

Implementación y evaluación de aplicaciones usando BERT,  
para los problemas de análisis de sentimiento y respuesta a preguntas.



**Universidad**  
Internacional  
de Valencia

Por el alumno  
**Luis Arturo Izaguirre Viera**

Dirigido por **Andrés Díaz Pinto**

De:  
 Planeta Formación y Universidades

Máster Universitario  
en Inteligencia Artificial

*Curso académico 2021-22*

*Matrícula abril 2021*

*Tercera convocatoria: mayo 2022*

# Índice

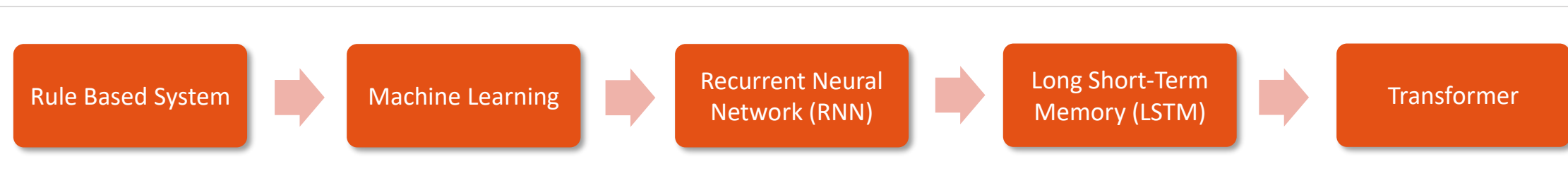
- ▶ Introducción y marco teórico.
- ▶ Objetivo general.
- ▶ Metodología.
- ▶ Experimentos: análisis de sentimiento y respuesta a preguntas.
  - ▶ Descripción del problema.
  - ▶ Descripción del conjunto de datos.
  - ▶ Estado del arte.
  - ▶ Solución propuesta.
  - ▶ Resultados del experimento.
- ▶ Conclusiones.
- ▶ Limitaciones.

# Procesamiento del Lenguaje Natural

- Habilita a las computadoras a comprender, entender y procesar el lenguaje humano.
- Se ha convertido en una de las tareas más complejas a resolver.

# Procesamiento del Lenguaje Natural

- Habilita a las computadoras a comprender, entender y procesar el lenguaje humano.
- Se ha convertido en una de las tareas más complejas a resolver.



# Procesamiento del Lenguaje Natural

- Habilita a las computadoras a comprender, entender y procesar el lenguaje humano.
- Se ha convertido en una de las tareas más complejas a resolver.

Rule Based System



Machine Learning



Recurrent Neural  
Network (RNN)



Long Short-Term  
Memory (LSTM)

