



Detección Automática de Misoginia en MEMES Mediante Técnicas de Deep Learning

Trabajo de Fin de Grado
Pablo Cordón Hidalgo

ÍNDICE

01

INTRODUCCIÓN

- ❖ Objetivos
- ❖ Definición de **misoginia**
- ❖ Definición de **“meme”**

02

COMPETICIÓN SEMEVAL-2022

- ❖ Descripción de la tarea a resolver
- ❖ Tecnologías utilizadas
- ❖ Análisis del conjunto de datos
- ❖ Soluciones propuestas y resultados

03

PROPUESTAS DE MEJORA

- ❖ Marco de experimentación
- ❖ Optimización de hiperparámetros
- ❖ Ensembles

04

CONCLUSIONES

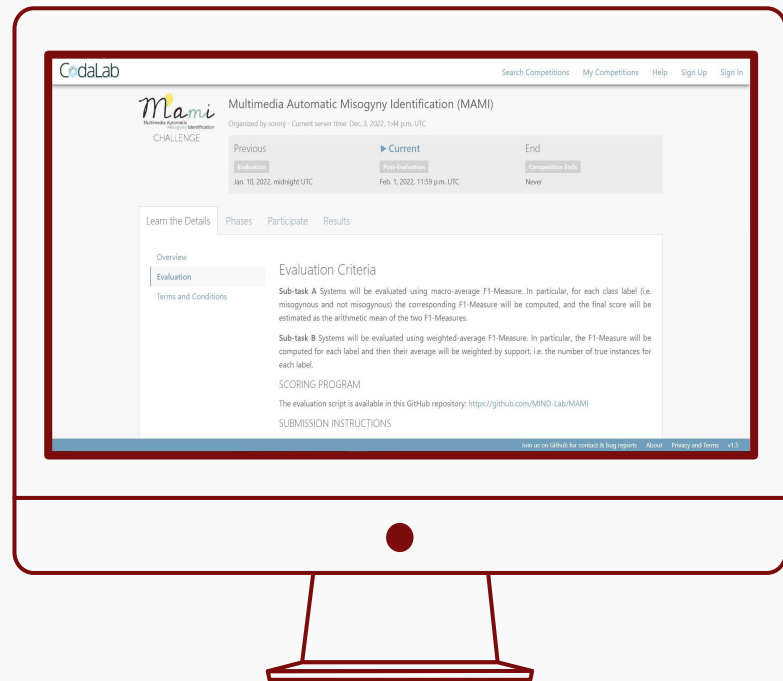
- ❖ Conclusión
- ❖ Trabajo futuro

Introducción

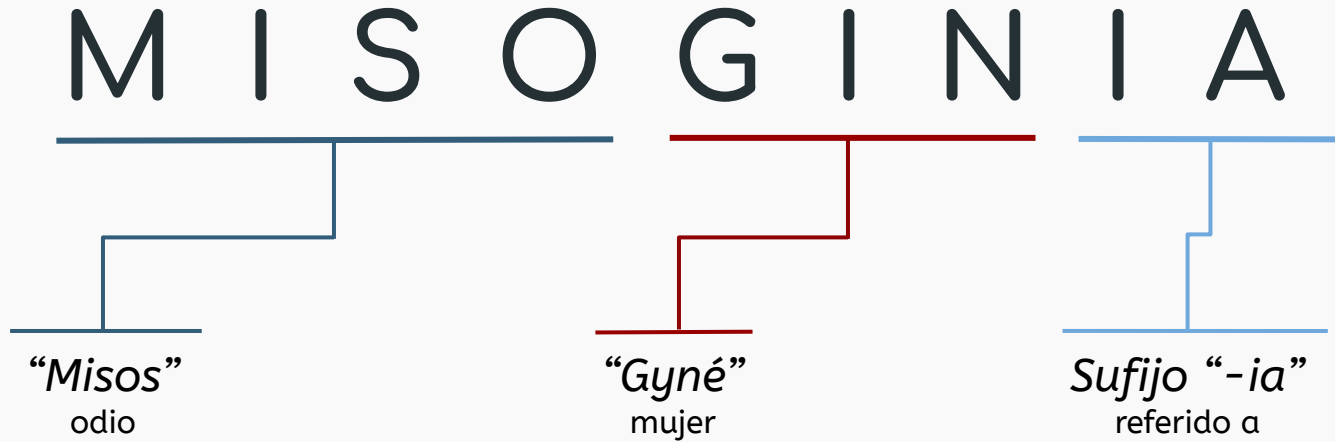


1.1 Objetivos

- ❖ Participación tarea 5 de **SemEval-2022**: detección de misoginia en memes en inglés
- ❖ Implementación modelos *Machine Learning* y *Deep Learning*
- ❖ Publicación de un *paper* científico



1.2 ¿Qué es la misoginia?



Del griego *misogynía*, **aversión**, **desprecio** u **odio** a las mujeres

1.2 ¿Qué es un meme?

- ❖ **Imagen** con **texto** superpuesto a posteriori con un significado conjunto
- ❖ Objetivo: ser divertido y/o irónico
- ❖ Gran valor como manifestación cultural
- ❖ Amplia difusión en entornos digitales



1.2 ¿Qué es un meme?

¿Cómo sabemos cuando un meme es misógino?

- Aunque en ocasiones hay características que nos dan pistas, **no siempre es tan sencillo**

NECESIDAD: Automatizar la detección de memes misóginos de la forma más precisa posible



Competición SemEval-2022



2.1 Descripción de la tarea a resolver

Subtarea A de la tarea 5 de SemEval2022

Clasificación binaria una sola etiqueta

Cada meme será misógino (1) o no misógino (0)

