

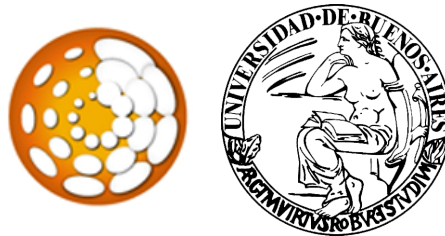
Aplicación de modelos de lenguaje para la identificación de emociones en Twitter durante las elecciones presidenciales de 2022 en Colombia

Tesis presentada para optar por el título de

**Magister en Explotación de Datos y
Descubrimiento del Conocimiento**

por

Juan Jose Iguaran Fernandez



Universidad de Buenos Aires

**Facultad de Ciencias Exactas y Naturales
Departamento de Ciencias de la Computación**

[Insert Month and Year]

AGRADECIMIENTOS

Agradezco

Resumen

Twitter ha sido analizado como un medio particularmente interesante para el estudio indirecto de fenómenos sociales a través del uso de técnicas del procesamiento del lenguaje natural (NLP) pues es capaz de captar a una gran cantidad de usuarios sobre una gran numero de tópicos en una cantidad limitada de palabras con estilo propio. Esto ha sido estudiado en el pasado en por ejemplo, como los análisis de texto provenientes de este, coincide con lo que arrojan otras aproximaciones de las realidades sociales tales como las encuestas de opinión, discursos políticos o eventos de relevancia popular.

Una tarea en particular dentro del NLP, la detección de emociones, resulta de particular interés al estudiar la respuesta individual a fenómenos sociales pues 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. En ese contexto, 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 contruidos 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 capacidad de tener en cuenta el contexto dentro del cual se encuentran las palabras en el texto, es decir, su relación con otras palabras. Una aplicación particular de los Transformers BERT, es particularmente útil pues consiste en un modelo pre entrenado que es capaz de desempeñar la tarea de NLP en particular para la cual se haga un entrenamiento final

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 Transformers 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.

Palabras Clave: [BERT, Colombia, Elecciones, Detección de emociones]

Índice general

Índice de figuras	6
Índice de cuadros	7
1. Introducción	1
1.1. Motivación	1
1.2. Marco Teórico	1
1.2.1. Análisis de sentimiento y emociones en el texto	1
1.2.2. Sentimientos y emociones en redes sociales	5
1.2.3. Redes neuronales para el análisis de texto	9
1.2.4. Detección de sentimientos y emociones en español	13
2. Datos	16

3. Metodología	18
Bibliografía	19

Índice de figuras

1.1. Red neuronal recurrente	10
1.2. Arquitectura de los Transformers	12
1.3. Arquitectura de BERT	13

Índice de cuadros

2.1. Cantidad de Tweets para cada emoción por Dataset	17
---	----

Capítulo 1

Introducción

1.1. Motivación

Este trabajo es importante por que

1.2. Marco Teórico

1.2.1. Análisis de sentimiento y emociones en el texto

Las seres humanos atravesamos la experiencia de la vida a través de nuestra cognición, es decir, el conjunto de procesos mentales que culminan en la representación interna del mundo que nos rodea. Una parte fundamental de esta, son las emociones. En un esfuerzo por intentar identificarlas de una manera generalizada,

Ekman [Ekman, 1993] realiza un estudio de la respuesta fisiológica en general y de las expresiones faciales en particular, del ser humano en diferentes culturas ante distintas circunstancias. Esto lo lleva a concluir que existen grandes grupos en donde las distintas expresiones faciales pueden ser agrupadas ya que estas reflejan el estado emocional interno de los individuos. A estos grupos los denomino emociones básicas y son los siguientes: Alegría, enojo, sorpresa, asco, miedo y tristeza. Este modelo de emociones básicas es comúnmente usado para los estudios relacionados con las emociones.

Debido la relevancia que presentan las emociones en la manera en como procesamos la informacion, en el libro [Picard, 2000] Picard pone de manifiesto que la búsqueda de una inteligencia artificial capaz de interactuar eficazmente con los seres humanos, debe traer consigo la capacidad de reconocer, entender, tener y expresar emociones. Es así como bautiza el capo de estudio que busca dicha interacción como computación afectiva, y expone los avances hechos hasta la fecha en el mismo.