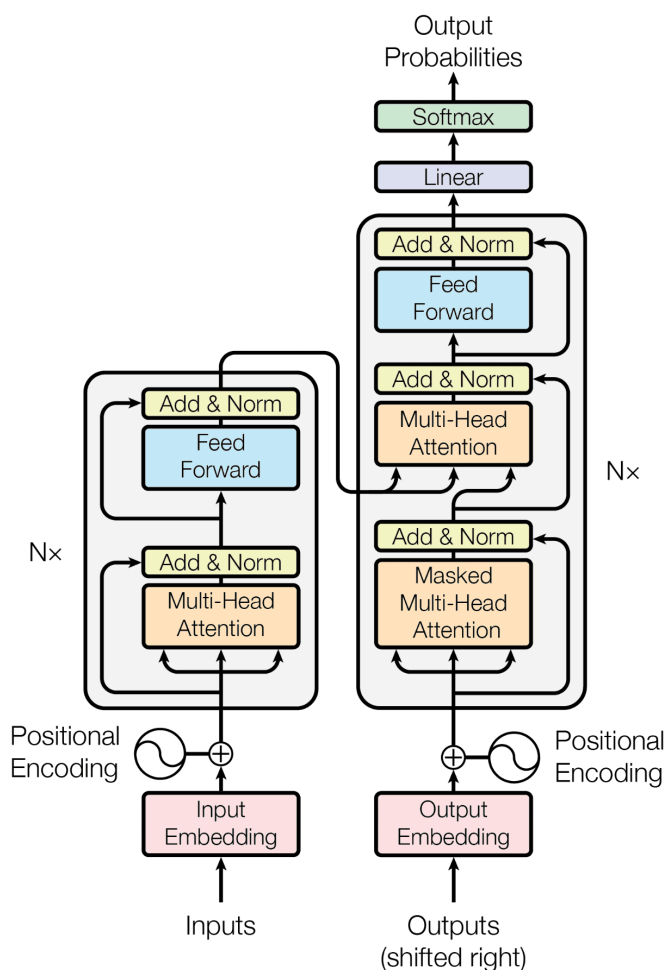


Transformer

**Attention is all you need!**



# Arquitectura Transformer

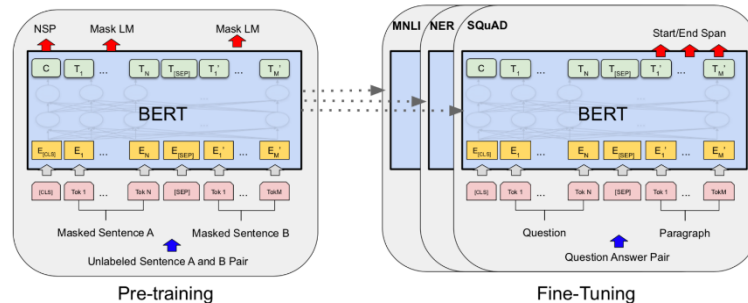
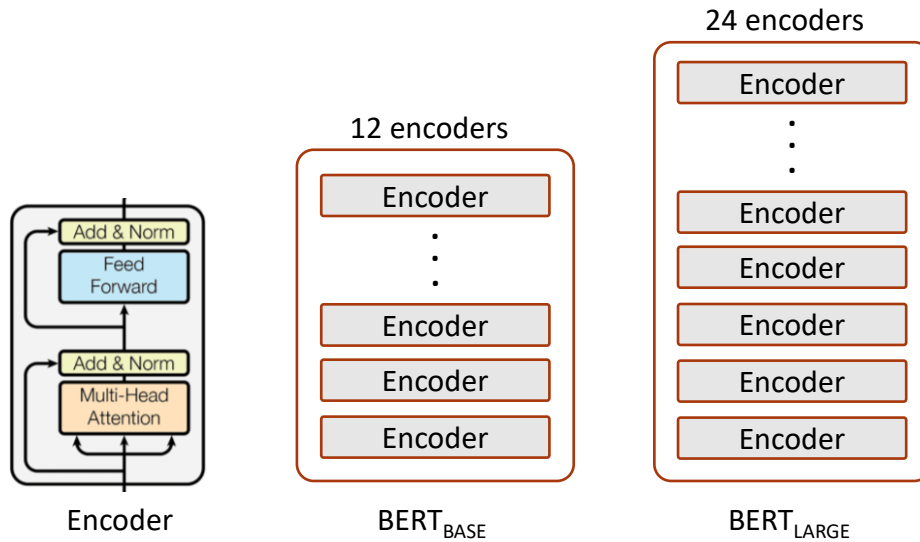


## Attention is all you need!

- En un paper publicado por Google a finales del año 2017.
- Se presentó la arquitectura del Transformer que sustituye las capas recurrentes por capas de atención.
- Capas de atención codifican cada palabra de una frase en función del resto de la secuencia.
- Transformer, a diferencia de las RNNs, procesa todos los elementos simultáneamente.
- Transformer gestiona completamente las dependencias entre la entrada y la salida con atención y recurrencia.

## BERT

### Bidirectional Encoder Representations from Transformer



- Modelo de lenguaje presentado en un paper publicado por investigadores de Google en octubre 2018.
- Ha causado revuelo en la comunidad de ML al presentar resultados que definen el estado del arte en una amplia variedad de tareas de NLP.
- Usa solo la capa Encoder de Transformer
- Preentrenamiento de BERT en base a 2 tareas no supervisadas:
  - Masked LM
  - Next Sentence Prediction
- Se basa en la arquitectura unificada del modelo, que permite a partir de la adecuación de las entradas y salidas realizar un proceso de finetuning para abordar distintos problemas.

# Objetivo

## Objetivo General

Evaluar el uso y desempeño de BERT (Bidirectional Encoder Representations for Transformers) como modelo en la resolución de tareas relacionadas al procesamiento del lenguaje natural, distintas a las que para en un principio fue preentrenado.



