



Artículo Original

Análisis automático de sentimiento en tuits de política de Honduras

Automatic sentiment analysis in Honduran political tweets

Nicole Rodríguez Alcántara^{a,1} , Angella Falck Durán^a , Sergio Antonio Suazo Barahona^b 

^aFacultad de Ingeniería, Universidad Tecnológica Centroamericana, UNITEC, San Pedro Sula, Honduras

^bFacultad de Ingeniería, Universidad Tecnológica Centroamericana, UNITEC, Tegucigalpa, Honduras

Historia del artículo:

Recibido: 29 octubre 2022

Revisado: 16 noviembre 2022

Aceptado: 24 noviembre 2022

Publicado: 8 diciembre 2022

Palabras clave

Análisis de sentimiento

Aprendizaje de maquina supervisado

Política de Honduras

Procesamiento de lenguaje natural

Twitter

Keywords

Sentiment analysis

Supervised machine learning

Honduran politics

Natural language processing

Twitter

RESUMEN. Introducción. Twitter se ha convertido en un medio de expresión política ciudadana, transmitiendo sentimientos y opiniones de los usuarios mediante tuits. Analizar estos datos permite conocer tendencias y puntos de inflexión en los criterios políticos. El objetivo del estudio fue desarrollar un proceso de análisis automático de sentimiento en tuits de política de Honduras, mediante técnicas de aprendizaje de máquina supervisado. **Métodos.** Se efectuó una recolección de 1,800 tuits de política hondureña mediante filtros basados en usuarios y hashtags en un periodo de enero a septiembre de 2022, seguido de un etiquetado manual de tuits. Se aplicaron las técnicas de procesamiento de lenguaje natural: Bolsa de Palabras (BOW, siglas en inglés) y frecuencia de término-frecuencia inversa de documento (TF-IDF, siglas en inglés). Los métodos considerados fueron SVM lineal, regresión logística y Naive Bayes multinomial (MNB). Las métricas de rendimiento para la comparación de clasificadores fueron el valor-F (F1), la exactitud y los tiempos de entrenamiento y de validación. **Resultados.** El modelo seleccionado fue el MNB por su mayor frecuencia de término (F1) (62.48%) y menor tiempo de entrenamiento. Mientras que SVM lineal obtuvo 61.80% y regresión logística 61.34%. El rendimiento final del MNB con tuits nuevos fue de 63.37% de F1. **Conclusión.** Para el conjunto de datos presentado, se obtuvo que el mejor clasificador fue MNB. Sin embargo, la brecha de rendimiento entre clasificadores es reducida, lo que implica que optimizaciones de preprocesamiento y debe considerarse una colección de datos a mayor escala.

ABSTRACT. Introduction. Twitter has become a medium for citizens to express in politics, transmitting feelings and opinions of users through tweets. Analyzing this data allows to discover trends and turning points in political criteria. The study aim was to develop an automatic sentiment analysis process in Honduran political tweets, through supervised machine learning techniques. **Methods.** A collection of 1,800 Honduran political tweets was carried out through filters based in users and hashtags in the period from January to September 2022, followed by a manual tweet tagging. The following techniques of natural language processing were applied: Bag of Words (BOW) and term frequency-inverse document frequency (TF-IDF). The considered methods were: linear SVM, logistic regression and multinomial Naive Bayes (MNB). The performance metrics used to compare classifiers were a term frequency (F1-score), accuracy and time (training and validation). **Results.** The selected model was the MNB due to its higher F1-score (62.48%) and shorter training time, while linear SVM obtained 61.80% and logistic regression 61.34%. The final performance of the MNB with new tweets was an F1-score of 63.37%. **Conclusion.** For the data set presented, it was found that the best classifier was MNB. However, the performance gap between classifiers is small, which implies that preprocessing optimizations and larger scale data collection should be considered.

1. Introducción

El uso de las redes sociales para compartir opiniones y experiencias ha aumentado recientemente, emergiendo un especial interés en la política y parte de un fenómeno denominado ciberpolítica (Cotarelo, 2013). Fue a partir de las elecciones estadounidenses del 2008 que se conoció el poder de los medios digitales y su influencia en las personas debido a la campaña del expresidente Barack

Obama, pionera en el uso de tecnologías de la información y comunicación (Robertson et al., 2010). Desde entonces la mayoría de los políticos se han beneficiado del uso de las redes sociales para conectar con sus audiencias. Twitter es la plataforma predilecta para ello con presencia oficial del 97% de los 193 estados miembros de la Organización de las Naciones Unidas (ONU) (Twiplomacy, 2018). Adicionalmente, en Twitter existe una predominancia de contenido político. La cuenta más popular es del expresidente Obama con 132.3 millones de

¹ Autor correspondiente: nicole.rodriguez@unitec.edu, Universidad Tecnológica Centroamericana, Campus San Pedro Sula

Disponible en: <http://dx.doi.org/10.5377/innovare.v11i3.15349>

© 2022 Autores. Este es un artículo de acceso abierto publicado por UNITEC bajo la licencia <http://creativecommons.org/licenses/by-nc-nd/4.0>

seguidores (Kepios, 2022).

