

Análisis de sentimientos de textos de Twitter utilizando aprendizaje profundo

Jessica Olivares L., Abraham Sánchez L., Rogelio González V.,
Abraham Maldonado G.

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación,
México

{olivares.lopez.jessica3p, rogelio.gzzvzz}@gmail.com,
{abraham.sanchez, abraham.maldonadoga}@correo.buap.mx

Resumen. En este trabajo se propone una serie de estrategias provenientes del área de aprendizaje máquina, incluyendo el aprendizaje profundo. El conjunto de datos utilizados está integrado con texto extraído de Twitter, específicamente con tweets relacionados a la inteligencia artificial, lo cual adicionalmente proporciona un pequeño panorama respecto a la opinión pública de esta área de la ciencia, a partir del análisis y la exploración de datos. Se realizaron experimentos para análisis de sentimientos desde distintos puntos de referencia, de manera supervisada y no supervisada. El desempeño o resultado de cada una de estas estrategias está variado por la integridad de datos, el problema o tema principal de los datos, la variación de parámetros, entre otros más factores.

Palabras clave: Análisis de sentimientos, Twitter, aprendizaje profundo, PLN.

Sentiment Analysis of Twitter Texts Using Deep Learning

Abstract. In this paper, a series of strategies from the machine learning area, including deep learning, are proposed. The data set used is integrated with text extracted from Twitter, specifically with tweets related to artificial intelligence, which additionally provides a small overview of public opinion in this area of science, based on data analysis and exploration. Experiments for sentiment analysis were carried out from different points of reference, in a supervised and unsupervised manner. The performance or result of each of these strategies is varied by data integrity, the main data problem or issue, parameter variation, among other factors.

Keywords: Sentiment analysis, Twitter, Deep learning, NLP.

1. Introducción

El análisis de sentimientos es una de las aplicaciones de la clasificación de textos del Procesamiento del lenguaje natural (PLN), básicamente asigna una categoría apropiada al contenido de una oración, texto o documento, a partir del procesamiento

de texto (previamente no estructurado). Esta clasificación se hace mediante la asignación de una polaridad de sentimiento: positivo, negativo o neutro a una oración o un documento, o a partir de la asignación de una emoción que se identifique en la oración [1].

El análisis de sentimientos es una de las técnicas de mayor relevancia de esta área, generalmente para la evaluación de todo el contenido en texto generado en la web [2]. Sin embargo, el análisis de sentimientos es un área multidisciplinar, pues además del procesamiento de lenguaje natural intervienen disciplinas como lingüística y psicología.

El hecho de definir las categorías en las que será clasificado un texto basado en la premisa de “sentimiento” es una cuestión psicológica muy ambigua y diversa, pues existe una variedad de posturas al respecto, hay cientos de estados emocionales que forman parte de la condición humana.

Una clasificación o representación de las emociones humanas popular, es la rueda de Plutchik. El psicólogo Robert Plutchik sugiere que hay 8 emociones evolutivas humanas, es decir, emociones que han formado parte de la supervivencia humana y que han sido transferidas de generación en generación. Estas emociones son las siguientes: Ira, Miedo, Tristeza, Repugnancia, Sorpresa, Expectación, Confianza y Alegría.

En la representación de Plutchik se muestra que cada una de estas emociones centrales puede intensificarse, atenuarse o incluso combinarse para producir cualquier estado emocional. Sin embargo, muchos de los trabajos realizados para análisis de sentimientos, están basados en la clasificación del texto de acuerdo con la polaridad del sentimiento, bueno, malo o neutral, en su mayoría.

Lo cual facilita considerablemente el problema de análisis de sentimientos, además, la polaridad puede ser más precisa porque solo hay dos o tres clases distintas, que hace más fácil de distinguirlas entre sí, mientras que una clasificación por emociones puede ser más ambigua, pues una oración puede involucrar más de una emoción. Generalmente el análisis de sentimientos hace la categorización de texto no procesado en polaridades de acuerdo con las siguientes categorías: Positivo, Negativo y Neutro.

El conjunto de estrategias que se proponen en este trabajo, puede ser aplicable o adaptable a cualquier conjunto de datos en formato de texto para problemas de análisis de sentimientos sin importar el dominio del texto, como se verá en el apartado de resultados obtenidos, el desempeño de cada una de las técnicas propuestas varía de acuerdo con la integridad de los datos, el problema o tema principal de los datos, la variación de los parámetros, entre otros factores.

