



# Técnicas de *Deep Learning* Para la Detección Automática de Discurso Irónico en Redes Sociales

Trabajo Fin de Grado

Escuela Técnica Superior de Ingeniería  
Grado En Ingeniería Informática

29 de septiembre 2022

Autor: Adrián Moreno Monterde

Tutores: Jacinto Mata Vázquez

Victoria Pachón Álvarez

# Índice

01

## Introducción

...

- Objetivos
- Sarcasmo en Redes Sociales

03

## Propuestas de mejora

...

- Descripción del marco de experimentación
- Modelo RoBERTa-base
- Modelo RoBERTa-large
- Mejores métricas

02

## Competición SemEval2022

...

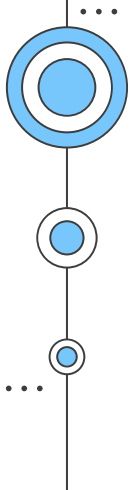
- Descripción de la tarea a resolver
- Análisis del conjunto de datos
- Soluciones propuestas
- Resultados de la etapa de evaluación

04

## Conclusiones y trabajo futuro

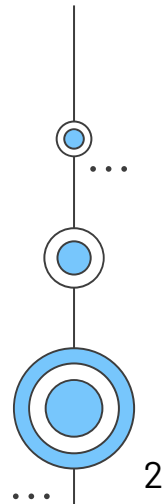
...

- Conclusiones
- Trabajo futuro



# 1

## Introducción

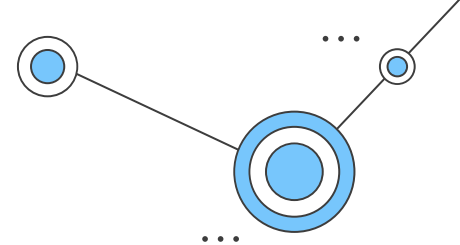


# 1.1 Objetivos

- Participación Tarea 6 *SemEval 2022*: Detección del sarcasmo intencionado en inglés.
- Implementación modelos *Machine Learning* y *Deep Learning*
- Publicación de un artículo científico



# 1.2 Sarcasmo en Redes Sociales

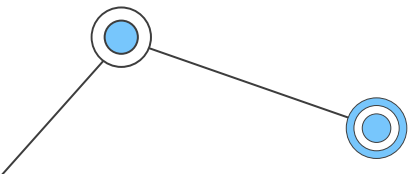
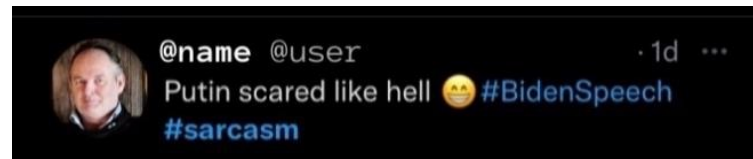


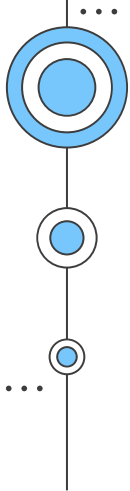
## ¿Qué es el Sarcasmo?

- Discrepancia entre el significado literal y pretendido
- Confuso sin previo contexto
- Omnipresencia en RRSS



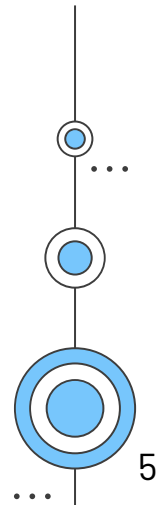
“Lo único que he aprendido en estos años de universidad es ... mi adicción a la cafeína”