En Twitter se publican en promedio 6000 tuits por segundo, lo que corresponde a 500 millones de tuits por día (Internet Live Stats, 2022). Esto equivale a un volumen masivo de mensajes relacionados a política. Analizar estos datos manualmente es una tarea tediosa y costosa para un humano. Sin embargo, este análisis es necesario para la detección de tendencias y la identificación de patrones de comportamiento humano y cambios en las corrientes de opinión de los usuarios (Jungheer, 2015). Consecuentemente, propicia una mejor toma de decisiones, respuestas tempranas y eficientes y capacidad de pronosticar acontecimientos. Por ejemplo, la predicción electoral donde Twitter es una fuente que ha demostrado resultados aceptables y flexibilidad y variedad de métodos (Chauhan et al., 2021).

Como tal, el análisis de sentimiento consiste en “el estudio computacional de las opiniones, valoraciones y emociones de las personas hacia entidades, eventos y sus atributos” (Liu, 2010). Las principales técnicas de análisis de sentimiento están basadas en aprendizaje automático (machine learning) y en léxico (Medhat et al., 2014). El aprendizaje automático se subdivide entre supervisado y no supervisado, siendo el primero el más utilizado debido a la posibilidad de evaluar el rendimiento de los algoritmos (Baviera, 2017). En adición, el aprendizaje supervisado se subdivide en regresión y clasificación. Para el análisis de sentimiento en un tuit se suele aplicar la clasificación, donde se asigna una categoría (positivo, negativo o neutral), entre los métodos populares se encuentran Naive Bayes, máquina de vector de soporte, redes neuronales, entre otros. Investigaciones han comparado diferentes modelos de Naive Bayes, utilizando el modelo multinomial y el multivariado de Bernoulli, obteniendo que el modelo multinomial brinda mejores resultados (Singh et al., 2019).

Una pluralidad de estudios relacionados al análisis de sentimientos ha proliferado desde los 2000, principalmente por el auge de las redes sociales (Appel et al., 2015). Ha atraído la atención tanto de la academia como de la industria debido a muchos problemas desafiantes y su amplia gama de aplicaciones (Indurkya & Damerou, 2010). Sus principales aplicaciones se encuentran en sitios web de reseñas, negocios y marketing, inteligencia gubernamental, sistemas de recomendación y detección de contenido sensible, entre otros (Pang & Lee, 2008). El área de la salud también ha conducido estudios de infodemiología a través de Twitter, como ser del brote del H1N1 (Chew & Eysenbach, 2010), para conocer inquietudes de los usuarios durante la pandemia del COVID-19 (Chang et al., 2021) y la aceptación de vacunación contra el COVID-19 (Marcec & Likic, 2022), este último evaluando el sentimiento específico hacia las vacunas de Pfizer/BioNTech, AstraZeneca/Oxford y Moderna a lo largo del tiempo y empleando un método simple basado en léxico.

El presente estudio se contextualiza en la política de Honduras, la cual atraviesa un periodo de inestabilidad y un estancamiento en su desarrollo democrático. Desde las elecciones generales de 2017, se generó un creciente sentimiento de insatisfacción ciudadana con el funcionamiento de la democracia (Rodríguez, 2019). Debido a esto, se propicia el desarrollo de nuevos métodos que permitan examinar automáticamente en la opinión pública. El objetivo de esta investigación fue desarrollar un proceso de análisis automático de sentimiento en tuits de política de Honduras, mediante técnicas de aprendizaje de máquina supervisado. Considerando los métodos de máquina de vector de soporte (SVM, por sus siglas en inglés) lineal, regresión logística y Naive Bayes multinomial (MNB, por sus siglas en inglés).

