THE ROLE OF USER POLARITY NOVELTY ON SARCASM DETECTION

Propuesta de Investigación

Maestría en Analítica para la Inteligencia de Negocios

JAIME ANDRÉS VARGAS CRUZ

DIRECTOR

JORGE ANDRÉS ALVARADO VALENCIA

Doctor en Ingeniería

PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE INGENIERÍA

BOGOTÁ

2018

Contenido

Resumen Ejecutivo

El propósito principal de esta investigación fue proponer un método innovador para la detección de sarcasmo, basado en la historia de la polaridad del usuario y técnicas de detección de anomalías. Se trabajó con información de Twitter y se definió una entidad principal para el caso de estudio, respecto a la cual se analizaron los casos de sarcasmo. Adicional a esto, se propuso hacer uso de entidades relacionadas a la entidad principal, con lo que fue posible obtener información histórica, incluso cuando la entidad principal no había sido mencionada previamente.

Posteriormente, se compararon diferentes técnicas para la detección de anomalías, lo cual permitió identificar la más adecuada. Durante este mismo proceso, se propuso una modificación a los datos de entrenamiento que permitió mejorar la calidad del algoritmo.

Finalmente se concluye que el uso de técnicas de detección de anomalías en los sentimientos históricos expresados por el usuario, permite una mejor detección del sarcasmo, que otros métodos que excluyan esta información o que usen aproximaciones más sencillas para trabajar con la información histórica. Otra conclusión importante fue que las metodologías basadas en analizar las características del texto, identifican casos diferentes a los casos identificados por medio de la polaridad histórica, por lo que ambos son enfoques complementarios, permitiendo abordar el problema de una manera más completa.

Introducción

El análisis de sentimiento es un área del Procesamiento del Lenguaje Natural (PLN) que en los últimos años ha atraído gran atención, en parte por los múltiples avances tecnológicos que han permitido implementar diferentes tipos de análisis en esta área, pero también, por la expansión de la web 2.0, lo cual ha generado un aumento en la cantidad de información disponible para analizar [1]. Otro factor importante ha sido que las empresas están cada vez más interesadas en conocer la opinión de sus clientes y en entender cómo sus marcas u organizaciones son percibidas por estos mismos [2], ya que además de permitirles a las empresas comprender mejor el mercado, los potenciales clientes también basan sus decisiones en la información presente en internet, sean comentarios o reseñas [3], [4].

El sarcasmo, al igual que otras formas del lenguaje figurado, ha representado un fuerte e importante desafío en el área de estudio del Procesamiento del Lenguaje Natural, debido a su importancia en el lenguaje no formal [5], su efecto sobre el discurso y la dificultad inherente para su detección. Por ello, en los últimos años múltiples investigadores han buscado diferentes maneras de solucionar el problema de su detección.

Para el caso del análisis de sentimientos, el efecto del sarcasmo es tal que con su uso se niega o invierte el sentido de su polaridad[6]. Por esto mismo, ignorar este tipo de lenguajes tiene la capacidad de perjudicar significativamente el desempeño de un sistema de análisis de sentimiento[7].

A pesar de su efecto en el desempeño, en muchos casos el sarcasmo no es considerado en los modelos de análisis de sentimiento, por la dificultad asociada a su detección [7]. Esto se debe en gran parte a que el sarcasmo, en muchos casos, requiere de una gran sutileza para ser detectado [8]. Soportando esta idea, en el 2014, Wallace *et al.* [9] comprobaron que los seres humanos no

somos buenos detectando el sarcasmo escrito cuando no tenemos a nuestra disposición el contexto de una opinión, por lo que aproximaciones algorítmicas que ignoren este elemento podrían tener inconvenientes para detectarlo, dado que se encuentran sujetas a estas mismas limitaciones. Tal es su dificultad, que en la cultura popular se habla de la 'ley de Poe' [10], la cual se interpreta entendiendo que sin pistas explícitas de que alguien está siendo irónico o sarcástico en un mensaje escrito, no solo un mensaje sarcástico podrá ser interpretado como una opinión sincera, sino que a su vez, dicho mensaje será idéntico en contenido a opiniones no sarcásticas [11]; por lo que no solo hace falta conocer sobre el contexto del tema, sino también, el contexto de la persona que transmite su opinión.

Debido a lo anterior, se hace evidente que es de gran interés desarrollar algoritmos que tomen en cuenta el contexto y especialmente, la manera en la que piensa el interlocutor sobre determinados temas, para así poder detectar casos de sarcasmo que normalmente no serían identificados haciendo uso único del texto. En este trabajo investigativo se propone una nueva metodología que aprovecha este contexto mediante un enfoque basado en detección de anomalías en las opiniones de un individuo. Así mismo, se realiza un análisis que permite profundizar y entender cómo este enfoque puede ser complementado por otras teorías y algoritmos de detección de sarcasmo.

El documento que delinea la investigación que busca responder a este objetivo está dividido en las siguientes secciones: 1. Estado del arte, 2. Formulación del Problema, 3. Metodología, 4. Resultados, 5. Discusión, 6. Conclusiones y 7. Recomendaciones y Trabajo Futuro.

Estado del Arte

El sarcasmo, el cual es ampliamente utilizado en comunicaciones informales, es definido por el diccionario de Oxford como una manera de usar palabras que significan lo opuesto, buscando ser antipático o burlarse de alguien [12]. El sarcasmo también es llamado ironía verbal [13], [14] y según la teoría lingüística que se escoja, las características de éste y los procesos cognitivos subyacentes pueden variar.

Una de estas teorías se conoce como *Echoic Reminder* [14] la cual declara que el reconocimiento del sarcasmo depende de conocer un contexto previo. Otra teoría conocida como *Allusional Pretense* [15] requiere de dos criterios, que lo que se diga sea de forma insincera y que haga mención a una expectativa o norma social incumplida. Por otro lado, Creusere [16] plantea a partir de la investigación en infantes y su desarrollo, que el sarcasmo debe contener un significado literal positivo, un significado real negativo y una víctima clara.

Complementando la existencia de múltiples teorías en los campos de la lingüística y psicología, varios autores en el área del PLN han tratado el problema de la identificación de sarcasmo en diferentes tipos de textos, siendo twitter [17]–[22], uno de los más populares, aunque también se han realizado varios estudios en reseñas de productos [13], [23], [24] e incluso en otras redes como Facebook y Reddit [9]. Twitter ha adquirido su popularidad en PLN debido a que es una de las redes más populares, con más de 328 millones de usuarios activos por mes alrededor del mundo [25], y debido a que permite el acceso a una muestra de su amplio volumen de textos sin ningún costo [26].

Por la relevancia de la detección del sarcasmo en análisis de sentimiento, varios autores han propuesto múltiples métodos para abordar el problema de detección de sarcasmo. Kumar *et al.*

[27] en su revisión de técnicas para detección de sarcasmo, agrupan los métodos existentes en tres categorías: una basada en reglas, otra basada en aproximaciones estadísticas y una última basada en aprendizaje de máquina; entre las aproximaciones estadísticas mencionan una subcategoría, que incluye un único artículo, realizado por Khattri et. al [28], el cual hace uso de la información histórica del usuario. Los autores concluyen respecto a la revisión de la literatura, que la mayoría de enfoques se concentran en el texto y adicionalmente comentan que existe una tendencia hacia el uso de métodos basados en aprendizaje de máquina, lo cual se debe a que están superando los resultados obtenidos a partir de enfoques estadísticos.

