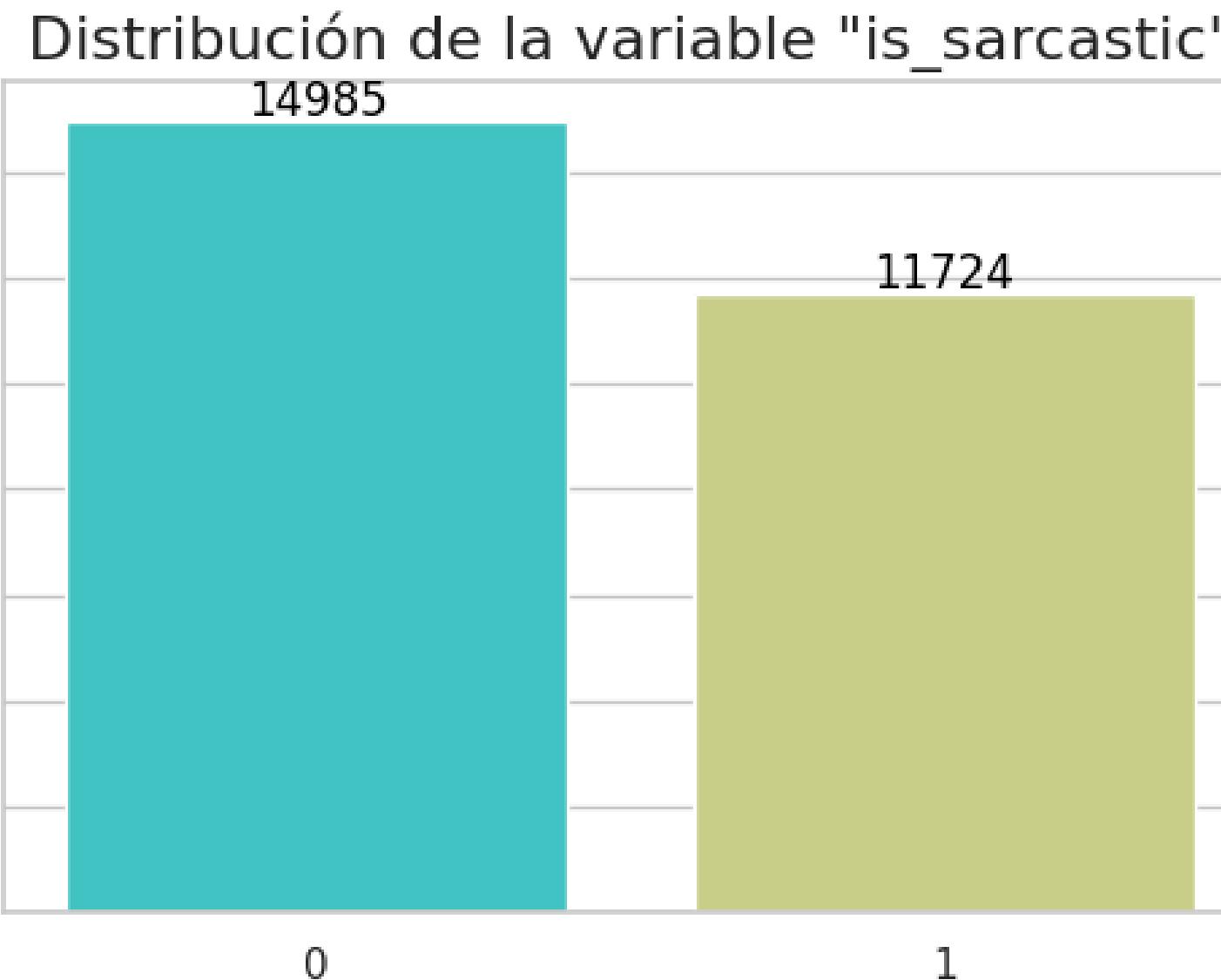
Practica y aprendizaje de modelos usando lenguaje natural (NLP).

## Objective n° 3

Explorar y entender las características del sarcasmo en los titulares de noticias teniendo en cuenta los diferentes contextos.