2. Métodos

El proceso diseñado para el análisis del sentimiento de la política hondureña consiste en técnicas de aprendizaje automático supervisado junto con procesamiento de lenguaje natural (NLP, por sus siglas en inglés). Está constituido por una fase de entrenamiento y una de inferencia (Figura 1). Primero se efectuó la recolección de datos de Twitter, mostrando las opiniones de los usuarios en cuanto al desempeño político. Para el entrenamiento, los datos fueron sometidos a un preprocesamiento manual y un etiquetado de sentimiento: positivo, negativo o neutral. Posteriormente, se utiliza una técnica de NLP llamada frecuencia de término – frecuencia inversa de documento (TF-IDF, por sus siglas en inglés) la cual genera un vector numérico de cada tuit. Con estos vectores, se entrenaron los algoritmos automáticos supervisados de clasificación y se seleccionó al de mejor rendimiento. La inferencia consistió en utilizar los modelos previamente entrenados del extractor y el clasificador selecto para la clasificación de nuevos tuits.

2.1. Recolección de datos

La recolección de tuits se elaboró mediante la librería de código abierto Twint para Python (Zacharias, 2020) que extrae tuits de perfiles en Twitter sin limitación de datos ni proceso de autenticación. Se filtraron los tuits por medio de hashtags y usuarios específicos en un periodo de enero a septiembre de 2022. Algunos de los usuarios utilizados fueron: SalvaPresidente, pjbarquero, XiomaraCastroZ, MelZelayaR, Lredondo, Congreso_HND, PartidoLibre y HCHTelevDigital. También hashtags como Libre, Nasralla, Perro Amarillo y Congreso Nacional de Honduras. En total se recolectaron 1,800 datos.

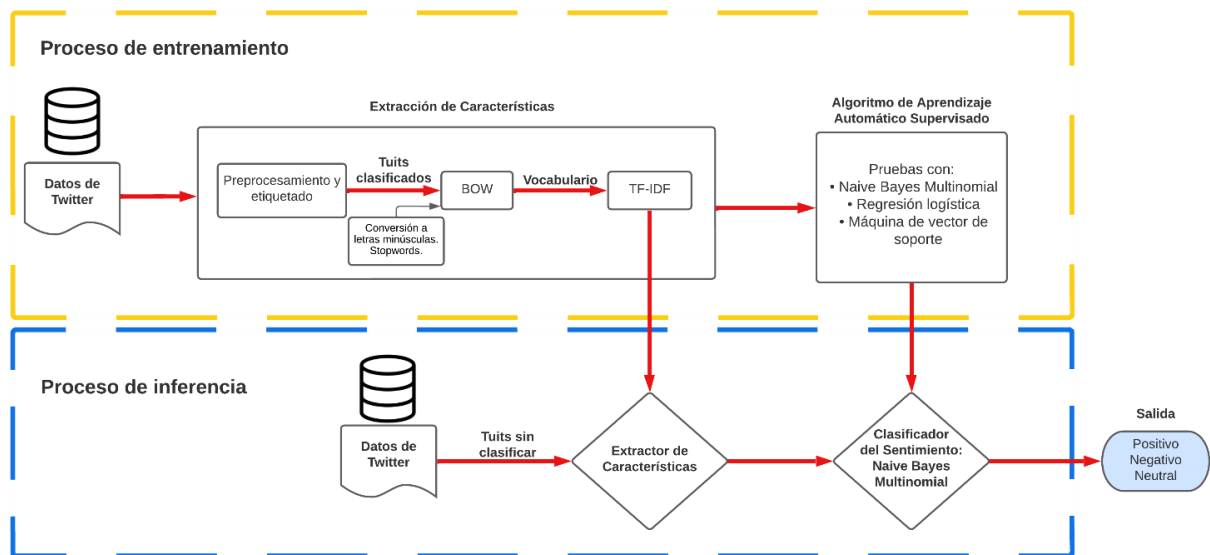


Figura 1. Diagrama del proceso de análisis de sentimiento.

2.2. Preprocesamiento de datos

Los métodos de preprocesamiento de texto mejoran la calidad de los datos para la clasificación de texto (HaCohen-Kerner et al., 2020). Existe una amplia variedad de métodos. Para esta investigación, se efectuaron manualmente los siguientes: eliminación de URLs, sustitución de emoticones por sus respectivas palabras, corrección de errores ortográficos, eliminación de texto objetivo y eliminación de tuits repetidos. Las jergas y el lenguaje soez fueron preservados, ya que se precisaba una representación genuina del habla coloquial hondureña.

Adicionalmente, se realizó el etiquetado manual de los datos para clasificar cada tuit con su respectivo sentimiento (positivo, negativo o neutral). Identificar el sentimiento de un mensaje consiste en la lista de palabras positivas o negativas que lo conforman refiriéndose hacia una entidad en particular (Hu & Liu, 2004). Por ello, el etiquetado se realizó como una clasificación global del sentimiento del tuit mediante sus palabras clave. Si un tuit tiene palabras como “malo”, “corrupto”, o alguna palabra soez, tiene tendencia a ser un tuit negativo. Mientras que palabras como “bien”, “excelente” o “viva” se inclinan a tuits positivos. Los tuits neutrales son textos objetivos, con escasez de palabras clave ni un sentimiento en particular (Figura 2).

