

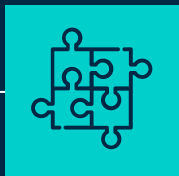
DETECCIÓN DE NOTICIAS FALSAS MEDIANTE TÉCNICAS DE DEEP LEARNING



Trabajo de Fin de Grado
Álvaro Esteban Muñoz



Índice



01

INTRODUCCIÓN

- Definición de “Fake New”
- Planteamiento del problema



02

EXPERIMENTACIÓN

- Preprocesamiento básico
- Ranking básico
- Preprocesamientos alternativos
- Ranking múltiple



03

MEJORAS

- Sintetización de texto
- Head & Tail embeddings
- Ensembles



04

CONCLUSIONES

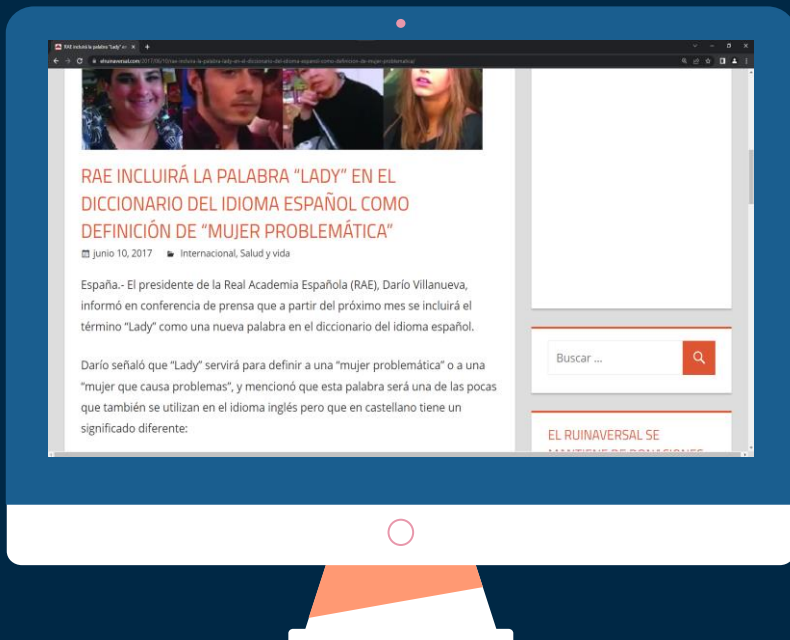
- Conclusión
- Trabajo futuro

Introducción

01

¿Qué es una noticia falsa?

- Información parcial/totalmente falsa
- Posee apariencia de noticia
- Objetivo: Provocar miedo, agitación, difamación o fraude.



¿Cómo sabemos si la noticia que tenemos delante es falsa o no?

- Aunque hay ciertas características que nos dan pistas, no siempre es tan sencillo.

NECESIDAD: **Automatizar** la detección de noticias falsas.

Planteamiento del problema

Descripción de la tarea

- Nos enfrentamos a FakeDeS, una tarea propuesta por IberLEF en su edición de 2021.
- “Decidir si una noticia es *real* o *falsa* analizando su representación textual.”

Planteamiento del problema

Datos

Train & Validation sets

	TRAIN	VALIDATION	TOTAL
VERDADERAS	338	153	491
FALSAS	338	142	480
TOTAL	676 (70%)	295 (30%)	971 (100%)

Planteamiento del problema

Datos

Test set

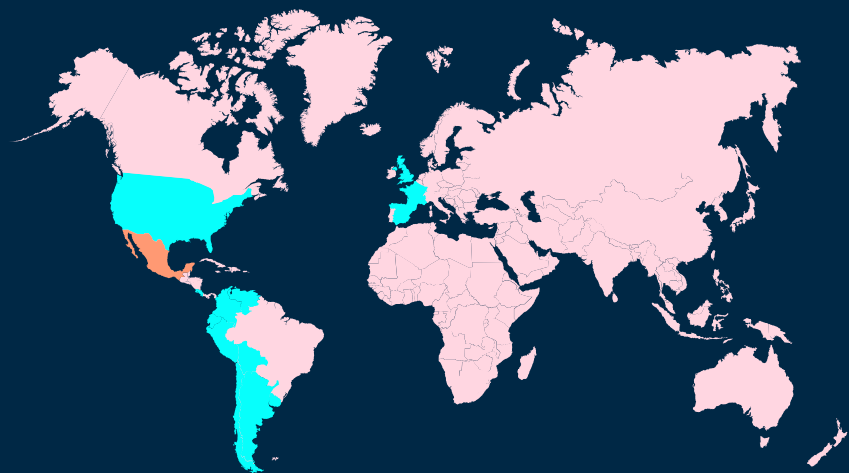
TEST

VERDADERAS	286
FALSAS	286
TOTAL	572

Periodo de extracción:

