
ANÁLISIS DE SENTIMIENTOS PARA TWEETS DE ELON MUSK

Carrillo Sánchez Ricardo
Facultad de Ingeniería, UNAM

García Miranda Athenas Marlene
Facultad de Ingeniería, UNAM

Rico López Kevin
Facultad de Ingeniería, UNAM

Sánchez Rojo Juan Pablo
Facultad de Ingeniería, UNAM

Torres Verastegui Jose Antonio
Facultad de Ingeniería, UNAM

31 de mayo de 2022

ABSTRACT

TextBlob, Afinn, Vader, etc. Son algunas de los métodos más populares para el análisis de sentimientos en textos, sin embargo, esta tarea tiene diferentes objetivos en la industria que hace que ciertos métodos resulten más eficientes que otros. En este texto, se abarca una investigación y análisis sobre Text Blob, que es una técnica basada en NLP y machine learning, sobre tweets hechos por Elon Musk.

1. Introducción

Para comenzar, es importante dar a conocer que el análisis de sentimientos se utiliza usualmente para conocer si nuestro documento representa algo positivo, neutral o negativo; sin embargo, también se puede hacer uso de estados particulares o, dicho de otra forma, sentimientos en específico. Aunado a lo anterior, en Twitter existe una enorme cantidad de tweets para poder analizar y conocer la orientación del tweet de manera más simplificada, tomando en consideración que, en las redes sociales particularmente, muchas personas suelen ser más emocionales al poder expresarse más libremente detrás de una pantalla.

De esta manera, para el presente trabajo se realiza un análisis de sentimientos de tweets para llegar a conocer si presentan una connotación positiva, negativa o neutral haciendo uso de TextBlob.

Análogamente, llevar a cabo esta implementación nos brinda la posibilidad de obtener una interpretación más certera acerca de lo que, para un documento particular como lo es un tweet, tiene mayor probabilidad de contener; en otras palabras, lograr para nuestro tweet, de acuerdo con Liu en 2012, un análisis de las opiniones, sentimientos, valoraciones, actitudes y emociones de las personas hacia determinadas entidades como productos, servicios, organizaciones, individuos, problemas, sucesos, temas y sus atributos.

Finalmente, en particular se lleva a cabo un análisis de los tweets de Elon Musk quien, debido a su popularidad en el mundo de la tecnología a nivel global y a su relevancia derivada de múltiples acciones entre las que destacan el intento de compra de la red social Twitter, resulta ser un candidato muy particular para el presente trabajo debido a que se desconoce con mayor certeza el impacto que está teniendo en los usuarios. De esta manera, con ayuda del API de Twitter y de la herramienta TextBlob, se examina el análisis de sentimientos para tweets (información subjetiva, es decir, que interfiere una interpretación) el impacto que ha tenido Musk como líder social en su área, así como su opinión.

El resto del artículo se organiza de la siguiente manera. En la sección del trabajo previo se hace mención acerca de la herramienta utilizada, por qué se seleccionó y el porqué se descartó otra alternativa. En la sección del Marco Teórico se habla en detalle del corpus usado para este experimento y de la metodología utilizada. En la sección experimental se presenta el experimento que se lleva a cabo, la obtención de resultados y la presentación de éstos. Por último, en la sección de conclusiones se analizan los resultados obtenidos y se propone un trabajo futuro para el que el presente sea de base.

2. Trabajo previo

2.1. AFINN

El primer acercamiento se realizó por medio de afinn que es un analizador de sentimientos que a partir de un diccionario de palabras con su significado, hace una reconstrucción de las oraciones que se le dan como entrada, otorgandonos un puntaje “score”, es decir, una asignación numérica, el cual, de acuerdo a ello, nos brinda un sentimiento “positivo, negativo o neutro”.

Tomando un dataset de tweets, seleccionados a partir de los trendings que ocasiona Apple se buscó analizar y comparar los resultados obtenidos para corroborar que afinn pueda ser un contendiente para analizar tweets. Lamentablemente en el dataset no se especifica una metodología para llevar a cabo el análisis por lo que se tomarán como una medición base.