2.3. Extracción de características

Se utilizó la implementación Tfidf Vectorizer de Scikit-Learn (Pedregosa et al., 2011), configurada para realizar una conversión de letras mayúsculas a minúsculas y eliminación de stopwords. La clase Tfidf Vectorizer convierte una colección de documentos sin procesar a un

vector de características TF-IDF. Esta es equivalente a realizar primero una bolsa de palabras (BOW, por sus siglas en inglés) y a este calcularle el TF-IDF.

2.3.1. Eliminación de stopwords

Un listado de 306 *stopwords* o palabras frecuentes del español (por ejemplo, “a”, “uno”, “el”) fueron tomadas de <https://kaggle.com/code/mpwolke/spanish-stopwords-w2v>. Estas se incluyeron en la configuración de BOW para su eliminación en los tuits.

2.3.2. BOW

El modelo tradicional de clasificación de textos se basa en la representación de bolsa de palabras. Se asocia un texto con un vector que indica el número de ocurrencias de cada palabra elegida en el conjunto de documentos de entrenamiento, aquí el orden de las palabras se pierde (Sebastiani, 2002). El resultado fue un vocabulario de 6348 palabras para el conjunto de entrenamiento.

2.3.3. TF-IDF

La frecuencia de término (TF) considera el número de ocurrencias f de un término t en un documento d :

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t \in d} f_{t,d}} \tag{1}$$

La frecuencia inversa del documento (IDF) es una medida de cuánta información proporciona un término t , teniendo un total de N documentos y $|\{d \in D: t \in d\}|$ número de documentos donde t aparece, se obtiene mediante:

$$idf(t, D) = \log \frac{N}{|\{d \in D: t \in d\}|} \tag{2}$$

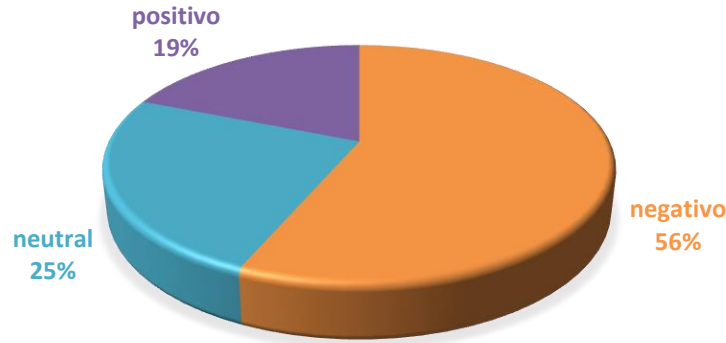


Figura 2. Gráfica de distribución original del sentimiento (positivo, negativo o neutral) de los datos recolectados.

Entonces TF-IDF se calcula como:

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (3)$$

Esta medida permite ciertas cualidades (Sebastiani, 2002). Primero, a mayor ocurrencia de un término en un documento, mayor peso tendrá. Segundo, a mayor número de documentos donde el término ocurre, el término se vuelve menos discriminativo. El resultado fue un vector de 6,348 características TF-IDF para cada documento del conjunto de entrenamiento.

2.3.4. TF-IDF

La frecuencia de término (TF) considera el número de ocurrencias f de un término t en un documento d :

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

La frecuencia inversa del documento (IDF) es una medida de cuánta información proporciona un término t , teniendo un total de N documentos y $|\{d \in D: t \in d\}|$ número de documentos donde t aparece, se obtiene mediante:

$$idf(t, D) = \log \frac{N}{|\{d \in D: t \in d\}|} \quad (2)$$

Entonces TF-IDF se calcula como:

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (3)$$

Esta medida permite ciertas cualidades (Sebastiani, 2002). Primero, a mayor ocurrencia de un término en un documento, mayor peso tendrá. Segundo, a mayor número de documentos donde el término ocurre, el término se vuelve menos discriminativo. El resultado fue un vector

de 6,348 características TF-IDF para cada documento del conjunto de entrenamiento.

2.4. Técnicas de aprendizaje automático

Se consideraron tres técnicas de aprendizaje automático supervisado: SVM lineal, regresión logística y MNB. Para las implementaciones de los tres algoritmos se utilizó la librería de Scikit-Learn (Pedregosa et al., 2011).