2. Trabajos relacionados y enfoques

El análisis de sentimientos es una de las áreas de investigación más vigorosas en el campo del procesamiento de lenguaje natural (PLN) que se centra en analizar las opiniones, sentimientos, actitudes y emociones de las personas hacia varias entidades, como productos, servicios, organizaciones, problemas, eventos y temas [1]. Recientemente se han publicado una gran cantidad de trabajos de investigación sobre análisis de sentimientos en diferentes idiomas.

Por lo tanto, para lograr esto, primero deberemos revisar las bases para la investigación sobre el análisis de sentimientos mediante la revisión de la literatura relevante sobre estudios anteriores realizados en este campo. En uno de estos trabajos, los autores proponen que el análisis de sentimientos se podría dividir en tres enfoques principales, a saber, un enfoque basado en el aprendizaje máquina, un enfoque basado en el conocimiento y un enfoque híbrido [2]. En este trabajo adoptamos el enfoque de aprendizaje máquina.

Un trabajo interesante y relacionado a nuestra propuesta, se propuso en [2] para el análisis automático de los comentarios de los usuarios de Internet que se publican en las páginas oficiales de los supermercados en Túnez en las redes sociales de Facebook. En su propuesta utilizaron CNN (Convolution Neural Networks), LSTM (Long Short Term Memory) y Bi-LSTM (Bi-directional Long Short Term Memory) indicando que los resultados obtenidos son satisfactorios, especialmente para los algoritmos LSTM y Bi-LSTM.

En [3] los autores señalan que el uso de la técnica de análisis de características juega un papel importante en el desarrollo y la mejora de un modelo de análisis de sentimientos. La propuesta de los autores coincide con esta propuesta en el hecho de poder utilizar datos ruidosos, palabras ajenas al vocabulario, etc.

Su propuesta es interesante pues utilizan un modelo BERT previamente entrenado que permite obtener características semánticas y contextuales a nivel de oración y con ello generar integridades. En sus resultados experimentales, destacan el uso de un algoritmo CNN dilatado para poder extraer información local y global.

Los autores en [4] presentan un estudio de varias arquitecturas de aprendizaje profundo y sus aplicaciones en el análisis de sentimientos. En el trabajo, ellos detallan cómo muchas de estas técnicas de aprendizaje profundo han demostrado su utilidad en muchas tareas del análisis de sentimientos y vislumbran en los años venideros, un aumento importante en el uso del aprendizaje profundo en investigaciones.

La investigación existente en el área ha producido muchísimas técnicas para diversas tareas del análisis de sentimientos, que incluyen algoritmos supervisados y no supervisados. Es claro que dentro de los enfoques supervisados se utilizan ampliamente las máquinas de vectores de soporte (SVM), entropía máxima, naïve Bayes, entre otros; incluso combinaciones de estos algoritmos han dado buenos resultados.

Por su parte, los algoritmos no supervisados incluyen métodos que explotan los léxicos de sentimiento, el análisis gramatical y los patrones sintácticos. Sería poco práctico detallar los libros y artículos, pero el lector interesado puede revisar la literatura existente en el área [1, 7, 8].

3. Extracción y exploración de datos

Uno de los objetivos de la propuesta, es recuperar datos de Twitter, obteniendo así datos no estructurados, a los cuales es necesario aplicar estrategias de preprocesamiento para poder realizar el análisis de sentimientos. El primer paso, es consolidar un conjunto de datos que contenga la información y estructura necesaria para realizar la tarea propuesta. El proceso de creación de un conjunto de datos consta de tres procesos: adquisición de datos, limpieza de datos y etiquetado de datos.



Fig. 1. Estructura de consulta de extracción de texto de Twitter.

3.1. Adquisición de datos

Para la adquisición de datos se hizo uso de la API de Twitter, bajo una cuenta de desarrollador. La consulta de extracción de datos se describe en los parámetros de la Fig 1.

Se realizó un filtro de los Tweets que en su contenido se encontrarán los términos de la Fig 1, inteligencia artificial, deep learning y machine learning esencialmente, al ser considerados unos de los temas más relacionados y populares de la inteligencia artificial en ese periodo de tiempo. Obteniendo así, un total de 12,000 observaciones, que en este contexto son Tweets.