Wicana *et al.* [29] dividen su clasificación en aprendizaje supervisado, semisupervisado, estructurado, basado en reglas y enfoques híbridos, en el cual incluyen los populares *Word Embeddings*; adicional a esto, los autores hacen mención de algunos artículos [17], [22], [30] que utilizan información contextual para generar sus predicciones, los cuales profundizaremos en los siguientes párrafos. Por otro lado, en la revisión realizada por Dave y Desai [31] no se realiza mención alguna de un enfoque basado en el usuario o en el histórico, y dividen los diferentes tipos de enfoques en análisis léxicos, técnicas difusas, negaciones de hechos y en extracción de conocimiento temporal.

Unos de los autores que han incluido el uso de contexto e historia para la detección de sarcasmo son Bamman y Smith [17]. En este trabajo los autores proponen usar un clasificador binario basado en regresión logística regularizada, la cual hace uso de un alto número de características, algunas perteneciendo al tuit que se analiza, otras a la audiencia, y otras relativas al autor. Respecto al autor se incluyen las 100 palabras con el mayor TF-IDF[32], agrupaciones de palabras según LDA[33], información extraída del perfil y el sentimiento del histórico. Para esta última característica hacen uso de una de las propuestas de Rajadesingan *et al.*[21], quienes

incluyen el tuit inmediatamente anterior al que se está calificando, como una de las características a usar en sus clasificadores.

En este trabajo realizado por Rajadesingan *et al.* [21], los autores buscaron crear un *framework* comportamental para detectar el sarcasmo, para lo cual generaron características a partir de múltiples teorías comportamentales sobre el sarcasmo, incluyendo la teoría del *Echoic Reminder* [14] y la teoría de *Allusional Pretense* [15], así como otras que buscan explicar el uso del sarcasmo en los seres humanos. Este *framework* hace uso de un alto número de características, incluso algunas relacionadas con la historia y el contexto del usuario, como el porcentaje de transiciones realizadas entre tuits positivos y negativos, la cantidad de tuits sarcásticos en el pasado y la frecuencia de uso de palabras positivas y negativas en los últimos tuits.

Otra aproximación bastante innovadora, pero enfocada en aprendizaje profundo, es la propuesta de Amir *et al.* [30] quienes hacen uso de redes neuronales convolucionales y de *user embeddings* para detectar el sarcasmo. En este artículo se comparan con Bamman y Smith [17], alcanzando un resultado similar al usar la misma base. El enfoque seguido por Amir *et al.* no hace uso de características explícitas del usuario ni de su historia, pero utiliza la totalidad de los textos del usuario para alimentar el *user embedding* y finalmente, un clasificador, lo cual implica que esta metodología hace uso de la historia del usuario, pero no de su polaridad histórica.

Por otro lado, el trabajo realizado por Khattri y Joshi [28], enfocado en twitter, hace uso de una arquitectura que integra dos elementos, el primero de ellos es el contraste de palabras en el tuit, es decir, si el tuit contiene palabras con diferentes polaridades; y el segundo, un predictor basado en el histórico. Este segundo enfoque detecta frases que contengan nombres propios, y si es el caso, clasifica la frase según su polaridad mediante un enfoque basado en lexicón y reglas semánticas,

tales como manejo de negaciones y conjunciones. Luego de esto, asigna la polaridad del tuit a los nombres propios que se encuentran dentro de éste. Posteriormente, para calcular la polaridad histórica del nombre propio, se asigna la polaridad según una votación entre las polaridades asignadas a los nombres propios, de tal manera, que, si un tuit nuevo tiene una polaridad opuesta al histórico, se considerará como sarcasmo. Finalmente, para integrar los dos métodos, aplican una regla donde, si ambos enfoques predijeron el tuit como sarcástico, así es catalogado, pero si no hay información histórica, confía en los resultados del enfoque de contrastes.

Los autores mencionan algunas dificultades con el método empleado, uno de ellos es que, si el primer tuit es sarcástico, el enfoque histórico puede tener problemas al predecir nuevos tuits, así mismo mencionan la dificultad de conseguir la *timeline* de algunos usuarios que pueden tener sus tuits privados. Un inconveniente que los autores no mencionan, pero que podría ser evidente, es la dificultad para capturar la polaridad sobre una entidad siguiendo el enfoque de polaridad por tuit, ya que la polaridad del tuit y la polaridad de las entidades dentro de éste no siempre concuerdan, y las entidades dentro de un texto pueden tener polaridades opuestas. Profundizando en esta idea, Jiang *et al.* [34] encontraron que aproximadamente el 40% de los errores en análisis de sentimiento, donde el objetivo es una entidad, se deben a no considerar explícitamente a la entidad durante el análisis, tal como ocurre en el trabajo de Khattri y Joshi [28].

Algo similar a lo realizado por Khattri y Joshi es realizado por Wallace *et al.* [35] quienes también usan los nombres propios detectados mediante etiquetado gramatical [36] y las polaridades de los textos asociados. A diferencia del trabajo anteriormente mencionado, al enfocarse en *Reddit*, un foro online que se divide en diferentes comunidades temáticas, no usa las polaridades históricas del usuario como características, sino la polaridad histórica de la

comunidad a la que pertenece. En la tabla 1 se puede observar una tabla que compara los artículos mencionados.

Artículo	**Clasificación según Algoritmo**	**Características**	**Precisión**	**Exhaustividad**	**F Score**
Khattri y Joshi (2015)	Reglas	Sentimiento del mensaje, Sentimiento por Entidades del autor	88,0%	81,0%	84,4%
Wallace et al. (2015)	Modelo Estadístico	Sentimiento Comunidad, Sentimiento del texto	14,1%	37,7%	20,5%
Amir et al. (2016)	Aprendizaje de Máquina	Textos del autor (*User Embedding*), texto del mensaje	87,2%	N.R	N.R
Bamman y Smith (2015)	Modelo Estadístico	Texto, sentimientos y tópicos del mensaje, del autor y de la audiencia	85,1%	N.R	N.R
Rajadesingan et al. (2015)	Aprendizaje de Máquina	Texto, sentimientos y tópicos del mensaje, del autor y de la audiencia	83,4%	N.R	N.R

Tabla 1. Cuadro Comparativo entre metodologías contextuales

A partir de todo lo anterior, podemos observar que la detección de sarcasmo es un campo en el cual hay bastante interés y un alto número de desarrollos recientes. A pesar de esto, muy pocos investigadores [17], [22], [30] han trabajado en la inclusión de información histórica del usuario y de su polaridad, así se haya demostrado que estas características permiten mejorar la calidad de los pronósticos, por lo que en este trabajo se buscó mejorar las aproximaciones existentes.

Formulación del Problema

Pregunta Principal

¿La presencia de sentimientos atípicos en los usuarios de twitter permite mejorar o complementar la detección de sarcasmos?

Preguntas Secundarias

¿Qué tan buen indicador del sarcasmo es la polaridad histórica del usuario?

¿Qué algoritmo se adecúa mejor para la detección de anomalías en polaridad?

¿El uso del histórico permite detectar casos no detectados por metodologías que ignoren el comportamiento histórico del usuario?

