

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

# Estudio del rendimiento de BERT frente a métodos clásicos de procesamiento de lenguaje natural

**Santiago González- Carvajal Centenera.**

Tutor: Eduardo César Garrido Merchán.

Ponente: Daniel Hernández Lobato.

Universidad Autónoma de Madrid.

June 16, 2020

Estado del arte ooo	Definición del proyecto o o o o	Diseño del proyecto o ooooooooo o ooo	Implementación oo	Experimentos o oo oo oo oo	Conclusiones y trabajo futuro oooooooo
------------------------	---	---	----------------------	---	---

# Contenidos

## Estado del arte

## Definición del proyecto

Objetivos

Hipótesis

Asunciones y restricciones

## Diseño del proyecto

Diseño del modelo

Diseño técnico

Planificación del proyecto

## Implementación

## Experimentos

Primer experimento

Segundo experimento

Tercer experimento

Cuarto experimento

## Conclusiones y trabajo futuro

Estado del arte	Definición del proyecto	Diseño del proyecto	Implementación	Experimentos	Conclusiones y trabajo futuro
●○○	○ ○ ○ ○ ○	○ ○○○○○○○○○ ○ ○○○	○○	○ ○○ ○○ ○○ ○○ ○○	○○○○○○○

# Contenidos

## Estado del arte

### Definición del proyecto

Objetivos

Hipótesis

Asunciones y restricciones

### Diseño del proyecto

Diseño del modelo

Diseño técnico

Planificación del proyecto

### Implementación

### Experimentos

Primer experimento

Segundo experimento

Tercer experimento

Cuarto experimento

### Conclusiones y trabajo futuro

## Motivación

- **NLP** una rama de la **AI** y de la **lingüística** cuyo objetivo es que los computadores entiendan el **lenguaje humano**.
- Aproximaciones **lingüísticas** y aproximaciones basadas en **ML**.
- Problemas como **multilingüismo**, forma de expresarnos **distinta** dependiendo de la situación, etc.

Estado del arte	Definición del proyecto	Diseño del proyecto	Implementación	Experimentos	Conclusiones y trabajo futuro
○○●	○ ○ ○ ○	○ ○○○○○○○○○ ○ ○○○	○○	○ ○○ ○○ ○○ ○○	○○○○○○○

## Estado del arte

- Enfoques **lingüísticos**: VADER, STRING (híbrido), etc.
- Enfoques basados en **ML**: GloVe, word embedding, ELMo, Transformer, LSTM, GPT, ULMFiT, **BERT**, GPT-2, XLNet, ERNIE, T-NLG, etc. Algunos utilizan **TF-IDF** o similares para construir el vocabulario.

Estado del arte  
ooo

Definición del proyecto  
●  
○  
○  
○  
○

Diseño del proyecto  
○  
○○○○○○○○○  
○  
○○○

Implementación  
○○

Experimentos  
○  
○○  
○○  
○○  
○○  
○○

Conclusiones y trabajo futuro  
○○○○○○○

# Contenidos

Estado del arte

Definición del proyecto

Objetivos

Hipótesis

Asunciones y restricciones

Diseño del proyecto

Diseño del modelo

Diseño técnico

Planificación del proyecto

Implementación

Experimentos

Primer experimento

Segundo experimento

Tercer experimento

Cuarto experimento

Conclusiones y trabajo futuro

Estado del arte  
ooo

Definición del proyecto  
o  
●  
o  
o

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Objetivos del proyecto

- O-1 Comparar el rendimiento de **BERT** con algoritmos clásicos en **Inglés**.
- O-2 Comparar el rendimiento de **BERT** con algoritmos clásicos en **otros idiomas**.
- O-3 Medir el rendimiento de **BERT** en alguna clase de **competición**.
- O-4 Demostrar la efectividad de **BERT** sobre **datasets pequeños**.
- O-5 **Optimización Bayesiana** de algunos parámetros de **BERT**.