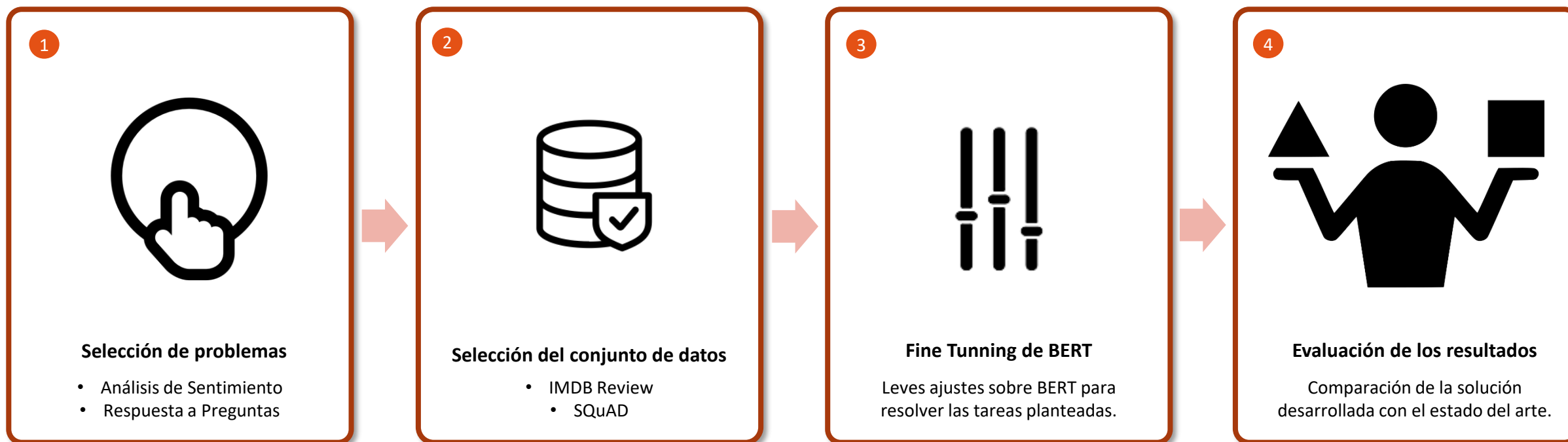
Análisis de sentimiento



Respuesta a preguntas



## Metodología



# Análisis de sentimiento, descripción del problema

El objetivo principal del problema es analizar una secuencia de texto y a partir de ella identificar o extraer información subjetiva que permita inferir si el sentimiento asociado a la frase es positivo o negativo.



## 4.

# Análisis de sentimiento conjunto de datos

## Large Movie Review Dataset IMDB

### Características

- Conjunto de datos con 50.000 reseñas de películas.
- Clasificación binaria de sentimientos.
- Completamente balanceado.
- Solo críticas altamente polarizadas,  $\text{rating} \leq 4$  ó  $\geq 7$

### IMDB Dataset of 50K Movie Reviews

[Data](#) [Code \(447\)](#) [Discussion \(6\)](#) [Metadata](#)

#### IMDB Dataset.csv (66.21 MB)

[Detail](#) [Compact](#) [Column](#)

<a href="#">review</a>	<a href="#">sentiment</a>
49582 unique values	2 unique values
Yup, that's right folks, this is undoubtedly the worst show in the history of television. If you wan...	negative
Yumiko (Wakana Sakai), the pretty adopted daughter of a recently-deceased Japanese businessman, inhe...	negative
Yul Brynner was a symbol of villein in the tine of 50,s , he play a role of Russian leader in Hungar...	positive
Yul Brynner is Major Surov, a singing, dancing, vodka-drinking Russian Officer stationed near the Au...	positive
Yuck. I thought it odd that their ancient book on curses was made using a common script font instead...	negative

# Análisis de sentimiento, estado del arte

## Large Movie Review Dataset IMDB

<b>Año</b>	2022
<b>Fuente</b>	NLP Progress <a href="https://nlpprogress.com/english/sentiment_analysis.html">https://nlpprogress.com/english/sentiment_analysis.html</a>
<b>Autor</b>	<ul style="list-style-type: none"> <li>Sebastian Ruder</li> </ul> Ruder, S. (2022). NLP-progress (Version 1.0.0) [Computer software]. <a href="https://doi.org/10.5281/zenodo.1234">https://doi.org/10.5281/zenodo.1234</a>
<b>Estudio</b>	Repositorio en línea que hace un seguimiento del progreso en distintos problemas en el procesamiento del lenguaje natural NLP, incluidos los datasets y el estado del arte actual para las tareas de NLP más comunes.
<b>Resultados</b>	<ul style="list-style-type: none"> <li>En este estudio dos de los primeros tres modelos están basados en una implementación de BERT</li> <li>El modelo que presenta el mejor performance, es un método de preentrenamiento autorregresivo generalizado similar a BERT que permite aprender contextos bidireccionales.</li> </ul>

Model	Accuracy	Paper / Source
XLNet (Yang et al., 2019)	96.21	<a href="#">XLNet: Generalized Autoregressive Pretraining for Language Understanding</a>
BERT_large+ITPT (Sun et al., 2019)	95.79	<a href="#">How to Fine-Tune BERT for Text Classification?</a>
BERT_base+ITPT (Sun et al., 2019)	95.63	<a href="#">How to Fine-Tune BERT for Text Classification?</a>
ULMFIT (Howard and Ruder, 2018)	95.4	<a href="#">Universal Language Model Fine-tuning for Text Classification</a>
Block-sparse LSTM (Gray et al., 2017)	94.99	<a href="#">GPU Kernels for Block-Sparse Weights</a>
oh-LSTM (Johnson and Zhang, 2016)	94.1	<a href="#">Supervised and Semi-Supervised Text Categorization using LSTM for Region Embeddings</a>
Virtual adversarial training (Miyato et al., 2016)	94.1	<a href="#">Adversarial Training Methods for Semi-Supervised Text Classification</a>
BCN+Char+CoVe (McCann et al., 2017)	91.8	<a href="#">Learned in Translation: Contextualized Word Vectors</a>