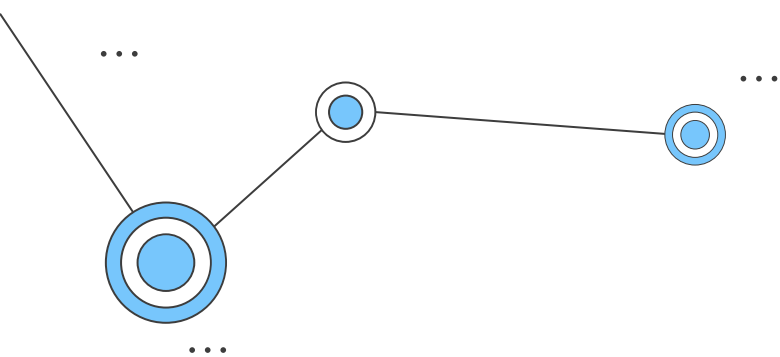




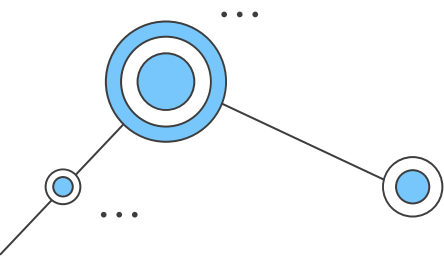
# 2

## Competición SemEval2022





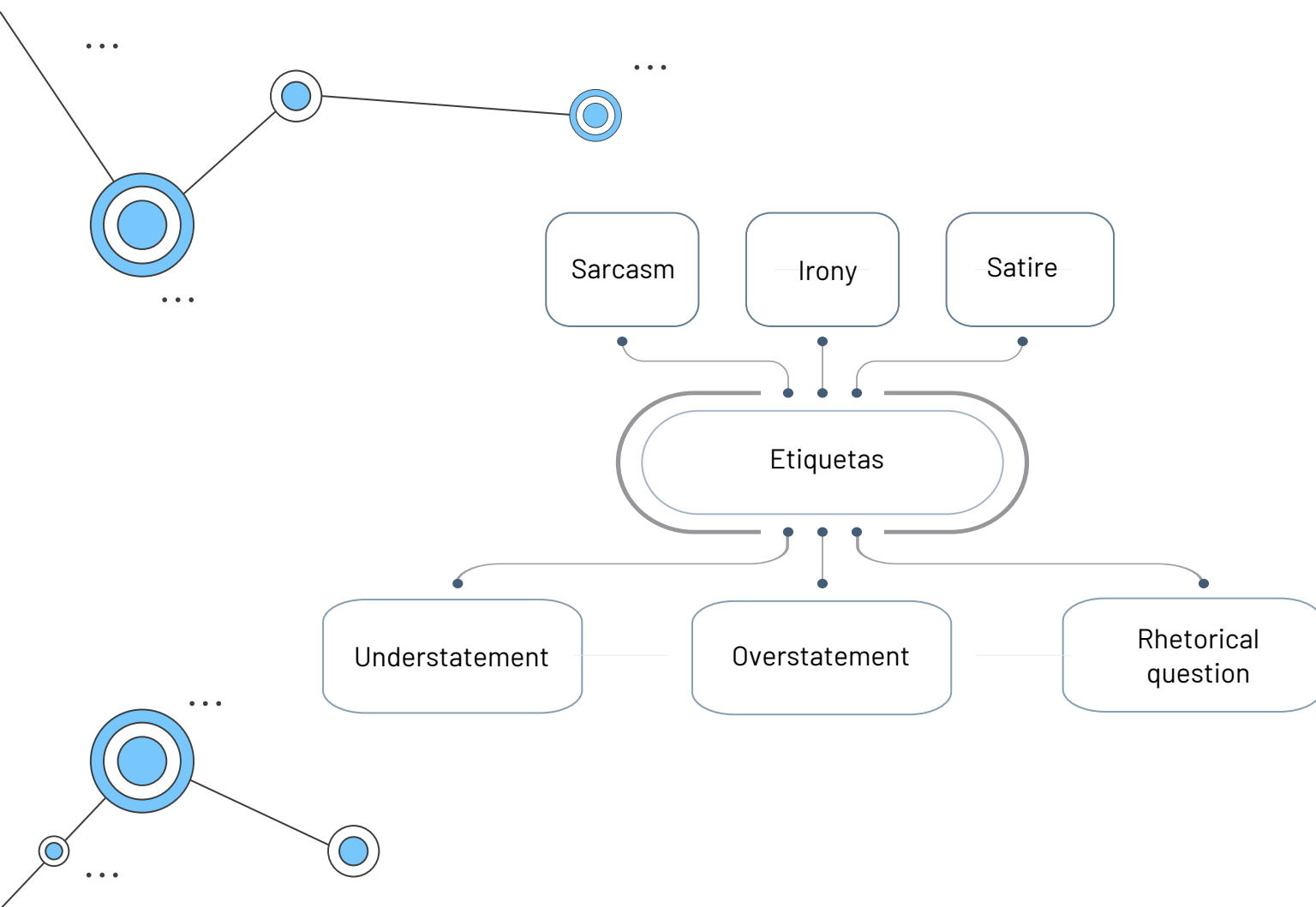
## 2.1 DESCRIPCIÓN DE LA TAREA A RESOLVER



Subtarea B de la Tarea 6 de SemEval2022

Clasificación binaria multi-etiqueta

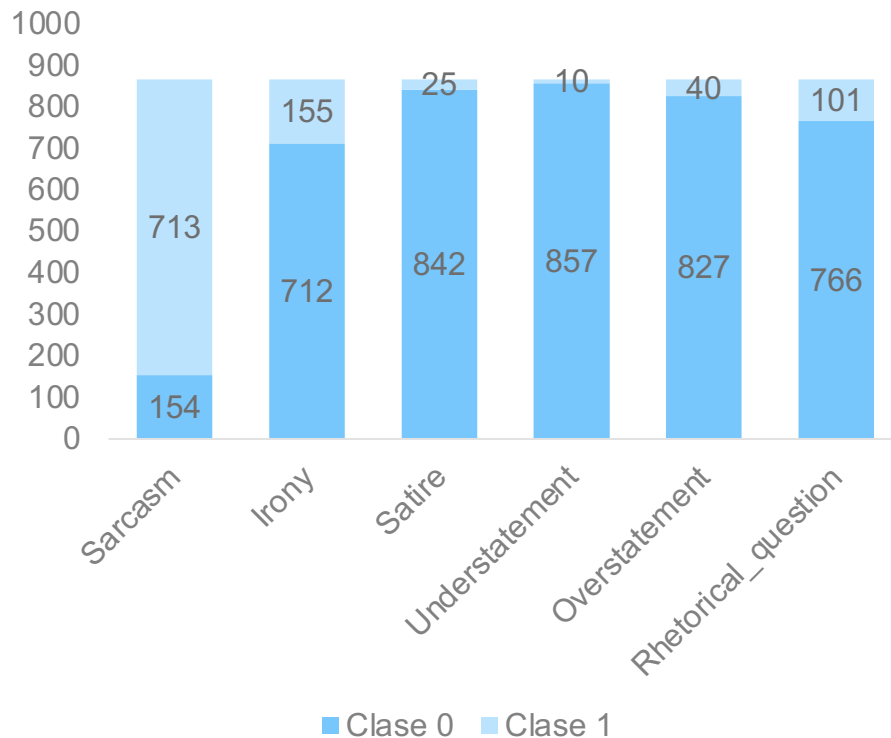
Dificultad de los sistemas para interpretar  
lenguaje figurativo





## 2.2 Análisis del conjunto de datos

Total → 867 Tweets



## 2.2 Análisis del conjunto de datos

Tweet	Sarcasm	Irony	Satire	Under-statement	Over-statement	Rhetorical_question
Today my pop-pop told me I was not “forced” to go to college 😞 okay sure sureeee	1	0	0	0	0	0
See Brexit is going well	0	1	0	0	0	0
Beautician and the Beast (1997). Best Timothy Dalton film of all time.	1	0	1	0	0	0
Billy Gunn sorta relevant for the first time in years...	1	0	0	1	0	0
90% of adulthood is just refilling your Brita pitcher.	0	1	0	0	1	0
do i just blast maneskin to get hyped for my osce or??	1	0	0	0	0	1

## 2.3 Soluciones propuestas

### Métricas utilizadas

- **Matriz de confusión**

		Prediction	
		Positives	Negatives
Real Values	Positives	True Positives (TP)	False Negatives (FN)
	Negatives	False Positives (FP)	True Negatives (TN)

- **Precision**

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

- **Recall**

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

- **F1-score**

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

## 2.3 Soluciones propuestas

### Balanceo de datos

#### Data Augmentation

- Aumentador de sinónimos
- Corpus WordNet
- Librería *nlpaug*
- Clases minoritarias x2

#### Base de datos externa

- *iSarcasm2020*
- *tweet\_id*, *sarcasm\_label* y *sarcasm\_type*
- [https://twitter.com/user/status/{tweet\\_ID}](https://twitter.com/user/status/{tweet_ID})
- De 867 a 905 tweets