3.2. Limpieza de datos

El proceso de limpieza de datos es uno de los procesos que define el éxito de un modelo de aprendizaje máquina (AM). Para tareas o problemas de PLN se realizan algunas estrategias básicas que facilitan la manipulación del texto. El proceso de limpieza de esta propuesta está compuesto de dos partes tal y cómo se ilustra en la siguiente Fig. 2.

La limpieza básica consiste en la eliminación de algunos elementos, pues al ser extraídas las observaciones de un medio social como Twitter, en el cual el texto lo integran elementos adicionales como: usuarios, menciones, hashtags, enlaces o URL y emoticones.

Algunos de estos elementos deben ser eliminados, pues pueden introducir ruido y reducir el desempeño o aprendizaje de un modelo. La segunda parte de la limpieza de datos es más general para cualesquiera tareas de Lenguaje de Procesamiento Natural y consiste en la realización de las siguientes tareas [9].

- **Convertir el texto a minúsculas.** Esto facilita algunos procesos de exploración que se muestran más adelante.
- **Eliminar espacios dobles.** Se hace la sustracción de espacios dobles, porque no aportan nada al contenido del texto.
- **Eliminar números.** La eliminación de números previene la errónea interpretación de números en los modelos.

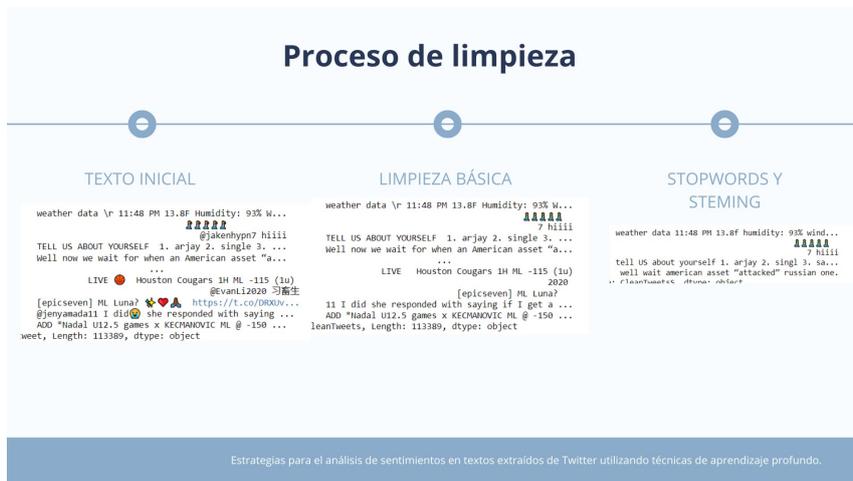


Fig. 2. Proceso de limpieza del conjunto de datos.

- **Eliminar StopWords.** Consiste en la eliminación de palabras gramaticales que regularmente se ocupan con mayor frecuencia en el idioma y que podrían no añadir valor al contexto, ejemplos de estas palabras son: conjunciones, artículos, preposiciones, etcétera.
- **Stemming.** Este proceso consiste en convertir una palabra a su palabra base (obtener la raíz de una palabra).

3.3. Etiquetamiento de datos

Finalmente, una vez recolectados y estructurados los datos, se procede a la asignación de una etiqueta, en particular, una polaridad de sentimiento. Esta asignación se realizó de manera semi-supervisada. Mediante un flujo de trabajo en Python que solicitó respuesta del servicio de Text Analysis de Azure, el cual, entre otras tareas, puede asignar una polaridad de sentimiento a un conjunto de documentos o textos.

Además de una supervisión manual sobre la etiqueta asignada por Text Analysis. Text Analysis, asigna una de las siguientes categorías o polaridad de texto: neutral, positive, negative y mixed.

También devuelve puntuaciones de confianza entre 0 y 1 para cada documento y oraciones dentro del documento. Una vez realizada la solicitud de las 12,000 observaciones recolectadas se obtiene la siguiente cantidad de observaciones para cada categoría de sentimiento: neutral; 6162, positive; 2965, negative; 2290 y mixed; 583, relación que se muestra en la tabla 1.