Metodología

1. Generalidades

La idea detrás de la metodología a implementar se centra en la hipótesis de que una adecuada manera de detectar si un texto es sarcástico, es conocer previamente la opinión de la persona sobre la entidad a la que hace alusión su mensaje [21], [28]. En otras palabras, es posible detectar el sarcasmo en la medida en que se tiene un contexto sobre la persona que lo emite. Por ejemplo, si conocemos que una persona es muy religiosa, probablemente no hablará positivamente del aborto, o que, si esta es vegetariana, probablemente no tendrá opiniones positivas sobre la tauromaquia. Como es de esperarse y como ocurre cuando nos referimos al comportamiento humano, no hay reglas absolutas, pero sí patrones que pueden servir como fuente de información. Por ello en este proyecto de investigación se decidió hacer uso de las opiniones pasadas, basándonos en la medida de su polaridad, para detectar el sarcasmo.

El trabajo se centró en detectar el sarcasmo en una entidad específica para el idioma inglés, esta decisión respecto al idioma se tomó debido al alto número de desarrollos existentes para esta lengua. Por otro lado, y buscando facilitar la generación de las entidades relacionadas, se decidió trabajar con una única entidad, a pesar de que en caso de que fuese necesario, se podrían escoger múltiples. La entidad escogida para el caso de estudio fue el presidente de los Estados Unidos de América, Donald Trump, el cuál es una entidad que genera fuertes polaridades y que tuvo bastante relevancia en las redes sociales durante el periodo de estudio.

En la primera etapa se definieron y generaron una serie de entidades relacionadas a la entidad principal. En la segunda etapa se procedió a realizar la Extracción de Información y Construcción del Conjunto de Datos. En la tercera etapa se procedió a identificar las entidades y a calcular las polaridades de los tuits relevantes. En la cuarta etapa, se calculó el puntaje de cercanía a la

entidad principal y se dividieron los datos Entrenamiento y Validación. En la quinta etapa se implementaron los algoritmos de detección de atípicos, y en la siguiente se calcularon las métricas. Finalmente, se comparó la metodología propuesta con otras metodologías.

En la figura 1 se puede observar el diagrama de flujo que ilustra los pasos seguidos para este proyecto.

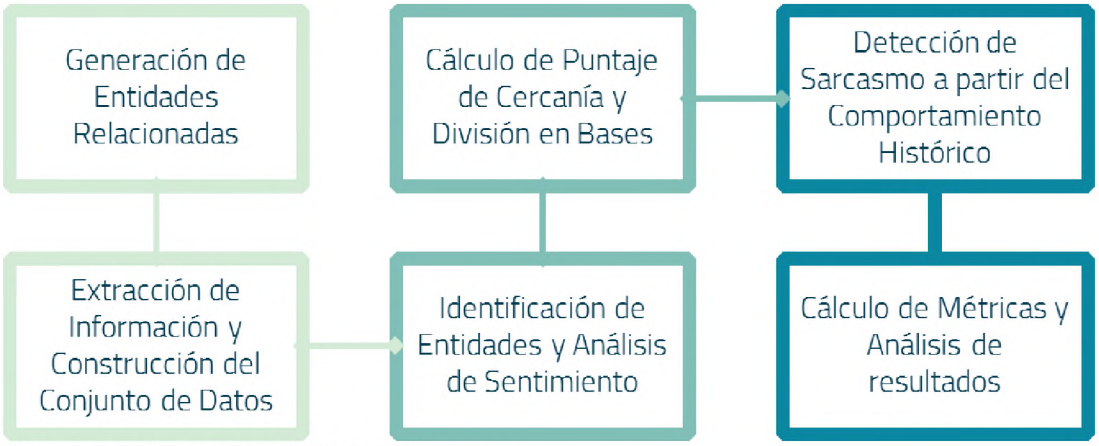

Figura 1. Diagrama del Flujo del Proyecto

Este proceso, si fuera a ser usado en un proyecto aplicado, requeriría de algunas variaciones en los pasos a seguir. Inicialmente, la generación de entidades relacionadas se debería realizar para todas las entidades de interés, o en su defecto, no relacionarlas y trabajar de manera individual con cada una de las entidades.

La segunda etapa podría ser muy similar, aunque cambiaría según las necesidades del proyecto, el origen de los datos y la necesidad de tener un *gold standard*. La tercera etapa y la cuarta etapa, relacionada con la identificación de entidades, análisis de sentimiento, cálculo de los puntajes de cercanía y la división de bases, se podría realizar de la misma manera propuesta.

Posteriormente, se realizaría la detección de sarcasmo, la cual, a diferencia de lo realizado en este proyecto investigativo, sólo requeriría usar el algoritmo de detección de atípico escogido como el

mejor. Finalmente, la metodología propuesta se podría usar como parte de un ensamble junto a un método que haga uso de las características dentro del texto.

2. Generación de Entidades Relacionadas

El enfoque usado se basó en la generación manual y supervisada de entidades relacionadas. Para la definición de estas entidades y su relación con Trump, se hizo una búsqueda en *google*, y se analizaron los primeros 5 resultados obtenidos. Las ecuaciones de búsqueda fueron *"Trump biggest critics"*, *"Trump allies and enemies"* y *"Trump allies in congress"*. De esta búsqueda se encontraron 58 entidades, 33 a favor de Trump, 25 en contra: esta clasificación se hizo con ayuda del contenido de los artículos escogidos. Una lista de las entidades elegidas se puede observar en el anexo 1. Inicialmente se buscó listas que fueran fruto de investigaciones, pero no se encontró ninguna que hiciera referencia directa a la entidad en estudio. Otra alternativa considerada y descartada, fue usar una lista con la totalidad de senadores y figuras importantes del partido republicano y demócrata, pero debido a que Donald Trump no mantiene buenas relaciones con todos los miembros de su partido [37], esta camino no fue viable y se requirió construir una lista según la metodología propuesta.

Para futuras investigaciones centradas en entidades diferentes, se podría generar esta lista de entidades relacionadas a partir de búsquedas similares o haciendo uso de otros métodos, tales como el uso de opiniones de expertos o inferir las relaciones a partir de estudios publicados y/o encuestas que permitan conocer la opinión del público respecto a diferentes entidades. Lo anterior, dependiendo de la disponibilidad de información sobre la entidad.

3. Extracción de Información y Construcción del Conjunto de Datos

Para realizar la extracción de la información, el primer paso fue construir una base de usuarios relevantes para el tema en cuestión. Para esto se realizó una búsqueda de tuits, mediante el buscador de Twitter, que contuviese simultáneamente los términos #sarcasm y Trump, esta búsqueda se realizó durante el periodo comprendido entre el primero de octubre de 2017 y el 23 de agosto de 2017.

Luego de descargar estos tuits, se empezaron a registrar en una lista, las cuentas de usuario de los autores de éstos. Se definió un criterio de parada que garantizara un número no muy alto de usuarios, siendo acorde con las limitaciones de las herramientas a ser usadas en etapas posteriores, especialmente la API de Google en la cual se profundiza en la siguiente sección. Por ello, se detuvo la selección de nuevos usuarios en el momento en que, luego de analizar 5 tuits, no aparecieran nuevos usuarios para registrar, es decir, que los tuits en revisión tuvieran autores previamente registrados en la lista. A partir de este proceso, fueron seleccionados un total de 407 usuarios.