## 2.3 Soluciones propuestas



### Balanceo de datos

	<i>Dataset original</i>	<i>Dataset original</i> + <i>BD externa</i>	<i>Dataset original</i> + <i>Aumento</i>	<i>Dataset original</i> + <i>BD externa</i> + <i>Aumento</i>
<b>Sarcasm</b>	Clase 0: 154 Clase 1: 713	Clase 0: 192 Clase 1: 713	Clase 0: 308 Clase 1: 713	Clase 0: 384 Clase 1: 713
<b>Irony</b>	Clase 0: 712 Clase 1: 155	Clase 0: 737 Clase 1: 168	Clase 0: 712 Clase 1: 310	Clase 0: 737 Clase 1: 336
<b>Satire</b>	Clase 0: 842 Clase 1: 25	Clase 0: 871 Clase 1: 34	Clase 0: 842 Clase 1: 50	Clase 0: 871 Clase 1: 68
<b>Understatement</b>	Clase 0: 857 Clase 1: 10	Clase 0: 889 Clase 1: 16	Clase 0: 857 Clase 1: 20	Clase 0: 889 Clase 1: 32
<b>Overstatement</b>	Clase 0: 827 Clase 1: 40	Clase 0: 860 Clase 1: 45	Clase 0: 827 Clase 1: 80	Clase 0: 860 Clase 1: 90
<b>Rhetorical_question</b>	Clase 0: 766 Clase 1: 101	Clase 0: 799 Clase 1: 106	Clase 0: 766 Clase 1: 202	Clase 0: 799 Clase 1: 212

## 2.3 Soluciones propuestas

### Preprocesamiento del texto

V1.0

- Caracteres a minúsculas
- Extensión de las contracciones (what's → what is)
- Emojis
- Caracteres especiales
- Espacios múltiples entre caracteres

V2.0

- Funcionalidades v1.0
- Emojis hechos con teclado
- Menciones
- Enlaces

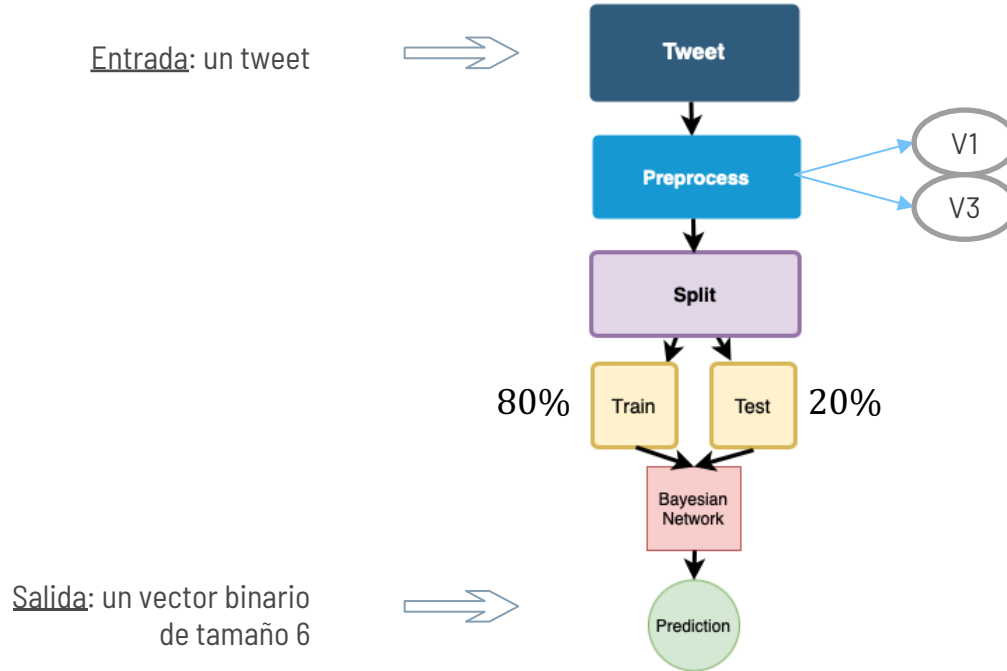
V3.0

- Funcionalidades v2.0
- Stopwords

## 2.3 Soluciones propuestas



### Entrenamiento Modelo Redes Bayesianas



## 2.3 Soluciones propuestas

### Entrenamiento Modelo Redes Bayesianas

Métricas	V1.0 Preprocesamiento de texto	V3.0 Preprocesamiento de texto
Macro F1-score Sarcasm	0.89	0.84
Macro F1-score Irony	0.26	0.20
Macro F1-score Satire	0.38	0.51
Macro F1-score Understatement	0.17	0.48
Macro F1-score Overstatement	0.47	0.47
Macro F1-score Rhetorical Question	0.12	0.20
Media Macro F1-score	0.38	0.45