2.4.1. SVM

Es un clasificador no probabilístico que clasifica los puntos mediante un hiperplano dimensional con el mayor margen posible. Es decir, el que mejor separa los conjuntos de objetos con diferentes clases de pertenencia (Devika et al., 2016). Se manejó un soporte lineal, variando el parámetro C que determina el margen de error permitido.

2.4.2. Regresión logística

Es un clasificador que mide la relación entre la variable dependiente categórica y una o más variables independientes al estimar la probabilidad de ocurrencia de un evento usando su función logística. Se realizaron variaciones de su parámetro de regularización C .

2.4.3. MNB

Es un clasificador probabilístico basado en el teorema de Bayes (Devika et al., 2016). La probabilidad condicional de que un tuit T sea de una clase C (positiva/negativa/neutral) está determinada por:

$$P(C|T) = \frac{P(C)P(T|C)}{P(T)} \quad (4)$$

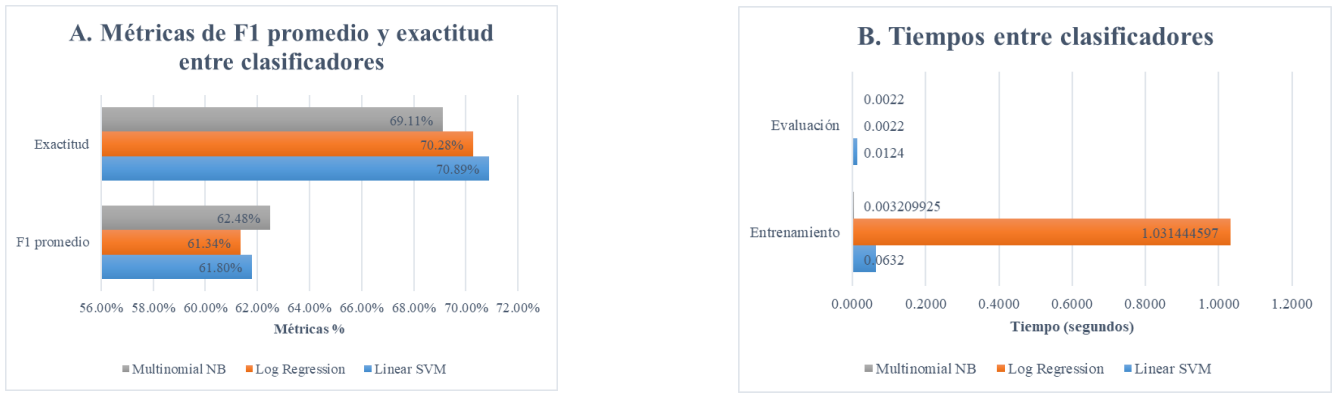


Figura 3. Gráfica de resultados de validación cruzada para la mejor variación de parámetro de cada clasificador.

La implementación empleó clasificador Naive Bayes con distribución multinomial, con variaciones en el parámetro de suavizado aditivo alfa.

3. Resultados

3.1. Experimentos entre clasificadores

3.1.1. Validación cruzada

Para la experimentación entre los tres clasificadores, se utilizó un procedimiento llamado validación cruzada. Consiste en dividir la muestra total en n particiones, n – 1 se utilizan como datos de entrenamiento y 1 como datos de validación para cada combinación posible de n (Russell & Norvig, 2021). Este método es útil cuando el tamaño de datos es muy reducido para dividirlo en entrenamiento y validación. Su implementación fue mediante el módulo cross_validate de Scikit-Learn (Pedregosa et al., 2011). Se utilizaron cinco particiones y el total de 1,800 datos recolectados, dividiendo cada partición en 360 datos.

3.1.2. Métricas de rendimiento

Para comparar el rendimiento entre los clasificadores se utilizaron las métricas de F1, exactitud, tiempo de entrenamiento y tiempo de validación. La métrica prioritaria en la selección de los clasificadores es el F1, puesto que la exactitud es insensible a la representación de cada clase dándole baja prioridad a las clases menos representativas (Sebastiani, 2022). El F1 es la media armónica de la precisión y sensibilidad, de tal forma que:

$$F1 = 2 \cdot \frac{Precisión \cdot Sensibilidad}{Precisión + Sensibilidad} \quad (5)$$

3.1.3. Configuraciones de clasificadores

Para el clasificador SVM lineal se probaron valores para su parámetro C de 0.01, 0.1, 0.3, 0.5, 0.7, 0.8, 0.9, 1, 5, 10 y 100. El mejor valor de C para el conjunto de datos

brindado fue de 0.8. Con el clasificador de regresión logística las variaciones de su parámetro C correspondieron a los valores 1, 10, 50, 70, 80, 90, 100 y 150. El valor de C=80 tuvo el mejor rendimiento. Para MNB se configuraron los valores 0.01, 0.09, 0.1, 0.2, 0.5 y 1 para su parámetro alfa. Fue el valor de 0.1 el que obtuvo mejores resultados. La Figura 3 muestra los resultados de los clasificadores con su mejor configuración.