Dado que el lenguaje en general es el medio principal mediante el cual los seres humanos transferimos informacion de manera general, y mediante las computadoras en particular, los esfuerzos orientados hacia esta computación afectiva tienen un componente fundamental en el procesamiento del lenguaje para dicho propósito, por lo que se genera la necesidad de entender la relación que existe entre lenguaje y emoción antes de poder automatizar este proceso. Dentro de este marco, [Ortony et al., 1987] se habla de como a pesar de que las emociones son procesos mentales internos que no residen en el lenguaje, este es el medio no fenomenológico mas conveniente a través del cual podemos acceder a ellas. Elabora entonces una serie condiciones que deben estar presentes en los términos para poder referirse de una manera acertada a los estados emocionales, y las aplica sobre términos presentes en la literatura con

respecto a las emociones, construyendo así un léxico emocional. De manera análoga, [Hatzivassiloglou and McKeown, 1997] construye a partir de un corpus extenso de adjetivos que vienen en parejas usando distintos conectores, y un etiquetado manual de algunos de ellos, un algoritmo para determinar lo que los autores denominan la orientación semántica de los mismos, esto es determinar si determinado adjetivo tiene una connotación negativa o positiva de la característica que describe. Este método permite la creación automática de un corpus extenso de adjetivos cuyo sentimiento se encuentra identificado.

Mas allá de la relación individual que existe entre palabras y estados emocionales, [Wiebe, 1994] a través del análisis de un tipo particular de texto, la ficción, expone que la narración puede ocurrir desde un punto de vista objetivo, es decir, una descripción de hechos comprobables, y también desde un punto de vista subjetivo, es decir, poniendo de manifiesto los hechos atravesados por los estados mentales internos del narrador, y que la distinción entre un tipo de narración y otra no esta siempre clara, por lo que propone un algoritmo capaz de hacer esta distinción de manera automática. Esto es retomado por [Yu and Hatzivassiloglou, 2003] en donde se busca hacer esta distinción a nivel de documentos, por ejemplo diferenciar editoriales de noticias, así como a nivel de frases mediante el uso modelos de aprendizaje supervisado para estas tareas en particular.

De una manera general, los estados emocionales internos de las personas que ocurren como respuesta a los eventos externos pueden ser agrupados según su polaridad, es decir, catalogar el sentimiento como positivo o negativo. Es así como [Pang et al., 2002] establece la importancia de desarrollar, ante textos que reflejen opiniones subjetivas, sistemas capaces de identificar si dicha opinión es negativa o positiva. Para este propósito, se emplea el dominio de las reviews online de

películas, que facilitan la tarea al contar con una valoración negativa o positiva por parte del usuario, construyendo algoritmos de aprendizaje supervisado que utiliza como features principalmente unigramas, que es un elemento individual del texto y como variable objetivo el sentimiento la valoración expresada. Por otro lado, [Turney, 2002] pretende establecer a través de aprendizaje no supervisado, si determinada review sobre diversos temas online, presenta un sentimiento negativo o positivo. Para ello, emplea el concepto de orientación semántica presente en [Hatzivassiloglou and McKeown, 1997] para determinar si una frase tiene orientación negativa o positiva para luego determinar si la review en su conjunto es positiva o negativa. En general, la determinación de la orientación negativa o positiva dentro del texto, es conocida como análisis de sentimiento

Así como en el análisis de sentimiento, contar con un corpus de términos que cuenten con una clasificación previa de su orientación es clave para poder llevar a cabo automáticamente la tarea, para los estados emocionales es igualmente importante, por lo que [Strapparava et al., 2004] realiza una anotación manual de estados emocionales basados en las categorías de [Ortony et al., 1987] sobre algunos términos encontrados en WordNet [Miller, 1995], que es una base de datos de términos en inglés agrupados por grupos de sinónimos y con relación semántica entre grupos. A partir de ahí se establece la categoría emocional de nuevos términos gracias a los sinónimos y las relaciones semánticas, construyendo así una base de datos de estados emocionales llamada WordNet-Affect. Dentro de esta misma línea, [Wiebe et al., 2005] elabora una anotación manual de los estados emocionales presentes en las oraciones de un gran volumen de noticias, en donde se tiene en cuenta el contexto.

Más allá de la relación entre términos particulares y estados emocionales, [Alm et al., 2005] utilizan los cuentos infantiles para desarrollar un modelo de aprendizaje supervisado

capaz de detectar emociones en las frases del texto. Para ello se elabora una anotación manual de las frases que constituyen el set de datos y luego, se generan un set de features para estas que pasaran a entrenar un clasificador lineal.

1.2.2. Sentimientos y emociones en redes sociales

Con el paso del tiempo, Internet se han convertido en un lugar de intercambio de informacion prevalente entre sus usuarios y dicha informacion se presenta, entre otras formas, en el texto, por ello, es una fuente de generación masiva del mismo que puede facilitar su estudio. En ese sentido, [Pang et al., 2008] se hace manifiesto la importancia que ha venido ganando el campo del análisis de sentimiento del texto en Internet, tanto para usuarios individuales como para la industria de la publicidad, el mercado financiero y la academia, por lo que se hace un recuento de las distintas técnicas y aplicaciones que son consideradas relevantes por los autores hasta la fecha.

