

Tarea 2

IRVING DANIEL ESTRADA LÓPEZ
Facultad de Ciencias Físico Matemáticas
Universidad Autónoma de Nuevo León
 Nuevo León, México
 irving.estradalo@uanl.edu.mx

I. INTRODUCCIÓN

El análisis de interacciones ha cobrado relevancia en los últimos años, ya que a la industria le interesa saber la perspectiva que tienen sus clientes acerca del desempeño de sus productos o servicios brindados. Hoy en día el reto es el análisis de los datos, ya que se genera una gran cantidad de datos diarios. El análisis de interacciones genera áreas de oportunidad para mejorar el producto brindado por la empresa.

El análisis de sentimientos nos ayuda a tener una referencia de la perspectiva que tienen los clientes de nuestro producto, nos brinda una retroalimentación para de esta forma detectar fallas dentro del producto o de lo contrario su buen desempeño.

En este artículo continuaremos el análisis acerca de las reseñas que tuvo el producto “Iphone SE”, en esta ocasión enfocándonos en el análisis de sentimientos que las reseñas nos proporcionan.

Comenzaremos con un análisis exploratorio dirigido a los sentimientos reales de nuestras reseñas. Continuando con el preprocesado de datos que nos ayudará a obtener la estructura necesaria para el análisis de sentimientos. Después utilizaremos 3 librerías de Python que llevan a cabo el análisis de sentimientos. Por último, compararemos los resultados de cada una de las librerías, analizando su desempeño para encontrar la mejor, también haremos una predicción utilizando un modelo de Random Forest comparándolo con los sentimientos reales.

II. DATOS

El conjunto de datos con el que se trabajó en este artículo es de dominio público y fue obtenido de Kaggle. Nuestro conjunto de datos contiene 9,713 registros los cuales fueron extraídos de Flipkart utilizando Selenium y BeautifulSoup. Como podemos ver en la figura 1 el conjunto de datos contiene la valoración que le dan al producto Iphone SE, un comentario y la reseña de los clientes en el e-commerce de la India Flipkart.

Fig. 1.

	Ratings	Comment	Reviews	Real_Sentiment
0	5	Super!	Great camera for pics and videos Battery life ...	Positive
1	5	Must buy!	Great device. Let me tell the Pros..1. Superb ...	Positive
2	5	Great product	Who all loves older size i.e., 4.7 inch type s...	Positive
3	5	Simply awesome	This iPhone SE is the best phone ever you get....	Positive
4	5	Classy product	This is my second iphone after iphone 4s. I've...	Positive

En la tabla 1 podemos revisar los atributos de nuestro conjunto de datos junto a una descripción, es importante

destacar que la variable Ratings va de 1 a 5. Siendo la calificación más baja 1 y siendo la mejor 5.

Variable	Representación
Ratings	La calificación que los clientes le dieron al producto.
Comment	Comentario, es el título de la reseña.
Reviews	La reseña del cliente hacia el producto.
Real_Sentiment	Sentimiento de la reseña.

TABLE I
 VARIABLES DEL CONJUNTO DE DATOS

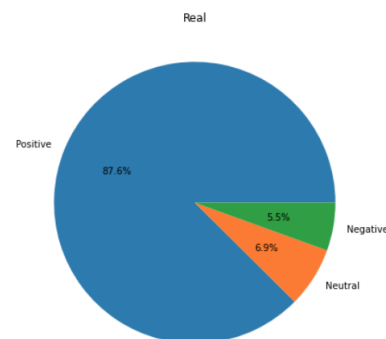
También se agregó una columna llamada “Real_Sentiment” que contiene el sentimiento verdadero según la calificación, asignándole “Positivo” si la calificación es 5 o 4, asignándole “Neutral” si la calificación es 3 y por ultimo asignándole “Negativo” si la calificación es 2 o 1. Estos intervalos se tomaron considerando el 3 como la mediana de las posibles calificaciones, la parte superior fue considerada como una buena calificación y la parte inferior como mala.

III. METODOLOGÍA

A. Análisis Exploratorio

Después de crear la columna “Real_Sentiment”, haremos un análisis exploratorio de dicha columna, ya que es importante identificar las proporciones de nuestras reseñas basadas en los sentimientos que generaron.