2.2 Tecnologías utilizadas



PYTHON



COLAB



TENSORFLOW



HUGGINGFACE



NUDENET

2.3 Análisis del conjunto de datos

Conjuntos de train y test

	Train	Test
Misógino	5.000	500
No Misógino	5.000	500
Total	10.000	1.000

Fecha de publicación

- Conjunto de train (periodo de entrenamiento)
1 de octubre de 2021
- Conjunto de test sin etiquetar (periodo de evaluación)
10 de enero de 2022
- Conjunto de test etiquetado (post-evaluación)
2 de Febrero de 2022

2.3 Análisis del conjunto de datos

file_name	misogynous	shaming	stereotype	objetification	violence	Text Transcription
10.jpg	1	0	0	0	1	ROSES ARE RED, VIOLETS ARE BLUE IF YOU DON'T SAY YES, I'LL JUST RAPE YOU quickmeme.com



2.3 Análisis del conjunto de datos

file_name	misogynous	shaming	stereotype	objetification	violence	Text Transcription
1.jpg	0	0	0	0	0	Milk Milk.zip



2.3 Análisis del conjunto de datos

Palabras Frecuentes

Clase 1



Clase 0



2.4 Soluciones Propuestas

Métricas utilizadas

Matriz de confusión

		Predicción	
		Negativo	Positivo
Valor real	Negativo	True Negative (TN)	False Positive (FP)
	Positivo	False Negative (FN)	True Positive (TP)

f1-score $f1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$

precisión $precision = \frac{TP}{TP + FP}$

recall $recall = \frac{TP}{TP + FN}$

accuracy $acc = \frac{TP + TN}{TP + TN + FP + FN}$

2.4 Soluciones Propuestas

Preprocesamiento del texto

V1

- ❖ Caracteres a minúsculas
- ❖ Eliminación de páginas web
- ❖ signos de puntuación
- ❖ números.
- ❖ palabras < 2 caracteres
- ❖ stopwords

V2

- ❖ Funcionalidades V1
- ❖ Eliminación de usuarios (@...)
- ❖ Hashtags (#...)
- ❖ Enlaces