3.2. Evaluación con MNB

Utilizando el clasificador MNB con alfa de 0.1 se dividió el conjunto total de datos, utilizando 1,488 como datos de entrenamiento y 312 como datos de validación. Para la evaluación del modelo final, se generó una matriz de confusión (Figura 4) y las métricas de precisión, sensibilidad, F1 y exactitud (Cuadro 1). Para esto se empleó el módulo sklearn.metrics de Scikit-Learn (Pedregosa et al., 2011) que contiene el método confusion_matrix, la cual definen como “una matriz de confusión C es tal que Ci,j es igual al número de observaciones que se sabe que están en el grupo i y predijo estar en grupo j”. También el método classification_report, el cual imprime un informe con las principales métricas de clasificación.

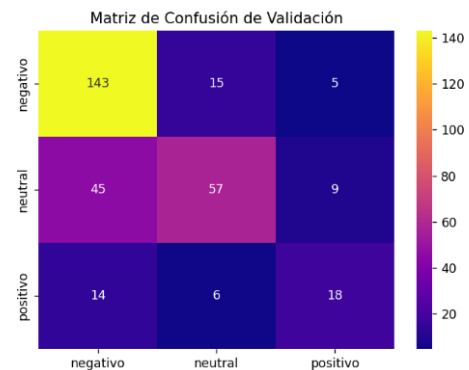


Figura 4. Matriz de confusión del clasificador de Naive Bayes multinomial en los datos de validación.

Cuadro 1

Métricas de rendimiento del clasificador Naive Bayes multinomial en datos de validación.

Clase	Precisión	Sensibilidad	F1	Exactitud
Negativo	70.79%	87.73%	78.36%	
Neutral	73.08%	51.35%	60.32%	
Positivo	56.25%	47.37%	51.43%	
Macro Promedio	66.71%	62.15%	63.37%	69.87%

3.3. Análisis

Los resultados de la experimentación entre clasificadores muestran el comportamiento de cada algoritmo con el conjunto de datos recolectado. Analizando las métricas de rendimiento, se pudo notar que la métrica de F1 y exactitud no convergen. Es decir, la mejor exactitud la tuvo el clasificador SVM con una diferencia de 1.78% del MNB, el peor resultado. Mientras que en el F1 el clasificador de MNB tuvo el mejor rendimiento con una brecha de 1.14% del clasificador de regresión logística. Ya que la métrica prioritaria es el F1, se obtuvo que el mejor clasificador para el conjunto de datos presentado fue el clasificador MNB con un parámetro de alfa de 0.1.

Adicionalmente, se compararon los tiempos de entrenamiento y validación entre los clasificadores. El tiempo de entrenamiento difirió notoriamente entre los clasificadores. El tiempo del MNB fue equivalente aproximadamente al 5% del tiempo del SVM, mientras que el tiempo de regresión logística fue aproximadamente 16 veces más que el tiempo de SVM. El tiempo de validación se mantuvo cercano, con un promedio de 0.0056 segundos.

La experimentación entre los clasificadores de SVM, regresión logística y MNB produjo dos consideraciones relevantes. Primero, la importancia de afinar los parámetros de cada modelo para obtener mejor rendimiento. Segundo, si bien el clasificador MNB fue el selecto, hay una reducida brecha entre los demás clasificadores. Esto concuerda con estudios que presentan técnicas alternativas de aprendizaje no supervisado, con exactitudes similares entre algoritmos de agrupamiento jerárquico (74.5%) y MNB (76.25%) (Bibi et al., 2022). Esto alienta que es posible seguir explorando diferentes métodos, acorde a los datos y al contexto específico de cada estudio.

La evaluación del clasificador MNB obtuvo una matriz de confusión con tendencia a acertar tuits negativos. Mientras que los tuits neutrales y positivos se reconocieron con menor precisión. Por cada tuit negativo, existió una posibilidad del 87.73% de darle una predicción como negativo. Por cada tuit positivo, existió una

posibilidad del 47.37% de darle una predicción como positiva. Y por cada tuit neutral, el 51.35% fue predicho como neutral. El F1 final de MNB fue de 63.37% para el conjunto de tuits obtenido y tuvo un 69.87% de clasificaciones correctas (exactitud). Se observó que la clase negativa tuvo mejor F1 (78.36%), seguida de la neutral (60.32%) y finalmente la positiva (51.43%). Esto se relacionó a la distribución original de los datos de entrenamiento. Para fines prácticos, obtener mejor rendimiento al clasificar negativos es preferible, puesto que alerta de mayores amenazas o posibles acontecimientos futuros.