Un espacio particular en donde los usuarios individuales pueden generar grandes volúmenes de informacion sobre distintos contenidos, y por consiguiente, una fuente rica de datos para el análisis de sentimiento son los blogs. Por esta razón, [Aman and Szpakowicz, 2007] utiliza texto proveniente de ellos para realizar detección de emociones presentes en las oraciones de los mismos. Para este propósito recurren primero a una anotación manual de las mismas y luego a la construcción de features para entrenar distintos modelos supervisados.

Twitter, un sitio de blogs en particular cuyo formato es el microblogging, es decir, publicaciones pequeñas a las que denomina tweets, ha cobrado una gran relevancia en el estudio del texto debido a su gran popularidad entre los usuarios de Internet

de todo tipo, desde marcas, usuarios individuales y políticos, para abarcar cualquier tema. En ese contexto, [Pak and Paroubek, 2010] extrae tweets de diferentes usuarios sobre distintos temas para realizar un análisis de sentimientos sobre estos. Para ello se procede a la identificación de tweets que contengan emoticones felices y tristes, los cuales etiqueta como reflejando un sentimiento positivo o negativo respectivamente, así como tweets provenientes de cuentas de medios de noticias los cuales etiqueta como neutrales. luego se procede a la construcción de features usando n-gramas, que son elementos que contienen la combinación de n palabras del texto, a partir de las palabras presentes en el tweets con estos se entrenaron varios clasificadores.

Un elemento característico de twitter, son los hashtags que consisten en palabras precedidas del signo numeral (#) y sirven para denotar una temática en particular relacionada con el tweet, por ejemplo ”#emociones”. Gracias a esta potencial para categorizar de los hashtags,[Davidov et al., 2010] parte del supuesto de que estos y los emoticones contienen información relevante el cuanto al sentimiento del tweet, por lo que partiendo de una base de datos de 475 millones de tweets publicados entre mayo del 2009 y enero del 2010, se hace una selección de 50 hashtags que tengan una asociación fuerte a sentimientos y se entrena un modelo supervisado a partir de los tweets que contengan estos hashtags, representándolos a través de vectores de features . El modelo es posteriormente empleado para clasificar otros tweets y jueces humanos verifican su eficacia. Una idea similar emplea [Wang et al., 2012], en donde este concepto sirvió ademas para la extracción de datos pues se utilizan hashtags que contengan términos claves provenientes de las 5 emociones básicas de propuestas por [Ekman, 1993] para realizar el llamado de la API (application programming interface) de Twitter, es decir, solicitando del sitio que se extraigan solo los tweets que contengan estos hashtags. Esta manera de utilizar los hashtags como un tipo de

clasificación, es conocida como supervisado distante.

Si bien el método de supervisado distante puede servir como una estimación de la relación entre el texto del tweet y su contenido emocional, es un método en el que no se puede afirmar con certitud dicha relación. Por ello, [Roberts et al., 2012] identifica la necesidad de contar con un corpus etiquetado que sirva de base para la tarea de identificación de emociones en twitter. Para ello, se seleccionan 14 temas que para los autores tienen un fuerte contenido emocional y las palabras clave asociados a estos para ser usados como hashtags en las extracción. A partir de ahí, manualmente se etiquetaron los tweets con su respectiva emoción. Esto sirvió de base para entrenar un modelo de aprendizaje supervisado y así verificar el poder predictivo de los modelos partiendo de datos etiquetados.

La identificación del sentimiento presente en Twitter puede servir como fuente de aproximación a la realidad social en la que los usuarios se desenvuelven debido a la naturaleza inmediata, relevante para los usuarios y descentralizada propia del flujo de información de esta red social. Por esto, [O'Connor et al., 2010] se plantea la pregunta si existe una correlación entre el sentimiento encontrado en twitter y las encuestas de opinión. Para ello, se toma una muestra de mil millones de tweets entre 2008 y 2009 y se toman aquellos tweets que contengan palabras claves asociadas a los temas que se están investigando. EL sentimiento de estos tweets se determina a partir de la proporción de palabras con asociación negativa o positiva presentes en el tema que se esta analizando en un día en particular. Los resultados dan una correlación alta entre el sentimiento encontrado a través del texto y las encuestas. Esta relación entre la realidad social y el sentimiento presente en los tweets es también abordada por [Bollen et al., 2011], donde se procede a realizar una medida del estado emocional de una muestra de tweets entre agosto y diciembre del 2008, en donde el mismo se mide

a través de la similaridad presente entre las palabras de los tweets y ciertos términos claves asociados a estados emocionales. Esto permite encontrar que determinados eventos relevantes generan un impacto emocional significativo y durante un periodo de tiempo en los usuarios.