V3

- ❖ Específico para modelos **Transformer**
- ❖ Funcionalidades V2 excepto:
- ❖ **NO** se eliminan stopwords
- ❖ **NO** se eliminan caracteres menores a cierta longitud

2.4 Soluciones Propuestas

Nudenet

Umbral de inseguridad: **0.45**

file_name	misogynous	Text Transcription	Unsafe
10.jpg	1	roses are red...	0.0005
1002.jpg	1	Its super rare...	0.691
10029.jpg	0	just call me harry...	0.003
10039.jpg	0	valentines days coming...	0.102

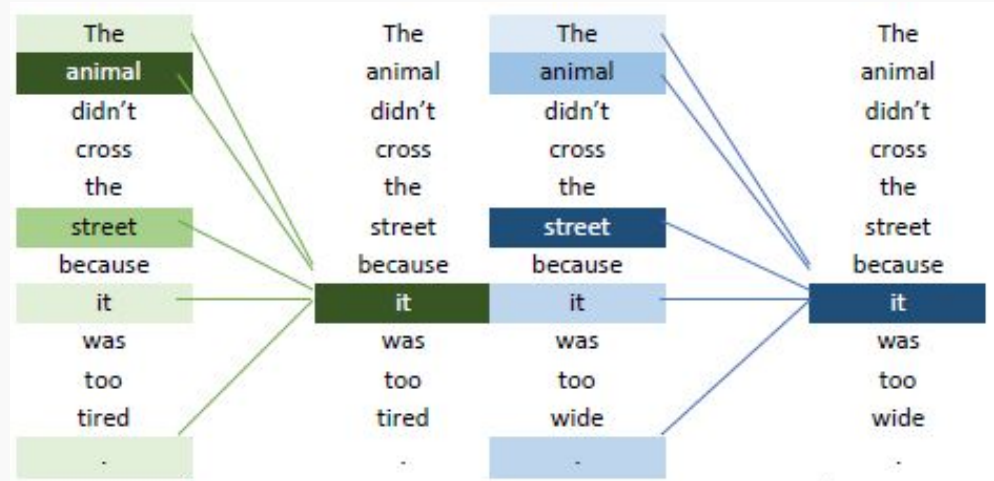
2.4 Soluciones Propuestas

Transformers

Técnica más usada actualmente para NLP

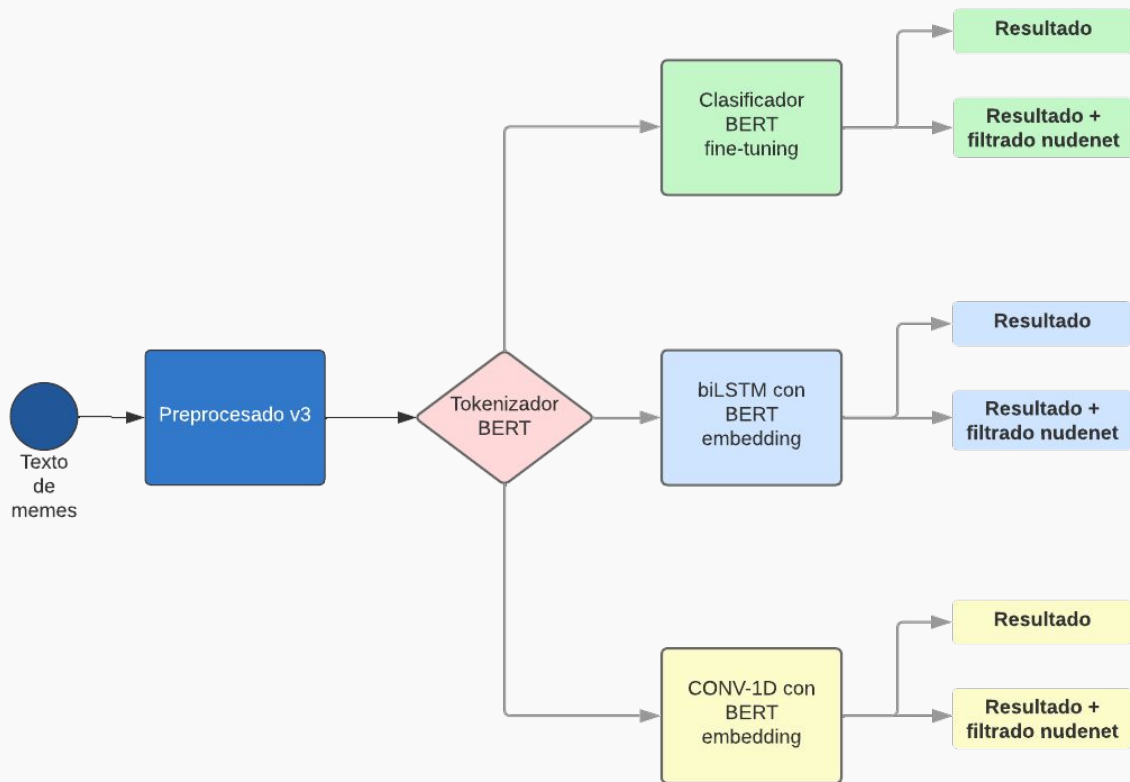
Modelos de Deep Learning que hacen uso de capas de atención