La recopilación y el análisis previo de los sentimientos de los tweets son resultados bastante interesantes, en este documento notamos que se estableció como 1 a un sentimiento que expresa algo negativo mientras que mayor sea el score este se asocia a un mensaje positivo. Por otro lado, afinn establece un rango de $[-5, 5]$ donde a mayor número se expresa un sentimiento positivo, caso contrario se expresa un sentimiento negativo; el cero funcionará como bandera y como un indicador de un sentimiento que llamaremos neutro.

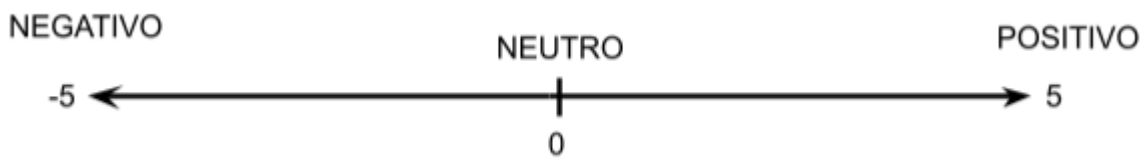


Figura 1: Escala de puntaje afinn

	tweet	affin_score	affin_veredict	csv_score
0	#AAPL:The 10 best Steve Jobs emails ever...htt...	3.0	POSITIVE	3
1	RT @JPDesloges: Why AAPL Stock Had a Mini-Flas...	-2.0	NEGATIVE	3
2	My cat only chews @apple cords. Such an #Apple...	0.0	NEUTRAL	3
3	I agree with @jimcramer that the #Individualln...	4.0	POSITIVE	3
4	Nobody expects the Spanish Inquisition #AAPL	-2.0	NEGATIVE	3
...
3881	(Via FC) Apple Is Warming Up To Social Media -...	0.0	NEUTRAL	3
3882	RT @MMLXIV: there is no avocado emoji may I as...	-1.0	NEGATIVE	3
3883	@marcbulandr I could not agree more. Between @...	4.0	POSITIVE	5
3884	My iPhone 5's photos are no longer downloading...	1.0	POSITIVE	1
3885	RT @SwiftKey: We're so excited to be named to ...	6.0	POSITIVE	5

Figura 2: Resumen de los resultados obtenidos del análisis de tweets

2.2. El problema de afinn

Algunos de los resultados que nos arroja a finn no son precisos en cuanto a lo que el usuario quería expresar. Tomando como ejemplo el tweet: “My iPhone 5’s photos are no longer downloading automatically to my laptop when I sync it.

@apple support is unhelpful. Any ideas?” el cual afinn lo describe como ligeramente positivo. Si analizamos lo que el usuario quería expresar hablamos de que este presentaba una falla en el sistema de sincronización y que al parecer no ha recibido ayuda del soporte técnico de Apple; por tanto, objetivamente no es factible bajo ningún contexto.

Las fallas de afinn pueden ser derivadas del “diccionario” con el que afinn reconstruye las oraciones y les agrega un peso ya que este está conformado de palabras en un contexto general donde sus significados pueden no coincidir con el tema al que estamos analizando. Esto podría tener una solución a partir de la creación de un lexicón con pesos basados en el objetivo que se trata de analizar; sin embargo, esto puede llegar a ser bastante ineficiente a causa de los cambios en el lenguaje que existen entre un tiempo determinado y sus regiones.

2.3. Textblob y un acercamiento de NLP

Un acercamiento basado en procesamiento de lenguaje natural y algoritmos de machine learning es la habilidad de crear y adaptar modelos para objetivos y contextos diversos. Por lo que la precisión en estos modelos no se basa específicamente en lo establecido en un diccionario sino que el contexto del texto cobra un mayor peso en los resultados.