Estado del arte  
ooo

Definición del proyecto  
o  
o  
●  
o

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Hipótesis del proyecto

- H-1 BERT tiene un **mejor rendimiento** que algoritmos clásicos.
- H-2 BERT es un modelo capaz de obtener buenos resultados en **cualquier idioma**.
- H-3 BERT es **fácil de implementar** comparado con los métodos tradicionales.
- H-4 BERT es capaz de conseguir buenos resultados incluso sobre **datasets pequeños**.



Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
●

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Asunciones y restricciones

- A-1 El modelo **BERT** utilizado ha sido **pre-entrenado correctamente**.
- A-2 El modelo **BERT** utilizado está implementado siguiendo la lógica explicada en el **artículo original**.
- A-3 Los paquetes de **AutoML** son fiables.
- R-1 Equipo disponible: **ordenador personal**.
- R-2 Tiempo disponible: **300** horas.
- R-3 La situación vivida con el **COVID-19**.

# Contenidos

## Estado del arte

## Definición del proyecto

## Objetivos

## Hipótesis

## Asunciones y restricciones

## Diseño del proyecto

## Diseño del modelo

## Diseño técnico

## Planificación del proyecto

## Implementación

## Experimentos

## Primer experimento

## Segundo experimento

### Tercer experimento

## Cuarto experimento

## Conclusiones y trabajo futuro

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
●ooooooo  
o  
oo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## TF-IDF

- *Inverse Document Frequency* (IDF):

$$idf(t_i) = \log \frac{N}{n_i}$$

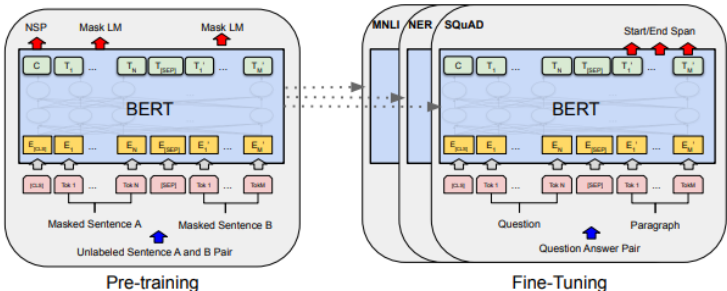
- *Term Frequency* (TF): frecuencia del término en el documento.
- *Term Frequency - Inverse Document Frequency* (TF-IDF):

$$tfidf(t_i) = tf_i \cdot idf(t_i)$$

# BERT

BERT tiene **dos pasos principales** en cuanto a su entrenamiento:

- *Pre-training*: gran corpus **sin etiquetar**.
- *Fine-tuning*: datasets **etiquetados** para tareas específicas.



Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
oo●ooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Arquitectura de BERT (Codificador) (1/3)

BERT es un **Transformer codificador bidireccional multicapa**.

- **6 capas** idénticas.
- Cada capa tiene **2 sub-capas**.
- La primera es una capa con un mecanismo de **auto-atención multi-cabezal**.
- La segunda es una **red neuronal pre-alimentada totalmente conectada**.
- Utiliza una **conexión residual**, seguida de una **normalización de capa**.

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
ooo●oooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Arquitectura de BERT (Codificador) (2/3)

- La **normalización** de capa se consigue mediante:

$$\bar{a}_i^l = \frac{g_i^l}{\sigma^l}(a_i^l - \mu^l)$$

- Por su parte, la **conexión residual** consiste en saltarse el entrenamiento en algunas capas.

Estado del arte  
○○○

Definición del proyecto  
○  
○  
○  
○

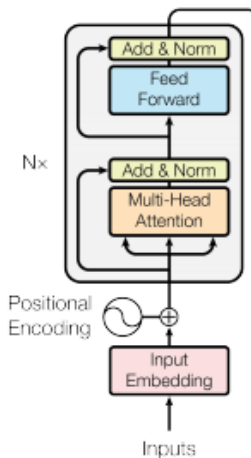
Diseño del proyecto  
○  
○○○○●○○○  
○  
○○○

Implementación  
○○

Experimentos  
○  
○○  
○○  
○○  
○○

Conclusiones y trabajo futuro  
○○○○○○○

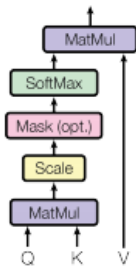
## Arquitectura de BERT (Codificador) (3/3)