Al ser el escenario político un caso particular de los fenómenos sociales cuyo impacto se traduce en el estado emocional de las personas, el sentimiento presente en tweets de contenido político puede dar un indicio de la percepción pública de dicho escenario. Con esto en mente, [Tumasjan et al., 2010] se planteó el uso de twitter como plataforma de medición de la percepción pública respecto a la política durante las elecciones parlamentarias en Alemania en el 2009. Una de sus preguntas de investigación estuvo relacionada con los sentimientos que se reflejan en los tweets que mencionan a los políticos que hacen parte de las campañas y para esto, se empleó sobre el texto proveniente de twitter, un software capaz de identificar palabras claves asociadas a estados emocionales y cognitivos en el texto. El resultado fue un perfil emocional para cada político que en líneas generales, parece estar de acuerdo con su discurso político. Esta relación entre política y sentimiento presente en twitter, fue también estudiada por [Cerón-Guzmán and León-Guzmán, 2016] en donde realizó un análisis de sentimientos sobre tweets relacionados con las elecciones presidenciales en Colombia en el 2014 y se compararon los resultados de este con las encuestas de opinión. Para ello, inicialmente obtiene tweets que tengan palabras claves y hashtags relacionados con las elecciones, luego hace un filtrado de spam y finalmente entrena un modelo supervisado con los features que genera para los tweets restantes. Los resultados no son consistentes por lo que se plantea en futuros trabajos una caracterización demográfica.

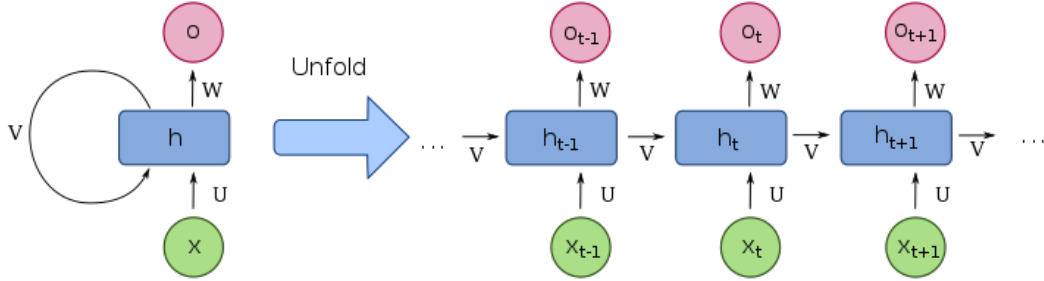
1.2.3. Redes neuronales para el análisis de texto

Las técnicas tradicionales de análisis de sentimiento y emociones están principalmente basadas en la relación que existe entre los términos que constituyen el texto y determinado estado emocional. Este tipo de análisis, si bien ha sido de gran utilidad, no dispone de la capacidad de tener en cuenta el contexto, es decir la relación que existe entre las palabras que constituyen el texto y el orden en que estas se presentan. Debido a esto, recientemente el campo de ido migrando hacia el uso del algoritmos capaces de captar esta relación contextual, específicamente las redes neuronales. Bajo perspectiva, [Acheampong et al., 2021] hace un recuento del estado del arte de la detección de emociones en el texto. Es aquí donde se pone en evidencia la conveniencia de usar RNN (redes neuronales recurrentes) debido a la naturaleza secuencial de su arquitectura.

Las RNN (redes neuronales recurrentes) son un tipo de arquitectura de red neuronal cuyo uso es especial para datos secuenciales, tales como las tareas de NLP. En estas, para un ejemplo nuevo, es posible utilizar el resultado del procesamiento de un dato anterior, para el del siguiente dato (como en una secuencia de caracteres por ejemplo), es decir la misma topología de pesos recibirá para cada ejemplo nuevo, el resultado anterior además de la nueva entrada. En la imagen 1.1 [Wikipedia contributors, 2022] se evidencia un esquema de esta arquitectura.

Sin embargo, debido a su arquitectura, la optimización de los pesos que constituyen la topología de las RNN, se vuelve compleja pues para el cálculo del gradiente del error, pues se deberá pasar tantas veces por los pesos de la red como pasos en el tiempo haya (como número de caracteres en una palabra) haciendo que en secuencias particularmente largas, este cálculo crezca o disminuya en demasía. Para enfrentar

Figura 1.1: Red neuronal recurrente



este problema, [Hochreiter and Schmidhuber, 1997] y [Chung et al., 2014] plantean nuevas arquitecturas de RNN conocidas como LSTM (Long-Short term memory) y GRU (Gated Recurrent Unit) respectivamente, en donde a través de nuevas unidades que permiten la activación/cancelación de las señales que constituyen la red, se puede realizar la optimización de manera directa sin pasar por los pesos.