2.4. Textblob

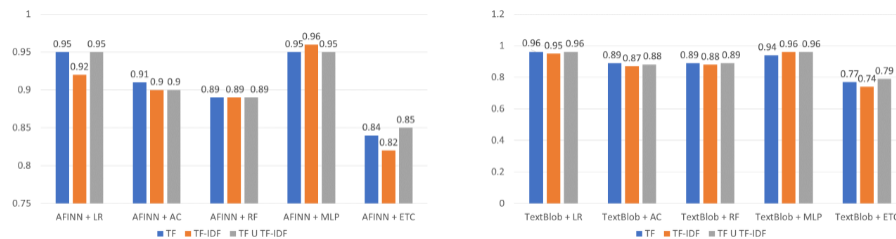


Figura 3: Comparación de la precisión en los resultados entre afinn y textblob en contexto médico

Los resultados anteriores¹, son parte del análisis de la eficacia de drogas en foros en internet, la gráfica de la izquierda corresponde a los resultados obtenidos con afinn mientras que los de la derecha utilizando textblob. El análisis es sencillo tomando en cuenta que el contexto médico es un contexto específico en el cual los modelos basados en NLP sacan ventaja sobre los basados en lexicon.

3. Marco Teórico

El procesamiento del lenguaje natural (NLP) surge como apoyo ante la necesidad que se tiene para el análisis de grandes cantidades de texto y lo atractivo que es desde hace años para múltiples empresas que buscan aprovechar dicha información. Además, debido al auge de internet, se tienen retos con gran dificultad, por lo que se busca hacerlo de manera automática, esto mediante técnicas que abarcan NLP.

De esta manera, el Análisis de sentimientos (AS) es un área del PLN cuyo objetivo es extraer una opinión, analizarla y determinar su inclinación, es decir, si es positiva, negativa o neutra. De acuerdo con Steinberger, Brychcín, y Konkol en 2014, casi todos los enfoques para el análisis de sentimientos lo realizan de manera general para una frase, un párrafo o un texto.

Por otro lado, Textblob es una librería de procesamiento de texto para Python. Ésta nos brinda la posibilidad de realizar tareas para PLN tales como el análisis morfológico, extracción de entidades, análisis de opinión, traducción automática, etc. Como parte de sus ventajas es que, mediante una interfaz más simple, nos permite hacer uso de dos librerías muy útiles y conocidas en el lenguaje Python, las cuales son NLTK y Pattern.

¹Saad, E., Din, S., Jamil, R., Rustam, F., Mehmood, A., Ashraf, I., & Choi, G. S. (2021). Determining the Efficiency of Drugs Under Special Conditions From Users' Reviews on Healthcare Web Forums. IEEE Access, 9, 85721–85737. <https://doi.org/10.1109/access.2021.3088838>

4. Experimental

4.1. Metodología

Para llevar a cabo la utilización de la herramienta textblob se dividió en tres partes el proceso completo para poder analizar los tweets de algún usuario.

4.1.1. Obtención y reformateo de los datos

Para poder llevar a cabo el análisis es necesario contar con una cantidad significativa de datos por lo que se decidió utilizar el api de Twitter la cual nos provee con bastantes endpoints para solicitar diferentes tipos de información, desde número de retweets hasta todos los tweets que ha emitido un usuario desde su registraron en la plataforma.

Para nuestro caso se solicitará cierta cantidad de tweets de un usuario a lo largo del tiempo, esto con el fin de analizar los sentimientos ocultos detrás de cada uno de ellos.

Es necesario tener un formato específico para procesar los datos con textblob, por lo que los objetos obtenidos del api de twitter son reformateados en forma de arreglo en el que cada unidad es una cadena que representa el texto escrito en el tweet. Una vez en este formato se necesitar descartar algunos caracteres de cada uno de los tweets, lo cual nos lleva a la siguiente etapa.

4.1.2. Limpiar el tweet

Para poder llevar a cabo un correcto procesamiento de la cadena se tienen que eliminar algunos caracteres en especial, se descartaran los caracteres siguientes:

- Links
- Menciones de Twitter
- Hashtags
- Algunos otros caracteres especiales

Eso se llevará a cabo utilizando una expresión regular, la cual nos permite establecer la gramática para encontrar un patrón en los tweets y descartar los caracteres no deseados.

4.1.3. Análisis con Textblob

Se usa la api de Textblob para procesar cada uno de los tweets con la finalidad de obtener cuál es el sentimiento predominante en el tweet, para ello se utiliza un clasificador bayesiano.