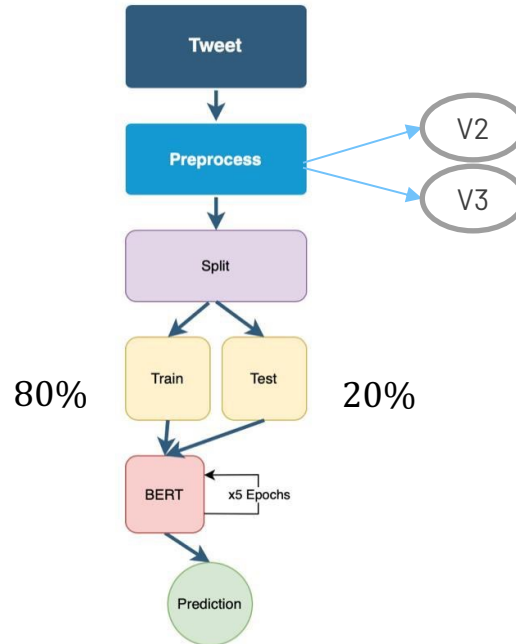


## 2.3 Soluciones propuestas



### Entrenamiento Modelo BERT

- Bert-base-uncased
- Estrategia de relevancia binaria
- 32 instancias
- 5 epochs



- V1.0. Baseline
- V2.0. Versión anterior + aumento clases minoritarias
- V3.0. Versión anterior + BD externa
- V4.0. Versión anterior + v2.0 preprocesamiento de texto
- V5.0. Versión anterior + v3.0 preprocesamiento de texto

## 2.3 Soluciones propuestas



### Entrenamiento Modelo BERT

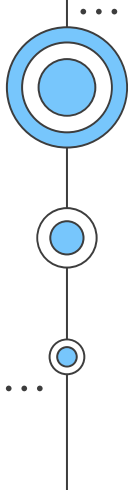
Métricas	V1.0	V2.0	V3.0	V4.0	V5.0
Macro F1-score Sarcasm	0.45	0.78	0.81	0.80	0.70
Macro F1-score Irony	0.45	0.72	0.77	0.77	0.64
Macro F1-score Satire	0.49	0.63	0.66	0.64	0.57
Macro F1-score Understatement	0.50	0.60	0.63	0.62	0.52
Macro F1-score Overstatement	0.49	0.65	0.69	0.71	0.60
Macro F1-score Rhetorical Question	0.84	0.90	0.91	0.89	0.87
Media Macro F1-score	0.54	0.71	<b>0.75</b>	0.74	0.65

## 2.4 Resultados de la etapa de evaluación

Modelo BERT-base-uncased  
V3.0

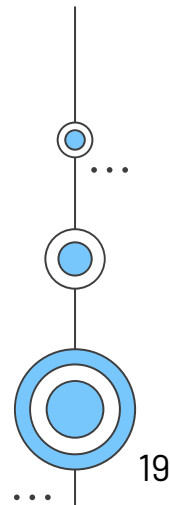
10ª Posición							2ª Posición	
Macro F1-score	F1 Sarcasm	F1 Irony	F1 Satire	F1 Under statement	F1 Over statement	F1 Rhetorical question		
0.0699	0.2430	0.0485	0.0000	0.0000	0.0000	0.1280		
		3ª Posición						

