

Clasificación de sentimientos IMBD con BERT

Por Juan Pablo Bernal Lafarga - A01742342

Accuracy de Regresión Logística			
Approach/Parameters	Bag Of Words	TF-IDF	BERT
maxtokens = 50, maxtokenlen = 20	70%	70.5%	0.755%
maxtokens = 100, maxtokenlen = 100	72.66%	76.16%	0.8064%
maxtokens = 200, maxtokenlen = 100	80.33%	82%	0.8393%
maxtokens = 200, maxtokenlen = 200	—	—	0.813%
maxtokens = 230, maxtokenlen = 200	—	—	0.8186%

Los cuadros verdes presentan el enfoque con mejor exactitud en un modelo de regresión logística por cada selección de parámetros (recuadros azul oscuro).

Como podemos observar, el uso de los embeddings contextuales otorgados por BERT nos es de provecho, pues obtenemos mejores resultados en todas las selecciones de parámetros contra enfoques como Bag of Words y el TF-IDF.

Y como observación adicional, notemos que el accuracy se ve afectado de manera un poco negativa al aumentar desmedidamente el maxtokenlen.