# EXPLORACIÓN DE DATOS



# ANALISIS DE SARCASMO

# PALABRAS MÁS FRECUENTES EN LOS TITULARES DE NOTICIAS

# NO SARCASICAS



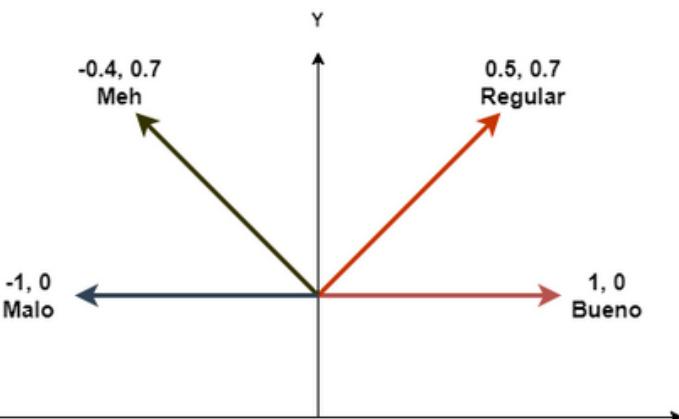
# SARCÁSTICAS



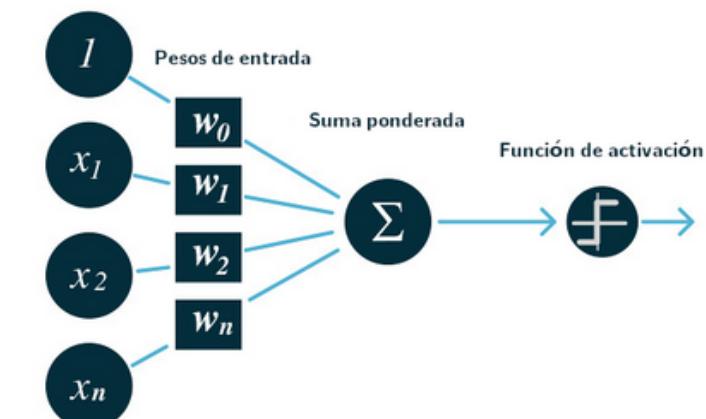
# MODELO

```
{'<OOV>': 1, 'to': 2, 'of': 3,  
'the': 4, 'in': 5, 'for': 6, '...'}  
[ 308, 15115, 679 ,3337,  
2298, 48, 382, 2576,  
15116, 6, 2577, 8434, 0, 0,  
0, 0, 0, 0, 0, 0, 0, 0, 0,  
0, 0, 0, 0, 0 ]
```