Fig. 2.



En la figura 2 tenemos un gráfico de pastel mostrándonos las proporciones de sentimientos de nuestras reseñas según su calificación. Predomina el sentimiento positivo dirigido hacia el producto. En el artículo anterior identificamos que las

calificaciones con 4 y 5 abarcan más del 75% de nuestros registros, reflejándose en la proporción del 87.6% de sentimientos positivos.

B. Preprocesado

Antes de comenzar con el análisis de sentimientos empleando las librerías de Python, debemos adecuar nuestros datos para las librerías que utilizaremos, en este caso nos concentraremos en la columna “Reviews” la cual son los registros de las reseñas y en base a ellas clasificaremos, obteniendo como resultado el sentimiento detectado.

A continuación, se enlistan los pasos de preprocesado que se llevaron a cabo, los cuales son muy similares a los del artículo pasado, de haber alguna aclaración se especificará en cada uno de los puntos:

- 1) **Remover patrones de ruido:** En nuestro conjunto de datos existe un patrón de ruido, contiene “READ MORE” al final de cada reseña.
- 2) **Remover Emojis**
- 3) **Remover URL**
- 4) **Remover signos de puntuación y números**
- 5) **Convertir a minúsculas**
- 6) **Revisión de Ortografía**
- 7) **Obtener el lemma de cada una de las palabras:** Esta parte del preprocesado se divide en dos partes. La primera parte consiste en obtener el POS (Part of Speech) el cual indica lo que representa esa palabra dentro de la oración. Se agregó una columna llamada “POS_tagged”, ya que la utilizaremos después para una de las librerías. En la segunda parte se genera el lemma de la palabra. Es importante destacar que para que la librería que estamos utilizando en Python (WordNetLemmatizer) transforme correctamente las palabras en el lemma, hay que indicarle el POS ya que de lo contrario existe un margen de error mayor.
- 8) **Tokenization**
- 9) **Remover Stop Words**

Una vez finalizado el preprocesamiento, tenemos dos nuevas columnas la anteriormente mencionada “POS_tagged” y la de “Cleaned_Reviews” la cual contiene las reseñas con el preprocesado aplicado. Teniendo estas dos nuevas columnas ya podemos pasar a probar las 3 librerías de Python que nos ayudarán en el análisis de sentimiento, las cuales son: TextBlob, VaderSentiment y sentiwordnet.

C. TextBlob

Es una librería para el procesamiento de texto, nos provee una API para hacer tareas comunes en el NLP como el obtener POS (Part of Speech) tagging, la extracción de frases nominales, el análisis de sentimiento, clasificación, traducción, entre otras. En esta ocasión la utilizaremos para el análisis de sentimientos, en donde podemos extraer dos propiedades: la polaridad y la subjetividad. La polaridad nos ayuda a indicar si el texto es positivo o negativo, este valor va de -1 a 1. La subjetividad va de 0 a 1, donde 0 sería que el texto es totalmente objetivo, mientras que 1 sería que es totalmente

subjetivo. En base a la polaridad generada con el texto, asignaremos el sentimiento identificado. Si es mayor a 0 será “positive”, si es menor a 0 será “negative” y si es igual a 0 será “neutral”

D. VaderSentiment

VADER (Valence Aware Dictionary and sEntiment Reasoner) es un diccionario y una herramienta de análisis de sentimientos basado en reglas el cual está diseñado específicamente para el análisis de texto en redes sociales. Las bases de VADER están inspiradas en bancos de palabras como LIWC (Linguistic Inquiry Word Count), ANEW (Affective Norms for English Words) y GI (General Inquirer). También detecta emoticones, junto a acrónimos y jergas comunes. Los valores de la polaridad de VADER van de -1 que equivale a negativo, 1 que equivale a positivo y 0 que equivale a neutral. Su clasificación de sentimientos se llevó a cabo de la misma manera que en la anterior librería, de acuerdo con su polaridad y el intervalo.