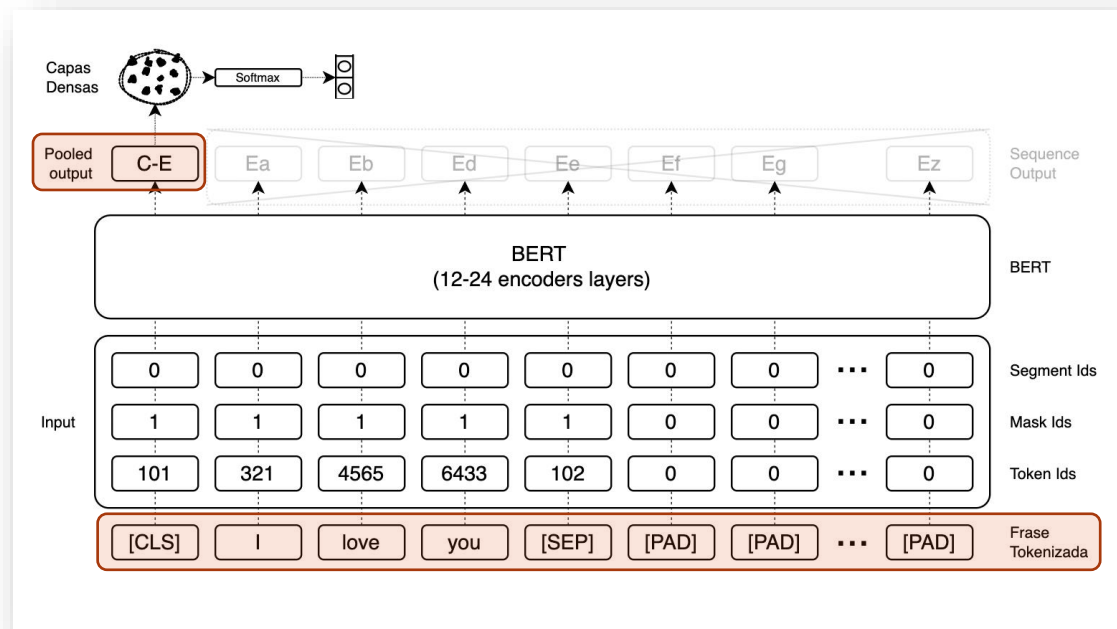
# Análisis de sentimiento, estado del arte

## Large Movie Review Dataset IMDB

Año	2022
Fuente	<a href="https://paperswithcode.com/">https://paperswithcode.com/</a>
Autor	Papers with Code
Estudio	Es un recurso de referencia para papers científicos que representan el más actual estado del arte en distintas casuísticas del machine learning.
Resultados	Es interesante observar que entre las 10 mejores soluciones encontramos 7 trabajos basados en Transformers y la mitad de ellos son modelos basados en BERT.

Rank	Model	Accuracy↑	Extra Training Data	Paper	Code	Result	Year	Tags
1	NB-weighted-BON + dv-cosine	97.4	✓	Sentiment Classification Using Document Embeddings Trained with Cosine Similarity	<a href="#">GitHub</a>	<a href="#">Result</a>	2019	
2	XLNet	96.21	✓	XLNet: Generalized Autoregressive Pretraining for Language Understanding	<a href="#">GitHub</a>	<a href="#">Result</a>	2019	<a href="#">Transformer</a>
3	EFL	96.1	×	Entailment as Few-Shot Learner	<a href="#">GitHub</a>	<a href="#">Result</a>	2021	<a href="#">Transformer</a>
4	GraphStar	96.0	✓	Graph Star Net for Generalized Multi-Task Learning	<a href="#">GitHub</a>	<a href="#">Result</a>	2019	
5	BERT large finetune UDA	95.8	✓	Unsupervised Data Augmentation for Consistency Training	<a href="#">GitHub</a>	<a href="#">Result</a>	2019	<a href="#">Transformer</a>
6	BERT_large+ITPT	95.79	✓	How to Fine-Tune BERT for Text Classification?	<a href="#">GitHub</a>	<a href="#">Result</a>	2019	<a href="#">Transformer</a>
7	L MIXED	95.68	✓	Revisiting LSTM Networks for Semi-Supervised Text Classification via Mixed Objective Function	<a href="#">GitHub</a>	<a href="#">Result</a>	2020	<a href="#">LSTM</a>
8	BERT_base+ITPT	95.63	✓	How to Fine-Tune BERT for Text Classification?	<a href="#">GitHub</a>	<a href="#">Result</a>	2019	<a href="#">Transformer</a>
9	BERT large	95.49	✓	Unsupervised Data Augmentation for Consistency Training	<a href="#">GitHub</a>	<a href="#">Result</a>	2019	<a href="#">Transformer</a>
10	ULMFiT	95.4	✓	Universal Language Model Fine-tuning for Text Classification	<a href="#">GitHub</a>	<a href="#">Result</a>	2018	<a href="#">LSTM</a>