- Train & Validation:
Enero (2018) – Julio (2018)
- Test:
Noviembre (2021) – Marzo (2022)

Planteamiento del problema



■ Test split

■ Train & Valid split

Nuevos temas:

- COVID-19
- INTERNACIONAL
- AMBIENTE

Retos: Variabilidad en el **lenguaje** y los **temas**.

Planteamiento del problema

ID	1, 2, 3, 4, ...
Category	True/Fake
Topic	Science, Sports, Economy, Education, Entertainment, Politics, Health, Security, Society
Source	"Nombre del periódico"
Headline	RAE INCLUIRÁ LA PALABRA "LADY" EN EL DICCIONARIO DEL IDIOMA ESPAÑOL...
Text	España.- El presidente de la Real Academia Española (RAE), Darío Villanueva,...
Link	http://www.elruinaversal.com/2017/06/10/rae-incluire-la-palabra-lady-en-el-dicc...

Planteamiento del problema

F1-Score

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

Precisión

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

Recall

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

Accuracy

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

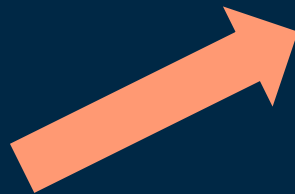
Confusion Matrix

		Predictions	
		Positive	Negative
Positive		True positive (TP)	False negative (FN)
Negative		False positive (FP)	True negative (TN)

Planteamiento del problema

Técnicas más utilizadas actualmente para
NLP → Transformers

- Modelos de Deep learning que hacen uso de mecanismos de atención.
- El más famoso es el conocido BERT, desarrollado por Google



Desarrollo de
“BERT-based
models”

Planteamiento del problema

¿Cómo sabemos que nuestros experimentos proporcionan un avance?

Tomaremos como línea base los resultados de los ganadores de IberLeF de 2021

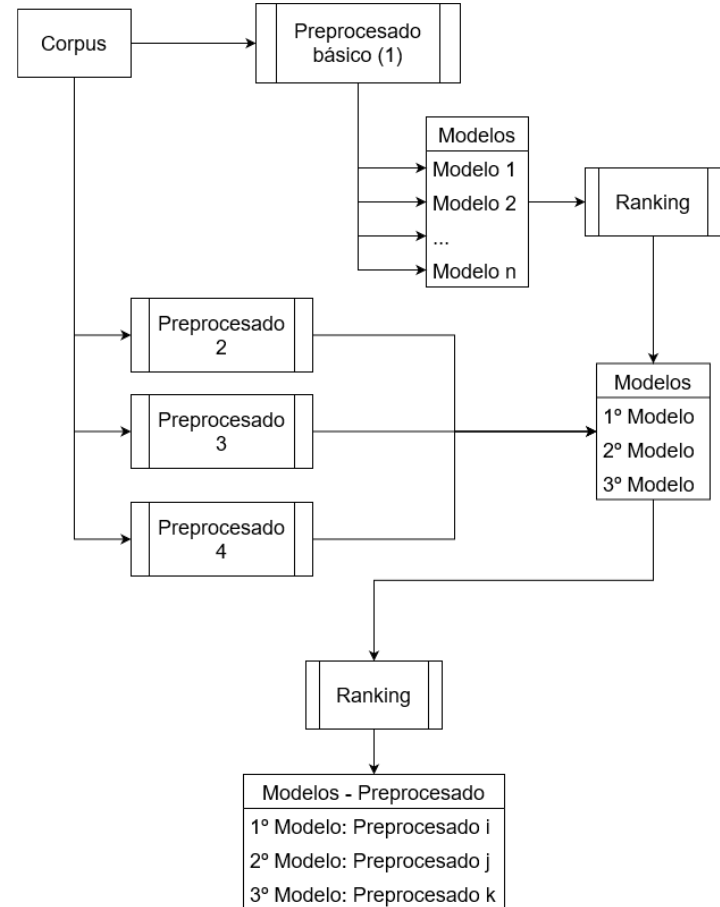
- Establecemos una línea base

	Validation set		Test set
bert-base-spanish-wwm-cased	Accuracy: 0,8644	F1: 0,8507	F1: 0,7666

