

Aplicación de modelos de lenguaje para la identificación de emociones presentes en twitter durante el periodo de elecciones presidenciales en Colombia.

Por: Juan Jose Iguaran Fernandez

Dentro del texto existe información objetiva, como hechos verificables e información subjetiva, que corresponde a los procesos internos que los individuos experimentan y son plasmados en el texto, tal como las opiniones.

Las emociones son parte de esta información subjetiva, y su clasificación en términos generales ha sido definida en seis emociones básicas: miedo, rabia, tristeza, alegría, sorpresa y disgusto.

La detección de las emociones presentes en el texto es un sub campo del análisis de sentimiento en el texto que busca determinar la polaridad y el grado de las distintas dimensiones de la subjetividad presentes en el texto.

El estudio del análisis de sentimientos en general y de emociones en particular se ha hecho usualmente a través tradicionales de NLP tales como el empleo de modelos de aprendizaje supervisado a partir de features construidos a partir del texto.

Durante los últimos años, estas técnicas están siendo remplazadas por modelos de lenguajes usando redes neuronales, en particular arquitecturas como los transformers debido a su mejor desempeño y robustez.

Un medio particularmente interesante para la detección de emociones son las redes sociales pues son capaces de captar a una gran cantidad de usuarios sobre una gran cantidad de tópicos en un numero limitado de palabras con estilo propio. Esto ha sido estudiado en el pasado en por ejemplo, como los análisis de texto provenientes de estas, coincide con lo que arrojan otros modelos de las realidades sociales tales como las encuestas de opinión.

En español existen pocos casos de detección de emociones en redes sociales, y no se conoce de ninguno que use modelos de lenguaje basado en redes neuronales para este fin en un contexto político.

EL presente trabajo tiene por objetivo el empleo de modelos de lenguaje, específicamente BERT que es una red neuronal pre entrenada con la wikipedia basada en transformers para detectar emociones presentes en twitter durante las elecciones en Colombia.