4. Discusión

El análisis de sentimientos en Twitter es un dominio novedoso y desafiante dada su naturaleza informal y limitaciones en los mensajes de la plataforma (Giachanov & Crestani, 2016). Los hallazgos principales del estudio consisten en la exploración de métodos de aprendizaje automático y NLP para el habla hondureña, provee una línea base para aplicaciones y estudios variados en el contexto nacional y comprueba el potencial de las técnicas automatizadas para el análisis masivo de datos en medios digitales.

Es importante denotar que los resultados de cada método dependen del conjunto de datos utilizado y su lenguaje. Actualmente, existe una escasez de investigaciones del habla español. Estudios han recalado problemas de portabilidad entre idiomas cuando el conjunto de datos es reducido y se utiliza lenguaje coloquial (Boiy & Moens, 2009). Por ello, se debe continuar desarrollando y optimizando el proceso de análisis de sentimiento para el contexto hondureño volviendo cada etapa más robusta, para beneficiar diversas áreas de interés social, como ser la política.

El presente trabajo es una línea base para investigaciones futuras, donde se espera incorporar mejoras. Primero, una recolección de datos intensiva para obtener un mayor tamaño de muestra y beneficiar al entrenamiento de los clasificadores. Seguido, optimizar el preprocesamiento con métodos más sofisticados como la lematización para mejor caracterización de los datos. En

esta misma etapa, se podría incorporar la preservación de emoticones puesto que encapsulan sentimientos y reacciones. Un problema frecuente es el uso del lenguaje figurativo que utiliza recursos como sarcasmo, ironía y metáforas, los cuales pueden afectar la clasificación del sentimiento (Ghosh et al., 2015). Como solución, se sugiere incorporar modelos basados en léxico de detección de ironía que consideran contenido emocional y psicológico (Hernández-Farías et al., 2016). Incorporando estos trabajos futuros, el proceso propuesto se vuelve más robusto y óptimo. Brindando un programa útil en diversas aplicaciones políticas como ser análisis de satisfacción ciudadana, predicciones electorales e identificación de amenazas.

La principal delimitación del enfoque propuesto es el requerimiento de un etiquetado manual de los datos de entrenamiento, lo que restringe el tamaño de muestra disponible. Existen bases de datos previamente generadas con este fin, pero no en el contexto de Honduras. Otra alternativa para ello es utilizar técnicas de aprendizaje no supervisado, sin embargo, esto no permitiría medir el rendimiento.

5. Conclusión

Se desarrolló un proceso de análisis automático de sentimiento en tuits de política de Honduras, mediante la comparación del rendimiento de técnicas de aprendizaje de máquina supervisado. Estos métodos incluyeron SVM lineal, regresión logística y MNB. El proceso presentado abarca la recolección de datos, el etiquetado y preprocesamiento de datos, el procesamiento del lenguaje natural con BOW y TF-IDF y los métodos automatizados. Concluyendo que el MNB fue el método más efectivo para el conjunto de datos recolectado por sus resultados de F1 y reducido tiempo de entrenamiento. Sin embargo, la brecha de rendimiento entre clasificadores fue reducida.

6. Contribución de los Autores

AF y SS realizaron la recolección, preprocesamiento y etiquetado de datos. NR hizo la experimentación y elaboración del manuscrito. Todos los autores leyeron y aprobaron la última versión del manuscrito.

7. Reconocimientos

Al Ph.D. Kenny Dávila por su asesoría en el desarrollo del proyecto y elaboración del manuscrito.

8. Conflictos de Interés

Los autores declaran no tener ningún conflicto de interés.