E. sentiwordnet

Esta librería es un recurso léxico para apoyar explícitamente a la clasificación de sentimientos y a la aplicación de minería de opiniones. Opera junto a la base de datos de WordNet y las funciones que provee proporcionan una métrica de lo positivo, negativo o neutral. Para esta librería utilizaremos la columna “POS_tagged” ya que con ella se obtendrán los valores de los sentimientos. Estos valores de los sentimientos se restan ($score = score_{positive} - score_{negative}$) para obtener un valor definitivo el cual podamos clasificar. Cabe mencionar que WordNet es una base de datos léxica que compone palabras en inglés agrupadas como sinónimos, en lo que se conoce como synset.

F. Predicción de Sentimientos con Random Forest

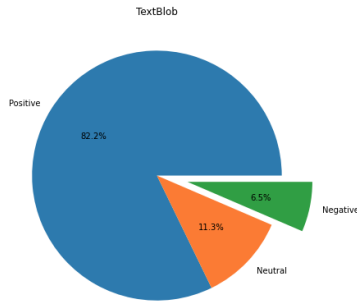
Random Forest es un algoritmo de Machine Learning que se puede utilizar tanto para clasificación como para regresión. En este caso lo utilizaremos para clasificar los sentimientos detectados en las reseñas. A diferencia de las anteriores 3 librerías, los modelos de Machine Learning aprenden en base a los registros que se tienen. En este caso se agregó otro paso en el preprocesado, para crear una matriz de palabras en donde en cada uno de los valores se asigna el TF-IDF (Term frequency – Inverse document frequency) el cual representa la significancia que tuvo cada palabra dentro de ese registro, representándolo como un ratio.

IV. RESULTADOS

A continuación, se presentan los resultados de cada una de las pruebas de las librerías de análisis de sentimientos en Python y la clasificación con el modelo Random Forest. Comenzamos con un gráfico de pastel, siguiendo de la matriz de confusión comparando los sentimientos generados en la columna “Real_Sentiment” contra los análisis de cada librería, en la parte final de cada una de las subsecciones se presenta un reporte de clasificación. En caso de que se hiciera alguna modificación a la función vista en clase se especificará en cada una de las librerías.

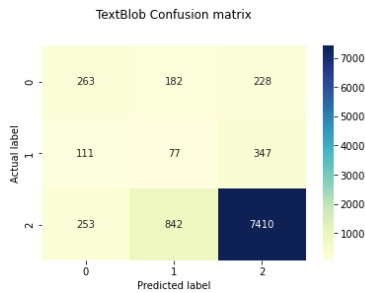
A. TextBlob

Fig. 3.



Como podemos ver en la figura 3 la librería TextBlob hizo un análisis de sentimiento muy parecido en proporción a los valores que tenemos en nuestra variable “Real_Sentiment”. Recordando que la mayoría de nuestras reseñas son positivas, aparentemente la librería de TextBlob se acerca a los sentimientos reales de nuestras reseñas, sin embargo, no podemos concluir en base a este gráfico, para evaluar el desempeño de cada una de las librerías utilizaremos la matriz de confusión, así como el reporte de clasificación.

Fig. 4.



En la figura 4 podemos darnos cuenta de que se llevaron a cabo una gran cantidad de clasificaciones positivas en nuestras reseñas, asemejándose a la clasificación real. Aparentemente la librería tuvo dificultad al clasificar las observaciones que tienen un sentimiento neutral, ya que en nuestra matriz de confusión podemos ubicar que el 65% de las observaciones totales con clasificación real neutral, son etiquetadas en su mayoría como positivas. Pasa lo mismo con la clasificación de reseñas negativas, sin embargo, es menor la cantidad.