Al calcular un sentimiento para una sola palabra, TextBlob utiliza la técnica de "promedio" que se aplica a los valores de polaridad para calcular una puntuación de polaridad para una sola palabra y, por lo tanto, se aplica una operación similar a cada palabra y obtenemos una polaridad combinada para textos más largos.

Los sentimientos de un texto son divididos en tres categorías:

- Negativo: Se utiliza un valor de -1
- Neutral: Es asignado el valor de 0
- Positivo: Se utiliza el valor de 1

4.2. Obtención de resultados

Para poder tener acceso a los datos de twitter, primero se necesita obtener las credenciales necesarias para comunicarse con la API.

4.2.1. Credenciales Twitter

Se necesita crear una cuenta en el portal de desarrolladores de twitter en el cual se solicitaban algunos datos personales. También se necesitaba especificar cual seria el uso del api de twitter en nuestra aplicación, una vez realizados estos pasos, en los próximos días llegaría un correo con las claves para tener acceso a la api.

Are you planning to analyze Twitter data? ☐ No

Will your app use Tweet, Retweet, like, follow, or Direct Message functionality? ☐ No

Do you plan to display Tweets or aggregate data about Twitter content outside of Twitter? ☒ Yes

Please describe how and where Tweets and/or data about Twitter content will be displayed outside of Twitter.

Please be thoughtful and thorough

Response must be at least 100 characters 100

Will your product, service or analysis make Twitter content or derived information available to a government entity? ☐ No

In general, schools, colleges, and universities **do not** fall under this category.

Figura 4: Configuración de Twitter

4.2.2. Instalación de paquetes necesarios

Toda la parte experimental sería realizada en el lenguaje de programación Python, el cual nos proporciona un instalador de paquetes llamado PIP. Se procedió a instalar algunos paquetes:

- Tweepy: Una librería que nos permite interactuar de una manera más sencilla con el api de Twitter, esto agiliza bastante el desarrollo de nuestra aplicación.
- Textblob: Esta nos permitirá aplicar algunos algoritmos para poder analizar los tweets que obtengamos del api de twitter.
- Tabulate: Se utilizó para mostrar de una mejor manera los resultados obtenidos.



Figura 5: Logo de Tweepy



Figura 6: Logo de TextBlob

4.2.3. Obtención y procesamiento de los tweets

Primero necesitamos establecer la clave secreta que nos proporcionó twitter para poder hacer solicitudes de información, para ello se utilizará el módulo tweepy que ya contiene funciones para agilizar este proceso, una vez realizado esto podremos obtener los tweets necesarios para su procesamiento.

Nos centraremos en analizar cierto número de tweets de un usuario a lo largo del tiempo que lleva registrado en la plataforma, para ello se seleccionarán algunos otros datos que consideramos importantes para el análisis de los tweets, como la cantidad de retweets, likes y plataforma.

El primer paso es quitar los caracteres que no son necesarios para el análisis, estos son links hacia otras páginas o menciones que pueden ser fácilmente identificadas por el carácter @ , para ello se estableció una expresión regular que nos permite establecer el patrón que queremos conservar, esto se logró gracias al módulo de expresiones regulares de Python.

El último paso es usar el api de textblob para procesar cada uno de los tweets y obtener su clasificación de sentimiento.

4.3. Presentación de resultados

Se mostrarán los datos en terminal, para mostrarlos de una forma más intuitiva se utilizó el módulo tabulate el cual nos permite imprimir de una manera más adecuada las tablas con los datos y su clasificación de sentimiento.