TOKENIZAR



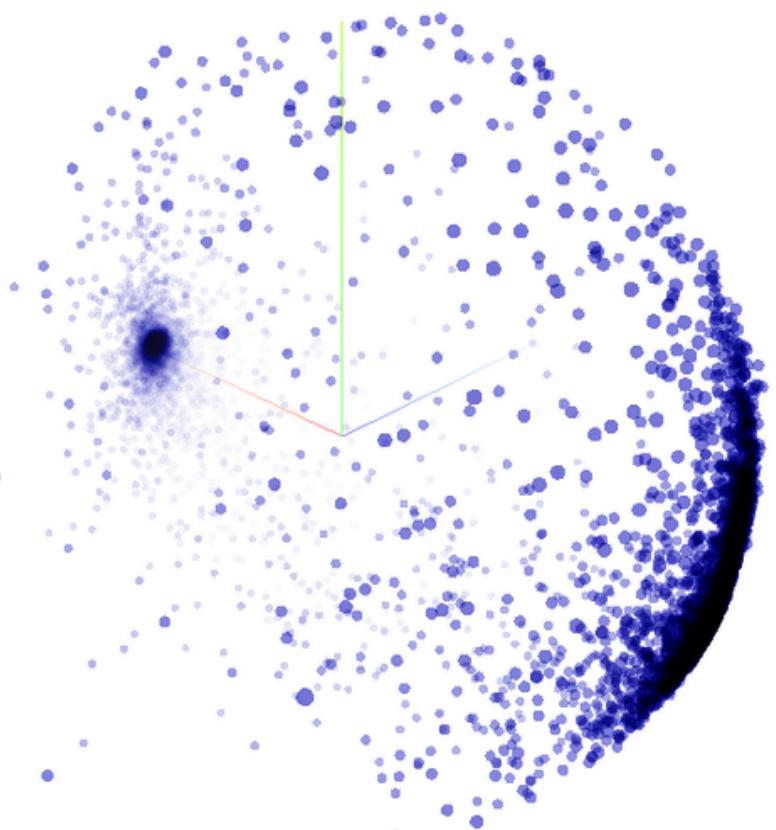
EMBEDDING