# Análisis de sentimiento, solución propuesta



“El primer token de cada secuencia es siempre un token de clasificación especial [CLS]. El estado oculto final correspondiente a este token se utiliza como representación de secuencia agregada para tareas de clasificación.” (Devlin et al. (2018))

## Análisis de sentimiento, resultados

En los benchmarks estudiados, en general, los modelos basados en BERT eran capaces de obtener un rendimiento ligeramente superior al 95%

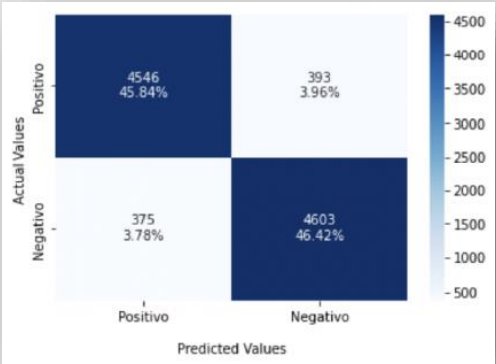
Durante el entrenamiento:

```
Epoch 1/3
4959/4959 [=====] - 3387s 680ms/step - loss: 0.4759 - accuracy: 0.7706 - val_loss: 0.3462 - val_accuracy: 0.8656
Epoch 2/3
4959/4959 [=====] - 3369s 679ms/step - loss: 0.3328 - accuracy: 0.8928 - val_loss: 0.3936 - val_accuracy: 0.8878
Epoch 3/3
4959/4959 [=====] - 3369s 679ms/step - loss: 0.2565 - accuracy: 0.9345 - val_loss: 0.4660 - val_accuracy: 0.8925
```

Durante la etapa de evaluación:

```
Accuracy: 0.9225572347640991
```

	precision	recall	f1-score	support
0	0.92	0.92	0.92	4939
1	0.92	0.92	0.92	4978
accuracy			0.92	9917
macro avg	0.92	0.92	0.92	9917
weighted avg	0.92	0.92	0.92	9917



# Respuesta a preguntas, descripción del problema

- La tarea de Respuesta a Preguntas o *Question Answering* (QA), se basa en intentar encontrar automáticamente la respuesta contextual y semánticamente correcta para una pregunta proporcionada en un texto





# Respuesta a preguntas conjunto de datos

## *The Stanford Question Answering Dataset (SQuAD)*

### Características

- Publicado en el año 2016 por investigadores de la Universidad de Stanford.
- Este conjunto de datos contiene un total de 107.785 preguntas extraídas de un conjunto de 536 artículos seleccionados de la Wikipedia, en un amplio rango de tópicos.
- Existe una segunda versión de SQuAD (v2.0) publicada por los mismos investigadores en el año 2018. Esta versión se diferencia de la primera en que en algunas oportunidades la respuesta no se encuentra en el texto original y está enfocada a ser utilizada en soluciones donde el modelo debe fabricar la respuesta de algún modo.
- Por simplicidad en este trabajo se utilizó la versión 1.1 de este conjunto de datos.

Tesla was the fourth of five children. He had an older brother named Dane and three sisters, Milka, Angelina and Marica. Dane was killed in a horse-riding accident when Nikola was five. In 1861, Tesla attended the "Lower" or "Primary" School in Smiljan where he studied German, arithmetic, and religion. In 1862, the Tesla family moved to Gospić, Austrian Empire, where Tesla's father worked as a pastor. Nikola completed "Lower" or "Primary" School, followed by the "Lower Real Gymnasium" or "Normal School."

**What language did Tesla study while in school?**

Ground Truth Answers: German German German

Prediction: German

**In what year did Tesla's family move to Gospić?**

Ground Truth Answers: 1862 1862 1862

Prediction: 1862

**What was Tesla's brother's name?**

Ground Truth Answers: Dane Dane Dane

Prediction: Dane

**What were Tesla's sisters' names?**

Ground Truth Answers: Milka, Angelina and Marica Milka, Angelina and Marica Milka, Angelina and Marica

Prediction: Milka, Angelina and Marica

**What happened to Dane?**

Ground Truth Answers: killed in a horse-riding accident killed in a horse-riding accident killed in a horse-riding accident

Prediction: killed in a horse-riding accident

# Respuesta a preguntas, estado del arte

Hay dos métricas dominantes utilizadas por muchos conjuntos de datos de respuesta a preguntas, incluido SQuAD:

- Coincidencia exacta o Exact match (EM).
- La puntuación F1 o F1 score.

## Coincidencia exacta o Exact Match

- Texto + Pregunta → Predicción
- Si, Predicción == Respuesta → EM=1
- Sino, EM=0

Esta es una métrica estricta de todo o nada, estar equivocado por un solo carácter da como resultado una puntuación de 0.