	tweets	id	length	date	platform	likes	retweets	sentiment
0	@inaZenik @finballReed @crikey_news @umjackson @cameronwilson For AI and most things, Python. JavaScript for web. https://t.co/zguSec28c	1531659985603727362	140	2022-05-31 15:32:52+00:00	Twitter for iPhone	1096	115	1
1	https://t.co/G83vCtHd0f	1531647840599051921	23	2022-05-31 14:44:30+00:00	Twitter for iPhone	264390	26804	0
2	@DogeCoinNorway @Louisbottle17 @teslaownersSV @Pathole @Gfliche @MadeInTheISANU @JoelBiden Yeah!	1531642481506108288	97	2022-05-31 14:23:10+00:00	Twitter for iPhone	1224	61	0
3	@Andst7 @Pathole @HamLinZeke @Orknoid1All16 @stevenmarkryan @garaga Maybe if they do the opposite, Twitter will b. https://t.co/sRBP0y0K	1531640532350513155	140	2022-05-31 14:15:34+00:00	Twitter for iPhone	1561	101	0
4	@Pathole @HamLinZeke @Orknoid1All16 @stevenmarkryan @garaga Yes	1531635250273177600	65	2022-05-31 13:54:36+00:00	Twitter for iPhone	1439	71	0
5	@Pathole @HamLinZeke @Orknoid1All16 @stevenmarkryan @garaga Checking	1531634975003682020	70	2022-05-31 13:53:29+00:00	Twitter for iPhone	1435	75	0
6	@finballReed @crikey_news @umjackson @cameronwilson My kids wrote better code when they were 12 than the nonsense. https://t.co/Blylp2D0uZ	1531632807866576904	139	2022-05-31 13:45:11+00:00	Twitter for iPhone	3683	375	1
7	@crikey_news @umjackson @cameronwilson You falsely claimed ur lane snippet of Python gets rid of bots. Ok buddy, L. https://t.co/9EgTt6SFQ7	1531626805400550912	140	2022-05-31 13:21:03+00:00	Twitter for iPhone	4188	374	-1
8	@HamLinZeke @Orknoid1All16 @stevenmarkryan What's going on @garaga?	1531624032272650097	68	2022-05-31 13:13:00+00:00	Twitter for iPhone	4376	521	0
9	@stevenmarkryan Very strange indeed!	1531539935106322433	36	2022-05-31 07:35:49+00:00	Twitter for iPhone	5551	362	-1
10	@Louisbottle17 @teslaownersSV @Pathole @Gfliche @MadeInTheISANU @JoelBiden Interesting	1531519444077490176	87	2022-05-31 06:14:24+00:00	Twitter for iPhone	2160	105	1
11	@holafarsBlog Yeah, maybe a snidge higher.	1531511274718339072	44	2022-05-31 05:41:56+00:00	Twitter for iPhone	12822	480	1
12	@egoyal00 @holafarsBlog @Pathole @Gfliche @MadeInTheISANU Yes. It's safe to say that Mackenzie [ahem] Scott is no. https://t.co/TWnKh3Pv0	1531508960622004993	140	2022-05-31 05:32:46+00:00	Twitter for iPhone	1747	128	1
13	@teslaownersSV @Pathole @Gfliche @MadeInTheISANU @JoelBiden Feels like a Looney Tunes episode	1531504159089909765	93	2022-05-31 05:13:40+00:00	Twitter for iPhone	3053	194	0
14	@teslaownersSV @Pathole @Gfliche @MadeInTheISANU @JoelBiden Yup	1531504072834703104	63	2022-05-31 05:13:19+00:00	Twitter for iPhone	2291	99	0
15	@ExplainsHisBob I support you, Bob! Some bots are good.	1531502041302409330	55	2022-05-31 05:00:25+00:00	Twitter for iPhone	2737	165	1
16	@SKYstarship @Pathole @Gfliche @MadeInTheISANU Yeah, although I hesitate to admit that, because there are a lot o. https://t.co/guwtCn04Y0	1531500365732583425	140	2022-05-31 04:50:21+00:00	Twitter for iPhone	4095	342	0
17	@teslaownersSV @Pathole @Gfliche @MadeInTheISANU The Democrat vs Republican tribalism among otherwise intelligent. https://t.co/cprp0gKv	1531498532268575232	139	2022-05-31 04:55:14+00:00	Twitter for iPhone	5691	640	1
18	@Pathole @Gfliche @MadeInTheISANU True. Tesla was attacked by Romney in 2 of 3 presidential debates. Those lines w. https://t.co/sNNtBffp	1531497009421139073	140	2022-05-31 04:40:26+00:00	Twitter for iPhone	3089	411	1
19	@kinotcon I would be shocked if I'm not being spied on haha.	1531492501573525504	140	2022-05-31 04:27:39+00:00	Twitter for iPhone	17144	1167	-1
20	My only ask is that anyone spying on me please no. https://t.co/9xypE26i1							
20	@Gfliche @MadeInTheISANU I'm just talking about the mid-term elections in November. Not sure that there is a third. https://t.co/TdcU0B10en	15314918302239531009	139	2022-05-31 04:24:40+00:00	Twitter for iPhone	4712	379	-1