3.4 Exploración de datos

La exploración de datos es el proceso en el cual se realiza la examinación de los datos para entender la composición o integración de estos y obtener las primeras observaciones e inferencias sobre ellos. Para el procesamiento del lenguaje natural

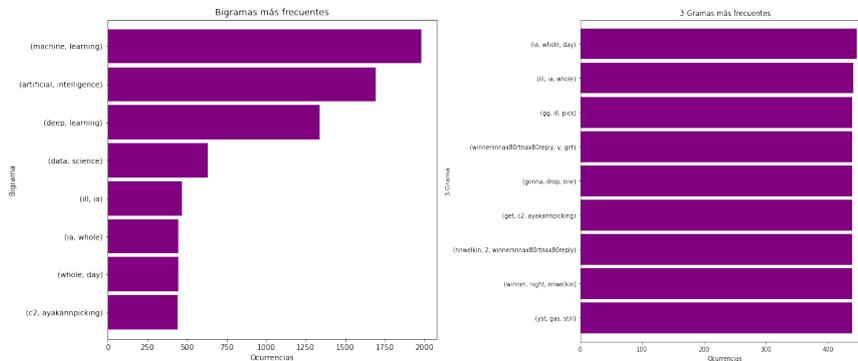


Fig. 3. Histogramas de bigramas y trigramas más frecuentes.

Tabla 1. Número de observaciones por sentimiento.

Sentimiento	Observaciones
Neutral	6162
Positivo	2965
Negativo	2290
Mezclado	583

existen varias técnicas que permiten visualizar y deducir conclusiones eficientemente. Se realizaron algunas de estas técnicas, con las que se obtuvieron peculiares observaciones que se detallan ampliamente en [10].

En primer lugar, se obtuvieron dos histogramas de la frecuencia de palabras a partir de la composición de n-gramas. Los n-gramas son una subsecuencia, donde n es el número de elementos de la subsecuencia; esta técnica es usualmente utilizada en el procesamiento del lenguaje natural para el tratamiento de textos.

Los histogramas de las siguientes figuras corresponden a la división del texto en 2 y 3 gramas, bigramas y trigramas, respectivamente. Para obtener este gráfico es necesario obtener la lista de bigramas y trigramas con su respectivo número de ocurrencias a partir del corpus.

Al comparar ambos histogramas, la implementación de bigramas favorece al comportamiento de los datos, pues retomando el contexto, se tiene mayor coherencia del contenido con respecto de la división en trigramas donde, los subconjuntos de palabras no tienen mucha relevancia al contexto del problema. Entre los bigramas más relevantes del histograma de la Fig. 3 se encuentran, machine-learning, artificial-intelligence, deep-learning y data-science, todos estos términos que son directamente relacionados con la “Inteligencia artificial”.

A partir de la lista de bigramas obtenidos, también se puede obtener una representación semántica, la cual muestra la relación que hay entre distintos bigramas establecidos en el paso anterior. La visualización de bigramas es de ayuda para confirmar que los datos si están constituidos en base al contenido deseado, es decir, con opiniones respecto a la Inteligencia artificial.

Se utilizaron representaciones de texto muy populares en aprendizaje profundo, entre las que se encuentran Word embeddings o incrustaciones de palabras. Sin

Esta herramienta permite a partir de una palabra, encontrar su representación vectorial y las palabras más cercanas, es decir, las palabras relacionadas a partir del cálculo de la distancia (coseno o euclidiana).

Otras deducciones interesantes abordadas en el proceso de exploración de datos fue la identificación del número de observaciones para cada clase, en el histograma de la Fig. 5 se muestran las observaciones que se tienen para cada una de las categorías de sentimiento.

De la cual a primera vista se puede ver una desproporción de observaciones. La distribución de datos está desbalanceada, y este problema podría afectar el desempeño de los modelos, al introducir un sesgo de información sobre las clases mayoritarias y minoritarias; neutral y mezclada, respectivamente. Para evitar problemas en fases futuras, se aplican técnicas de balanceo de clases [12].

Para no reducir considerablemente el número de instancias de cada clase se eliminaron las muestras de la clase mixed, reduciendo el número de observaciones totales a 11,417. La eliminación de observaciones o instancias para el balanceo de clases es lo que se conoce como undersampling, y ha sido aplicado a la distribución de datos, de forma aleatoria para al final tener 2,290 observaciones de cada clase.

Para reducir la dimensionalidad del conjunto de datos, solo se seleccionaron algunas columnas sobre las cuales trabajaremos para realizar la tarea de análisis de sentimientos con diversos algoritmos, como se verá a continuación.

4. Implementación de las estrategias

Una característica importante para la realización de alguna tarea no supervisada es que las observaciones no están etiquetadas previamente, como es el caso del conjunto de datos propuesto. Sin embargo, se debía implementar una forma de tener datos etiquetados que ayude a evaluar el comportamiento de los modelos, y aunque no sean entrenados con estas etiquetas en los modelos no supervisados, permita comparar su desempeño y además generar modelos de manera supervisada.