Los usuarios fueron seleccionados de esta manera ya que se buscó que por lo menos el usuario tuviera un tuit sarcástico relacionado con el presidente Trump, y adicionalmente se supuso que esto podría ser un buen predictor de la preferencia por el uso de sarcasmo en política, lo cual nos permitiría tener más tuits adecuados para el análisis.

La extracción se realizó mediante la API Rest de Twitter [38] De esta descarga se tuvo un total de 936.239 tuits para analizar, pertenecientes a 407 usuarios, junto a sus descripciones de perfil. Es importante recordar que, a pesar del criterio de selección, no hay garantía de que todo usuario

tendrá por lo menos un uso de sarcasmo en sus tuits, debido a que la API Rest no garantiza la descarga de todos los tuits publicados por el usuario. [26]

Luego de esto, se realizó una limpieza de los tuits. Para ello se hizo uso de expresiones regulares que permitieron eliminar algunos elementos tales como URLs y "RT @User", las cuales se incluyen en el Anexo 2. Así mismo se eliminaron aquellos tuits que, según Twitter, fueron escritos en un idioma diferente al inglés, o que tuvieran menos de 4 palabras, los cuales fueron criterios basados en lo propuesto por Rajadesingan *et al.* [21].

En esta etapa también se procedió a eliminar de los tuits, los hashtags #sarcasm y #sarcastic, y a su vez, anotando estos mismos tuits como sarcásticos. La razón para ello fue evitar que el algoritmo de polaridad tomara en cuenta el hashtag para definir la polaridad, y adicionalmente con esto se genera una base anotada que nos permite evaluar la calidad del algoritmo en detectar sarcasmo. El uso de estos hashtags para anotar la base ha sido utilizado por frecuencia por otros investigadores al trabajar en el tema de detección de sarcasmo [17], [39].

Luego de esto, se procedió a convertir las menciones de algunas cuentas como @realDonaldTrump por el nombre de la entidad (Donald Trump), al igual que el de otras entidades de interés para la investigación.

4. Identificación de Entidades y Análisis de polaridad

La siguiente etapa consistió inicialmente en la pre-identificación de entidades en los tuits, con lo que se buscó acotar el número de tuits a analizar en etapas posteriores. Esta pre-identificación se realizó mediante la búsqueda de coincidencias exactas entre la lista de entidades y los tuits; lo cual, a pesar de permitir cumplir el objetivo de filtrar la mayoría de tuits no relevantes y de ser más exhaustivo al capturar variaciones posibles de las entidades, incluso dentro de Hashtags,

tuvo la deficiencia de identificar algunas entidades de manera inadecuada, por ejemplo, podría encontrar la cadena de caracteres "trump" en un tuit que use el hashtag "#trumpets" (trompetas).

En esta parte del proceso se procedió a hacer uso de un algoritmo de análisis de sentimiento enfocado en una entidad [40]. Esta subclase de Análisis de Sentimiento fue requerida, ya que luego de algunas pruebas con algoritmos que usan como unidad de análisis el texto completo [41], [42] se determinó que un enfoque que no realizara el análisis sobre la entidad, tendría resultados inapropiados, que afectarían el algoritmo en etapas posteriores. Este resultado fue respaldado por un estudio realizado por Jiang *et al.* [34] en el cual encontraron que aproximadamente el 40% de los errores en Análisis de Sentimiento Enfocados en la Entidad se deben a no considerar explícitamente a la entidad durante el análisis.

A causa de esto se prosiguió a usar un algoritmo que cumpliera con estas características. A pesar del alto interés que ha tomado este sub-problema en la comunidad científica [34], [43], [44], no se encontró ninguna herramienta gratuita o código disponible que permitiera su implementación, por lo que para esta investigación se usó mediante Python, la API de Google Cloud Platform, específicamente la herramienta de Análisis de Sentimiento por Entidad, que ofrece en su módulo de Lenguaje Natural, la cual realiza detección de entidades nombradas y a su vez, *Targeted Entity Sentiment Analysis* [40]. El algoritmo ofrecido por Google tiene la ventaja de que se encarga por sí mismo del preprocesamiento del texto y por lo tanto puede trabajar con los textos de los tuits sin procesar, aunque el uso de esta API influyó en la decisión de trabajar con una muestra no muy alta de usuarios. Por limitaciones de llamados máximos a la API, se redujo el tamaño de la base a 167.000 tuits y a 350 usuarios.

Un paso adicional fue requerido debido a que, por el tipo de análisis realizado, se obtuvo más de una polaridad por tuit en aquellos tuits con más de una entidad. El criterio para reducir nuevamente a una polaridad por tuit, fue dejar aquel que hiciera referencia a Donald Trump, y en caso de que no se pudiera usar este criterio, se seleccionó aquel que hiciera referencia a una entidad preseleccionada y que tuviera la polaridad, en valor absoluto, más alta.

5. Cálculo de Puntaje de Cercanía y División en Bases

En esta etapa se generó un Puntaje de cercanía para cada usuario, que busca representar qué tan acorde está con la entidad principal (Trump). Para esto se hizo uso de la lista de entidades, la cual estaba anotada con un valor que representa la cercanía, con 1 para un aliado, -1 para una entidad opuesta y 0 para el resto, y representado en la fórmula 1 como $C(e_i)$. En la ecuación 1 se puede observar el cálculo usado para inferir la cercanía a la entidad principal, el subíndice i hace referencia una de las entidades encontradas y el subíndice j al tuit, y $Pol_{i,j}$ es el resultado obtenido por Google Cloud Platform y que puede tomar valores entre -1 y 1.

$$\boldsymbol{P_{i,j} = C(e_i) * Pol_{i,j}}$$

Ecuación 1.

Una vez calculados los puntajes individuales para cada entidad presente en los tuits, se realizó una partición de la base en Entrenamiento y Validación. Los tuits en la base de entrenamiento fueron usados para conocer a los usuarios, y los tuits en la base de validación fueron usados para realizar predicciones sobre el uso de sarcasmo. Esta división requirió de un trato especial, ya que se debía garantizar que cada usuario mantuviera, de ser posible, la proporción 70/30.

La base de entrenamiento obtuvo cerca al 70% de los tuits (52296) de todos los usuarios, excluyendo aquellos tuits marcados como sarcasmo, y como se mencionó previamente, con el

objetivo de identificar cuál era el comportamiento esperado para el usuario. De esta base se excluyeron aquellos tuits que tuvieran una polaridad con un valor igual a 0. Esta decisión se tomó debido a que luego de analizar los resultados del algoritmo de Google, se evidenció que era común ver este valor cuando el usuario no daba su opinión sobre una entidad, sino que simplemente la mencionaba. Siguiendo esta línea de pensamiento, si un usuario menciona una entidad, pero no da un juicio de valor sobre ésta, no podemos concluir si está a favor o en contra. Por ejemplo, decir que Trump está de visita en china, no nos dice si el usuario aprueba o desaprueba su gestión o la acción en particular y, por lo tanto, este elemento no será de utilidad para conocer la opinión de la persona sobre la entidad, sino que nos inclinará a creer que está dando una opinión "neutral", cuando en realidad no está dando ninguna opinión. Con este filtro, se eliminó más de la mitad de los tuits, dejando sólo 20529 como insumo para conocer a los usuarios.