BERT es el modelo pre-entrenado más famoso, se desarrollan los modelos **BERT-based**



2.4 Soluciones Propuestas

Transformers



2.4 Soluciones Propuestas

Transformers

Modelo	accuracy	precision	recall	F1-score
BiLSTM BERT embedding + nudenet	0.670	0.66	0.65	0.665
BiLSTM BERT embedding	0.662	0.65	0.64	0.652
Clasificador BERT Fine-tuning + nudenet	0.658	0.69	0.66	0.649
Clasificador BERT Fine-tuning	0.654	0.67	0.65	0.645
Conv1D BERT embedding + nudenet	0.641	0.69	0.64	0.639
Conv1D BERT embedding	0.635	0.68	0.62	0.630

2.4 Soluciones Propuestas

Transformers

Modelo	accuracy	precision	recall	F1-score
BiLSTM BERT embedding + nudenet	0.670	0.66	0.65	0.665

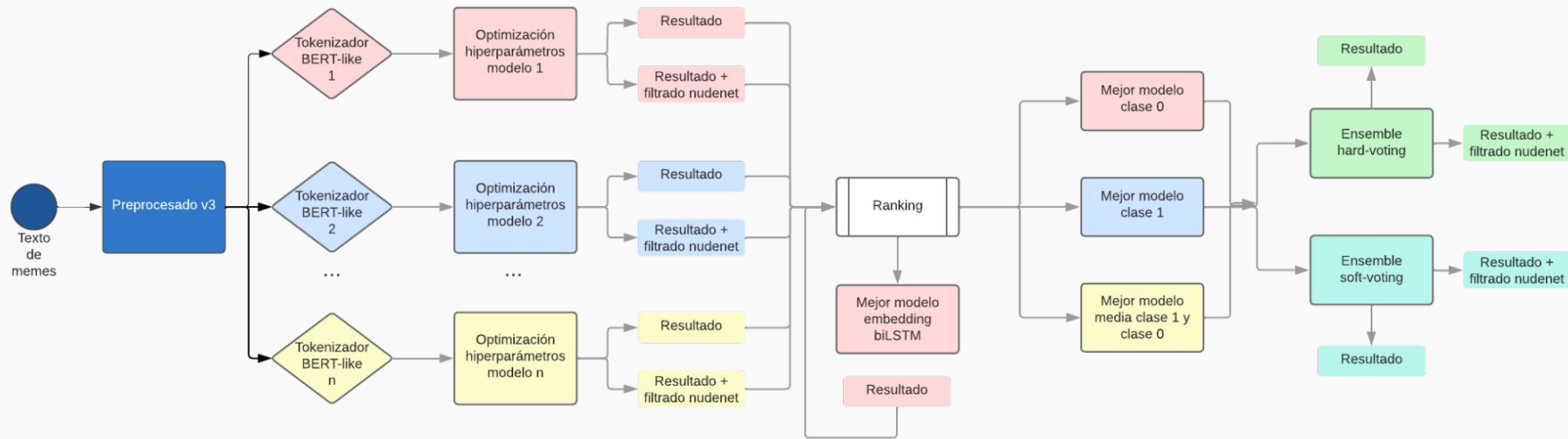
Posición **43** de 83 participantes

Propuestas de mejora



3

3.1 Marco de Experimentación



3.2 Optimización de hiperparámetros

Parámetros elegidos

Parámetro	Valores
número de épocas	[2, 3, 4]
batch size	[16, 32]
learning rate	[4e-5, 3e-5, 2e-5]

3.2 Optimización de hiperparámetros

Modelos elegidos

bert-base-uncased

bert-large-uncased

xlnet-base-cased

roberta-base

roberta-large-misogyny

distilbert-base-uncased

3.2 Optimización de hiperparámetros

Mejores Resultados

Modelo	Num epochs	Batch size	Learning rate	Acc nudenet	F1 Score nudenet	Acc	F1 Score
roberta-misogyny	3	32	3e-05	0.692	0.680	0.681	0.673
roberta-misogyny	4	32	4e-05	0.684	0.672	0.670	0.662
roberta-misogyny	3	16	3e-05	0.679	0.668	0.665	0.658
bert-base-uncased	3	32	2e-05	0.678	0.667	0.67	0.663
roberta-misogyny	4	16	4e-05	0.680	0.665	0.669	0.659

3.2 Optimización de hiperparámetros

Mejores Resultados

Modelo	Num epochs	Batch size	Learning rate	Acc nudenet	F1 Score nudenet	Acc	F1 Score
roberta-misogyny	3	32	3e-05	0.692	0.680	0.681	0.673



Mejora 2.26%



Posición 38
(+ 5)

