ISBN: 978-628-96613-1-6. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2025.1.1.1502

Unmasking False News in the Twitterverse: Precision-Optimized Classification Algorithms for Verifying Information in Peru

Sandro Castillo[1]; Paul Tocto[2]
[1,2]*Universidad Nacional de Ingeniería, Perú, sandro.castillo.a@uni.pe, ptocto@uni.edu.pe*

***Abstract*– The Internet has the problem known as the spread of fake news. Based on this, the authors have chosen the social network Twitter to be studied because it is known as the medium in which false news is frequently disseminated. Therefore, this investigation elaborates on a data set using natural language processing. It comprises 1600 tweets classified as true or false according to their content and based on news verification articles from Perú. With this data set, four classification models are designed with high precision to identify if a tweet is true or false, using first Natural Language Processing, Logistic Regression, Support Vector Machine, Dense Neural Network, and Random Forests algorithms. Then, the hyperparameters of all algorithms are tuned. Finally, after the performance evaluation of the classification models, the authors recommend the Support Vector Machine as the best algorithm.**
***Keywords*-- Detection, Fake News, LR, NLP, SVM..**

Desenmascarando noticias falsas en el Twitterverso: Algoritmos de clasificación optimizados con el fin de verificar la información en el Perú

Sandro Castillo[1]; Paul Tocto[2]
[1,2]Universidad Nacional de Ingeniería, Perú, *sandro.castillo.a@uni.pe, ptocto@uni.edu.pe*

Resumen– ***Internet presenta el problema conocido como la difusión de noticias falsas (o 'fake news'). En base a esto, los autores han escogido la red social Twitter para su estudio, ya que es conocida como un medio en el que se diseminan frecuentemente noticias falsas. Por lo tanto, esta investigación trabaja con un conjunto de datos utilizando procesamiento de lenguaje natural. Comprende 1600 tweets clasificados como verdaderos o falsos de acuerdo a su contenido y contrastándolos con artículos de verificación de noticias de Perú. Con este conjunto de datos, se diseñaron cuatro modelos de clasificación con alta precisión para identificar si un tweet es verdadero o falso. Estos modelos aplican procesamiento de lenguaje natural y algoritmos de: regresión logística, máquinas de soporte vectorial (SVM), redes neuronales densas y bosques aleatorios (Random Forests). Luego, se optimizaron los hiperparámetros de todos los algoritmos. Finalmente, tras una evaluación de rendimiento de los modelos de clasificación, los autores recomiendan las máquinas de soporte vectorial (SVM) como el mejor algoritmo.***

Palabras clave-- Detección, Noticias falsas, LR, NLP, SVM..

I. Introducción

Las consecuencias de las noticias falsas se han hecho evidentes en países como Paraguay, donde la propia sociedad ha generado la transmisión de información falsa que ha contribuido a que gran parte de la comunidad nacional caiga en un estado de neurosis, provocando todo tipo de acciones, como compras compulsivas, discriminación hacia sujetos de otras nacionalidades, xenofobia, incitación al vandalismo, consumo inadecuado de medicamentos, prácticas inusuales e indeseables de atención médica, entre otros. En Paraguay se encontraron noticias falsas sobre el COVID-19 en 2020; algo similar ocurrió en Bolivia en 2019 en el contexto de sus elecciones generales. Así, la desinformación digital, como fenómeno nuevo y negativo de los medios digitales, puede dañar aún más nuestra democracia.[1][2]

La problemática incluye al periodismo como una de sus variables, en el que sus actores podrían ser sancionados por promover su crecimiento. Sin embargo, supongamos que la imputación de responsabilidad civil no va precedida de una ponderación reflexiva de los bienes jurídicos en un conflicto, que permita decidir si la empresa periodística actuó en la difusión de información a pesar de no haber corroborado su veracidad. En ese caso, ello implicará un riesgo mucho más grave para nuestra democracia: desalentar la actividad periodística en asuntos de indudable interés público. En otras palabras, la publicación de noticias debe ir acompañada de un marco regulatorio que permita identificar la intención con la que se llevan a cabo estas acciones para permitir que la prensa complete su trabajo sin temor a medidas excesivas. Desde otra perspectiva, la crisis del periodismo se agravó durante la pandemia que aún vivimos y se ha magnificado con la intensificación de la digitalización de la información. Ante esta afirmación, las redes sociales pueden considerarse parte de la segunda razón, ya que son plataformas digitales de difusión de información.[3][4]

Así, en respuesta a las noticias falsas, nacieron los proyectos de fact-checking, que consisten en desmentir rumores o noticias no fácticas y crear conciencia en la ciudadanía. Sin embargo, el ciudadano es consciente del entorno actual de desinformación y noticias falsas pero lo sitúa en el contexto de la corrupción y la mala calidad del periodismo, la abundancia de información, los políticos que mienten y la polarización ciudadana. Por lo tanto, el fact-checking debe acercarse al público para obtener mejores resultados.[5]

Además, los verificadores de datos destacan el trabajo colaborativo de las plataformas de redes sociales para abordar la desinformación y señalar afirmaciones engañosas de individuos específicos a través de las mismas redes sociales. Ante esta afirmación, es fundamental destacar que las redes sociales pueden ser una fuente de noticias falsas, por lo que es un entorno propicio para el estudio de esta cuestión.[6]