Adicionalmente, si un usuario no contenía por lo menos dos tuits válidos en la base de entrenamiento, se eliminaba de las dos bases, ya que se consideró que no se tenía suficiente información para "conocer" al usuario. Debido a este proceso se pasó de 350 usuarios a 336.

6. Detección de Sarcasmo a partir del Comportamiento Histórico

En esta etapa se procedió a implementar diferentes metodologías para la detección de sarcasmo, las cuales fueron ejecutadas sobre la base de validación. En esta etapa se probaron 4 diferentes algoritmos de detección de atípicos: z-score, rango intercuartil, *double median absolute deviation* y votación. Los tres primeros, escogidos por su uso en la detección de atípicos univariados [45], y el último, la votación, para realizar una comparación con el método usado por Kathri *et al.* para predecir el sarcasmo [28]. Adicionalmente, no se implementaron algoritmos más complejos

debido a la restricción de realizar el análisis por usuario, por lo que en la mayoría de casos no se tenía suficientes datos para aprovechar técnicas basadas en aprendizaje semisupervisado.

Para esto fue requerido el cálculo, por usuario, de estadísticos basados en $P_{i,j}$, tales como el promedio, la desviación estándar, cuartiles y otros insumos requeridos para implementar las técnicas previamente mencionadas, incluyendo el método de votación, el cual requiere conocer la polaridad predominante del usuario, y considerará como sarcasmo todo aquel tuit que no se encuentre en ese lado del umbral.

Las técnicas de *z-score* y *double median absolute deviation (dMad)* [45], [46] requirieron como parámetro el número de desviaciones estándar necesarias para que un valor se considere atípico. Este número se definió haciendo uso del teorema de Chebyshev, ampliamente utilizado en la detección de datos atípicos [47], y que nos permite deducir, que sin importar la distribución, por lo menos el 89% de los datos se encontraran a menos de 3 desviaciones estándar.

Por ello, para el caso del z-score, si el puntaje de cercanía se encontraba por fuera del rango delimitado por las tres desviaciones estándar a partir de la media del usuario, el tuit se consideró como sarcástico. Para el caso de dMad, también se calcularon umbrales superiores e inferiores, aunque a diferencia del z-score, dMad no se basa en la media, sino en la mediana, y adicionalmente presenta robustez frente a distribuciones asimétricas debido a que toma en cuenta la asimetría de la distribución al calcular los umbrales.

Para el caso de detección mediante rango intercuartil, se procedió a utilizar el método de Tukey, también conocido como BoxPlot [45]. Para hacer uso de este método es necesario calcular el rango intercuartil (IQR), que es la distancia entre el tercer y primer cuartil. Posteriormente, se

considera como dato atípico todo lo que se encuentre por encima de 1.5*IQR + Q3, o por debajo de Q1-1.5*IQR.

Adicionalmente, se consideró la hipótesis de que polaridades cercanas a cero son más propensas a haber sido calculadas inadecuadamente o a que estas hacen referencia a tuits que no conllevan una opinión. Para esto, se definieron como neutros aquellos tuits con valores entre -0.3 y 0.3, lo cual llevó a trabajar con el 55% de la base, dejando un total de 11.411 tuits para el entrenamiento. El corte se definió en 0.3 para que los rangos de las tres categorías, positivo, neutro y negativo, tuvieran un tamaño similar. Finalmente, se volvieron a ejecutar los algoritmos, pero ignorando para el entrenamiento aquellos tuits que se definieron como neutrales.

7. Métricas y Validación

Con el objetivo de conocer el desempeño de cada una de las técnicas y conocer cuáles fueron las mejores, se procedió a realizar esta etapa donde se calculó la precisión, la exhaustividad y el *f1 score*. Para ello, como se había explicado con anterioridad, se consideró que un tuit era sarcástico si en su texto original contenía #sarcasm o #sarcastic

A continuación, y haciendo uso de las etiquetas de sarcasmo provenientes de los hashtags, se procedió a evaluar los resultados de las diferentes técnicas de detección de atípicos mediante las métricas de precisión, exhaustividad y *f1 score*. Debido a las intenciones del proyecto, en el cual se busca que esta metodología sea utilizable en conjunto con otras técnicas que aprovechen otras características del tuit, tales como la discordancia entre palabras [39] o los *emojis* usados [48], se prefiere a la precisión sobre la exhaustividad.

8. Análisis de Detecciones

Debido al interés de conocer si el algoritmo planteado puede detectar casos de sarcasmo que no serían identificados por otras metodologías, se realizó una comparación con el algoritmo desarrollado por Mathieu Cliche de Cornell University [49], el cual entrenó un SVM [50] a partir de características extraídas del texto, tales como las palabras, el tema y la polaridad de diferentes partes del tuit.

Esta comparación también se realizó con el método de contraste explícito de polaridad, basado en lo realizado por Kattri *et. al* [28] y por Joshi *et. al* [8]. Este método consiste en hacer uso de diferencias en la polaridad entre las palabras de un mismo tuit para definir si éste es sarcástico.

En la sección siguiente se hace un recuento de los resultados obtenidos al seguir la metodología planteada.

Resultados

1. Consideraciones Iniciales

Al finalizar el proceso se midió el desempeño de los algoritmos planteados para predecir el uso de sarcasmo. La base de validación tiene un total de 21.374 tuits, de los cuales tan solo 295 son sarcasmos, lo cual conlleva a que un pronóstico ingenuo tenga una precisión de 1.38%

En esta sección se usará el asterisco para denotar que el entrenamiento del algoritmo se hizo sin usar los casos denominados como neutrales, de tal manera que IQR* sólo se diferencia de IQR por la base que usó para conocer al individuo.

2. Comparación entre Metodologías

Las medidas de desempeño para la clasificación de los 21.374 tuits presentes en la base de validación se pueden observar en la tabla 2 y tabla 3. La tabla 1 hace énfasis en la clase principal, mientras la tabla 3 muestra los resultados para textos no sarcásticos.

	Z-Score*	Z Score	IQR	IQR*	Votación	dMad	dMad*	Contrast	Cliche
Precisión (S)	**19,1%**	**11,4%**	5,7%	4,7%	2,2%	1,9%	1,8%	1,2%	2,0%
Exhaustividad (S)	16,4%	3,6%	9,2%	15,0%	17,3%	5,9%	16,5%	**20,3%**	**60,7%**
F1 Score (S)	**17,6%**	5,4%	7,0%	**7,2%**	4,0%	2,8%	3,3%	2,3%	3,8%

Tabla 2. Resultados predicción de Sarcasmo

	Z-Score*	Z Score	IQR	IQR*	Votación	dMad	dMad*	Contrast	Cliche
Precisión (NS)	98,9%	98,7%	98,7%	98,8%	98,7%	98,9%	99,0%	98,6%	99,1%
Exhaustividad (NS)	99,1%	99,6%	97,9%	95,8%	89,4%	96,5%	90,2%	77,3%	58,1%
F1 Score (NS)	99,0%	99,2%	98,3%	97,3%	93,8%	97,7%	94,4%	86,7%	73,2%

Tabla 3. Resultados predicción de No Sarcasmo

A partir de las tablas anteriores, se puede observar que los mejores resultados fueron los obtenidos por el algoritmo basado en el z-score, pero entrenado sin usar tuits neutrales. Por otro

lado, se puede observar que el algoritmo de Cliche y el de contrastes tienen alta exhaustividad, pero son propensos a falsos positivos.