## La puntuación F1 o F1 score

- Se calcula sobre las palabras individuales de la predicción frente a las de la respuesta verdadera

# Respuesta a preguntas, estado del arte

## The Stanford Question Answering Dataset (SQuAD)

Año	2019
Fuente	"A BERT Baseline for the Natural Questions"
Autor	<ul style="list-style-type: none"> <li>Chris Alberti, Google</li> <li>Kenton Lee, Google</li> <li>Michael Collins, Google</li> </ul>
Estudio	En este mismo estudio los autores comparan el desempeño de BERT con otros modelos que resolvían la tarea en ese momento y además lo compararon con el desempeño de la tarea realizada por un humano común y un humano experto.
Resultados	<ul style="list-style-type: none"> <li>los autores concluyen que BERT debe ser la nueva línea base y que constituye un buen punto de partida para los modelos de preguntas y respuestas y datasets con características similares.</li> <li>Los autores concluyen que los resultados obtenidos por los modelos de respuesta a preguntas basados en BERT también se están acercando rápidamente al rendimiento humano informado para estos conjuntos de datos.</li> </ul>

	Long Answer Dev			Long Answer Test			Short Answer Dev			Short Answer Test		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
DocumentQA	47.5	44.7	46.1	48.9	43.3	45.7	38.6	33.2	35.7	40.6	31.0	35.1
DecAtt + DocReader	52.7	57.0	54.8	54.3	55.7	55.0	34.3	28.9	31.4	31.9	31.1	31.5
<b>BERT<sub>joint</sub> (this work)</b>	<b>61.3</b>	<b>68.4</b>	<b>64.7</b>	<b>64.1</b>	<b>68.3</b>	<b>66.2</b>	<b>59.5</b>	<b>47.3</b>	<b>52.7</b>	<b>63.8</b>	<b>44.0</b>	<b>52.1</b>
Single Human	80.4	67.6	73.4	-	-	-	63.4	52.6	57.5	-	-	-
Super-annotator	90.0	84.6	87.2	-	-	-	79.1	72.6	75.7	-	-	-

# Respuesta a preguntas, estado del arte

## The Stanford Question Answering Dataset (SQuAD)

Año	2019
Fuente	“ALBERT: A Lite BERT for Self-supervised Learning of Language Representations”
Autor	<ul style="list-style-type: none"> <li>Zhenzhong Lan, Google</li> <li>Sebastian Goodman, Google</li> <li>Piyush Sharma, Google</li> <li>Radu Soricut, Google</li> <li>Mingda Chen, Toyota Technological Institute at Chicago</li> <li>Kevin Gimpel, Toyota Technological Institute at Chicago</li> </ul>
Estudio	El paper presenta ALBERT es una versión de BERT mucho más ligera con 12 millones de parámetros en vez de 110 millones como los que contiene BERT.
Resultados	En el estudio comparan el desempeño de ALBERT con BERT en las tareas de SQuAD obteniendo resultados bastante similares. ALBERT es una versión de BERT mucho más ligera con 12 millones de parámetros en vez de 110 millones como los que contiene BERT.

	Model	Parameters	SQuAD1.1	SQuAD2.0	MNLI	SST-2	RACE	Avg	Speedup
BERT	base	108M	90.4/83.2	80.4/77.6	84.5	92.8	68.2	82.3	4.7x
	large	334M	92.2/85.5	85.0/82.2	86.6	93.0	73.9	85.2	1.0
ALBERT	base	12M	89.3/82.3	80.0/77.1	81.6	90.3	64.0	80.1	5.6x
	large	18M	90.6/83.9	82.3/79.4	83.5	91.7	68.5	82.4	1.7x
	xlarge	60M	92.5/86.1	86.1/83.1	86.4	92.4	74.8	85.5	0.6x
	xxlarge	235M	<b>94.1/88.3</b>	<b>88.1/85.1</b>	<b>88.0</b>	<b>95.2</b>	<b>82.3</b>	<b>88.7</b>	0.3x