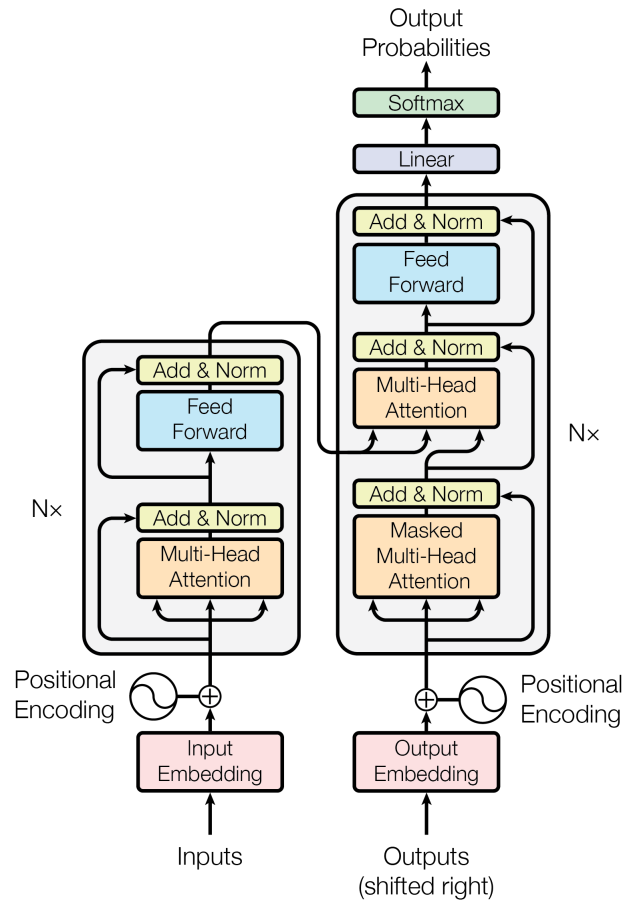
Luego se describe como el uso de Transformers permitió paralelizar el procesamiento, así como la capacidad de tener información contextual para intervalos secuenciales más amplios. Dentro del uso de Transformers se resalta el uso de modelos pre entrenados que puedan ser utilizados para tareas específicas a través de un entrenamiento final. Aquí se resalta BERT como el modelo más popular para la tarea de detección de emociones recientemente.

Si bien las RNN permiten preservar de alguna manera la relación contextual del texto, al tener una estructura secuencial, se impide la paralelización de su cómputo, pues se necesitan las salidas de ejemplos anteriores para llevar a cabo el siguiente paso en el tiempo. Además, debido a este mismo funcionamiento secuencial, si existe un

gran numero de pasos en el tiempo, es poco probable que la red tenga en cuenta información presente al inicio. Para solucionar estos inconvenientes, [Vaswani et al., 2017] desarrollan los Transformers, que son un tipo de arquitectura en la que múltiples datos de entrada son ingresados a la red de manera simultanea, como todas las palabras en una frase por ejemplo, mediante la representación de dichas palabras en la forma de un vector (Embedding) y a través de una arquitectura que definen como atención, se realiza una transformación de los datos de manera que la representación de cada dato, en este caso cada palabra, tenga en cuenta que tan importante es la misma para todas las demás palabras de la oración. Esta transformación, aplicada tanto en los datos de entrada como en los de salida, es luego usada para entrenar los pesos de la red. De esta manera la red es capaz de procesar datos en simultaneo así como tener en cuenta toda la información disponible. Un esquema de esta arquitectura se puede apreciar en la imagen 1.2 [Vaswani et al., 2017].