EXPERIMENTACIÓN

02

Pipeline de experimentación



Preprocesado básico

- Objetivo: Construir una representación textual lo más **informativa** posible
 - Modificar el texto lo menos posible. Prescindimos de métodos de limpiado de texto como lowercasing, stopwords o punctuation removal.
 - Elegir los campos cuidadosamente.

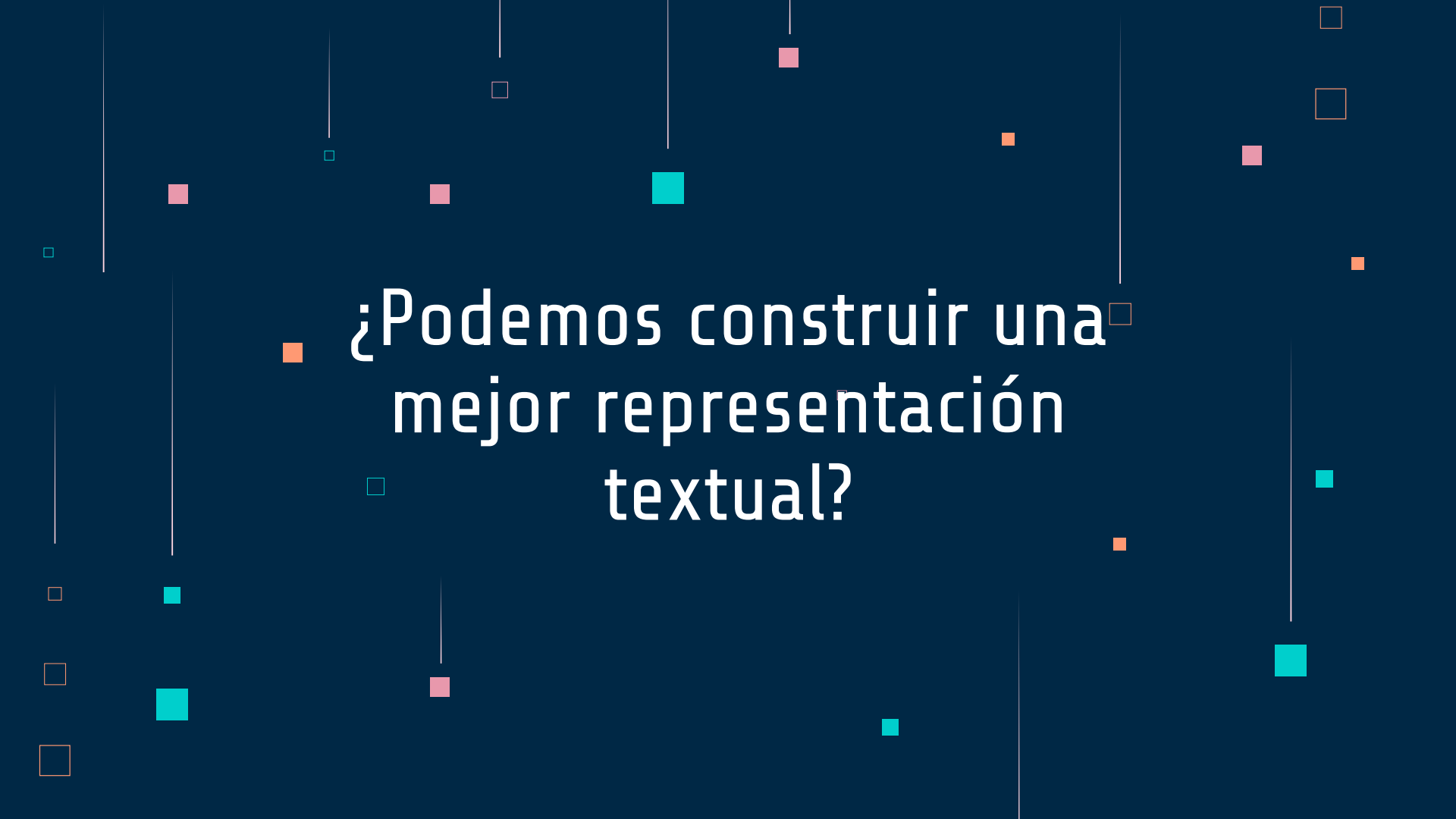
Primera aproximación: Campos **fundamentales**.