Además de esto, la aparición de nuevos medios, como Facebook o Twitter, y su inmediatez (compartir o retuitear) no permiten a las personas evaluar críticamente la información que se publica porque no hay un período necesario para la reflexión o revisión de las fuentes. Es así como el concepto de inmediatez puede considerarse un factor relevante que conduce a las noticias falsas en las redes sociales. Además, Twitter es una de las redes sociales que puede ser espacio para la difusión de noticias falsas. Un estudio mostró que los picos más altos de tuits que contenían información errónea se encontraron a mediados de marzo de 2020, mientras que los más bajos se encontraron en julio de 2020 .[7][8]

Asimismo, en el estado de Kano, Nigeria, después de realizar un estudio, el número de personas que estuvieron expuestas a noticias falsas sobre COVID-19 en las redes sociales fue significativamente alto en el estado de Kano. Además, la mayoría de los usuarios de las redes sociales utilizaron la plataforma de medios Facebook para obtener información sobre el Covid-19. Afirmaron que es seguro argumentar que la reciente pandemia exacerbó el problema de la desinformación y que hubo casos de uso de las redes sociales para difundir noticias falsas sobre la pandemia.[9]

Por otro lado, es fundamental tener en cuenta que el problema de las noticias falsas no termina con la verificación de las noticias falsas; se extiende a cuestiones más profundas.

Debemos establecer una cultura que equilibre a los creadores de noticias falsas, tanto en número como en competencia, a través de un estándar moral. Por otra parte, la construcción de criterios para evaluar la información, lejos de ser una fórmula precisa e infalible, es solo una orientación inicial en un proceso mucho más amplio que involucra la construcción del conocimiento a través de la lectura y el pensamiento crítico.[10][11]

Finalmente, los usuarios peruanos requerían herramientas adicionales, ya que abordar el problema de la desinformación exige esfuerzos renovados y conjuntos de diferentes campos de la informática y otras disciplinas relacionadas, como las ciencias sociales, políticas y psicológicas, que puedan ser utilizadas para detectar de manera inmediata la verdad o falsedad de las noticias en una de las redes sociales más utilizadas, Twitter.[12]

Algunos investigadores proponen un modelo basado en el algoritmo Naive-Bayes para crear un clasificador de noticias en falso y real utilizando un conjunto de datos público. Realizaron dos experimentos y concluyeron que el uso de la técnica de n-grama mejora el rendimiento del modelo, ya que obtuvieron un aumento en la métrica del área bajo la curva (AUC) del 91,2% al 93,1%. Otros investigadores utilizan Naive-Bayes como otros modelos, con un trasfondo adicional en el que se realizaron una serie de pruebas en ocho conjuntos de datos diferentes que contenían diferentes atributos. Para cada uno de ellos, evaluaron tres modelos: un modelo basado en el algoritmo Naive-Bayes, un modelo basado en el clasificador pasivo-agresivo y un modelo de red neuronal densa. Finalmente, se concluyó que el machine learning aplicado a la detección de noticias falsas apenas está comenzando y es el camino hacia un internet libre de noticias falsas. Así, algunos estudios apoyan el uso de modelos de clasificación, como el modelo de redes neuronales, para resolver el problema de la detección de noticias falsas.[13][14]

Se elaboró un modelo multiclase a partir de múltiples fuentes para la detección de noticias falsas basado en redes neuronales convolucionales y memoria prolongada a corto plazo. El resultado fue una precisión del 38,81% en las pruebas realizadas, superando así a otros modelos como Support Vector Machine, Random Forests y Neural Networks. En segundo lugar, las noticias no tienen por qué clasificarse necesariamente como falsas o reales, ya que algunas pueden ser discutibles, o sus titulares no están relacionados con el tema en cuestión. Por un lado, se puede conceptualizar la detección de postura como el proceso de detección automática de la relación entre dos fragmentos de texto. Del mismo modo, su trabajo se centró en la relación entre los titulares de las noticias y su contenido para clasificar los titulares como agradables, discutibles, desagradables y no relacionados. De esta forma, realizaron un modelo basado en redes neuronales densas con una precisión del 94,21%. Por otro lado, se propuso una arquitectura sencilla basada en codificación condicional bidireccional para codificar un artículo dividido en "n" oraciones. Además, se utilizó el conjunto de datos "Fake News Challenge", que clasifica los titulares de las noticias como agradables, discutibles, desagradables y no relacionados. Los resultados muestran que el mejor rendimiento de su modelo es cuando utilizan el modelo de atención coincidente aprovechando solo las incrustaciones de palabras, obteniendo así un valor de precisión del 64,14%, un valor de completitud del 58,53% y un valor F1 del 59,01%. En tercer lugar, una forma de mejorar la precisión de los modelos es aumentar los datos que se van a utilizar. En este sentido, se desarrolló un modelo basado en redes neuronales densas basado en tres dimensiones: creadores, contexto social y contenido noticioso, que logra mejores precisiones que trabajos similares, alcanzando niveles superiores al 85%. Y, por último, también se desarrolló el modelo predictor de noticias falsas utilizando un conjunto de datos extraído de la red social Twitter durante las protestas de Hong Kong de 2019. Este modelo basado en redes neuronales logra una precisión del 100 % y del 98 % para el conjunto de datos descrito. Por lo tanto, para obtener una solución a la detección de noticias falsas, requerimos un conjunto de datos extenso y la selección de modelos de clasificación. De la misma manera, métricas como la precisión, la integridad, el valor F1 y la exactitud se utilizan para medir el rendimiento de la clasificación. [15] [16][17][18][19]