22 Participantes en total

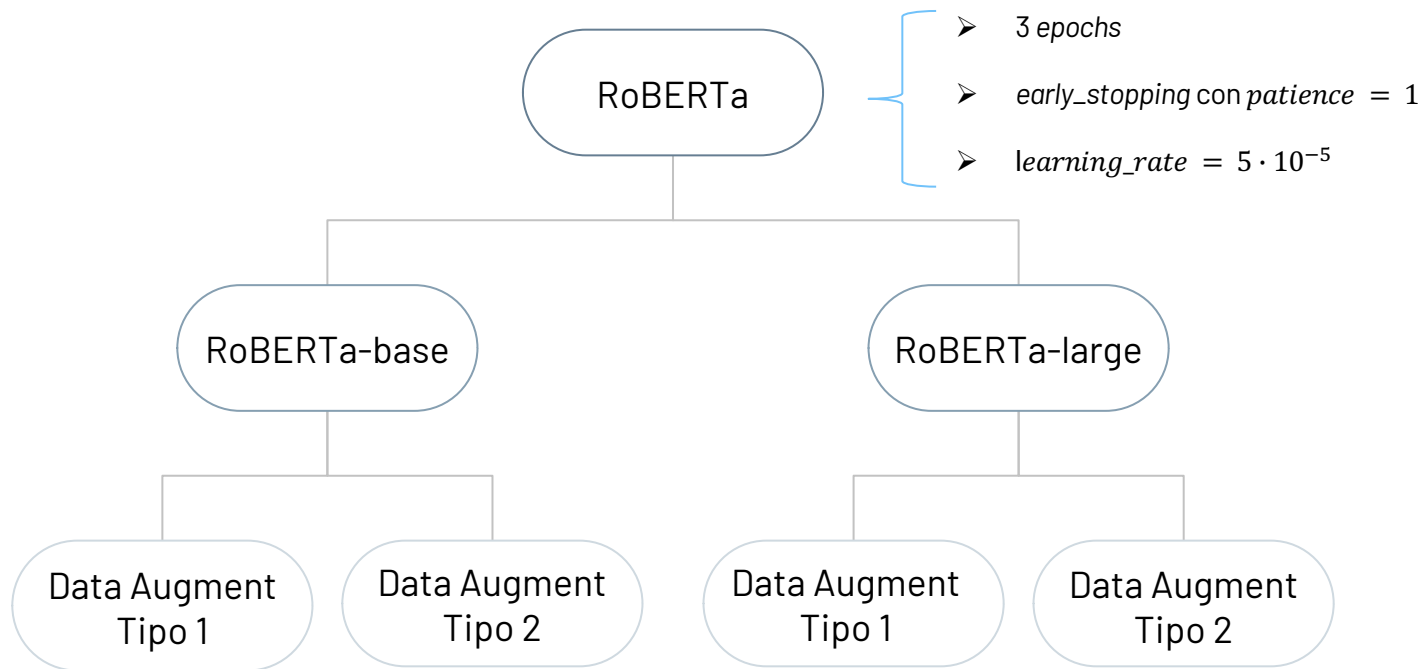


# 3

## Propuestas de mejora



### 3.1 Descripción del marco de experimentación



## 3.2 Modelo RoBERTa-base

Métricas	Competición	Baseline	AugmentTipo1	AugmentTipo1 + BD externa	AugmentTipo1 + BD externa + Preprocesamiento
F1 Sarcasm	0.2430	0.2278	0.2464	0.2306	0.2403
F1 Irony	0.0485	0.0000	0.0526	0.0737	0.0000
F1 Satire	0.0000	0.0000	0.0000	0.0930	0.0000
F1 Understatement	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Overstatement	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Rhetorical question	0.1280	0.0952	0.1224	0.0808	0.0727
Macro F1-score	0.0699	0.0538	0.0702	<b>0.0797</b>	0.0522

## 3.2 Modelo RoBERTa-base

Métricas	Competición	Baseline	AugmentTipo2	AugmentTipo2 + BD externa	AugmentTipo2 + BD externa + Preprocesamiento
F1 Sarcasm	0.2430	0.2278	0.2377	0.2434	0.2431
F1 Irony	0.0485	0.0000	0.0000	0.0000	0.0246
F1 Satire	0.0000	0.0000	0.0000	0.0628	0.0566
F1 Understatement	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Overstatement	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Rhetorical question	0.1280	0.0952	0.1238	0.1085	0.0681
Macro F1-score	0.0699	0.0538	0.0603	<b>0.0691</b>	0.0654

### 3.3 Modelo RoBERTa-large

Métricas	Competición	Baseline	AugmentTipo1	AugmentTipo1 + BD externa	AugmentTipo1 + BD externa + Preprocesamiento
F1 Sarcasm	0.2430	0.2278	0.2278	0.2278	0.2278
F1 Irony	0.0485	0.0000	0.0000	0.0000	0.0000
F1 Satire	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Understatement	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Overstatement	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Rhetorical question	0.1280	0.0000	0.0000	0.0000	0.0000
Macro F1-score	0.0699	<b>0.0380</b>	<b>0.0380</b>	<b>0.0380</b>	<b>0.0380</b>