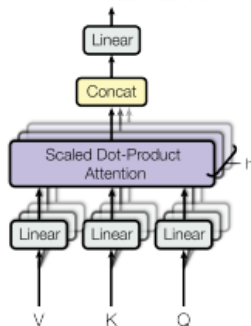
## Arquitectura de BERT (Auto-atención multi-cabezal)

- $Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$
- $MultiHead(Q, K, V) = Concat(head_1, ..., head_h)W^O$

Scaled Dot-Product Attention



Multi-Head Attention





## Arquitectura de BERT (Redes pre-alimentadas posicionales)

- **Red pre-alimentada totalmente conectada** aplicada a cada posición.
- Consiste en dos **transformaciones lineales** con activación **ReLU** entre ambas.

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2.$$

- Transformaciones lineales **idénticas** entre las posiciones.  
Parámetros **varían de una capa a otra**.

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
oooooooo●o  
ooo

Implementación  
oo

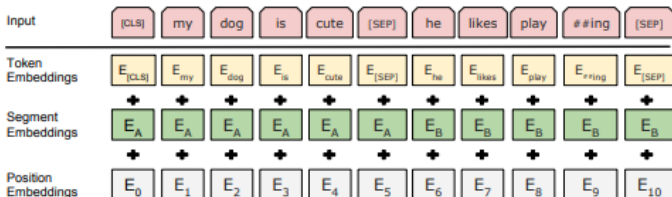
Experimentos  
o  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Funcionamiento de BERT (1/2)

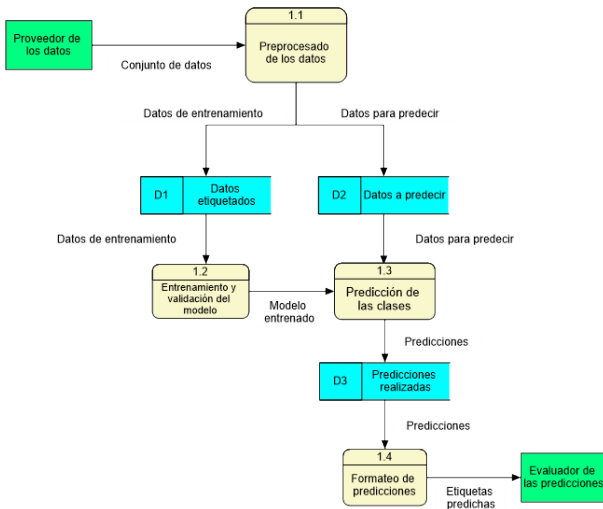
- **WordPiece** embeddings.
- El primer token es [CLS].
- Dos frases separadas por el token [SEP]. Embedding para indicar a qué frase pertenece cada token.
- La representación de un token se calcula sumando los embeddings del **token**, **segmento** y **posición**.

## Funcionamiento de BERT (2/2)



- *Pre-training*: **Masked LM** y **Next Sentence Prediction** (NSP). Wikipedia y BooksCorpus.
- *Fine-tuning*: muy sencillo gracias a la **auto-atención**. Datos con entrada y salida correspondiente.

## DFD nivel 1 general



Estado del arte  
○○○

Definición del proyecto  
○  
○  
○  
○

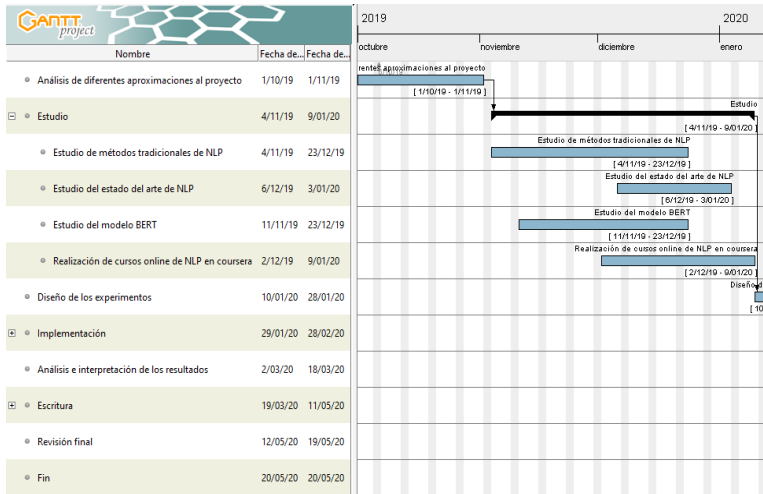
Diseño del proyecto  
○  
○  
○  
○  
○  
○  
●○○