Resumiendo todo lo anterior, utilizando técnicas de inteligencia artificial, la solución a la detección de noticias falsas puede darse durante el proceso de detección o como una clasificación de estas. Se requiere un conjunto de datos para construir el modelo y relacionar las operaciones sobre él en función de su contenido, ya que puede ser textual o visual. La clasificación se realiza a través de modelos como Naive Bayes, árboles de decisión y redes neuronales densas. Por último, se determinan las métricas para entender si la detección se logra con un alto rendimiento.

En consideración a lo anterior, esta investigación tiene como objetivo diseñar cuatro modelos de clasificación, entrenados con un conjunto de datos de noticias en español, para asignar si una noticia es falsa o verdadera utilizando el procesamiento del lenguaje natural y los siguientes algoritmos de aprendizaje automático: LR, SVM, DNN y RF.

Finalmente, los hallazgos más relevantes del proceso son que el modelo de máquina de vectores de soporte (SVM) tuvo el mejor rendimiento con una exactitud del 91%, superior a los modelos LR, DNN y RF. Por otro lado, se halló que la eliminación del carácter numérico (#) mejora la precisión del modelo. Además, se demostró que la retención de palabras con una frecuencia superior a cinco no mejora la precisión del modelo en al menos un 1%.

II. METODOLOGÍA

A. Flujo de trabajo

El flujo de trabajo utilizado en este trabajo se muestra en la Fig. 1. Como primer paso, se recogerán datos sobre los tuits con contenido relacionado con las noticias, clasificándolos como verdaderos o falsos. Posteriormente, los datos recopilados pasarán por la etapa de preparación, que será normalizada y tokenizada. A continuación, se seleccionarán los modelos

utilizados en la experimentación. Los cuatro modelos de clasificación se construirán como penúltimo paso, y se seleccionarán sus mejores parámetros. Finalmente, se evaluará el rendimiento de los cuatro modelos propuestos para encontrar el mejor.

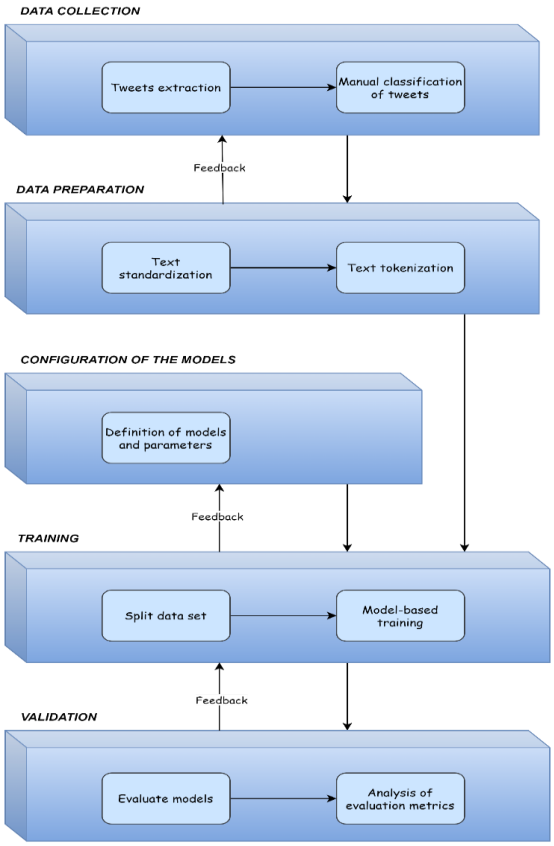

Fig. 1 Flujo de trabajo.

1) Recogida de datos: En esta fase, los tweets se extraen utilizando artículos de sitios de verificación de noticias como VerificadorLR y el motor de búsqueda Twitter.

Estos artículos se caracterizan por anunciar que un artículo o afirmación es falsa o real. Además, fundamentan por qué afirman algo verdadero o falso basándose en información existente o inexistente. Por ejemplo, en la Figura 2, el artículo menciona que no hubo declaraciones oficiales de Joe Biden o del gobierno de Estados Unidos dando una posición sobre la situación económica en el Perú. Sin embargo, hubo una reunión entre el ex candidato presidencial Hernando De Soto y el senador Ted Cruz en la que no se mencionó la preocupación seria por la economía peruana.

Fig. 2 Verificación: Es falso que Joe Biden haya anunciado un "cataclismo económico" en el Perú, como sugiere el viral

Conociendo estos artículos, las búsquedas se realizan utilizando palabras clave extraídas del artículo para encontrar tweets relacionados. Una vez que se encuentra este grupo, los tweets se extraen utilizando la API de Twitter y la biblioteca de Tweepy, junto con su información relacionada, como el contenido, los enlaces y la fecha. Al final, se obtiene un conjunto de tweets que serán clasificados como naturales o falsos.