Fig. 5.

	precision	recall	f1-score	support
Negative	0.42	0.39	0.40	673
Neutral	0.07	0.14	0.09	535
Positive	0.93	0.87	0.90	8505
accuracy			0.80	9713
macro avg	0.47	0.47	0.47	9713
weighted avg	0.85	0.80	0.82	9713

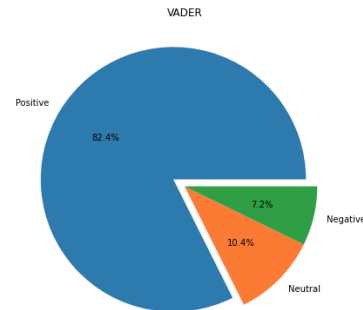
0.7978997220220323

En la figura 5 tenemos el reporte de clasificación el cual nos confirma que la librería TextBlob tuvo problemas clasificando las observaciones con sentimiento neutral, ya que el reporte

de clasificación nos indica que obtuvo 7% de precisión. Lo mejor que clasificó nuestra librería fueron las reseñas positivas con un 93%, la librería con las reseñas negativas no tuvo un buen desempeño ya que está por debajo del 50% la precisión. A pesar de lo mencionado, la librería alcanza el 80% de asertividad, esto debido a que tuvo un buen desempeño con las reseñas positivas y nuestras reseñas en su mayoría son de esta categoría.

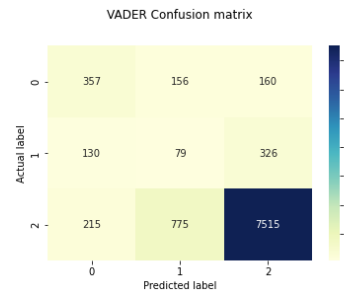
B. VaderSentiment

Fig. 6.



En la figura 6 tenemos las proporciones de la clasificación de nuestras reseñas con VaderSentiment, aparentemente las proporciones son parecidas a las reales. A diferencia de la función utilizada en clase, los cambios que se realizaron fueron los intervalos de cada una de las categorías con relación a su polaridad, de este modo se podrán hacer las comparaciones entre librerías sin tener diferencia en ellos, afectando su desempeño.

Fig. 7.



En la figura 7 identificamos que se llevaron a cabo de forma correcta en su mayoría las clasificaciones positivas, VaderSentiment tuvo problemas clasificando las observaciones neutrales ya que solamente el 14% de las clasificaciones neutrales que llevó a cabo fueron correctas. En la clasificación de reseñas negativas aparentemente obtuvo resultados aceptables ya que la mayoría de las observaciones fueron clasificadas correctamente.

Fig. 8.

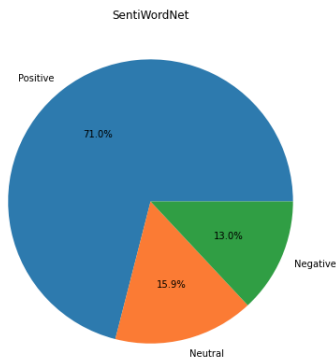
	precision	recall	f1-score	support
Negative	0.51	0.53	0.52	673
Neutral	0.08	0.15	0.10	535
Positive	0.94	0.88	0.91	8505
accuracy			0.82	9713
macro avg	0.51	0.52	0.51	9713
weighted avg	0.86	0.82	0.84	9713

0.8185936373931844

Por ultimo tenemos el reporte de clasificación en la figura 8 que confirma su buen desempeño con la clasificación positiva, el problema que tuvo con la clasificación de reseñas neutrales alcanzando solamente el 8% de precisión y la clasificación negativa que no es buena, sin embargo, supera a la precisión de la clasificación de reseñas neutrales con un 51% de precisión.

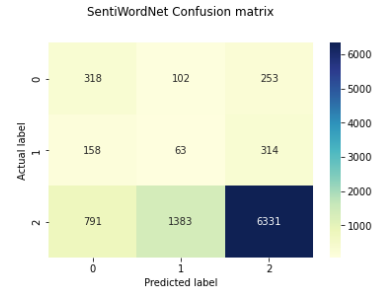
C. sentiwordnet

Fig. 9.