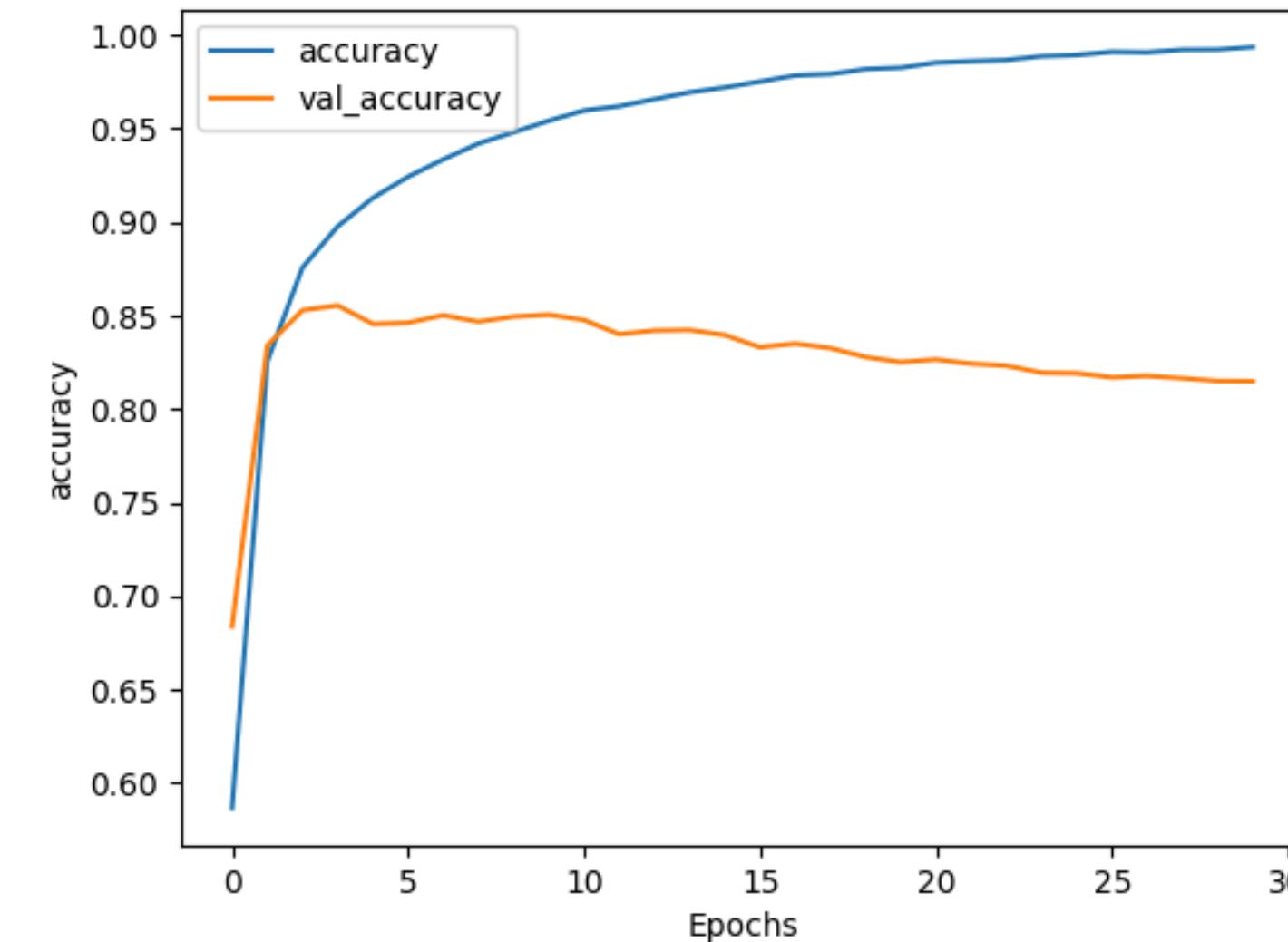
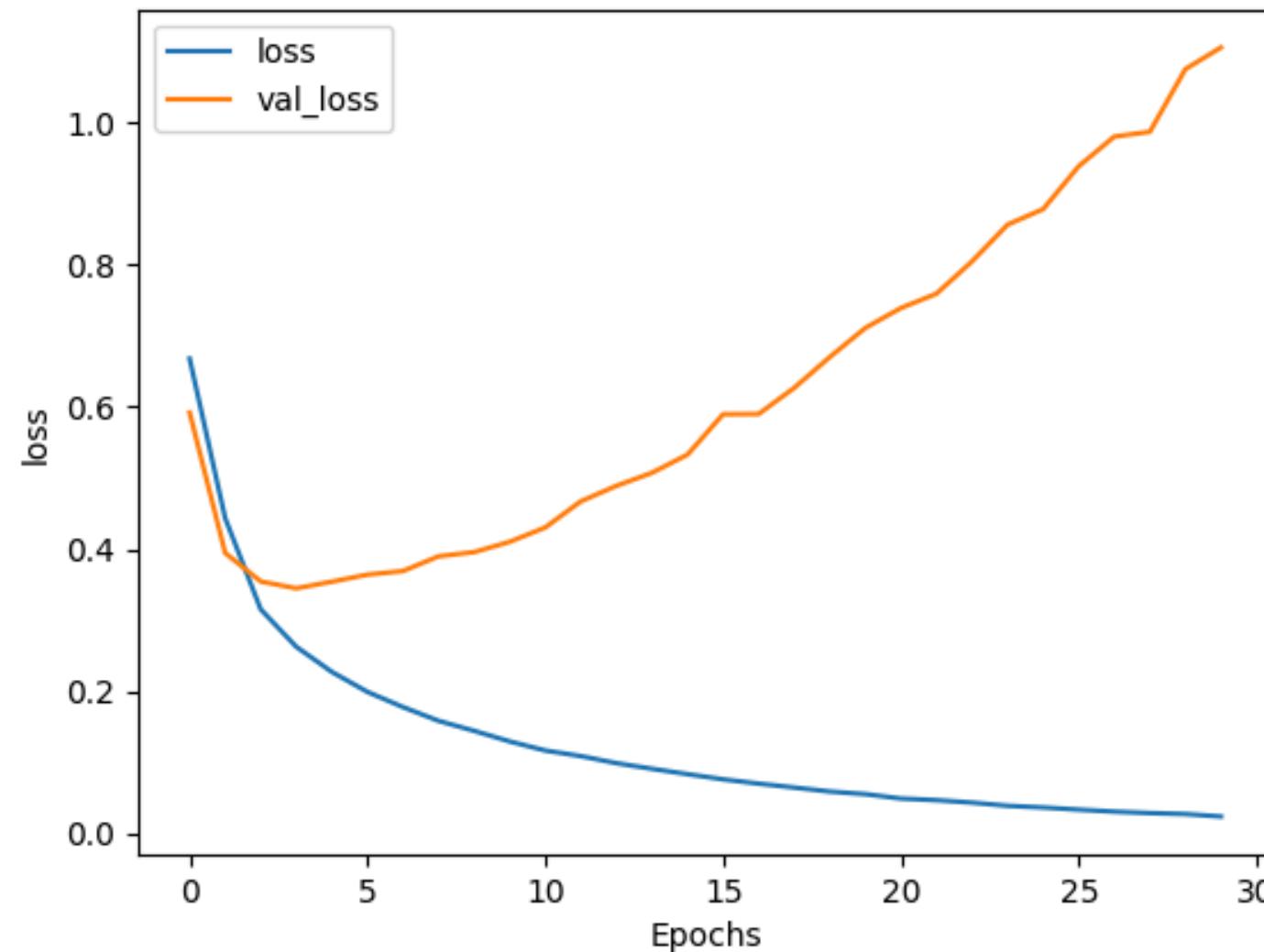
CNN