Implementación  
○○

Experimentos  
○  
○  
○  
○  
○  
○

Conclusiones y trabajo futuro  
○○○○○○○

## Diagrama de Gantt (1/3)



Estado del arte  
○○○

Definición del proyecto  
○  
○  
○  
○

Diseño del proyecto  
○  
○  
○○○○○○○○○  
○  
○●○

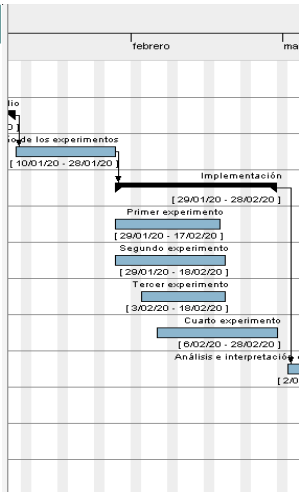
Implementación  
○○

Experimentos  
○  
○○  
○○  
○○  
○○

Conclusiones y trabajo futuro  
○○○○○○○

## Diagrama de Gantt (2/3)

Gantt project		
Nombre	Fecha de...	Fecha de...
• Análisis de diferentes aproximaciones al proyecto	1/10/19	1/11/19
⊞ • Estudio	4/11/19	9/01/20
• Diseño de los experimentos	10/01/20	28/01/20
⊞ • Implementación	29/01/20	28/02/20
• Primer experimento	29/01/20	17/02/20
• Segundo experimento	29/01/20	18/02/20
• Tercer experimento	3/02/20	18/02/20
• Cuarto experimento	6/02/20	28/02/20
• Análisis e interpretación de los resultados	2/03/20	18/03/20
⊞ • Escritura	19/03/20	11/05/20
• Revisión final	12/05/20	19/05/20
• Fin	20/05/20	20/05/20



Estado del arte  
○○○

Definición del proyecto  
○  
○  
○  
○

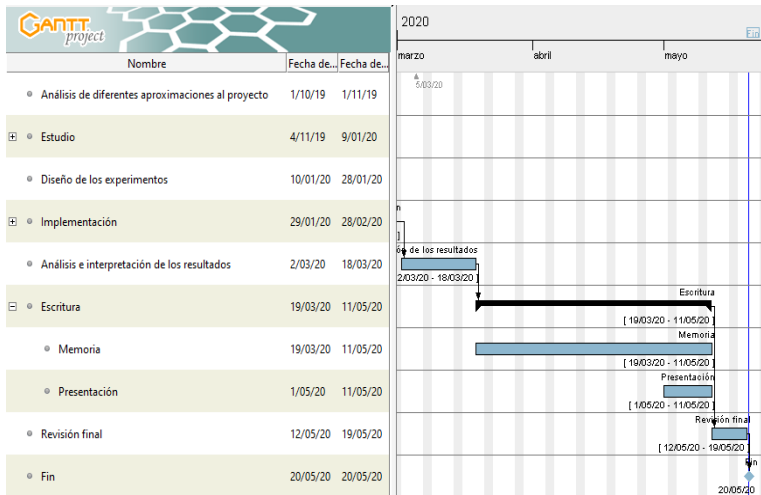
Diseño del proyecto  
○  
○  
○  
○  
○  
○  
○  
○  
○●