### 3.3 Modelo RoBERTa-large

Métricas	Competición	Baseline	AugmentTipo2	AugmentTipo2 + BD externa	AugmentTipo2 + BD externa + Preprocesamiento
F1 Sarcasm	0.2430	0.2278	0.2278	0.2278	0.2278
F1 Irony	0.0485	0.0000	0.0000	0.0000	0.0000
F1 Satire	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Understatement	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Overstatement	0.0000	0.0000	0.0000	0.0000	0.0000
F1 Rhetorical question	0.1280	0.0000	0.0000	0.0000	0.0000
Macro F1-score	0.0699	<b>0.0380</b>	<b>0.0380</b>	<b>0.0380</b>	<b>0.0380</b>

## 3.4 Mejores métricas

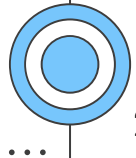
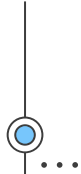
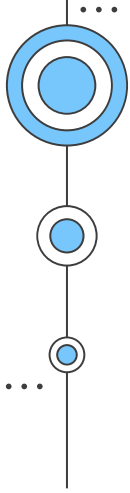
	<i>Macro F1-score</i>	<i>F1 Sarcasm</i>	<i>F1 Irony</i>	<i>F1 Satire</i>	<i>F1 Understatement</i>	<i>F1 Overstatement</i>	<i>F1 Rhetorical question</i>
<b>Competición</b>	0.0699	0.2430	0.0485	0.0000	0.0000	0.0000	0.1280
<b>Mejoras</b>	0.0902 (+29.04%)	0.2464 (+1.40%)	0.0737 (+51.96%)	0.0930 (+∞)	0.0000 (+0.00%)	0.0000 (+0.00%)	0.1280 (+0.00%)

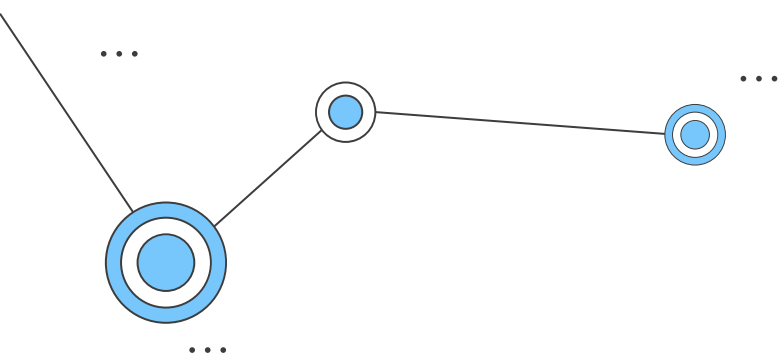
$10^a \rightarrow 2^a$        $11^a \rightarrow 8^a$        $11^a \rightarrow 3^a$

<b>Etiqueta</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>Matriz de confusión</b>
<b>Sarcasm</b>	0.1438	0.8611	0.2464	$\begin{bmatrix} 297 & 923 \\ 25 & 155 \end{bmatrix}$
<b>Irony</b>	0.0406	0.4000	0.0737	$\begin{bmatrix} 1191 & 189 \\ 12 & 8 \end{bmatrix}$
<b>Satire</b>	0.1081	0.0816	0.0930	$\begin{bmatrix} 1318 & 33 \\ 45 & 4 \end{bmatrix}$

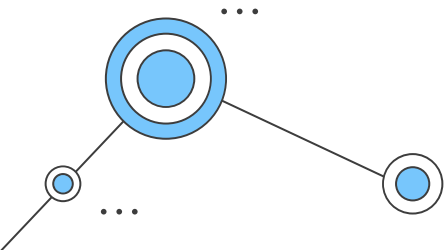
# 4

## Conclusiones y trabajo futuro





# 4.1 Conclusiones



- Paper científico
- Importancia del balanceo de datos
- Uso de una BD externa
- Eficacia de los Transformers

	Competición	Mejoras
Macro F1-score	0.0699	0.0902
Mejora del 29.04%		

2<sup>a</sup>  
Posición

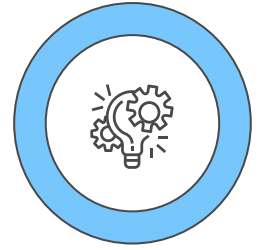


## 4.2 Trabajo futuro

Incrementar uso BD externa

UnderBagging

Modelos pre-entrenados con contenido sarcástico





# Gracias por vuestra atención

Dudas y preguntas

Adrián Moreno Monterde  
[adrian.moreno521@alu.uhu.es](mailto:adrian.moreno521@alu.uhu.es)