Respecto a los resultados obtenidos por el código desarrollado por Cliche, es interesante observar que estos tienen un *f1 score* del 3%, mientras los resultados expuestos en su trabajo tienen un valor del 60%. Esta gran diferencia entre los dos resultados nos podría sugerir que la base generada durante este proyecto es especialmente compleja de analizar, lo cual se podría deber a un alto número de entidades "opuestas" mencionadas en un mismo tuit, lo cual podría disminuir la utilidad de usar contrastes para detectar el sarcasmo. Algo similar ocurre al utilizar el método de contraste y el de votación, que obtienen resultados muy lejanos a lo que se ve en el artículo desarrollado por Khattri et al.[28].

Por lo que, a pesar de los no muy altos valores alcanzados con la propuesta realizada, los resultados son alentadores ya que la detección de sarcasmo es un problema que aún se encuentra sin resolver, y a que se superan los resultados de los algoritmos escogidos como base de comparación.

También es interesante resaltar que la mayoría de los algoritmos para detección de atípicos tuvieron un mejor resultado que el método basado en una votación simple. Por lo que la aproximación realizada por Khattri [28] se podría beneficiar de procesar el histórico de una manera diferente.

3. Casos detectados

El objetivo de este ejercicio fue observar si los casos de sarcasmo que se detectan con la detección de atípicos, son los mismos que se detectan por metodologías basadas en el texto o si, por el contrario, se complementan. Por lo que el análisis en esta sección se centrará en aquellos

tuits que realmente son sarcásticos, y se dejará de lado el resto de la base. Para este ejercicio la comparación se realizó entre el mejor de los algoritmos basados en detección de atípicos en la polaridad, Z-Score*, contra el algoritmo desarrollado por Cliche y el método basado en contrastes.

En la tabla 4 se puede observar los casos detectados por Z-Score* y aquellos detectados por el algoritmo desarrollado por Cliche. En la tabla 5, la comparación por casos entre el Z-Score* y Contraste; y en la tabla 6, una comparación entre Cliche y el método de Contraste.

	Z-Score*: Incorrecto	**Z-Score*: Correcto**	**Total**
Cliche: Incorrecto	94	23	117
Cliche: Correcto	156	22	178
Total	250	45	295

Tabla 4. Comparación por casos entre Z-Score* y Cliche

	Z-Score*: Incorrecto	**Z-Score*: Correcto**	**Total**
Contraste: Incorrecto	201	34	235
Contraste: Correcto	49	11	60
Total	250	45	295

Tabla 5. Comparación por casos entre Z-Score* y Contraste

	Cliche: Incorrecto	**Cliche: Correcto**	**Total**
Contraste: Incorrecto	99	136	235
Contraste: Correcto	18	42	60
Total	117	178	295

Tabla 6. Comparación por casos entre Cliche y Contraste

En la tabla 4 se puede observar que a pesar de que ambos algoritmos detectan en simultaneo 22 de los casos, más de la mitad de los casos identificados por el algoritmo de detección de atípicos en el histórico no fueron detectados por el algoritmo desarrollado por Cliche. De manera similar, en la tabla 5, se puede observar que la complementariedad es clara para el caso de contrastes, ya que solo en 11 de los 45 casos detectados por Z-Score* fueron detectados por la metodología de contrastes.

Por otro lado, en la tabla 6 podemos ver que el algoritmo de Cliche y el de contrastes no se complementan entre ellos y están detectando casos similares. El efecto es tal, que tan solo 18 de los 60 casos detectados por Contraste no son detectados por Cliche. Este resultado no sorprende ya que la aproximación de Cliche, basada en SVM, divide en tres secciones el tuit, e incluye cada una de estas polaridades como una de sus características.

Luego de revisar las detecciones correctas por parte del algoritmo de Cliche, el cual usa atributos ampliamente comunes en otros algoritmos de detección de sarcasmo, y las detecciones realizadas por el método de Contrastes, se encuentra que las detecciones ocurren en aquellos tuits que hacen uso de un sarcasmo explícito y que podrían ser identificados con facilidad, incluso en la ausencia de contexto. A continuación, se encuentra un ejemplo:

"Basically Trump is saying screw the environment. I'm sure that community is enjoying the fumes."

El cual hace uso de palabras con alta disparidad en sus polaridades, como el caso de "*enjoying*", un verbo asociado a algo gratificante, y, por otro lado, tenemos "fumes", un sustantivo con connotaciones negativas. Por otro lado, ejemplos como los siguientes son correctamente

identificados por la metodología propuesta, pero incorrectamente clasificados por el algoritmo desarrollado por Cliche y el de contrastes:

"Trump is a racist and black people hate him"

"@user Mmmm..... but Trump is the real enemy"

Con lo anterior se evidencia que una aproximación basada en la polaridad histórica es capaz de detectar casos que incluso un ser humano no podría detectar sin el contexto adecuado, y por lo tanto tampoco por algoritmos basados en aproximaciones que sólo usen el texto y no al usuario, ni a su contexto.

Discusión

Uno de los puntos claves que se trató en este proyecto investigativo fue la exploración de nuevas metodologías para hacer uso de la polaridad histórica del usuario. La inclusión de técnicas de detección de atípicos permitió aprovechar de mejor manera la información presente en la historia del usuario, sobre todo si lo comparamos con técnicas más sencillas como lo es una votación simple.

Adicionalmente, durante esta exploración, se llegó a un resultado inesperado, ya que las tres técnicas de detección de atípicos tuvieron un incremento importante en su f1-score y en su desempeño general al ignorar aquellos tuits con polaridades débiles. Una explicación a esto podría ser que, al eliminar los tuits con polaridades débiles, se excluyeron tuits que el algoritmo de polaridad no catalogó adecuadamente e incluso tuits factuales que no contenían una opinión, y por lo tanto no aportarían para conocer al usuario.

Siguiendo esta idea, otro elemento importante de esta metodología y uno de los aprendizajes del proyecto fue la determinar el algoritmo de polaridad adecuado para la tarea. En las pocas aproximaciones donde se había trabajado con la polaridad histórica, se hacía uso de la polaridad del tuit. Pero, como se descubrió durante este proyecto, si se desea obtener buenos resultados hace falta hacer uso de Análisis de Sentimiento Enfocado en una Entidad, ya que es común la presencia de múltiples entidades, con diferentes polaridades, en un mismo tuit.

Aunque lo anterior plantea una dicotomía interesante, ya que deseamos que el análisis de polaridad sea de la más alta calidad, debido a que un alto número de calificaciones incorrectas afectaría la distribución, y por lo tanto dificultando la capacidad de realizar predicciones certeras.

Pero, si el algoritmo de polaridad es lo suficientemente bueno como para detectar alguno o la totalidad de sarcasmos según su polaridad no literal, la metodología presentada podría tener inconvenientes.

Otro elemento interesante para analizar, fueron las diferencias entre las medidas de desempeño obtenidas y las encontradas en la literatura. Por un lado, a pesar de que Cliche ostentaba un F Score del 60%, al usar su código disponible en GitHub [51] se obtuvo un valor de 3.6% para esta medida. Esto sugiere que la base de tuits con la que se trabajó fue especialmente compleja, lo cual podría deberse a la entidad escogida y al hecho de incluir sarcasmos contextuales. Otro caso que soporta esta idea, fueron los malos resultados obtenidos al usar el método de votación y el método de contrastes, los cuales a pesar de no ser implementados de la misma manera que Khattri et al. [28], están muy lejos de los valores reportados por este.