F1/EM para SQuAD

# Respuesta a preguntas, estado del arte

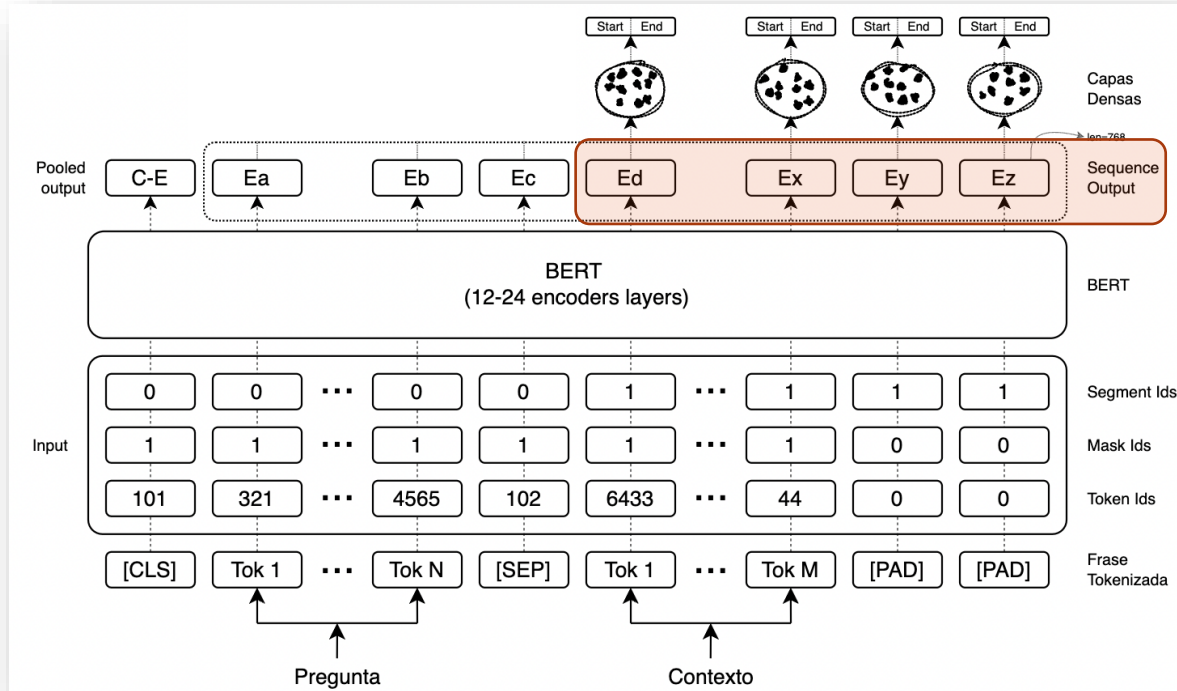
## The Stanford Question Answering Dataset (SQuAD)

<b>Año</b>	2022
<b>Fuente</b>	Ranking de la página oficial de SQuAD
<b>Autor</b>	
<b>Estudio</b>	Ranking en línea que mantiene el progreso en soluciones al dataset SQuAD v 1.1.
<b>Resultados</b>	Se pueden observar en las primeras posiciones observar muchos modelos basados en ALBERT.

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Jun 04, 2021	IE-Net (ensemble) RICOH_SRCB_DML	90.939	93.214
2 Feb 21, 2021	FPNet (ensemble) Ant Service Intelligence Team	90.871	93.183
3 May 16, 2021	IE-NetV2 (ensemble) RICOH_SRCB_DML	90.860	93.100
4 Apr 06, 2020	SA-Net on Albert (ensemble) QIANXIN	90.724	93.011
5 May 05, 2020	SA-Net-V2 (ensemble) QIANXIN	90.679	92.948
5 Apr 05, 2020	Retro-Reader (ensemble) Shanghai Jiao Tong University <a href="http://arxiv.org/abs/2001.09694">http://arxiv.org/abs/2001.09694</a>	90.578	92.978
5 Feb 05, 2021	FPNet (ensemble) YuYang	90.600	92.899
6 Apr 18, 2021	TransNets + SFVerifier + SFEnsembler (ensemble) Senseforth AI Research <a href="https://www.senseforth.ai/">https://www.senseforth.ai/</a>	90.487	92.894
6 Dec 01, 2020	EntitySpanFocusV2 (ensemble) RICOH_SRCB_DML	90.521	92.824
6 Jul 31, 2020	ATRLP+PV (ensemble) Hithink RoyalFlush	90.442	92.877
7 Mar 12, 2020	ALBERT + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	90.386	92.777
8 Feb 05, 2021	MixEnsemble (ensemble) Anonymous	90.194	92.594
9 Jan 10, 2020	Retro-Reader on ALBERT (ensemble) Shanghai Jiao Tong University <a href="http://arxiv.org/abs/2001.09694">http://arxiv.org/abs/2001.09694</a>	90.115	92.580
10 Jan 12, 2021	Answer Dependent Classify (single model) YITU	90.059	92.517