# VISUALIZACION Y EXPLORACION TUS INCRUSTACIONES

Se genera dos archivos, vecs.tsv y meta.tsv, que contienen los vectores de incrustación y las palabras correspondientes para visualizar y explorar las incrustaciones en <https://projector.tensorflow.org/>



# SOBREAJUSTE



# DESEMPEÑO DEL MODELO

```
[ ] sentence = ["granny starting to fear spiders in the garden might be real", "Harry Potter final movie showing this sunday night"]
sequences = tokenizer.texts_to_sequences(sentence) # Tokenizar
padded = pad_sequences(sequences, maxlen=max_length, padding=padding_type, truncating=trunc_type) # Rellenar

print('Oracione dadas:\n', sentence[0], '\n', sentence[1])
print(model.predict(padded)) # Utilizar el modelo entrenado para predecir.

Oracione dadas:
granny starting to fear spiders in the garden might be real
Harry Potter final movie showing this sunday night
1/1 [=====] - 0s 19ms/step
[[9.3817365e-01]
 [3.5251573e-09]]
```

El resultado de la predicción indica que:

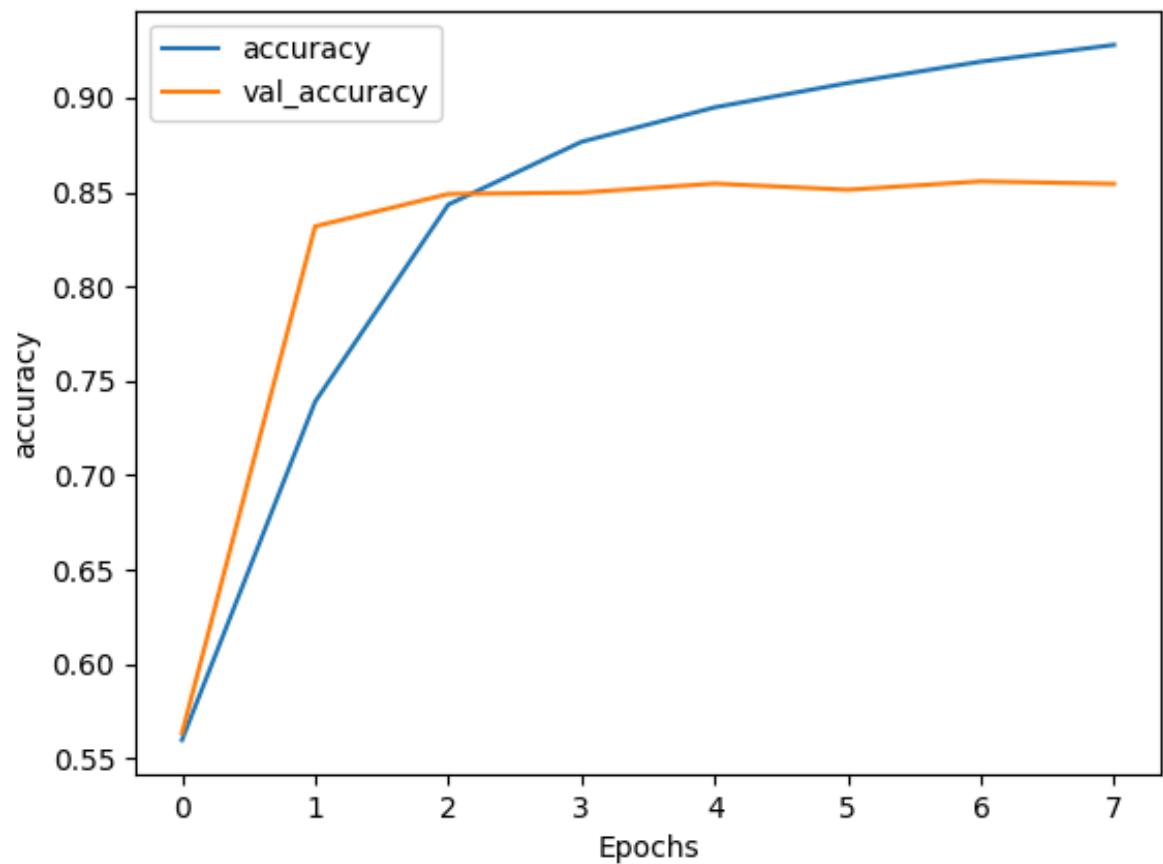
- La primera oración "**La abuela empieza a temer que las arañas en el jardín puedan ser reales**" ("granny starting to fear spiders in the garden might be real") tiene una **alta probabilidad** de ser sarcástica, con una probabilidad cercana a 0.94 (aproximadamente 93.8%).
- La segunda oración "**La última película de Harry Potter se proyecta este domingo por la noche**" ("Harry Potter final movie showing this sunday night") tiene una probabilidad **extremadamente baja** de ser sarcástica, con una probabilidad muy cercana a cero (aproximadamente 0.000000035%).