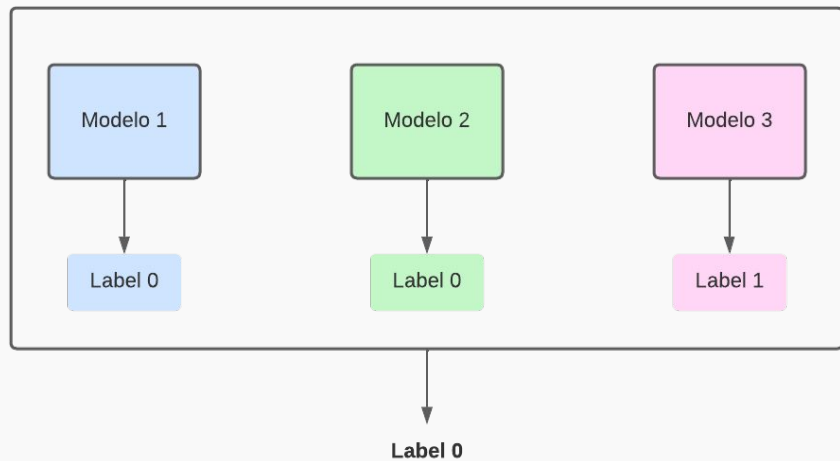
3.2 Optimización de hiperparámetros

Peores Resultados

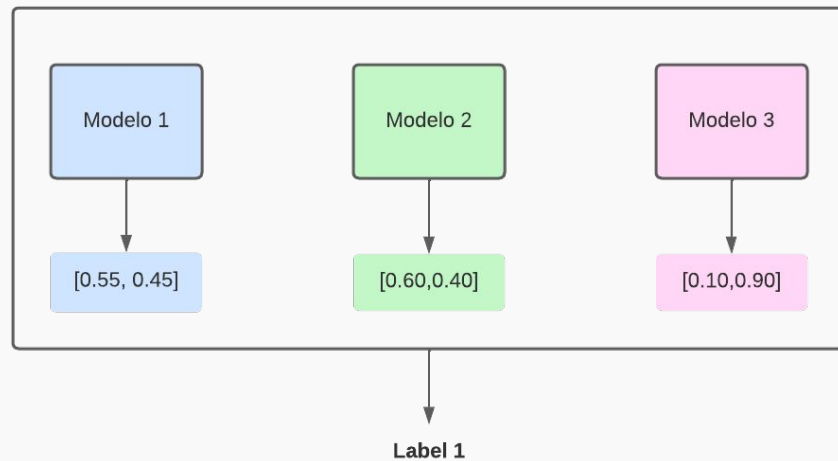
Modelo	Num epochs	Batch size	Learning rate	Acc nudenet	F1 Score nudenet	Acc	F1 Score
bert-large	2	16	4e-05	0.500	0.333	0.500	0.333
bert-large	2	32	4e-05	0.563	0.524	0.549	0.517
distilbert	2	16	3e-05	0.638	0.617	0.630	0.615
xlnet	2	32	4e-05	0.638	0.621	0.627	0.616
bert-large	2	16	3e-05	0.643	0.622	0.637	0.622

3.3 Ensembles

Hard voting

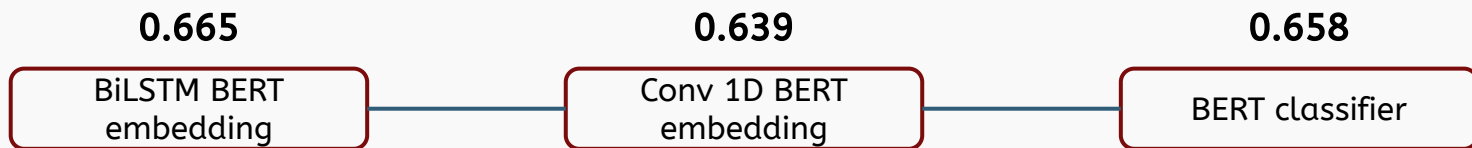


Soft voting



3.3 Ensembles

1ª aproximación – Mejores modelos etapa de evaluación



Modelo	accuracy	F1-score
Hard Voting	0.674	0.667
Soft Voting + nudenet	0.659	0.645
Soft Voting	0.651	0.642

3.3 Ensembles

2ª aproximación - F1-scores complementarias

F1 media - 0.69

BiLSTM RoBERTa
misogyny

F1 clase 1 - 0.74

RoBERTa misogyny
fine-tuning

F1 clase 0 - 0.72

RoBERTa misogyny
puro

Modelo	accuracy	F1-score
Soft Voting + nudenet	0.717	0.717
Hard Voting	0.709	0.710
Soft Voting	0.674	0.669

3.3 Ensembles

2ª aproximación

F1-scores complementarias

Modelo	accuracy	F1-score
Soft Voting + nudenet	0.717	0.717



Mejora 5.44%



Posición 22
(+ 21)

