

Análisis de Sentimiento.

David Eduardo Gallardo Fernández 0931556

Monterrey, N.L., México

Email: david.gallardofd@uanl.edu.mx

Abstract—Análisis de sentimiento se refiere al uso de la inteligencia artificial, particularmente el procesamiento de lenguaje natural, análisis de texto y lingüística computacional para identificar y extraer información subjetiva de los textos. El análisis de sentimientos es una tarea de clasificación masiva de documentos de manera automática, en función de la connotación positiva o negativa del lenguaje empleado en el documento. Estos tratamientos generalmente se basan en relaciones estadísticas y de asociación, no en análisis lingüístico.

Keywords: text, sentiment analysis.

I. INTRODUCCIÓN

En términos generales, el análisis de sentimiento intenta determinar la actitud de un interlocutor o usuario con respecto a algún tema o la polaridad contextual general de un documento. La actitud puede ser su juicio o evaluación, estado afectivo (o sea, el estado emocional del autor al momento de escribir), o la intención comunicativa emocional (o sea, el efecto emocional que el autor intenta causar en el lector).

Esta herramienta es muy útil para obtener información importante como por ejemplo la opinión sobre un producto, servicio, una persona, etc.

II. FUENTE DE DATOS

Para el presente trabajo se obtuvieron los datos de Kaggle. Se utilizaron un par de datasets uno relacionados con opiniones de libros de kindle [1] y otro sobre opiniones de teléfonos celulares [2]. El código fuente está en Github [3].

III. PROCEDIMIENTO

Se utilizó el mismo código que se vio en clase con algunas pequeñas variaciones las cuales se verán mas adelante.

A. Dataset: Amazon Kindle Book Review for Sentiment Analysis

Se ejecutaron dos pruebas, una con 1,000 datos y otra con 5,000 datos, los resultados son muy parecidos con las tres librerías, si se revisa la Figura 1 y la Figura 2 se puede ver que las proporciones varían muy poco sin embargo algo pasa cuando se usan mas datos con la librería SentiWordNet cuando se usan mas de 1,000 datos porque además de Positivo, Negativo y Neutral clasifica algunos datos como '0'. Esto desde mi punto de vista parece un error o un bug en la librería que necesita corregirse. Cabe señalar que para poder mostrar la gráfica de pie para este caso se tuvo que modificar el parámetro explota para que funcionara con estas cuatro clasificaciones Tabla I.

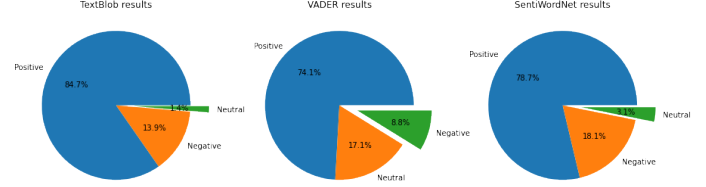


Fig. 1: Dataset: Amazon Kindle Book Review for Sentiment Analysis con 1,000 datos.

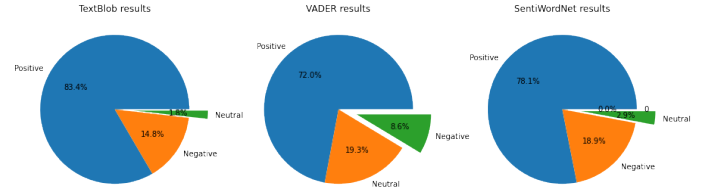


Fig. 2: Dataset: Amazon Kindle Book Review for Sentiment Analysis con 5,000 datos.

B. Dataset: Amazon Cell Phones Reviews

Al igual que con el dataset anterior también se ejecutaron dos pruebas con este dataset, una con 1,000 datos y otra con 5,000 datos, los resultados son muy parecidos con las tres librerías, si se revisa la Figura 3 y la Figura 4 se puede ver que las proporciones varían muy poco sin embargo algo pasa con la librería SentiWordNet ya que desde 1,000 datos además de Positivo, Negativo y Neutral clasifica algunos datos como '0'. A diferencia del dataset de los libros que con 1,000 datos SentiWordNet funcionaba bien, con este dataset de los celulares empieza a fallar desde los 1,000 datos e incluso se observó que también falla con menos de 1,000 datos Tabla II.

	TextBlob	VADER	SentiWordNet
Positive	4171	3602	3906
Negative	738	966	947
Neutral	91	432	146
0	-	-	1

TABLE I: Resultados para el dataset Amazon Kindle Book Review for Sentiment Analysis con 5,000 datos.

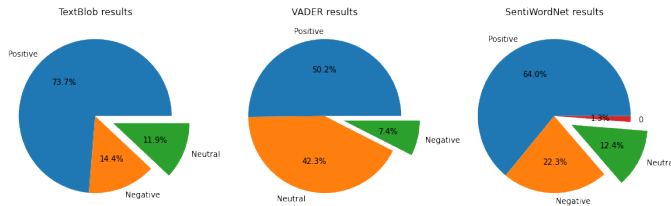


Fig. 3: Dataset: Amazon Cell Phones Reviews con 1,000 datos.

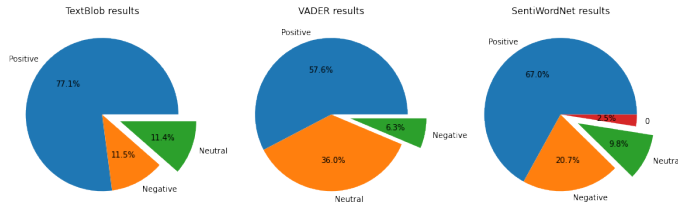


Fig. 4: Dataset: Amazon Cell Phones Reviews con 5,000 datos.