TABLA 1
TAMAÑO DEL CONJUNTO DE DATOS PARA EL FONDO DE REFERENCIA

Título	Autores	Tamaño del conjunto de datos
Técnica de aprendizaje semisupervisada mejorada para la detección automática de lenguaje abusivo sudafricano en Twitter	Oriola y Kotzé, 2020	10,245
Aprendizaje profundo para la detección de noticias falsas en un esquema de entrada textual por pares.	Mouratidis, Nikiforos y Kermanidis, 2021	2,363
Cross-SEAN: Modelo de atención neuronal semisupervisada en punto de cruz para la detección de noticias falsas sobre COVID-19.	Paka, Bansal, Kaushik, Sengupta y Chakraborty, 2021	45,261
Análisis y detección de información falsa sobre la pandemia en Twitter	Jeyasudha, Seth, Usha y Tanna, 2022	1,517
Detección multimodal de noticias falsas a través de redes de fusión progresiva	Jing, Wu, Sol, Colmillo y Zhang, 2023	11,000

Finalmente, se compara el contenido de cada tweet con el contenido de verificación para determinar si es verdadero o falso y así obtener la etiqueta final. En la Tabla 1 se muestra el tamaño del conjunto de datos para el fondo de referencia. En este artículo, el conjunto de datos utilizado tiene un tamaño de 1600 tweets.

2) Preparación de los datos: En esta fase, los textos extraídos de tweets y artículos de verificación de noticias se transforman en un formato que se puede utilizar para la vectorización y el posterior entrenamiento del modelo.

Los tweets suelen contener, además de texto, multimedia, emoticonos, enlaces, etiquetas y otros elementos. En el presente trabajo se requieren los textos, por lo que se eliminarán en esta actividad los elementos que no se vayan a utilizar.

A continuación se muestra un ejemplo de este tipo de tweet:

"@KatrinaDeCuba @DeZurdaTeam @Reylope13 @Ivancoll4 @reyes_medinilla @Ivettelvarez5 @PabloIvanF Días muy tristes pero orgullosos del deber cumplido y la satisfacción de que cómo dijo nuestro #Apostol: "La Muerte no es verdad cuando se ha cumplido bien la obra de la vida", por eso hoy, mañana y siempre #FidelViveEntreNosotros
#IzquierdaUnida
#FidelVive
https://t.co/g2EVZkvg9L"

Después del proceso de normalización, obtenemos el siguiente resultado:

"Días tristes orgullosos, deber, cumplido, satisfacción, dijo, muerte, verdad, cumplido, bien obra, vida, hoy, mañana, siempre"

El siguiente paso consiste en vectorizar el conjunto de palabras a través de la tokenización TF-IDF. Las columnas constan de una para el texto de cada tweet y n para el número de palabras en todos los textos. La presente matriz encontró 612 palabras distintas como columnas y sesenta y nueve textos como filas. Asimismo, los valores mostrados representan los valores de TF-IDF para cada término en relación con su texto. Además, no se requiere normalización de datos, ya que los valores de frecuencia oscilan entre 0 y 1.

3) Configuración de los modelos: En esta fase se definen los modelos a utilizar para la experimentación. Con base en trabajos relacionados, se obtuvieron los siguientes modelos de la investigación más cercana a la presente. La cercanía se definió en función de la similitud de los datos utilizados. Las tres investigaciones utilizaron como conjuntos de datos un grupo de tuits de tamaño comprendido entre mil y diez mil registros.

El primer modelo encontró que para el clasificador LogReg, las funciones de penalización L1 y L2 se evaluaron con valores np.logspace mayores que -4, 4 y 20.

Además, en esta investigación se obtienen resultados relacionados con el modelo anterior utilizando la máquina de vectores de soporte (SVM). Se evaluaron diferentes valores de regularización de C que van de 0,001 a 1000 para el clasificador SVM. [20]

Los parámetros utilizados para el modelo de redes neuronales densas se encuentran en la Tabla 2. [19] Por otro lado, en el documento relacionado al modelo de bosque aleatorio, no mencionan qué parámetros se utilizaron. Bajo esta premisa, se determinó un rango de valores para evaluar algunos de los parámetros más relevantes para este modelo.

En la Tabla 2 se describen los parámetros y sus rangos de valores para cada uno de los modelos.

TABLA 2
MATRIZ DE PARÁMETROS Y VALORES

Modelo	Parámetro	Rango de valores	Número de valores
Regresión logística	Solucionador	Liblineal	1
	Pena	L1, L2	2
	Parámetro de regularización, C	[-4, 4, 20]	20
Máquina de soporte de vectores	Núcleo	Lineal	1
	Parámetro de regularización, C	[-3, 3, 20]	20
Redes neuronales densas	Tamaño de las capas	Densa 1 y 2 con 128 unidades ocultas. Densa 3 con 1 unidad oculta.	1
	Funciones de activación de capas	Densa 1 y 2, Relu Densa 3, Sigmoidal	1
	Puntera de capa densa	0.4	1
	Optimizador	Adán	1
	Tasa de aprendizaje	0.005	1
	Función de pérdida	Entropía cruzada binaria	1
Bosques aleatorios	Número de estimadores	[100, 200, 300, 400, 500, 600, 700, 800, 900, 1000]	10
	Criterio	Gini, entropía	2
	Profundidad máxima	[Ninguno, 3, 5, 10, 20]	5

Para cada combinación se harán tres corridas de forma que, sumando las tiradas de los tres algoritmos, tengamos un total de 483.