Estos resultados sugieren que el modelo considera que la primera oración es muy probablemente sarcástica, mientras que la segunda oración es casi definitivamente no sarcástica.

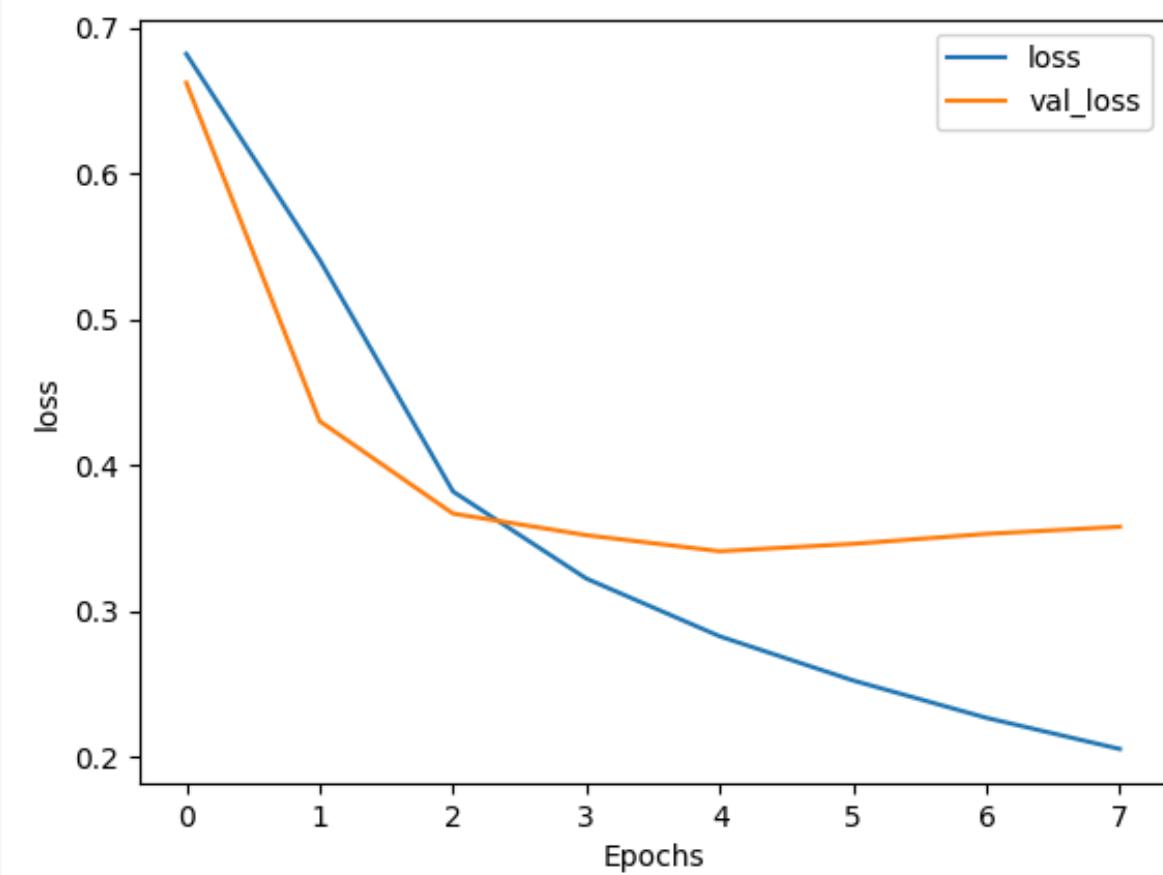
# OPTIMIZACION

- Regularización con Dropout
- Early Stopping
- Ajustar la Tasa de Aprendizaje del Optimizador

Accuracy



Loss



# CITACIÓN

1. MISRA, RISHABH AND PRAHAL ARORA. "SARCASM DETECTION USING NEWS HEADLINES DATASET." AI OPEN (2023).
2. MISRA, RISHABH AND JIGYASA GROVER. "SCULPTING DATA FOR ML: THE FIRST ACT OF MACHINE LEARNING." ISBN 9798585463570 (2021).

ENLACE A [RISHABHMISRA.GITHUB.IO/PUBLICATIONS](https://RISHABHMISRA.GITHUB.IO/PUBLICATIONS) COMO FUENTE DE ESTE CONJUNTO DE DATOS.  
¡GRACIAS!