Para lo anterior, se realizó el etiquetado del conjunto de datos mediante el apoyo de una herramienta en la nube, que permite categorizar texto, mediante el análisis de sentimientos, presentada en el punto anterior. Esta función de análisis de sentimientos proporciona etiquetas de polaridad de sentimiento (como "negativo", "neutral" y "positivo") basadas en la puntuación de confianza más alta encontrada por el servicio a nivel de oración y documento.

Es importante recordar que los procesos de limpieza y preproceso de datos es parte esencial para la realización de los algoritmos que se estarán valorando y/o evaluando. Entre los algoritmos seleccionados de clustering se eligieron los siguientes: K-means y algunas de sus variaciones.

Los algoritmos de clustering son fundamentalmente métodos de aprendizaje no supervisados, por lo que, las métricas de evaluación del modelo resultante se deben adaptar a este enfoque. Para las comparaciones de los desempeños se realizaron cuatro configuraciones distintas, con las siguientes especificaciones: i) K-means con TF-IDF, ii) K-means con LSA y TF-IDF, iii) K-means con LSA y vectores hashed, iv) MinibachKmeans con LSA [10].

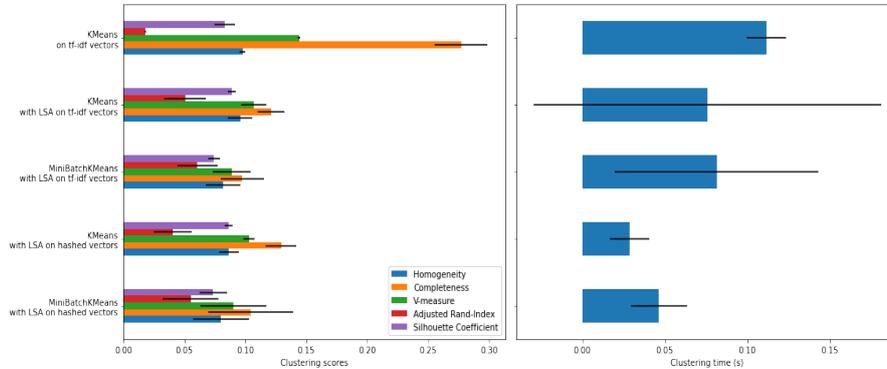


Fig. 6. Resultados de las métricas para la evaluación de los modelos de clustering con 3 clústeres.

Tabla 2. Relación de palabras clave creadas con el léxico de Sentiment Analysis VADER.

	Positive	Neutral	Negative
Num. Palabras Clave	2986	703	3813

Tabla 3. Parámetros para Lbl2Vec.

Parámetro	Valor
similarity_threshold	0.43
min_num_docs	2000
epochs	200

Se puede observar que el comportamiento del clustering con K-means no es efectivo, incluso haciendo métodos de reducción de dimensionalidad. El modelo que mejor comportamiento alcanzó fue K-means con TF-IDF, sin embargo, los valores están por debajo del valor medio óptimo.

Esto puede ser por la consistencia de los datos, donde de cierta manera existe una agrupación diferente a la que se desea, por polaridad de sentimiento y en cambio se tiene algo relacionado por los temas con los que fue recuperada la información o simplemente porque el algoritmo no es el adecuado para este problema.

Las cuatro configuraciones anteriores de clustering también fueron realizadas para 5 clústeres con el fin de ver el comportamiento de los modelos, los tiempos de ejecución son más grandes para la mayoría. En cuanto a las métricas, también hay un aumento para algunas de ellas en ciertas configuraciones, lo que podría corroborar la idea anterior, de que no necesariamente las etiquetas estas relacionadas con el comportamiento de los clústeres.

4.1. Lbl2Vec

Lbl2Vec es un algoritmo no supervisado para problemas de clasificación y recuperación de documentos [13]. Genera vectores de etiquetas, documentos y vectores de palabras automáticamente mediante la definición de palabras claves

Tabla 4. LSTM accuracy.