11 Nov 06, 2019	ALBERT + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	90.002	92.425
12 Apr 04, 2022	LANet (ensemble) 2digit-david	89.923	92.425
13 Sep 18, 2019	ALBERT (ensemble model) Google Research & TTIC <a href="https://arxiv.org/abs/1909.11942">https://arxiv.org/abs/1909.11942</a>	89.731	92.215
13 Feb 25, 2020	Albert_Verifier_AA_Net (ensemble) QIANXIN	89.743	92.180
13 Jun 27, 2020	ELECTRA+ATRLP+PV (single model) Hithink RoyalFlush	89.551	92.366
13 Jan 10, 2021	Span Extract + Classify (single model) Anonymous	89.562	92.226
13 Mar 28, 2020	Retro-Reader on ELECTRA (single model) Shanghai Jiao Tong University <a href="http://arxiv.org/abs/2001.09694">http://arxiv.org/abs/2001.09694</a>	89.562	92.052
13 Mar 27, 2020	albert+KD+transfer (ensemble) Anonymous	89.461	92.134
14 Nov 18, 2020	ROaD-Electra single model	89.449	92.118
15 Feb 02, 2021	ELECTRA + E-Verifier (ensemble) Midea NLP Team	89.348	91.985
15 Jan 03, 2021	ELECTRA + ROBERTA + ALBERT (ensemble) Midea NLP Team	89.325	91.994
15 Jan 12, 2021	2task (single model) Ted	89.325	91.939
16 Apr 21, 2020	albert+KD+transfer+twopass (single) SPPD	89.111	91.877
16 Apr 18, 2020	ALBERT + MTDA + SFVerifier (ensemble model) Senseforth AI Research <a href="https://www.senseforth.ai/">https://www.senseforth.ai/</a>	89.235	91.739
17 Apr 15, 2020	ALBERT + SFVerifier (ensemble model) Senseforth AI Research <a href="https://www.senseforth.ai/">https://www.senseforth.ai/</a>	89.133	91.666
17 Apr 22, 2020	ELECTRA+RL+EV (single model) Hithink RoyalFlush	89.021	91.765
18 Sep 17, 2021	AE-TEST ensemble	88.998	91.635
18 Dec 08, 2019	ALBERT+Entailment DA (ensemble) CloudWalk	88.761	91.745

# Respuesta a preguntas, solución propuesta



- Se usa el *Sequence Output*, el cual está representado por una lista de vectores (embedding), uno para cada una de las palabras, de esta forma se podrá localizar las más probables de ser la el inicio y el final de la respuesta.
- Se añade una capa densa al final. Estas capas densas se le aplica a cada uno de los elementos, a cada vector y nos retorna dos posibles respuestas. La primera nos representará una puntuación de qué tan probable es que esa palabra sea la posición inicial en la que empieza la respuesta y la segunda neurona representará una puntuación de qué tan probable es que esa palabra sea la que finaliza la respuesta

## Respuesta a preguntas, resultados

En los benchmarks y trabajos revisados los valores de EM y F1 varían considerablemente, sin embargo:

- F1 Score: se mantiene en un rango que va desde ligeramente inferior a 70% hasta ligeramente mayor a 90%.
- EM: Entre 80% y ligeramente superior a 90%

### Experimento 1:

bert\_multi\_cased\_L-12\_H-768\_A-12

- Exact Match (EM) = 29,33 %
- F1 Score = 41 %

### Experimento 2:

bert\_en\_uncased\_L-12\_H-768\_A-12

- Exact Match (EM) = 71,34 %
- F1 Score = 80,58 %

# Conclusiones

- BERT es una línea de base sólida para el desarrollo de soluciones a distintas tareas de procesamiento del lenguaje natural.
- Con configuraciones y ajustes básicos es suficiente para conseguir un rendimiento aceptable.
- La arquitectura unificada de BERT le permite al modelo poder adaptarse fácilmente a la solución de distintos problemas, obteniendo resultados cercanos a los que define el estado del arte.
- Se debe hacer un preprocesamiento del conjunto de datos adaptándolo a lo que el modelo espera recibir como entrada. Es importante considerar el proceso de limpieza de los datos para obtener un buen resultado.
- Es importante seleccionar una versión apropiada de BERT según las capacidades de cómputo y según los datos de entrenamiento con los que se cuente.



# Limitaciones y oportunidades de mejora.

- **Capacidades de cómputo:** La simplicidad de las soluciones propuestas se debe en gran parte a las capacidades de cómputo disponible para la implementación. El entrenamiento y fine-tuning de los modelos propuestos se realizó utilizando la versión gratuita de Google Colab, bajo un entorno donde los recursos no están garantizados y son limitados. Por este motivo se optó por usar la versión BERT<sub>BASE</sub> y no la versión BERT<sub>LARGE</sub> en las soluciones propuestas.
- **Escalar el enfoque:**
  - El enfoque de este trabajo se realizó sobre BERT, pero a la fecha existen **otros modelos de lenguaje preentrenados** basados en la arquitectura *Transformer* que están teniendo mucha relevancia en el campo del procesamiento del lenguaje natural como **XLNet, T5 o GPT-3**.
  - En el mismo sentido, existen distintas **variaciones de BERT** que también están teniendo excelentes resultados en la resolución de problemas específicos. Algunas de estas variaciones prometen mejoras, otras simplifican y hacen el entrenamiento aún más liviano.
  - Adicionalmente, en este trabajo se desarrollan dos aplicaciones que resuelven **dos tareas bien concretas y específicas dentro del NLP, el análisis de sentimiento a través de un problema de clasificación y el problema de preguntas y respuestas**. Se recomienda ampliar el estudio con otras tareas.
- **Mejorar fine-tuning:** Las soluciones planteadas en este trabajo se basan en un fine-tuning muy simple. Es muy probable que se puedan realizar mejoras sustanciales a los resultados obtenidos utilizando técnicas de fine-tuning más complejas.

**¡GRACIAS!**