4) Entrenamiento: Antes de comenzar el entrenamiento, el conjunto de datos se separa en entrenamiento y prueba. El conjunto de entrenamiento se utilizará para optimizar los modelos, y el conjunto de prueba se utilizará para comprobar la precisión de los modelos. El método utilizado para esto es el método de división de porcentajes.

Para el conjunto de datos de entrenamiento se utilizaron un total de 1281 registros, entre los cuales 522 pertenecen a tweets catalogados como reales y 759 catalogados como falsos. Por otra parte, el conjunto de datos de prueba está formado por 319 de registros, 114 de tweets reales y 205 de tweets falsos.

Los modelos con los que se llevará a cabo la experimentación son los siguientes: Regresión logística, máquina de vectores de soporte, redes neuronales densas y bosques aleatorios.

Después de las pruebas de configuración del modelo de regresión logística, se encontró que los mejores parámetros se

dan cuando el solver es "Liblinear", la penalización es L2 y el parámetro de regularización es 1438.45.

Luego de las pruebas de configuración del modelo de máquina de vectores de soporte, se encontró que los mejores parámetros se dan cuando el kernel es lineal y el parámetro de regularización es 2.97.

Pasadas las pruebas de configuración del modelo de bosque aleatorio, se encontró que los mejores parámetros se dan cuando el número de estimadores es 200, la profundidad máxima es ninguno y el criterio es Gini.

5*) Validación*: El rendimiento de los modelos propuestos se evaluó utilizando Precisión, Sensibilidad, Valor F1 y Exactitud. Posteriormente, se compararon los valores de las métricas de cada modelo para determinar el mejor modelo para identificar si un tweet es verdadero o falso.

$$Precision = VP/(VP + FP) \quad (1)$$

$$Sensibilidad = VP/(VP + FN) \quad (2)$$

$$Valor\ F1 = (2 * Precision * Sensibilidad)/(Precision + Sensibilidad) \quad (3)$$

$$Exactitud = (VP + VN)/(VP + VN + FP + FN) \quad (4)$$

Donde, VP es la cantidad de verdaderos positivos, FP es la cantidad de falsos positivos, VN es la cantidad de verdaderos negativos y FN es la cantidad de falsos negativos.

En resumen, los pasos explicados pueden expresarse mediante el siguiente pseudocódigo:

```
#Recogida de datos
tweets_df = leer_conjunto_de_datos(ruta)
insertar_enlaces_tweet(cliente, base_de_datos, tweets_df)
tweets_URL = extraer_enlaces_tweet(cliente, base_de_datos, colección)
actualizar_datos_tweet(cliente, base_de_datos, colección, tweets, api)

#Preparación de datos
tweets = conjunto_de_tweets
desde i hasta longitud(tweets):
tweets[i] = tweets[i].aminusculas()
tweets[i] = tweets[i].eliminar_etiquetas()
tweets[i] = tweets[i].eliminar_hashtags()
tweets[i] = tweets[i].eliminar_enlaces()
tweets[i] = tweets[i].eliminar_caracteres_especiales()
tweets[i] = tweets[i].eliminar_tildes()
tweets[i] = tweets[i].eliminar_palabras_vacias()
tweets[i] = tweets[i].corregir_ortografia()
vector_tweets = tokenizar(tweets)

#Entrenamiento
x_entrenamiento, x_prueba, y_entrenamiento, y_prueba = dividir_conjunto_de_datos(vector_normalizado_tweets)
modelo_rl = rl()
modelo_msv = msv()
modelo_ba = ba()
mejores_parametros_rl= modelo_rl.calcular_parametros(x_entrenamiento, y_entrenamiento)
mejores_parametros_msv= modelo_msv.calcular_parametros(x_entrenamiento, y_entrenamiento)
mejores_parametros_ba= modelo_ba.calcular_parametros(x_entrenamiento, y_entrenamiento)

#Validacion
prediccion_rl = modelo_rl.predecir(x_entrenamiento, y_entrenamiento, mejores_parametros_rl)
métricas_rl = obtener_metricas(y_prueba, prediccion_rl)
prediccion_msv = modelo_msv.predecir(x_entrenamiento, y_entrenamiento, mejores_parametros_msv)
métricas_msv = obtener_metricas(y_prueba, prediccion_msv)
prediccion_rnd = modelo_rnd.predecir(x_entrenamiento, y_entrenamiento, mejores_parametros_rnd)
métricas_rnd = obtener_metricas(y_prueba, prediccion_rnd)
prediccion_ba = modelo_ba.predecir(x_entrenamiento, y_entrenamiento, mejores_parametros_ba)
métricas_ba = obtener_metricas(y_prueba, prediccion_ba)
comparar(métricas_rl, métricas_svm, métricas_rnd, métricas_ba)
```

III. Resultados y discusión

En esta fase, los datos de evaluación se utilizan para evaluar la efectividad de los modelos a través de métricas. Se realizan tres evaluaciones: uso del carácter "#", uso de palabras con una frecuencia superior a 5 y uso de mejores parámetros. Estas tres evaluaciones buscan mejorar la precisión de los modelos que predicen si una noticia es real o falsa.