	Train	Test
Accuracy	0.70	0.62

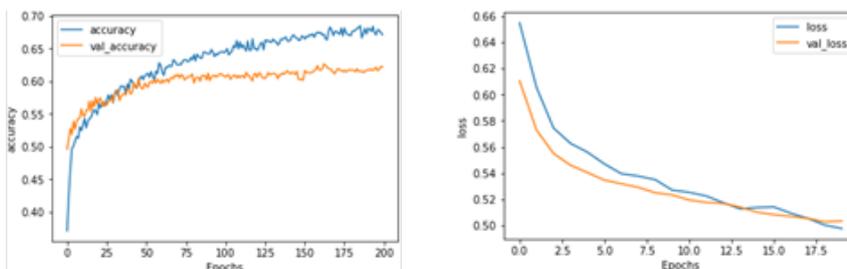


Fig. 7. Accuracy del modelo vs accuracy del conjunto de prueba y la pérdida del modelo vs la pérdida del conjunto de prueba.

(keywords) predefinidas manualmente. Está basado en la idea de que muchas keywords semánticamente similares pueden representar un tema o clase, en este caso en particular, un sentimiento.

En el primer paso, el algoritmo crea una incrustación conjunta de documentos y vectores de palabras. Una vez que los documentos y las palabras se incrustan en un espacio vectorial, el objetivo del algoritmo es aprender vectores de etiquetas a partir de palabras clave, estas previamente definidas, ocasionalmente de manera manual, donde cada conjunto de palabras clave representan un tema o clase. Finalmente, el algoritmo puede predecir la afiliación de documentos a una clase del vector de documento.

Para realizar el entrenamiento de este algoritmo se realizó una división del conjunto de datos, con el fin de evaluar el comportamiento del algoritmo mediante F1 Score. Por lo que, se definió el 30% de observaciones del conjunto de datos de prueba (Test Set) y el 70% para el conjunto de entrenamiento (Train Set).

La lista de palabras claves fue construida a partir del Lexicón de Sentiment Analysis VADER (Valence Aware Dictionary and Sentiment Reasoner) [14], mediante el indexado de la categorización de palabras: positivas, neutras y negativas. VADER es una herramienta de análisis de sentimientos basada en reglas y léxico, específicamente para los sentimientos expresados en las redes sociales, pero que incluso puede funcionar bien en textos de otros dominios.

El algoritmo se entrenó con una lista de palabras clave que contiene la relación de la tabla 2, con un número determinado de palabras clave por clase, o polaridad de sentimiento, entre las palabras también se encuentran, conjuntos de caracteres asociados a emojis, como ':)', ':|', ':/', '(-%',')' ':', ')-!:', entre muchos otros más, de igual forma están etiquetados según la polaridad de sentimiento que representan.

Para los otros parámetros del algoritmo, se definieron los valores de la tabla 3. Los resultados de este algoritmo se evaluaron con la métrica, F1 score obteniendo un valor de 0.7403697617091208, que se puede considerar un valor aceptable tomando en cuenta, que es un algoritmo no supervisado, y que el conjunto de datos fue recuperado de internet, es decir, que se ha tratado con los datos desde inicio a fin.

4.2. LSTM

LSTM es un algoritmo de aprendizaje supervisado, por lo que, se trabajó con el conjunto de datos etiquetado previamente con Azure. El primer paso por realizar para generar el modelo de LSTM es: cargar los datos, cómo estos ya han sido previamente procesados, están listos para ser utilizados en el modelo de LSTM. Primero es necesario hacer la división del conjunto de datos, en train y test.

Esta división también corresponde al 70% para el conjunto de train y 30% para el conjunto de test [15]. Antes de ingresar al modelo LSTM, los datos deben pasar por el proceso de padding y tokenización. Para más detalles del algoritmo, ver los detalles propuestos en [10].

Después de realizar pruebas con el desempeño de diferentes modelos de LSTM, este modelo fue el final, se realizó el entrenamiento con 200 épocas sobre el conjunto de datos de entrenamiento, y así mismo se realizó la validación con el conjunto de prueba, obteniendo cómo resultado los valores de la tabla 4, usando como métrica a Accuracy.

Se puede apreciar que el desempeño es mejor para el conjunto de entrenamiento, la diferencia no es mucha entre ambos valores, por lo que no se podría considerar como un problema de overfitting.

En la siguiente Fig.7 se muestra el comportamiento del accuracy y la pérdida a través del transcurso de las épocas en el entrenamiento del modelo. La curva de aprendizaje del conjunto de entrenamiento muestra que tan bien el modelo aprendió, mientras que la curva de aprendizaje del conjunto de validación representa que tan bien el modelo se comporta con nuevas observaciones.