Respecto al tema anterior, es importante tomar en cuenta que la detección de sarcasmos es una tarea especialmente compleja, y que, a pesar de que los resultados del algoritmo no presentaron una muy alta precisión, lo propuesto en este proyecto tiene el potencial de mejorar la detección de sarcasmo de una manera substancial. Esto se debe a que las propuestas de este proyecto abren la puerta para detectar casos de sarcasmo contextual, que, bajo la literatura actual, pasarían desapercibidos.

Adicionalmente, una propuesta interesante que se realizó en este proyecto fue la idea de trabajar con entidades relacionadas a la entidad principal, y no sólo con una entidad única. El beneficio para este proyecto fue contar con más información histórica de los usuarios, que a su vez permitió conocerlos mejor y poder predecir un sarcasmo a pesar de que la persona estuviera mencionando a la entidad principal por primera vez.

Finalmente, un resultado interesante, aunque esperado, fue mostrar que las técnicas basadas en texto (ya sean reglas o aprendizaje de máquina) son complementarias a enfoques basados en la historia del usuario. Lo cual a su vez resalta la importancia de aprovechar las múltiples fuentes de información que pueden estar disponibles.

Conclusiones

El uso de técnicas de detección de anomalías en los sentimientos históricos expresados por el usuario permite una mejor detección del sarcasmo, que al compararlo con métodos que excluyan esta información o que usen aproximaciones más sencillas.

El comportamiento histórico del usuario es un elemento importante y valioso para indicar el sarcasmo de un usuario, aunque debido a sus limitaciones, este no debería ser usado como único elemento, sino como una característica del modelo o como parte de un ensamble.

El algoritmo que tuvo el mejor desempeño para la detección de anomalías fue el tradicional z-score, el cual tuvo un desempeño superior al de los demás. Aunque los resultados mejoraron notablemente al excluir polaridades cercanas a cero a la hora de entrenar el algoritmo.

Las metodologías basadas en analizar las características del texto, detectan casos diferentes a los casos detectados por medio de la polaridad histórica, por lo tanto, ambos son enfoques complementarios que permiten abordar el problema de una manera más completa.

Recomendaciones y Trabajo Futuro

Para trabajos futuros un tema interesante para profundizar sería la generación automática de entidades relacionadas, lo cual facilitaría el uso de esta metodología e incluso permitiría manejar escalas y no solo valores enteros. Esta generación de listas se podría abordar desde varias perspectivas, incluyendo el uso de técnicas de detección de tópicos o implementar conceptos de teoría de grafos en la red social.

También sería valioso comparar los casos detectados por la metodología propuesta con lo casos detectados por algún algoritmo basado en *deep learning* o que haga uso de *user embeddings*, ya que permitiría conocer si estos algoritmos implícitamente están haciendo uso de esta información.

Finalmente, sería interesante profundizar en las razones por las cuales eliminar polaridades cercanas a cero conlleva a mejorar los resultados de la detección de anomalías. Para esto se podría comparar este efecto con el de filtrar mediante un clasificador que detecte opiniones, o incluso hacer uso de otro algoritmo de polaridad.

Referencias Bibliográficas

[1] B. Pang, L. Lee, y others, «Opinion mining and sentiment analysis», *Found. Trends® Inf. Retr.*, vol. 2, n.º 1–2, pp. 1–135, 2008.

[2] J. Zabin y A. Jefferies, «Social media monitoring and analysis: Generating consumer insights from online conversation», *Aberd. Group Benchmark Rep.*, vol. 37, n.º 9, 2008.

[3] J. A. Horrigan, «Online shopping», *Pew Internet Am. Life Proj. Rep.*, vol. 36, pp. 1–24, 2008.

[4] Z. Zhang, Q. Ye, R. Law, y Y. Li, «The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews», *Int. J. Hosp. Manag.*, vol. 29, n.º 4, pp. 694–700, 2010.

[5] R. W. Gibbs, «On the psycholinguistics of sarcasm.», *J. Exp. Psychol. Gen.*, vol. 115, n.º 1, p. 3, 1986.

[6] C. C. Liebrecht, F. A. Kunneman, y A. P. J. van Den Bosch, «The perfect solution for detecting sarcasm in tweets# not», 2013.

[7] D. Maynard y M. A. Greenwood, «Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis.», en *LREC*, 2014, pp. 4238–4243.

[8] A. Joshi, P. Bhattacharyya, y M. J. Carman, «Automatic sarcasm detection: A survey», *ArXiv Prepr. ArXiv160203426*, 2016.

[9] B. C. Wallace, L. Kertz, E. Charniak, y others, «Humans require context to infer ironic intent (so computers probably do, too)», en *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2014, vol. 2, pp. 512–516.

[10] N. Poe, «Poe's Law», 2005.

[11] S. F. Aikin, «Poe's law, group polarization, and the epistemology of online religious discourse», 2009.

[12] «sarcasm noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com». [En línea]. Disponible en: https://www.oxfordlearnersdictionaries.com/us/definition/english/sarcasm. [Accedido: 11-ene-2018].

[13] E. Filatova, «Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing.», en *LREC*, 2012, pp. 392–398.

[14] R. J. Kreuz y S. Glucksberg, «How to be sarcastic: The echoic reminder theory of verbal irony.», *J. Exp. Psychol. Gen.*, vol. 118, n.º 4, p. 374, 1989.

[15] S. Kumon-Nakamura, S. Glucksberg, y M. Brown, «How about another piece of pie: The allusional pretense theory of discourse irony.», *J. Exp. Psychol. Gen.*, vol. 124, n.º 1, p. 3, 1995.

[16] M. A. Creusere, «Theories of adults' understanding and use of irony and sarcasm: Applications to and evidence from research with children», *Dev. Rev.*, vol. 19, n.º 2, pp. 213–262, 1999.

[17] D. Bamman y N. A. Smith, «Contextualized Sarcasm Detection on Twitter.», en *ICWSM*, 2015, pp. 574–577.

[18] F. Barbieri, H. Saggion, y F. Ronzano, «Modelling Sarcasm in Twitter, a Novel Approach.», en *WASSA@ ACL*, 2014, pp. 50–58.

[19] R. González-Ibánez, S. Muresan, y N. Wacholder, «Identifying sarcasm in Twitter: a closer look», en *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2*, 2011, pp. 581–586.

[20] T. Ptáček, I. Habernal, y J. Hong, «Sarcasm detection on czech and english twitter», en *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 213–223.

[21] A. Rajadesingan, R. Zafarani, y H. Liu, «Sarcasm detection on twitter: A behavioral modeling approach», en *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 2015, pp. 97–106.

[22] Z. Wang, Z. Wu, R. Wang, y Y. Ren, «Twitter sarcasm detection exploiting a context-based model», en *International Conference on Web Information Systems Engineering*, 2015, pp. 77–91.

[23] D. Davidov, O. Tsur, y A. Rappoport, «Semi-supervised recognition of sarcastic sentences in twitter and amazon», en *Proceedings of the fourteenth conference on computational natural language learning*, 2010, pp. 107–116.