La primera evaluación evalúa si los modelos son más precisos eliminando solo el carácter # en lugar de todo el hashtag. En la Tabla 3 se muestran los resultados.

Tabla 3
Comparación de la exactitud mediante el uso del carácter "#"

Modelo	Sin hashtags	Sin "#"
Regresión logística	83%	85%
Máquina de vectores de soporte	88%	89%
Redes neuronales densas	88%	90%
Bosques aleatorios	86%	88%

Se observa una mejora del 2 % en la métrica de precisión para el modelo de regresión logística, las redes neuronales densas y los bosques aleatorios. Por otro lado, el modelo de máquina de vectores de soporte también mejora, pero solo en un 1%.

La segunda evaluación valora si los modelos tienen una mayor precisión a la hora de corregir o retener palabras con una frecuencia superior a 5. En la Tabla 4 se muestran los resultados de la comparación utilizando palabras con una frecuencia superior a cinco como condición.

TABLA 4
COMPARACIÓN DE LA PRECISIÓN MEDIANTE EL USO DE PALABRAS CON UNA FRECUENCIA SUPERIOR A 5

Modelo	Sin condición	Con condición
Regresión logística	83%	83%
Máquina de vectores de soporte	88%	87%
Redes neuronales densas	88%	86%
Bosques aleatorios	86%	86%

La métrica de precisión no mejora en ninguno de los casos. La métrica se mantiene para el modelo de regresión logística y los bosques aleatorios, mientras que este valor se reduce para el modelo de máquina de vectores de soporte y las redes neuronales densas.

En la tercera evaluación, comprobaremos los resultados con múltiples métricas para los cuatro modelos explicados en la sección 3.3. También consideraremos eliminar el carácter "#" y corregir palabras vacías con una frecuencia superior a 5.

Después de haberse realizado la comparación de los resultados obtenidos por el modelo de regresión logística con los parámetros por defecto y los mejores parámetros, se halló una mejora del 5% en la métrica de precisión utilizando los mejores parámetros. Del mismo modo, el valor F1 aumenta tanto para las clases falsas como para las reales.

En la figura 3, se muestra la curva de aprendizaje para el modelo de regresión logística.

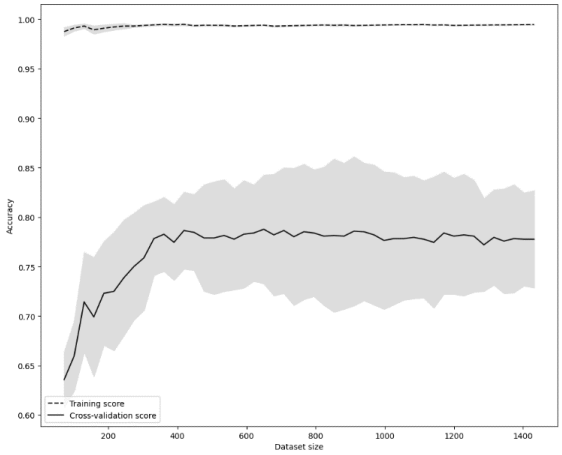

Fig. 3 Curva de aprendizaje con regresión logística

Así mismo, para el modelo de máquina de soporte de vectores, se encontró una mejora del 2% en la métrica de precisión utilizando los mejores parámetros. Del mismo modo, el valor F1 aumenta tanto para las clases falsas como para las reales.

En la figura 4, se muestra la curva de aprendizaje para el modelo de máquina de vectores de soporte.

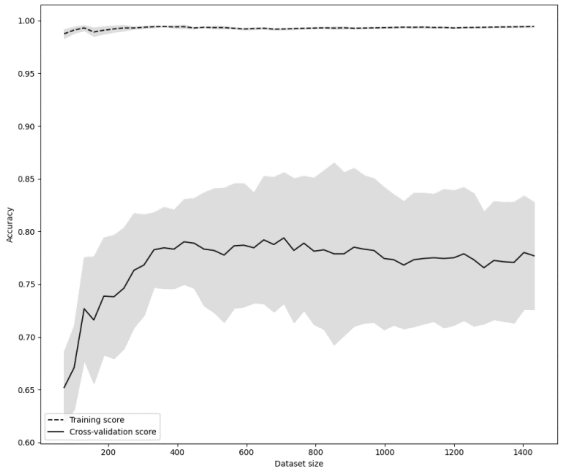

Fig. 4 Curva de aprendizaje con máquina de vectores de soporte

En cambio, para el modelo de redes neuronales densas hay una mejora del 11% en la métrica de precisión mediante la tokenización de frecuencia de términos. Del mismo modo, el valor F1 aumenta tanto para las clases falsas como para las reales.

En la figura 5, se muestra el modelo de precisión para el modelo de redes neuronales densas.