IV. CONCLUSIONES

Revisando los casos particulares en los cuales las librerías dan un resultado diferente para el mismo texto pude ver que en general TextBlob y VADER dan resultados similares mientras que SentiWordNet parece ser la que se equivoca en algunas ocasiones. Por ejemplo, hay una opinion donde el usuario habla de otros celulares que ha tenido en el pasado y da una opinion negativa de esos celulares pero da una opinion positiva sobre el celular en cuestion, en este caso TextBlob y VADER fueron capaces de detectar eso y clasificar la opinión como positiva mientras que SentiWordNet la clasificó como negativa.

El texto de la opción es este: "I had the Samsung A600 for awhile which is absolute doo doo. You can read my review on it and detect my rage at the stupid thing. It finally died on me so I used this Nokia phone I bought in a garage sale for \$1. I wonder y she sold it so cheap?... Bad: ===¿ I hate the menu. It takes forever to get to what you want because you have to scroll endlessly. Usually phones have numbered categories so u can simply press the and get where you want to go. ===¿ It's a pain to put it on silent or vibrate. If you're in class and it rings, you have to turn it off immediately. There's

	TextBlob	VADER	SentiWordNet
Positive	3855	2882	3348
Negative	574	1801	1036
Neutral	571	317	491
0	-	-	125

TABLE II: Resultados para el dataset Amazon Cell Phones Reviews con 5,000 datos.

no fast way to silence the damn thing. Always remember to put it on silent! I learned that the hard way. ===¿ It's so true about the case. It's a mission to get off and will break ur nails in the process. Also, you'll damage the case each time u try. For some reason the phone started giving me problems once I did succeed in opening it. ===¿ Buttons could be a bit bigger. Vibration could be stronger. Good: ===¿ Reception is not too shabby. I was using it in the elevator which is a remarkable feat considering my old phone would lose service by simply putting it in my pocket. ===¿ Compared to my old Samsung, this phone works quite well. The ring tones are loud enough to hear and the phone actually charges quickly and has great battery life. It doesn't heat up like a potatoe in the oven either during long phone convos. ===¿ Nice bright, large screen. ===¿ Cute ways to customize it. Scroll bar can be set to purple, pink, aqua, orange, etc. Overall: Okay phone. It serves its purpose but definitely pales in comparison to these new phones coming out from Sprint. Why get so so when you can get great?"

El lemma es: " Samsung awhile absolute doo doo read review detect rage stupid thing finally die use Nokia phone buy garage sale wonder sell cheap Bad hate menu take forever get want scroll endlessly Usually phone number category u simply press get want go pain put silent vibrate class ring turn immediately fast way silence damn thing Always remember put silent learn hard way true case mission get break ur nail process Also damage case time u try reason phone start give problem succeed open Buttons could bit big Vibration could strong Good Reception shabby use elevator remarkable feat consider old phone would lose service simply put pocket Compared old Samsung phone work quite well ring tone loud enough hear phone actually charge quickly great battery life heat like potatoe oven either long phone convos Nice bright large screen Cute way customize Scroll bar set purple pink aqua orange etc Overall Okay phone serve purpose definitely pale comparison new phone come Sprint get get great"

Después hay otro caso en donde la opinion es claramente positiva y así lo clasificaron TextBlob y VADER pero por alguna razón SentiWordNet la clasificó como neutral.

El texto de la opinión es: "GREAT PRODUCT THAT IS AS GREAT FOR NEXTEL AS IT IS FOR BOOST INSERT YOUR SIM GO!".

Y el lemma es: " GREAT PRODUCT GREAT NEXTEL BOOST INSERT SIM GO".

Para este comentario "I have had this model phone for 4-1/2 years and only now the battery is beginning to fail. FANTASTIC little phone that doesn't drop calls, can stick in a pocket or purse easily, does exactly what it is supposed to do at an unbeatable price. Refill for \$100 and get a bonus \$5." VADER y SentiWordNet lo clasificaron como positivo mientras que TextBlob lo clasificó como negativo. El comentario es positivo ya que dice que es fantastico y que hace lo que promete a un precio inmejorable asi que en este caso TextBlob falló.

En este otro caso "this is a good little inexpensive phone - basic, but that's what I wanted! Just a phone for my pay

as you go plan!” TextBlob y SentiWordNet clasificaron el comentario como positivo y VADER lo clasificó como neutral. Al revisar el comentario se puede ver que no esta hablando de manera negativa del celular pero tampoco se muestra mucho entusiasmo por el mismo, sólo dice que esta bien y es un celular básico por lo que creo que esta opinión en efecto es mas neutral, así que VADER lo hizo mejor.

Al analizar estos casos opino que VADER es la mejor librería para el análisis de sentimientos.

REFERENCES

[1] “Amazon Kindle Book Review for Sentiment Analysis”. <https://www.kaggle.com/meetnagadia/amazon-kindle-book-review-for-sentiment-analysis> (consultado el 26 de mayo de 2022).

[2] “Amazon Cell Phones Reviews”. <https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews> (consultado el 26 de mayo de 2022).

[3] degallardo, MCD-PCD. 2022. Consultado: el 19 de mayo de 2022. [En línea]. Disponible en: <https://github.com/degallardo/MCD-PCD/>

[4] S. Gupta, “Sentiment Analysis: Concept, Analysis and Applications”, Medium, el 19 de enero de 2018. <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17> (consultado el 26 de mayo de 2022).

[5] “Análisis de sentimiento”, Wikipedia, la enciclopedia libre. el 20 de mayo de 2022. Consultado: el 26 de mayo de 2022. [En línea]. Disponible en: <https://es.wikipedia.org/w/index.php?title=An>