Este comportamiento del modelo se puede deber a distintos factores, desde la integridad de los datos hasta la configuración del mismo modelo. Se realizaron pruebas incorporando más capas al modelo, pero se presentaron problemas de overfitting, lo mismo sucedió al aumentar el número de épocas.

El proceso de análisis y exploración de datos es uno de los procesos más relevantes de cualquier trabajo de aprendizaje máquina, la consistencia e integridad de los datos, es lo que define el funcionamiento de los modelos.

Para el enfoque supervisado, solo se exploró el uso de LSTM, El uso de un modelo pre-entrenado para la capa de incrustaciones de palabras o embedding layer proporcionó mayor estabilidad al modelo, esta representación permite obtener mejores relaciones semánticas de las palabras durante el entrenamiento.

Al hacer este pre-entrenamiento, se obtuvieron los vectores de incrustaciones resultantes del vocabulario del conjunto de datos, junto con las relaciones semánticas que GloVe proporciona, lo que adiciona mayor robustez al modelo.

Además, la integración del modelo en la capa oculta, con la capa bidireccional de LSTM y las dos capas densas lograron obtener un Accuracy aceptable, es cierto que los valores óptimos podrían ser más altos, pero también es interesante destacar que este conjunto de datos es completamente nuevo y no hay trabajos realizados sobre el mismo tópico, como pudiera ser para cualquier otro conjunto de datos obtenido de algún repositorio dedicado a la investigación.

5. Conclusiones y trabajo futuro

El análisis de sentimientos es un problema del PLN, que como muchos otros más problemas retoman el estado del arte de otras disciplinas, lo que hace que sea una de las áreas desafiantes de la inteligencia artificial.

El principal propósito del desarrollo de este trabajo fue la exploración de distintas alternativas de aprendizaje máquina para el análisis de sentimientos en textos no etiquetados. Esto porque la mayoría de los datos disponibles se encuentran no procesados, sin etiquetas o clasificación, generalmente distribuidos en grandes bases de datos, o cómo es el caso de este trabajo, información recuperada desde medios digitales, como redes sociales.

Por lo que, es necesario realizar estos procesos manualmente, el etiquetamiento muchas de las veces con ayuda de expertos, siendo una tarea ardua y costosa. Con el fin de buscar alternativas para realizar este proceso de una manera automatizada, se propusieron algunas soluciones basadas en aprendizaje máquina para realizar el proceso de análisis de sentimientos de manera no supervisada.

Sin embargo, para tener un punto de evaluación mediante métricas más precisas fue necesario hacer una clasificación mediante el etiquetamiento de una polaridad de sentimiento con el uso de una herramienta de la nube, que además de permitir hacer una comparación con el comportamiento de los algoritmos no supervisados, también abriera la posibilidad de emplear técnicas supervisadas, con el fin de evaluar el comportamiento de los datos en ambos escenarios.

Los algoritmos no supervisados explorados en este trabajo de tesis fueron K-means y Lbl2Vec, los cuales permitieron explorar dos tipos de técnicas no supervisadas, Clustering analysis y Self-supervised learning, respectivamente. Las métricas empleadas para la evaluación de k means y MiniBachKmeans fueron al igual que los algoritmos, no supervisadas, para obtener comparaciones justas de los resultados.

Métricas de las cuales apenas se alcanzó un resultado por debajo de la media de los resultados óptimos, de lo que se puede concluir que el clustering no es una solución óptima para el análisis de sentimientos en concreto, puede ser una estrategia interesante para descubrir conocimiento de los datos, en la etapa de análisis y exploración de datos, e incluso en el proceso de data mining, o para otros problemas semejantes como la categorización de documentos, en casos donde las etiquetas no han sido asignadas, es decir, a partir de los clústeres obtenidos, se puede definir que etiquetas se les asigna a los datos.

Después de realizar pruebas con distintas configuraciones de parámetros para Lbl2Vec, se obtuvieron buenos resultados, según el valor alcanzado con F1 Score, obteniendo un valor aproximado de 0.74, considerando que el valor máximo es 1, los resultados fueron parcialmente buenos.