Representación textual



Ranking básico

Ranking	Model	Accuracy	F1-Score
1º	BERTIN	0,85139	0,84403
2º	RoBERTa	0,84650	0,84403
3º	BETO-uncased	0,83391	0,82632
4º	BASELINE	-	0,76660
5º	BETO-cased	0,74475	0,74475
6º	RuPERTa	0,60489	0,64687



■ ¿Podemos construir una
mejor representación
textual?

Preprocesamientos alternativos

2

Source

Topic

Link

Text

3

Source

Link

Text

4

Source

Link

Headline

Text

◀ ▶ Hoja de datos +

Preprocesamientos alternativos

F1-Score

Ranking	Model	Preprocess 1	Preprocess 2	Preprocess 3	Preprocess 4
---------	-------	---------------------	--------------	--------------	--------------

1º	BERTIN	0,84403	0,86260	0,81081	0,85714
----	--------	---------	---------	---------	---------

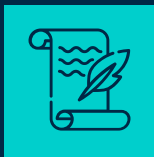
2º	RoBERTa	0,84403	0,85620	0,86330	0,77755
----	---------	---------	---------	---------	---------

3º	BETO	0,82632	0,81935	0,81717	0,82389
----	------	---------	---------	---------	---------

MEJORAS

03

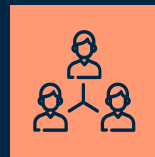
Mejoras



Sintetización de
texto



Head & Tail
embeddings

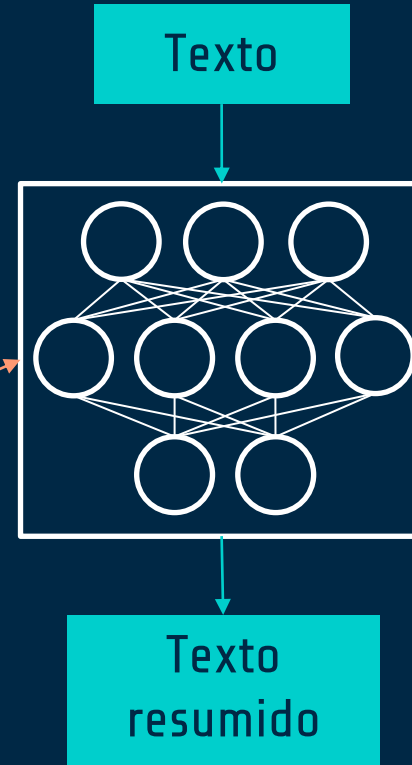


Ensembles

Sintetización de texto

- Objetivo: Evitar la redundancia en el cuerpo de la noticia.