9. Referencias Bibliográficas

- Appel, O., Chiclana, F., & Carter, J. (2015). Main concepts, state of the art and future research questions in sentiment analysis. *Acta Polytechnica Hungarica*, 12(3). <https://dx.doi.org/10.12700/APH.12.3.2015.3.6>
- Baviera, T. (2017). Técnicas para el análisis del sentimiento en Twitter: aprendizaje automático supervisado y sentistrength. *Dígitos*, 1(3), 33-50. <https://revistadigitos.com/index.php/digitos/article/view/74>
- Bibi, M., Abbasi, W. A., Aziz, W., Khalil, S., Uddin, M., Iwendi, C., & Gadekallu, T. R. (2022). A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis. *Pattern Recognition Letters*, 158. <https://dx.doi.org/10.1016/j.patrec.2022.04.004>
- Boiy, E., & Moens, M.-F. (2009). A machine learning approach to sentiment analysis in multilingual Web texts. *Information Retrieval*, 12, 526-558. <https://dx.doi.org/10.1007/s10791-008-9070-z>
- Chang, C.-H., Monselise, M., & Yang, C. C. (2021). What are people concerned about during the pandemic? Detecting evolving topics about COVID-19 from Twitter. *Journal of Healthcare Informatics Research*, 5, 70-97. <https://dx.doi.org/10.1007/s41666-020-00083-3>
- Chauhan, P., Sharma, N., & Sikka, G. (2021). The emergence of social media data and sentiment analysis in election prediction. *Journal of Ambient Intelligence and Humanized Computing*, 12, 2601-2627. <https://dx.doi.org/10.1007/s12652-020-02423-y>
- Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLOS ONE*, 5(11), e14118. <https://dx.doi.org/10.1371/journal.pone.0014118>
- Cotarelo, R. (2013). *Ciberpolítica. Las nuevas formas de acción y comunicación políticas* (1er. ed.). Tirant Humanidades.
- Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: a comparative study on different approaches. *Procedia Computer Science*, 87, 44-49. <https://dx.doi.org/10.1016/j.procs.2016.05.124>
- Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., & Reyes, A. (2015). Sentiment analysis of figurative language in Twitter. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 470-478. <https://dx.doi.org/10.18653/v1/S15-2080>
- Giachanou, A., & Crestani, F. (2016). Like it or not: a survey of Twitter sentiment analysis methods. *ACM Computing Surveys*, 49(2), 1-41. <https://dx.doi.org/10.1145/2938640>
- HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLOS ONE*, 15(5), e0232525. <https://dx.doi.org/10.1371/journal.pone.0232525>
- Hernández-Farías, D. I., Patti, V., & Rosso, P. (2016). Irony detection in Twitter: the role of affective content. *ACM Transactions on Internet Technology*, 16(3), 1-24. <https://dx.doi.org/10.1145/2930663>
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 168-177. <https://dx.doi.org/10.1145/1014052.1014073>
- Indurkha, N., & Damerou, F. J. (2010). *Handbook of natural language processing* (2do. ed.). Taylor & Francis.
- Internet Live Stats. (2022). *Twitter Usage Statistics*. <https://www.internetlivestats.com/twitter-statistics/>
- Jungherr, A. (2015). *Analyzing political communication with digital trace data: The role of twitter messages in social science research*. Springer.
- Kepios. (2022, julio). *The k Twitter statistics: everything you need to now*. DataReportal – Global Digital Insights. <https://datareportal.com/essential-twitter-stats>

- Liu, B. (2010). Sentiment analysis: A multi-faceted problem. *IEEE Intelligent Systems*, 25(3), 76-80.
- Marcec, R., & Likic, R. (2022). Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgraduate Medical Journal*, 98, 544-550. <https://dx.doi.org/10.1136/postgradmedj-2021-140685>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: a survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113. <https://dx.doi.org/10.1016/j.asej.2014.04.011>
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1-2), 1-135. <https://dx.doi.org/10.1561/15000000011>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- Robertson, S. P., Vatrappu, R. K., & Medina, R. (2010). Off the wall political discourse: Facebook use in the 2008 U.S. presidential election. *Information Polity*, 15(1-2), 11-31. <https://dx.doi.org/10.3233/IP-2010-0196>
- Rodríguez, C. G. (2019). *Panorama político hondureño. (In)satisfacción con el funcionamiento de la democracia y (des)confianza en las instituciones*. Instituto Español de Estudios Estratégicos. Gobierno de España. https://www.ieee.es/publicaciones-new/documentos-de-opinion/2019/DIEEEO06_2019CECROD-Honduras.html
- Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: a modern approach* (4ta ed.). Pearson.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1-47. <https://dx.doi.org/10.1145/505282.505283>
- Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, 593-596. <https://dx.doi.org/10.1109/ICACTM.2019.8776800>
- Twiplomacy. (2018, 10 de julio). *Twiplomacy Study 2018*. <https://www.twiplomacy.com/twiplomacy-study-2018>
- Zacharias, C. (2020). *Twint: an advanced Twitter scraping & OSINT tool. (2.1.20) [Python; Twint]*. MIT License. <https://github.com/twintproject/twint>