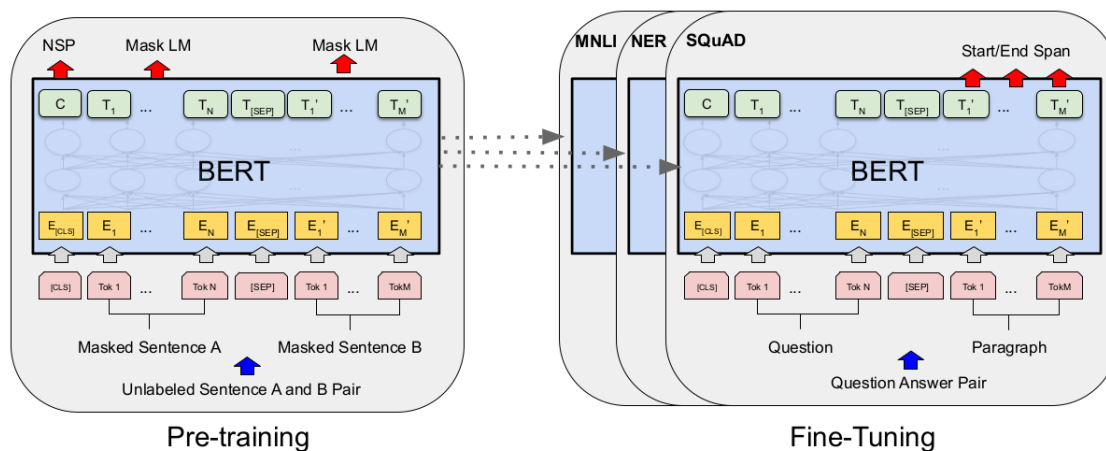
Gracias a la capacidad de paralelizacion y de conservar la relacion entre palabras distantes del texto que presentan los transformers, es posible entrenar un modelo del lenguaje con grandes volúmenes de información para de esta manera, contar con una red con un gran poder predictivo. Es así como [Devlin et al., 2018] desarrolla un modelo basado en Transformers capaz de aprender el contexto del lenguaje de una manera general y luego utilizar lo aprendido para distintas tareas de NLP denominado BERT (Bidirectional Transformers for Language Understanding). Para ello se utilizó una arquitectura en donde las capas de entrada son las representaciones vectoriales de las palabras (Embeddings) y a partir de ahí, contiene múltiples capas de Transformers. Para su entrenamiento, esta arquitectura tuvo dos tareas simultaneas: se suministran dos oraciones consecutivas con palabras faltantes y las tareas son saber que palabra podría ser la faltante, además de identificar el orden de las

Figura 1.2: Arquitectura de los Transformers



oraciones. Esto le permite al modelo aprender del contexto del lenguaje a nivel de palabras usando las que vienen antes y después como fuente de información, así como del contexto de las oraciones al identificar su orden. Para ello se utilizó toda la wikipedia. Este modelo pre entrenado es luego capaz de utilizar esta representación contextual del lenguaje para realizar distintas tareas de NLP tales como análisis de sentimiento, al agregar una capa final a la red que produzca un posible resultado y al comparar este con su respectivo valor esperado realizar un entrenamiento. El esquema de esta arquitectura se puede apreciar en 1.3 [Devlin et al., 2018].

Figura 1.3: Arquitectura de BERT



1.2.4. Detección de sentimientos y emociones en español

Las técnicas y desarrollos relativos al análisis de emociones y sentimientos, provienen inicialmente del inglés. Una vez se cuenta con un marco metodológico en funcionamiento para este idioma, se desarrollan esfuerzos para su aplicación en otros, como el español. En este contexto, debido a la necesidad de contar con términos en español que puedan ser relacionados a determinadas emociones para propósitos de la detección automática, [Sidorov et al., 2012] provee un léxico con tal propósito, que obtienen a través de la traducción de léxicos semejantes en inglés, que son luego evaluados por comentadores individuales, para filtrar aquellas palabras que presenten poca asociación con las emociones, así como dar un porcentaje de asociación del término a la emoción, dependiendo de cuantos comentadores estuvieron de acuerdo en dicha asociación. En un esfuerzo similar por contar con texto en español, así como para árabe, etiquetado para el dominio de detección emocional, [Mohammad et al., 2018]

se identifica términos claves en ingles asociados a las distintas emociones básicas. Estos términos fueron luego traducidos, verificados por hablantes nativos y expandidos utilizando sus sinónimos, para luego ser utilizados como parte de la query a la API de twitter y a partir de ahí, se procede a un etiquetado manual de los tweets a través de una plataforma de crowdsourcing. Esfuerzos similares para llevar a cabo el etiquetado manual de los tweets fueron llevados a cabo por [Sidorov et al., 2016] en donde los terminos emocionales fueron usados como hashtags a la hora de hacer los llamados a la api, para tweets de distintos paises y luego fueron verificados a mano. de igual manera, [Plaza-del Arco et al., 2020b] un set de datos de tweets con contenido emocional usando hashtags relacionados a temas con alto contenido emocional para hacer el llamado a la API y luego se utilizo la plataforma Amazon Mechanical Turk (MTurk) para el etiquetado manual.