En el gráfico de pastel de la figura 9 visualizamos que las proporciones de las clasificaciones de sentimientos elaborada por sentiwordnet es diferente a las clasificaciones de las otras librerías. La función vista en clase fue modificada, debido a que en nuestro conjunto de datos contiene reseñas con errores de ortografía a pesar de la revisión del preprocesado. Recordando cómo funciona sentiwordnet, esta librería trabaja con los POS de las palabras, en caso de no detectar la palabra, al igual que las demás librerías se le da de polaridad el valor de 0. La función vista en clase en el caso de que ninguna de las palabras dentro de cada registro de la columna "POS_tagged" esté dentro del synset de sentiwordnet, nuestra función no retornaba ninguno de los sentimientos sino retornaba el valor de 0. Sin embargo, en el caso de no ser posible identificar las palabras de la reseña en el synset, su polaridad es 0 de esta manera se clasificaría como una reseña neutral. El conjunto de datos contine reseñas de pocas palabras y con errores ortográficos, este es un caso por lo cual la librería no podría ubicar las pocas palabras con errores ortográficos en el synset y no podría continuar con su proceso.

Fig. 10.



En la matriz de confusión de la figura 10 sentiwordnet obtuvo buenos resultados en las clasificaciones positivas, tuvo dificultad al clasificar las reseñas neutrales de la misma manera que con las clasificaciones de reseñas negativas. Al menos con este conjunto de datos aparentemente no obtuvo los mejores resultados.

Fig. 11.

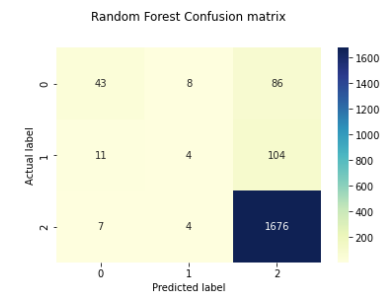
	precision	recall	f1-score	support
Negative	0.25	0.47	0.33	673
Neutral	0.04	0.12	0.06	535
Positive	0.92	0.74	0.82	8505
accuracy			0.69	9713
macro avg	0.40	0.44	0.40	9713
weighted avg	0.82	0.69	0.75	9713

0.6910326366725008

La figura 11 nos confirma lo anteriormente mencionado, con las clasificaciones negativas y neutrales tuvo resultados bajos en la precisión de ambas clases. En la parte de clasificación de reseñas positivas obtuvo un 92% de precisión que es un buen resultado favoreciéndole a la asertividad. La asertividad está por debajo del 70%, de esta manera teniendo la peor asertividad de las 3 librerías.

D. Random Forest

Fig. 12.



Por último, tenemos el modelo de Machine Learning Random Forest, el cual en su matriz de confusión en la figura 12 aparentemente obtuvo un buen desempeño con las clasificaciones positivas, sin embargo, revisando las otras categorías podemos darnos cuenta de que tuvo problemas clasificando las demás, confundiéndolas con la categoría positiva. Lo anteriormente mencionado se debe a que los modelos de Machine Learning aprenden a través de registros, el modelo fue entrenado en su mayoría con registros clasificados como

positivos. El modelo no tuvo oportunidad de aprender correctamente la categoría neutral y negativa, debido a su falta de registros de dichas categorías, dificultándole su clasificación.

Fig. 13.

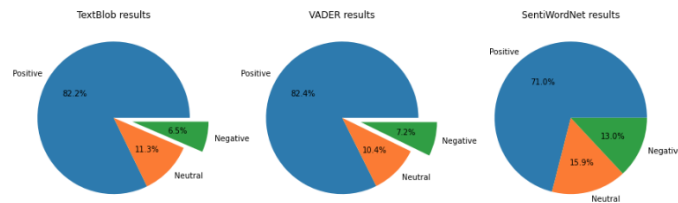
	precision	recall	f1-score	support
Negative	0.70	0.31	0.43	137
Neutral	0.25	0.03	0.06	119
Positive	0.90	0.99	0.94	1687
accuracy			0.89	1943
macro avg	0.62	0.45	0.48	1943
weighted avg	0.84	0.89	0.85	1943

0.8867730313947504

En la figura 13 identificamos que el modelo obtuvo una excelente asertividad con 88%, el reporte de clasificación nos confirma la dificultad que tuvo el modelo en la clasificación de reseñas neutrales. La clasificación negativa tuvo buenos resultados. Es importante recordar que el Random Forest se entrenó con el 80% de los datos y se probó con el 20%.

E. Comparativa entre librerías

Fig. 14.