Narrativa/Roberta2Roberta



Sintetización de texto

F1-Score

0,86330

Ranking

Model

Preprocessing

F1-Score

1º

BERTIN

2

0,84467

2º

RoBERTa

3

0,81849

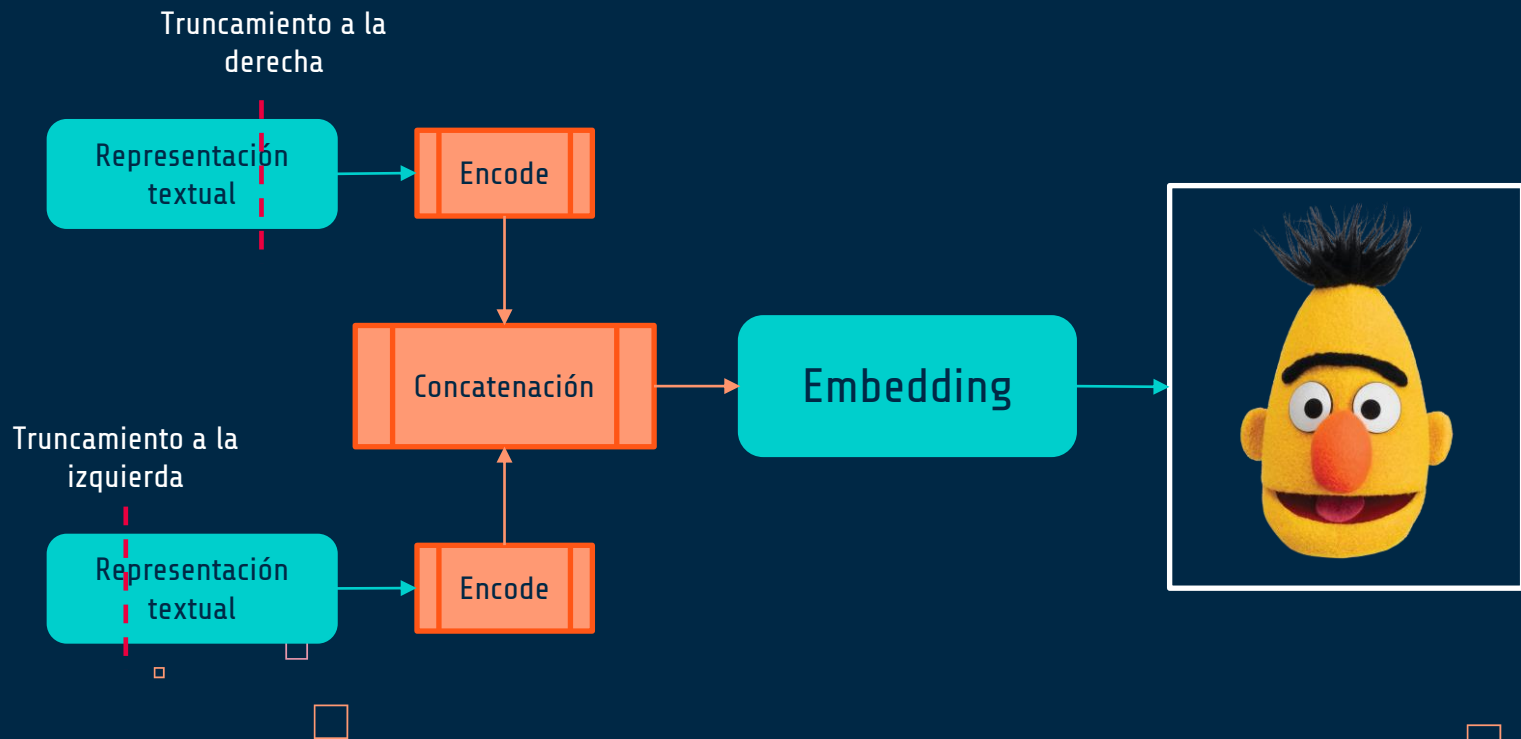
3º

BETO

1

0,81684

Head & Tail embeddings



Head & Tail embeddings

F1-Score

0,86330

Ranking

Model

Preprocessing

F1-Score

1º

BERTIN

2

0,86260

2º

RoBERTa

3

0,83968

3º

BETO

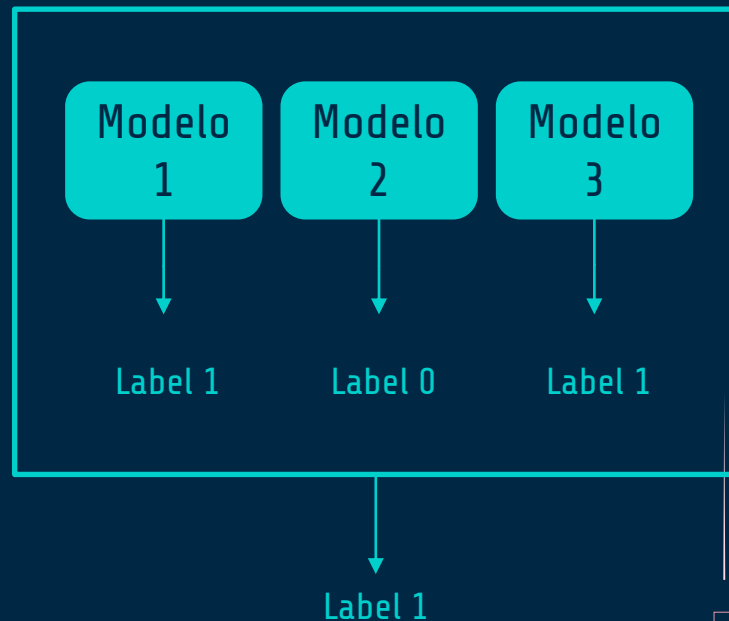
1

0,77544

Ensembles

Decidir la etiqueta mediante **votación** de varios modelos

- Aproximación 1: Mejores modelos
- Aproximación 2: Matrices de confusión complementarias



Ensembles

Predictions		
	Positive	Negative
Positive	264	22
Negative	76	210

Predictions		
	Positive	Negative
Positive	208	78
Negative	14	272

+ Modelo de apoyo
RoBERTa (3)

Ensembles

F1-Score

0,86330

Ranking

Model

F1-Score

1º

Ensemble 1

0,87931

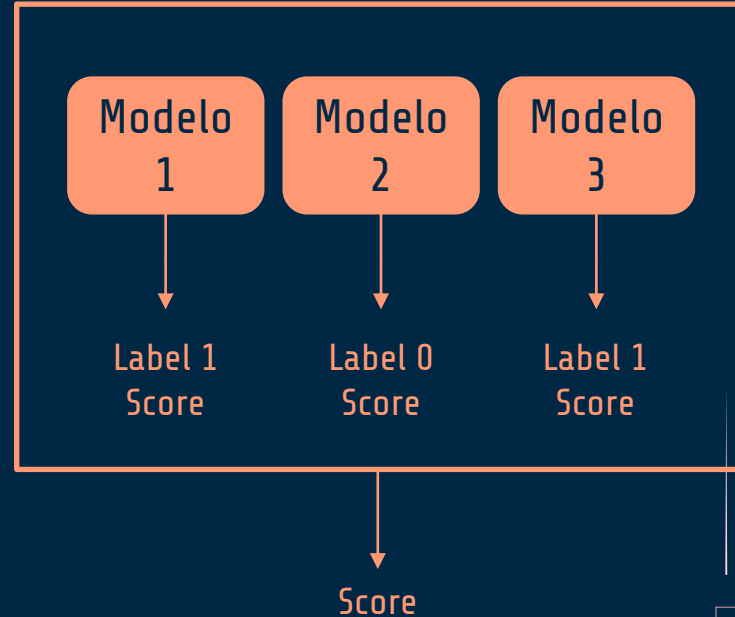
2º

Ensemble 2

0,87762

Ensembles (Ponderado)

- Normalizar la puntuación obtenida por cada modelo
- Suma ponderada de las puntuaciones