Estos esfuerzos previos son aprovechados por [Plaza-del Arco et al., 2020a] donde se identifica que existe poca literatura al rededor de la clasificación de emociones en texto en español. Para ello, se plantea la evaluación de distintos modelos de aprendizaje supervisado usando el corpus de tweets en español con categorías emocionales provenientes de [Mohammad et al., 2018]. A partir de ahí, se procede a generar representaciones vectoriales de los tweets. Se evalúan luego estos modelos de para tener un desempeño de base y luego se añade una variable que indique la presencia de algunos términos claves asociados con las emociones básicas, proveniente del trabajo realizado por [Sidorov et al., 2012]. Esto mejoro considerablemente el desempeño del modelo. Un esfuerzo similar realizo [Gil et al., 2013] quien lleva a cabo una clasificación con supervisado distante de las emociones. Para dicho propósito recurre al empleo de términos relacionados con algunas de las emociones básicas como hashtags a la hora de realizar la query a la API y se construye un dataset en donde los tweets

que contengan dichos hashtags son pertenecientes a determinada emoción. A partir de ahí se entrenan distintos algoritmos de aprendizaje supervisado.

Si bien estos esfuerzos sirvieron para sentar las bases del desarrollo del area de detección de emociones en español, así como en ingles, carecen de métodos que les permitan tener en cuenta la informacion contextual del texto. En ese sentido, con el desarrollo de las técnicas de NLP basadas en redes neuronales en ingles, surgieron así mismo esfuerzos para aplicar dichas técnicas en el dominio del texto en español. Es así como en [Canete et al., 2020] se destaca la importancia de los modelos pre entrenados basados en Transformers por su superior desempeño en tareas de NLP ademas de su practicidad de uso. Sin embargo, resaltan que no existía hasta la fecha un modelo de este tipo entrenado específicamente para español, ademas del BERT para múltiples lenguajes. Por ello, se proponen entrenar dicho modelo usando ademas de la Wikipedia en español, texto de publicaciones de las naciones unidas, gobiernos y charlas TED. El resultado fue un modelo que supera al BERT en múltiples idiomas para el español en casi todas las tareas evaluadas. Un es fuerzo similar llevo a cabo [Gonzalez et al., 2021] donde se resaltan las ventajas de la arquitectura de BERT, sin embargo se pone de manifiesto que el idioma español en general y el dominio de twitter en particular, poseen características propias que al ser tenidas en cuenta, podrían mejorar el desempeño para esta tarea en especifico con respecto al modelo de múltiples lenguaje de BERT. Para ellos, se recurre a un corpus de 41 millones de tweets en español con el que se entrena un modelo con la arquitectura de BERT al que denominan TwilBert, obteniendo resultados mejores que BERT para algunas tareas de NLP

Capítulo 2

Datos

En [Pak and Paroubek, 2010] los autores utilizan distintos emoticones en la query de la API para traer tweets que tengan sentimientos positivos y negativos.

En [Davidov et al., 2010] se utiliza un dataset de 450 millones de tweets se hace calcula la frecuencia de los hashtags hasta encontrar los mas comunes y de estos se seleccionan aquellos que tengan una asociación mas fuerte con los sentimientos para luego traer los tweets que los contengan.

En [Wang et al., 2012] se utilizan distintos términos relacionados con emociones que serán usados como hashtags para hacer el llamado a la API. luego se hace una preseleccion de los tweets en donde se verifique el hashtag este al final, que no contenga URL, que tenga mas de 5 palabras y que no tenga mas de 3 hashtags. Además, se agregaron variaciones lingüísticas a los términos: esperanza, esperanzador. inicialmente eran 5 millones de tweets que quedaron reducidos a 2 y medio

Con el objetivo de contar con un dataset lo suficientemente robusto para la ta-

rea, una vez descargados los tweets, estos fueron filtrados utilizando un modelo de lenguaje. Este modelo, fue entrenado con los datos provenientes de [Mohammad et al., 2018], [Sidorov et al., 2016], [Plaza-del Arco et al., 2020b]

Cuadro 2.1: Cantidad de Tweets para cada emoción por Dataset

	ira	miedo	tristeza	alegria	asco	sorpresa	otra
EmoEvent	840	95	992	1756	160	340	4010
SemEval	914	396	849	1741	168	138	1016
Sidorov	1830	1873	282	1952	1862	1976	0

Capítulo 3

Metodología

En [O'Connor et al., 2010] se utiliza una media móvil para atenuar el sentimiento percibido diariamente debido a lo ruidoso que puede ser

Bibliografía

- [Acheampong et al., 2021] Acheampong, F. A., Nunoo-Mensah, H., and Chen, W. (2021). Transformer models for text-based emotion detection: a review of bert-based approaches. *Artificial Intelligence Review*, 54(8):5789–5829.
- [Alm et al., 2005] Alm, C. O., Roth, D., and Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 579–586.
- [Aman and Szpakowicz, 2007] Aman, S. and Szpakowicz, S. (2007). Identifying expressions of emotion in text. In *International Conference on Text, Speech and Dialogue*, pages 196–205. Springer.
- [Bollen et al., 2011] Bollen, J., Mao, H., and Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 450–453.