Después de llevar a cabo de forma individual el desempeño que tuvo cada una de las librerías con respecto a la clasificación de sentimientos real, podemos compararlas para definir con el conjunto de datos que se trabajó en este artículo cuál fue la que obtuvo mejor desempeño basándonos en los reportes de clasificación y tomando en cuenta la versatilidad de cada librería. Es importante tomar en cuenta que las clases de nuestro conjunto de datos no están balanceadas, esto quiere decir que tenemos más observaciones de una clase en específico que de otras, en este caso positivas. Sin embargo, esto no afecta a ninguna de las librerías ya que la clasificación que se lleva a cabo en cada una de ellas es basada en reglas, a diferencia de los modelos de Machine Learning que se verían afectados al momento de aprender en la clasificación.

Las tres librerías tuvieron problemas clasificando las reseñas neutrales, existe sospecha que la dificultad de la clasificación se debe a que la polaridad debe ser exactamente 0, a diferencia de las demás que tienen un rango. Esta es un área de oportunidad para aplicar algún tipo de rango en la clasificación neutral, evitando ser tan absolutos.

Basándonos en la asertividad (accuracy), ya que esta métrica engloba todas las demás, la mejor es la librería de vader-Sentiment. Esta librería tuvo un buen desempeño tanto con los positivos como con los negativos, recordando que está diseñada para las redes sociales. De lo contrario sentiwordnet fue la que obtuvo los peores resultados, además de tener cierta

rigidez al momento de encontrar palabras, debido a que trabaja con el POS. Cabe destacar que la librería TextBlob es una buena opción para trabajar con texto, ya que no solamente se enfoca en la clasificación de sentimientos, tiene diversas herramientas que nos pueden ayudar en nuestros análisis. Además, TextBlob obtuvo buenos resultados con un 80% de asertividad.

V. CONCLUSIÓN

El análisis de interacciones ha cobrado relevancia en los últimos años ya que es un área de oportunidad para la industria al momento de valorar sus productos y servicios. Uno de los análisis más comunes dentro de las interacciones que realiza el usuario, es el análisis de sentimientos. El poder identificar si las reseñas que los clientes redactan del producto son buenas, neutrales o malas en su mayoría, es un indicador para llevar a cabo acciones, ya sean correctivas o acciones necesarias para obtener los mejores resultados. La clasificación de las librerías vistas en este artículo está basada en reglas. Algunas de las ventajas que se tiene al hacer una clasificación basada en reglas son que no se tiene una fase de entrenamiento, por lo tanto, no nos afecta la cantidad de registros del conjunto de datos, a diferencia de los modelos de Machine Learning que en la fase de entrenamiento tenemos que tener una cantidad considerable de registros para que se lleve a cabo un buen desempeño. Otra de las ventajas, es que no es necesario tener la etiqueta del sentimiento real, ya que las funciones que se utilizan dentro de las librerías no trabajan con ella, las librerías nos dan el valor de la polaridad y en base a ella clasificamos el sentimiento. Sin embargo, una de las desventajas que tiene la clasificación basada en reglas es que, de no identificar la palabra dentro de su diccionario, le asigna de polaridad 0. Por otra parte, los modelos de Machine Learning aprenden las palabras relevantes del sentimiento de esa reseña, de acuerdo con los registros vistos en el entrenamiento.

REFERENCES

- [1] KAMAL DAS. (2021). Apple iPhone SE reviews ratings. Mayo del 2022, de Kaggle Sitio web: <https://www.kaggle.com/datasets/kmlas/apple-iphone-se-reviews-ratings>
- [2] Irving Estrada. Github. 2022, Sitio web: <https://github.com/Irving-Estrada/Procesamiento>
- [3] Steven Loria. (2020-04-26). TextBlob: Simplified Text Processing. 26 de mayo del 2022, de TextBlob Sitio web: <https://textblob.readthedocs.io/en/dev/>
- [4] C.J. Hutto, Eric Gilbert. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Association for the Advancement of Artificial Intelligence.
- [5] Bruno Ohana, Brendan Tierney. (2009). Sentiment Classification of Reviews Using SentiWordNet. 9th. IT T Conference.