Conclusiones



4.1 Conclusiones

Paper científico

Eficacia de los transformers

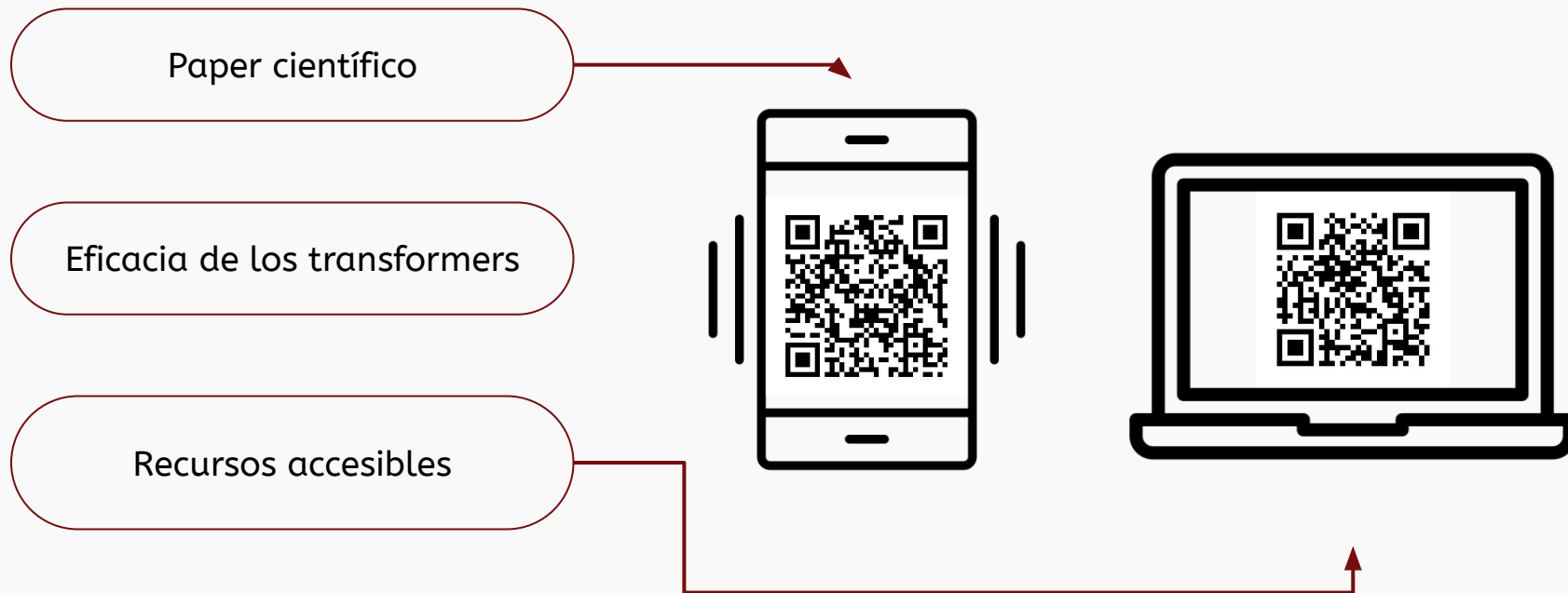
Recursos accesibles

	Competición	Mejoras
F1-score media	0.665	0.717
MEJORA DEL 7.82%		



Posición 22

4.1 Conclusiones



4.2 Trabajo Futuro

Experimentos con datasets
desbalanceados

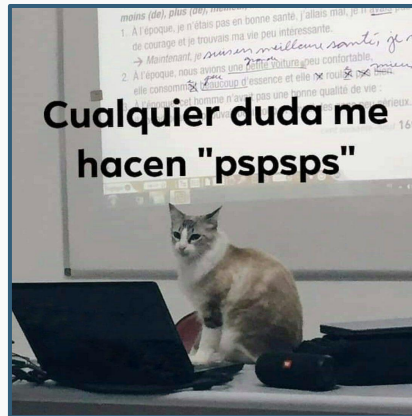
Uso de BD externas

Exploración de nuevas arquitecturas
(Multimodales)

GRACIAS

Dudas y preguntas

Pablo Cerdón Hidalgo
pablo.cordon113@alu.uhu.es





GRACIAS

Dudas y preguntas

Pablo Cordón Hidalgo
pablo.cordon113@alu.uhu.es

CREDITS: This presentation template was created by Slidesgo,
including icons by **Flaticon**, infographics & images by **Freepik**