[24] O. Tsur, D. Davidov, y A. Rappoport, «ICWSM-A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews.», en *ICWSM*, 2010, pp. 162–169.

[25] * All products require an annual contract Prices do not include sales tax, «Twitter: number of active users 2010-2017», *Statista*. [En línea]. Disponible en: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/. [Accedido: 13-oct-2017].

[26] «Getting started — Twitter Developers». [En línea]. Disponible en: https://developer.twitter.com/en/docs/basics/getting-started. [Accedido: 13-oct-2017].

[27] L. Kumar, A. Somani, y P. Bhattacharyya, «Approaches for Computational Sarcasm Detection: A Survey».

[28] A. Khattri, A. Joshi, P. Bhattacharyya, y M. Carman, «Your sentiment precedes you: Using an author's historical tweets to predict sarcasm», en *Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis*, 2015, pp. 25–30.

[29] S. G. Wicana, T. Y. İbisoglu, y U. Yavanoglu, «A Review on Sarcasm Detection from Machine-Learning Perspective», en *Semantic Computing (ICSC), 2017 IEEE 11th International Conference on*, 2017, pp. 469–476.

[30] S. Amir, B. C. Wallace, H. Lyu, y P. C. M. J. Silva, «Modelling context with user embeddings for sarcasm detection in social media», *ArXiv Prepr. ArXiv160700976*, 2016.

[31] A. D. Dave y N. P. Desai, «A comprehensive study of classification techniques for sarcasm detection on textual data», en *Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on*, 2016, pp. 1985–1991.

[32] J. Ramos, «Using tf-idf to determine word relevance in document queries», en *Proceedings of the first instructional conference on machine learning*, 2003, vol. 242, pp. 133–142.

[33] D. M. Blei, A. Y. Ng, y M. I. Jordan, «Latent dirichlet allocation», *J. Mach. Learn. Res.*, vol. 3, n.º Jan, pp. 993–1022, 2003.

[34] L. Jiang, M. Yu, M. Zhou, X. Liu, y T. Zhao, «Target-dependent twitter sentiment classification», en *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 2011, pp. 151–160.

[35] B. C. Wallace, D. K. Choe, y E. Charniak, «Sparse, Contextually Informed Models for Irony Detection: Exploiting User Communities, Entities and Sentiment.», en *ACL (1)*, 2015, pp. 1035–1044.
[36] «The Stanford Natural Language Processing Group». [En línea]. Disponible en: https://nlp.stanford.edu/software/tagger.shtml. [Accedido: 11-ene-2018].
[37] «Where GOP senators stand on Trump - Washington Post». [En línea]. Disponible en: https://www.washingtonpost.com/graphics/politics/senate-trump-support/?utm_term=.0d22c9e56f69. [Accedido: 09-nov-2017].
[38] «Overview — Twitter Developers». [En línea]. Disponible en: https://developer.twitter.com/en/docs/tweets/timelines/overview. [Accedido: 10-dic-2017].
[39] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, y R. Huang, «Sarcasm as Contrast between a Positive Sentiment and Negative Situation.», en *EMNLP*, 2013, vol. 13, pp. 704–714.
[40] «API Natural Language de Cloud», *Google Cloud Platform*. [En línea]. Disponible en: https://cloud.google.com/natural-language/?hl=es. [Accedido: 11-dic-2017].
[41] R. Socher *et al.*, «Recursive deep models for semantic compositionality over a sentiment treebank», en *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1631–1642.
[42] C. J. Hutto y E. Gilbert, «Vader: A parsimonious rule-based model for sentiment analysis of social media text», en *Eighth international AAAI conference on weblogs and social media*, 2014.
[43] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, y K. Xu, «Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification.», en *ACL (2)*, 2014, pp. 49–54.
[44] D.-T. Vo y Y. Zhang, «Target-Dependent Twitter Sentiment Classification with Rich Automatic Features.», en *IJCAI*, 2015, pp. 1347–1353.
[45] S. Seo, «A review and comparison of methods for detecting outliers in univariate data sets», PhD Thesis, University of Pittsburgh, 2006.
[46] «Using the Median Absolute Deviation to Find Outliers». [En línea]. Disponible en: http://eurekastatistics.com/using-the-median-absolute-deviation-to-find-outliers/. [Accedido: 11-ene-2018].
[47] B. G. Amidan, T. A. Ferryman, y S. K. Cooley, «Data outlier detection using the Chebyshev theorem», en *Aerospace Conference, 2005 IEEE*, 2005, pp. 3814–3819.
[48] D. Thompson y R. Filik, «Sarcasm in written communication: Emoticons are efficient markers of intention», *J. Comput.-Mediat. Commun.*, vol. 21, n.º 2, pp. 105–120, 2016.
[49] «The sarcasm detector». [En línea]. Disponible en: http://www.thesarcasmdetector.com/about/. [Accedido: 12-dic-2017].
[50] T. Joachims, «Making large-scale SVM learning practical», Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, 1998.
[51] M. Cliche, «GitHub: Sarcasm detector», 19-ene-2018. [En línea]. Disponible en: https://github.com/MathieuCliche/Sarcasm_detector. [Accedido: 31-ene-2018].

Anexo 1

Lista de Entidades

Entidad	Relación	Entidad	Relación
Chris Collins	1	Affordable Care Act	-1
David Perdue	1	Al Franken	-1
Deb Fischer	1	Andrew Cuomo	-1
Donald	1	Barack	-1
Duncan Hunter	1	Bernie Sanders	-1
Fox and Friends	1	Clinton	-1
Fox News	1	Cory Booker	-1
GOP	1	Dan Malloy	-1
James M. Inhofe	1	Democratic Party	-1
Joe Manchin	1	Democrats	-1
John Barrasso	1	Elizabeth Warren	-1
John Boozman	1	Hillary	-1
John Cornyn	1	James B. Comey	-1
Joni Ernst	1	Joe Biden	-1
Kevin Brady	1	John Hickenlooper	-1
Kevin Cramer	1	Kamala Harris	-1
Lou Barletta	1	Kirsten Gillibrand	-1
MAGA	1	Martin O'Malley	-1
Make America Great Again	1	Obama	-1
Marsha Blackburn	1	Obamacare	-1
Mike Crapo	1	Rand Paul	-1
Mike Enzi	1	Terry McAuliffe	-1
Mike Lee	1		
Mitch McConnell	1		
potus	1		
Republicans	1		
Ron Johnson	1		
Roy Blunt	1		
Steve Daines	1		
Ted Cruz	1		
Tim Scott	1		
Tom Cotton	1		
Tom Marino	1		
Trump	1		

Anexo 2

Expresiones Regulares usadas durante el preprocesamiento:

Estas fueron las expresiones regulares usadas para eliminar o reemplazar elementos en los tuits.

- Eliminación de RT @user:

```
text = re.sub(r"^(RT @)\w*\:", '', text)
```

- Eliminación de hashtags de sarcasm:

```
text = re.sub(r"(?i)(#sarcasm)|(#sarcastic)", '', text)
```

- Reemplazo de URL por [url]

```
text = re.sub(r'\w+:\/{2}[\d\w-]+(\.[\d\w-]+)*(?:(?:\/[^\s/]*))*', '[url]', text)
```