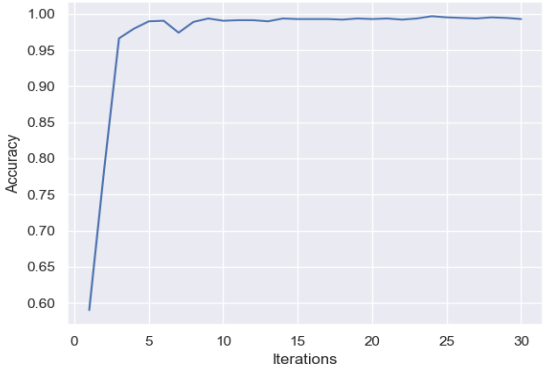

Fig. 5 Modelo de precisión para redes neuronales densas con tokenización de frecuencia de términos.

Finalmente, para el modelo de bosques aleatorios, existe una mejora del 1% en la métrica de precisión utilizando los mejores parámetros. Del mismo modo, el valor F1 aumenta tanto para las clases falsas como para las reales.

En la figura 6, se muestra la curva de aprendizaje para el modelo de bosques aleatorios.

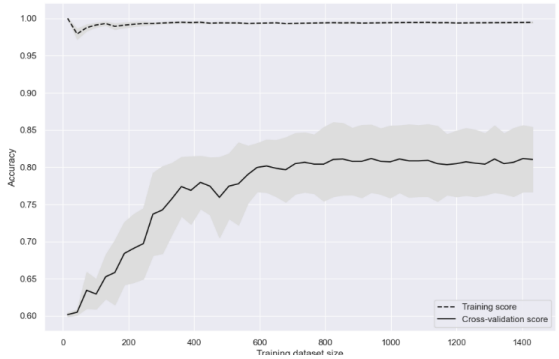

Fig. 6 Curva de aprendizaje para el modelo de bosques aleatorios.

A continuación, se selecciona el mejor modelo encontrado en función de las métricas del paso anterior. En la Tabla 5 se muestran los resultados obtenidos para cada uno de los cuatro modelos.

A partir de los datos obtenidos, se elige el modelo de máquina de vectores de soporte, que tiene una precisión del 91%, como el mejor modelo para predecir noticias falsas en Twitter.

TABLA 5
COMPARACIÓN DE MODELOS BASADOS EN MÉTRICAS

Clase	Métricas	LR	SVM	DNN	RF
Real	Precisión	78%	79%	82%	70%
Real	Sensibilidad	94%	94%	84%	98%
Real	Valor F1	85%	86%	83%	82%
Falso	Precisión	97%	97%	91%	99%
Falso	Sensibilidad	89%	89%	90%	86%
Falso	Valor F1	93%	93%	91%	92%
	Exactitud	90%	91%	88%	89%

IV. CONCLUSIONES

Este estudio diseñó cuatro modelos para identificar si un tweet es verdadero o falso, utilizando algoritmos RL, SVM, DNN y RF. El modelo de máquina de vectores de soporte tuvo el mejor rendimiento medido por precisión, mejor que RL y RF, Recall mejor que DNN e igual a RL, F1 Value mejor que RL, DNN y RF, y precisión mejor que RL, DNN y RF. Dados los resultados obtenidos, se concluyó que la eliminación del carácter numérico (#) incrementa la precisión del modelo de procesamiento del lenguaje natural en al menos un 1%. Además, se demostró que retener palabras con una frecuencia superior a cinco no aumenta la precisión del modelo de procesamiento del lenguaje natural en al menos un 1%.

En el proceso de investigación se detectaron algunas oportunidades de mejora para futuros trabajos. En primer lugar, aumentar el número de registros en el conjunto de datos. Esta mejora presenta una oportunidad para obtener un modelo con más entrenamiento y permite el uso de modelos como BERT, que requiere una gran cantidad de datos. En segundo lugar, evaluar el entrenamiento con otros modelos. Durante la investigación, se encontraron varios modelos que podrían ser utilizados para tareas de clasificación. Estos modelos incluyen BERT, redes neuronales recurrentes, redes neuronales convolucionales y Naive Bayes. Por último, obtener noticias a partir de las imágenes asociadas a los tuits. Algunos tuits utilizados para desarrollar la presente investigación encontraron imágenes asociadas que contenían texto. Estos textos se pueden extraer con una herramienta de OCR y agregar al texto del tweet.

REFERENCIAS

[1] O. E. Moreno, «La divulgación de la información en la encrucijada de la crisis del COVID-19 en Paraguay. Reacciones y trasmisión de datos falsos y científicos a través de las redes sociales y los medios masivos.» Revista de la Sociedad Científica del Paraguay. Vol. 25, pp. 58-85, 2020. Disponible: https://doi.org/10.32480/rscp.2020-25-1.58-85

[2] A. Ojeda y V. Peredo, «Convergencia entre desinformación política y social en el conflicto electoral de 2019 en Bolivia», Temas Sociales. Vol. 46, pp. 98-126, 2020. Disponible: http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S0040-29152020000100005&lng=es&tlng=es

[3] C. Banfi, «Por una responsabilidad civil de los Medios de comunicación coherente con el riesgo de divulgación de noticias falsas sobre personas públicas.», Revista Ius et Praxis. Vol. 24, pp. 255-308, 2018. Disponible: http://dx.doi.org/10.4067/S0718-00122018000300255