Adicionalmente a esta métrica también se realizó una comparación de los resultados obtenidos con este modelo respecto de los resultados de Azure, con lo que sorprendentemente se descubrió que las etiquetas de sentimiento si cambian en un 35% de las observaciones del conjunto de datos, pero esto es claramente comprensible, pues para empezar Lbl2Vec hace la clasificación con 3 polaridades de sentimiento, mientras que Azure lo realiza con 4. Además, muy posiblemente el algoritmo de Azure este entrenado con un vocabulario considerablemente más grande al creado en la implementación de Lbl2Vec.

Una LSTM, proporciona buenos resultados para un análisis de sentimientos basado en polaridad de sentimiento, y se puede decir que también sería efectivo para un análisis de sentimientos basado en emociones, en ese caso, el modelo se modificaría en la capa de salida, aumentando el número de neuronas de acuerdo con la relación de clases en las que se clasificarían las emociones, pero dado la integridad de este conjunto de datos hasta el momento, este experimento aun no es posible, pues sería necesario tener los datos etiquetados de acuerdo a emociones.

El aprendizaje no supervisado tiene muchos desafíos importantes en la investigación, sin embargo, hay muchos modelos y algoritmos desarrollados que permiten modelar los datos y obtener buenos resultados. Alternativas como el uso de Transformers, para entrenar modelos de análisis de sentimientos que pudieran ser más robustos, sin necesidad de tener datos etiquetados, al ser modelos pre-entrenados, por lo que, también se podrían obtener deducciones importantes y sería una buena implementación a futuro, que enriquezca este trabajo.

Por su parte el aprendizaje supervisado, provee resultados más precisos, además de que hay mucha información e investigación disponible sobre la cual trabajar. A diferencia del aprendizaje no supervisado para tareas de PLN. Por lo que, las mejores estrategias para tareas de PLN en general hasta el momento se derivan de técnicas supervisadas o semi- supervisadas.

Referencias

1. Shah, C.: A hands-on introduction to data science. Cambridge University Press (2020)
2. Masmoudi, A., Hamdi, J., Belguith, L. H.: Deep learning for sentiment analysis of Tunisian dialect. *Computación y Sistemas*, vol. 25, no. 1, pp. 129–148 (2021) doi: /10.13053/cys-25-1-3472
3. Kokab, S. T., Asghar, S., Naz S.: Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, Elsevier, vol. 14 (2022) doi: 10.1016/j.array.2022.100157
4. Zhang, L., Wang, B., Liu, B.: Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4 (2018) doi: 10.1002/widm.1253
5. Gao, L. Liu X., Yin, J.: Improved deep embedded clustering with local structure preservation. *IJCAI'17*: In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 1753–1759 (2017)
6. Moolayil, J.: Learn Keras for deep neural networks: A fast-track approach to modern deep learning with Python, Apress Berkeley, CA (2018) doi: 10.1007/978-1-4842-4240-7
7. Kantardzic, M.: Data mining: Concepts, models, methods and algorithms. Wiley-IEEE Press, Chapters 1–18, pp. 360 (2003)
8. Olson, D. L.: Data mining models. Second Edition 2nd edition. Business Expert Press (2018)
9. Kwartler, T.: Text mining in practice with R. First Ed. John Wiley & Sons (2017) doi: 10.1002/9781119282105
10. Olivares L. J.: Estrategias para el análisis de sentimientos en textos extraídos de Twitter utilizando técnicas de aprendizaje profundo. Tesis de Licenciatura, FCC-BUAP (2022)
11. Embedding projector. Visualization of high-dimensional data. <https://projector.tensorflow.org/>. (2022)
12. Garcia A. J.: Comparativa de técnicas de balanceo de datos.: Aplicación a un caso real para la predicción de fuga de clientes. Tesis de Maestría. Universidad de Oviedo (2021)

Jessica Olivares L., Abraham Sánchez L., Rogelio González V., Abraham Maldonado G.

13. Schopf, T., Braun, D. Matthes, F.: Lvl2Vec: An embedding-based approach for unsupervised document retrieval on predefined topics. In: Proceedings of the 17th International Conference on Web Information Systems and Technologies WEBIST'21, pp. 124–132 (2021) doi: 10.5220/0010710300003058
14. Hutto, C., Gilbert, E.: VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International AAAI Conference on Weblogs and Social Media, vol. 8, no. 1, pp. 216–225 (2014) doi: 10.1609/icwsm.v8i1.14550
15. Lindemann, B., Müller, T., Vietz, H., Jazdi, N., Weyrich, M.: A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, vol. 99, pp. 650–655 (2021) doi: 10.1016/j.procir.2021.03.088