Implementación  
○○

Experimentos  
○  
○  
○  
○  
○  
○

Conclusiones y trabajo futuro  
○○○○○○○

## Diagrama de Gantt (3/3)



Estado del arte ooo	Definición del proyecto o o o o o	Diseño del proyecto o ooooooooo o ooo	<b>Implementación</b> ●o	Experimentos o oo oo oo oo	Conclusiones y trabajo futuro oooooooo
------------------------	--	---	-----------------------------	---	---

# Contenidos

## Estado del arte

## Definición del proyecto

Objetivos

Hipótesis

Asunciones y restricciones

## Diseño del proyecto

Diseño del modelo

Diseño técnico

Planificación del proyecto

## Implementación

## Experimentos

Primer experimento

Segundo experimento

Tercer experimento

Cuarto experimento

## Conclusiones y trabajo futuro



Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o

Diseño del proyecto  
o  
oooooooo  
o  
ooo

Implementación  
o●

Experimentos  
o  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Implementación

- *Windows 10, Anaconda, Jupyter, Python 3.7 y kaggle.*
- **Entorno virtual** con ktrain, keras, tensorflow, pandas, sklearn, numpy, nltk, etc.
- **BERT** del paquete ktrain.
- **TF-IDF** con TfidfVectorizer de sklearn. **Palabras de parada** nltk.
- `train_test_split()` de sklearn. Manejo de los datos con `numpy.array` y `pandas.DataFrame`.

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
oooooooo  
o  
ooo

Implementación  
oo

**Experimentos**  
●  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

# Contenidos

Estado del arte

Definición del proyecto

Objetivos

Hipótesis

Asunciones y restricciones

Diseño del proyecto

Diseño del modelo

Diseño técnico

Planificación del proyecto

Implementación

**Experimentos**

Primer experimento

Segundo experimento

Tercer experimento

Cuarto experimento

Conclusiones y trabajo futuro

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
●o  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Descripción del primer experimento

- **Dataset:** *IMDB* dataset. 50000 movie reviews.
- **Idioma** del dataset: Inglés.
- **Clases:**
  - **pos:** movie reviews positivas.
  - **neg:** movie reviews negativas.
- **Train/Test:** 50% para train y 50% para test.

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
o●  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Resultados del primer experimento

Modelo	Precisión validación
<b>BERT</b>	<b>0.9387</b>
Voting Classifier	0.9007
Logistic Regression	0.8949
Linear SVC	0.8989
Multinomial NB	0.8771
Ridge Classifier	0.8990
Passive Aggressive Classifier	0.8931

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
●o  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Descripción del segundo experimento

- **Dataset:** *Real or Not? NLP with Disaster Tweets* de kaggle.
- **Idioma** del dataset: Inglés.
- **Clases:**
  - 1: tweets sobre desastres naturales reales.
  - 0: tweets que no son sobre un desastre natural real.
- **Train/Test:** 75% para train y 25% para validation. Después, predicción de los *datos de la competición*.
- “#anything” por “hashtag” y “@anyone” por “entity”.

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
ooooooooo  
o  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
o  
o●  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oooooooo

## Resultados del segundo experimento

Modelo	Precisión validación	"Score" competición
<b>BERT</b>	<b>0.8361</b>	<b>0.83640</b>
H2OAutoML	0.7875	0.77607

- Top 18% en la competición con BERT!

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
●o  
oo

Conclusiones y trabajo futuro  
oooooooo

## Descripción del tercer experimento

- **Dataset:** *FASAM - NLP Competition - Turma 4 de kaggle.*
- **Idioma** del dataset: Portugués.
- **Clases:** ambiente, equilibrioesaude, sobretudo, educacao, ciencia, tec, turismo, empreendedorsocial y comida.
- **Train/Test:** 75% para train y 25% para validation. Después, predicción de los *datos de la competición.*

## Resultados del tercer experimento

Modelo	Precisión validación	"Score" competición
<b>BERT</b>	<b>0.9093</b>	<b>0.91196</b>
Predictor	0.8480	0.85047

- Top 2% en la competición con BERT!