[4] V. de Sousa, E. Capoano, P. Rodrigues e I. Paganotti, «¿La Covid-19 ha infectado las noticias? Cómo los periodistas, las audiencias y los procesos de producción son alterados por las pandemias», Universitas-XXI. Vol. 37, pp. 19-41, 2022. Disponible: http://dx.doi.org/10.17163/uni.n37.2022.01

[5] R. A. Galarza, «Fact-Checking en México. Análisis de la percepción del público de Verificado 2018.,» Revista Mexicana de Opinión Pública. Vol. 29, pp. 41-53, 2020. Disponible: https://doi.org/10.22201/fcpys.24484911e.2020.29.70352

[6] M. Luengo y D. García, «La performance de la verdad: los políticos, el periodismo fact-checking y la lucha para hacer frente a la desinformación sobre la COVID-19», Revista Americana de Sociología Cultural. Vol 8, pp. 405-427, 2020. Disponible: https://doi.org/10.1057/s41290-020-00115-w

[7] A. Estrada, K. Alfaro y V. Saavedra, «Desinformación y desinformación, posverdad y fake news», Información, cultura y sociedad. Vol. 42, pp. 93-106, 2020. Disponible: https://doi.org/10.34096/ics.i42.7427

[8] G. K. Shahi, A. Dirkson y T. A. Majchrzak, «Un estudio exploratorio de la desinformación sobre la COVID-19 en Twitter», Redes sociales y medios en línea. Vol. 22, 2021. Disponible: https://doi.org/10.1016/j.osnem.2020.100104

[9] M. O. Ahmed y A. E. Msughter, «Evaluación de la difusión de noticias falsas de Covid-19 entre los usuarios de redes sociales en el estado de Kano, Nigeria», Computers in Human Behavior Reports. Vol. 6., 2022. Disponible: https://doi.org/10.1016/j.chbr.2022.100189

[10] Z. Ziegler, «El programa fiduciario de Michael Polányi contra las noticias falsas y los deepfakes en la era digital», AI & Society. Vol. 36, 2021. Disponible: https://doi.org/10.1007/s00146-021-01217-w

[11] L. Ripoll y J. C. Matos, «Fiabilidad de la información: criterios para identificar la desinformación en el entorno digital.», Investigación Bibliotecológica. Vol. 34, pp. 79-101, 2019. Disponible: https://doi.org/10.22201/iibi.24488321xe.2020.84.58115

[12] E. Ferrara, S. Cresci y L. Luceri, «Desinformación, manipulación y abuso en las redes sociales en la era de la COVID-19», Journal of Computation Social Science. Vol. 3, pp. 271-277, 2020. Disponible: https://doi.org/10.1007/s42001-020-00094-5

[13] A. Jain y A. Kasbe, «Detección de noticias falsas», 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science. Vol. 1, 2018. Disponible: https://doi.org/10.1109/SCEECS.2018.8546944

[14] R. Mandical, N. Mamatha, N. Shivakumar, R. Monica y A. N. Krishna, «Identificación de noticias falsas mediante el aprendizaje automático», Conferencia Internacional IEEE 2020 sobre Tecnologías Electrónicas, Informáticas y de Comunicación. Vol. 1, 2020. Disponible: https://doi.org/10.1109/CONECCT50063.2020.9198610

[15] H. Karimi, P. C. Roy, S. Saba y J. Tang, «Detección de noticias falsas multi-fuente multiclase», Actas de la 27ª Conferencia Internacional de Lingüística Computacional. Vol. 1, p. 1546-1557, 2018. Disponible: https://aclanthology.org/C18-1131
[16] A. Thota, P. Tilak, S. Ahluwalia y N. Lohia, «Fake News Detecion: A Deep Learning Approach», SMU Data Science Review. Vol. 1, pp. 1-20, 2018. Disponible: https://scholar.smu.edu/datasciencereview/vol1/iss3/10
[17] C. Conforti, M. T. Pilehvar y N. Collier, «Hacia la detección automática de noticias falsas: detección de postura cruzada en artículos periodísticos», Actas del Primer Taller sobre Extracción de Hechos y VERIFICACIÓN. Vol. 1, pp. 40-49, 2018. Disponible: http://dx.doi.org/10.18653/v1/W18-5507
[18] R. K. Kaliyar, A. Goswami y P. Narang, «EchoFakeD: mejora de la detección de noticias falsas en las redes sociales con una red neuronal profunda eficiente», Neural Computing and Applications. Vol. 32, 2020. Disponible: https://doi.org/10.1007/s00521-020-05611-1
[19] D. Mouratidis, M. N. Nikiforos y K. L. Kermanidis, «Deep Learning for Fake News Detection in a Pairwise Textual Input Schema», Computation. Vol. 9, 2021. Disponible: https://doi.org/10.3390/computation9020020
[20] O. Oriola y E. Kotzé, «Técnica de aprendizaje semisupervisada mejorada para la detección automática de lenguaje abusivo sudafricano en Twitter», South African Computer Journal. Vol. 32, p. 56–79, 2020. Disponible: https://doi.org/10.18489/sacj.v32i2.847