- [Canete et al., 2020] Canete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., and Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. *Pml4dc at iclr*, 2020:1–10.
- [Cerón-Guzmán and León-Guzmán, 2016] Cerón-Guzmán, J. A. and León-Guzmán, E. (2016). A sentiment analysis system of spanish tweets and its application in colombia 2014 presidential election. In *2016 IEEE international conferences on big data and cloud computing (BDCloud), social computing and networking (socialcom), sustainable computing and communications (sustaincom)(BDCloud-socialcom-sustaincom)*, pages 250–257. IEEE.
- [Chung et al., 2014] Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- [Davidov et al., 2010] Davidov, D., Tsur, O., and Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In *Coling 2010: Posters*, pages 241–249.
- [Devlin et al., 2018] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [Ekman, 1993] Ekman, P. (1993). Facial expression and emotion. *American psychologist*, 48(4):384.
- [Gil et al., 2013] Gil, G. B., Jesús, A. B. d., and López, J. M. M. (2013). Combining machine learning techniques and natural language processing to infer emotions

- using spanish twitter corpus. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 149–157. Springer.
- [Go et al., 2009] Go, A., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009.
- [Gonzalez et al., 2021] Gonzalez, J. A., Hurtado, L.-F., and Pla, F. (2021). Twilbert: Pre-trained deep bidirectional transformers for spanish twitter. *Neurocomputing*, 426:58–69.
- [Hatzivassiloglou and McKeown, 1997] Hatzivassiloglou, V. and McKeown, K. (1997). Predicting the semantic orientation of adjectives. In *35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics*, pages 174–181.
- [Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- [Kouloumpis et al., 2011] Kouloumpis, E., Wilson, T., and Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 538–541.
- [Miller, 1995] Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- [Mohammad, 2012] Mohammad, S. (2012). # emotional tweets. In ** SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 246–255.

- [Mohammad et al., 2018] Mohammad, S., Bravo-Marquez, F., Salameh, M., and Kiritchenko, S. (2018). Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th international workshop on semantic evaluation*, pages 1–17.
- [O’Connor et al., 2010] O’Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth international AAAI conference on weblogs and social media*.
- [Ortony et al., 1987] Ortony, A., Clore, G. L., and Foss, M. A. (1987). hatzivassiloglou1997predicting. *Cognitive science*, 11(3):341–364.
- [Pak and Paroubek, 2010] Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*.
- [Pang et al., 2008] Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135.
- [Pang et al., 2002] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.
- [Picard, 2000] Picard, R. W. (2000). *Affective computing*. MIT press.
- [Plaza-del Arco et al., 2020a] Plaza-del Arco, F. M., Martín-Valdivia, M. T., Ureña-López, L. A., and Mitkov, R. (2020a). Improved emotion recognition in spanish social media through incorporation of lexical knowledge. *Future Generation Computer Systems*, 110:1000–1008.
- [Plaza-del Arco et al., 2020b] Plaza-del Arco, F. M., Strapparava, C., Lopez, L. A. U., and Martín-Valdivia, M. T. (2020b). Emoevent: A multilingual emotion

- corpus based on different events. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1492–1498.
- [Roberts et al., 2012] Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., and Harabagiu, S. (2012). Empatweet: Annotating and detecting emotions on twitter. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 3806–3813.
- [Sidorov et al., 2016] Sidorov, G., Haro, S. N. G., and Vázquez, V. A. C. (2016). Construcción de un corpus marcado con emociones para el análisis de sentimientos en twitter en español. *Revista Escritos BUAP*, 1(1).
- [Sidorov et al., 2012] Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Trevino, A., and Gordon, J. (2012). Empirical study of machine learning based approach for opinion mining in tweets. In *Mexican international conference on Artificial intelligence*, pages 1–14. Springer.
- [Strapparava et al., 2004] Strapparava, C., Valitutti, A., et al. (2004). Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, page 40. Lisbon, Portugal.
- [Tumasjan et al., 2010] Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 4, pages 178–185.
- [Turney, 2002] Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.

- [Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [Wang et al., 2012] Wang, W., Chen, L., Thirunarayan, K., and Sheth, A. P. (2012). Harnessing twitter”big data”for automatic emotion identification. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, pages 587–592. IEEE.
- [Wiebe et al., 2005] Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2):165–210.
- [Wiebe, 1994] Wiebe, J. M. (1994). Tracking point of view in narrative. *arXiv preprint cmp-lg/9407019*.
- [Wikipedia contributors, 2022] Wikipedia contributors (2022). Recurrent neural network — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Recurrent_neural_network&oldid=1109264340. [Online; accessed 26-September-2022].
- [Yu and Hatzivassiloglou, 2003] Yu, H. and Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136.