- Calcular la puntuación final
- Obtener la etiqueta asociada a dicha puntuación



Ensembles (Ponderado)

F1-Score

0,87931

Ranking

Model

Pesos (W_1, W_2, W_3)

F1-Score

1º

Ensemble 1

(0.5, 0.3, 0.2)

0,88421

2º

Ensemble 2

(0.5, 0.35, 0.15)

0,88421

3º

Ensemble 2

(0.5, 0.25, 0.25)

0,87830

Conclusiones

04

Mejor estrategia
Ensemble ponderado

F1-Score

0,88421

Baseline

bert-base-spanish-wwm-cased

F1-Score

0,76660

Modelo más eficaz
RoBERTa-large-bne

F1-Score

0,86330

Modelo más eficiente

bertin-roberta-based-spanish

F1-Score

0,86260

Trabajo futuro

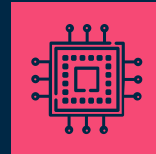
Técnicas de **data augmentation** (GANs)



Experimentos con datasets **desbalanceados**



Exploración de otras **arquitecturas** (GPT)



Dudas y preguntas

Álvaro Esteban Muñoz
alvaro.esteban775@alu.uhu.es

Gracias

CREDITS: This presentation template was created by [Slidesgo](#),
including icons by [Flaticon](#), and infographics & images by [Freepik](#)
Please keep this slide for attribution