Figura 7: Tabla con resultados de análisis

5. Discusión

Sin duda el análisis de de sentimientos para definir su polaridad con la ayuda de textblob nos arroja resultados bastante buenos, sin embargo no es infalible y llega a hacer ponderaciones erróneas como el tweet 18 “True. Tesla was attacked in 2 of 3 presidential debates. ...”. Ciertamente, si bien se trata de un debate y se usa la palabra verdad, la palabra ataque y que sea en 2 ocasiones de 3 no reflejaba para todos algo positivo, incluso que fuese neutral sonaba más razonable pero el algoritmo lo polarizó como una emoción positiva. Con lo anterior tenemos que recalcar que el análisis de emociones no siempre es fácil, incluso los seres humanos al ser subjetivos podemos tener emociones contrarias que son resultado de un mismo mensaje, aunque evidentemente habrá ciertos mensajes en los que negar la polaridad o neutralidad de los mismos de forma racional resulta improbable y gracias a esto podemos definir si un algoritmo de PLN orientado al análisis de emociones es efectivo. Para este trabajo en particular tenemos resultados satisfactorios que nos proporcionan un panorama general de la polaridad de los sentimientos que se transmiten a partir de los tweets de Elon Musk.

6. Trabajo futuro

Consideramos que si bien el PLN nos es muy útil a la hora de ponderar un sentimiento como positivo o negativo proporcionado por un tweet, naturalmente estamos omitiendo características que acompañan al tweet que podrían significar una diferencia, como el uso de mayúsculas, emojis como caracteres especiales, etc. Incluso las imágenes o links podrían analizarse para tener un mayor contexto sobre dicha polaridad, lo cual implicaría un análisis de parte de la información filtrada lo cual dependiendo de la interpretación de nuestro algoritmo podría no siempre resultar favorable. Por otra parte, si quisiéramos incluir imágenes ya tendríamos que implementar técnicas diferentes a las de procesamiento de texto. Por lo que la búsqueda de un mejor análisis de sentimientos nos arroja varios caminos a seguir.

Por último, el producto de esta investigación puede ser encaminado a sectores de la industria en las cuales puedan recabar datos sobre alguna persona de la que se esté haciendo tendencia y aplicar este conocimiento para marketing, ventas, etc.

7. Bibliografía

- Hazarika, D., Konwar1, G., Deb, S., & Bora, D. J. (2020, enero). Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing. ICRMAT, 24. <http://dx.doi.org/10.15439/2020KM20>
- Lam Díaz, Rosa María. (2016). La redacción de un artículo científico. Revista Cubana de Hematología, Inmunología y Hemoterapia, 32(1), 57-69. Recuperado en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-02892016000100006&lng=es&tlng=es
- Liu, B. 2012. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- Henríquez, Carlos; Pla, Ferran; Hurtado, Lluís-F.; Guzmán, Jaime Análisis de sentimientos a nivel de aspecto usando ontologías y aprendizaje automático Procesamiento del Lenguaje Natural, núm. 59, 2017, pp. 49-56 Sociedad Española para el Procesamiento del Lenguaje Natural Jaén, España
- Gallego, J. R. (2018, septiembre). Cómo se construye el marco teórico de la investigación. Scielo. <https://doi.org/10.1590/198053145177>
- Steinberger, J., T. Brychcín, y M. Konkol. 2014. Aspect-Level Sentiment Analysis in Czech. En Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, páginas 24–30.
- Saad, E., Din, S., Jamil, R., Rustam, F., Mehmood, A., Ashraf, I., & Choi, G. S. (2021). Determining the Efficiency of Drugs Under Special Conditions From Users' Reviews on Healthcare Web Forums. IEEE Access, 9, 85721–85737. <https://doi.org/10.1109/access.2021.3088838>