Estado del arte	Definición del proyecto	Diseño del proyecto	Implementación	Experimentos	Conclusiones y trabajo futuro
ooo	o o o o	o ooooooooo o ooo	oo	o oo oo oo oo ●o	oooooooo

## Descripción del cuarto experimento

- **Dataset:** *Chinese Hotel Reviews* dataset. 6000 hotel reviews.
- **Idioma** del dataset: Chino peninsular con caracteres simplificados, *zh-CN*.
- **Clases:**
  - **pos:** hotel reviews positivas.
  - **neg:** hotel reviews negativas.
- **Train/Test:** 85% para train y 15% para test.

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o

Diseño del proyecto  
o  
oooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo  
o●

Conclusiones y trabajo futuro  
oooooooo

## Resultados del cuarto experimento

Modelo	Precisión validación
<b>BERT</b>	<b>0.9381</b>
Predictor	0.7399

# Contenidos

## Estado del arte

## Definición del proyecto

## Objetivos

## Hipótesis

## Asunciones y restricciones

## Diseño del proyecto

## Diseño del modelo

## Diseño técnico

## Planificación del proyecto

## Implementación

## Experimentos

## Primer experimento

## Segundo experimento

## Tercer experimento

## Cuarto experimento

## Conclusiones y trabajo futuro

## Conclusiones

- O-1 Comparar el rendimiento de **BERT** con algoritmos clásicos en **Inglés**. 100%.
- O-2 Comparar el rendimiento de **BERT** con algoritmos clásicos en **otros idiomas**. 100%.
- O-3 Medir el rendimiento de **BERT** en alguna clase de **competición**. 100%.
- O-4 Demostrar la efectividad de **BERT** sobre **datasets pequeños**. 100%.
- O-5 **Optimización Bayesiana** de algunos parámetros de **BERT**. 20%.

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
oo●oooo

## Conclusiones

- H-1 BERT tiene un **mejor rendimiento** que algoritmos clásicos. *Contrastada.*
- H-2 BERT es un modelo capaz de obtener buenos resultados en **cualquier idioma.** *Contrastada.*
- H-3 BERT es **fácil de implementar** comparado con los métodos tradicionales. *Contrastada.*
- H-4 BERT es capaz de conseguir buenos resultados incluso sobre **datasets pequeños.** *Contrastada.*

Estado del arte  
ooo

Definición del proyecto  
o  
o  
o  
o  
o

Diseño del proyecto  
o  
ooooooooo  
o  
ooo

Implementación  
oo

Experimentos  
o  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro  
ooo●ooo

## Trabajo futuro

- **Optimización Bayesiana** de algunos parámetros de **BERT**.
- Realizar experimentos en más idiomas con **distintos alfabetos**.
- Probar más **métodos tradicionales**.
- Participar en nuevas competiciones de **kaggle**.
- Utilizar **BERT** para otros problemas de **NLP**.

Estado del arte	Definición del proyecto	Diseño del proyecto	Implementación	Experimentos	Conclusiones y trabajo futuro
ooo	o o o o	o ooooooooo o ooo	oo	o oo oo oo oo	oooo●oo

## Referencias

- J. L. Ba, J. R. Kiros, and G. E. Hinton, *Layer normalization*, arXiv preprint arXiv:1607.06450, (2016).
- J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, *Bert: Pre-training of deep bidirectional transformers for language understanding*, arXiv preprint arXiv:1810.04805, (2018).
- S. Robertson, *Understanding inverse document frequency: on theoretical arguments for idf*, Journal of documentation, (2004).

Estado del arte	Definición del proyecto	Diseño del proyecto	Implementación	Experimentos	Conclusiones y trabajo futuro
ooo	o o o o o	o ooooooooo o ooo	oo	o oo oo oo oo oo	ooooo●o

- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, *Attention is all you need*, in Advances in neural information processing systems, 2017, pp. 5998–6008.



Estado del arte

ooo

Definición del proyecto

o  
o  
o  
o  
o

Diseño del proyecto

o  
oooooooo  
o  
ooo

Implementación

oo

Experimentos

o  
oo  
oo  
oo  
oo  
oo

Conclusiones y trabajo futuro

oooooooo●

¡Muchas gracias  
por vuestra atención!