DESDE 2013
https://repository.uaeh.edu.mx/revistas/index.php/icbi/issue/archive
Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI
ISSN: 2007-6363

Publicación Semestral Pädi Vol. 10 No. Especial 3 (2022) 120–127

Detección automática de noticias falsas usando representaciones textuales tradicionales y soluciones basadas en aprendizaje profundo
Automatic detection of fake news using traditional textual representations and solutions based on deep learning

M. A. Espejel-Rivera [a,*], R. Calderón-Suárez [a], R. M. Ortega-Mendoza [a], C. J. Camacho-Bello [a], M. A. Márquez-Vera [b]

[a] *División de Investigación y Posgrado, Universidad Politécnica de Tulancingo, 43629, Tulancingo, Hidalgo, México.*
[b] *Departamento de Mecatrónica, Universidad Politécnica de Pachuca, 43830, Zempoala, Hidalgo, México.*

Resumen

Las noticias falsas se crean con el objetivo de manipular, dañar o desinformar. En los últimos años, este tipo de noticias ha impactado negativamente en diferentes sectores de la sociedad, como son: política, salud y movimientos sociales. El problema aumenta debido a su amplia y veloz propagación, que supera con creces la rapidez en la que un humano puede controlar o contener este fenómeno. La búsqueda de soluciones para detener la difusión de información falsa ha motivado el desarrollo de métodos computacionales para su detección automática. Comúnmente, tales enfoques son diseñados desde una perspectiva del procesamiento de lenguaje natural. Particularmente, en este trabajo se estudia el impacto del uso de diversas representaciones de características del contenido de la noticia para la detección de noticias falsas en Español con técnicas de aprendizaje automático, incluyendo arquitecturas profundas.

Palabras Clave: Noticias falsas, clasificación de textos, aprendizaje automático, procesamiento de lenguaje natural, aprendizaje profundo.

Abstract

Fake news is created with the aim of manipulating, harming or misinforming. In recent years, this type of news has impacted negatively on different sectors of society, such as politics, health, and social movements. Its severity increases due to its wide and fast propagation, which far exceeds the speed at which a human can control or contain this phenomenon. The search for solutions to combat the spread of false information has motivated the development of computational methods for automatic detection. Commonly, such approaches are designed from a natural language processing perspective. In particular, this paper studies the impact of using various representations of news content features to detect fake news in Spanish with machine learning techniques, including deep architectures.

Keywords: Fake news, text classification, machine learning, natural language processing, deep learning.

1. Introducción

Si bien, el término *noticias falsas* (del inglés, *fake news*) se refiere a la divulgación de información no verídica, es difícil acotarlo a una sola definición debido a sus diferentes características (e.g., intención, estilo de escritura o perfil del autor) (Santiago et al., 2019). En este contexto, Ireton and Posetti (2018) establecen tres expresiones relacionadas con estas propiedades: *información errónea*, que es incorrecta o engañosa, pero quien la difunde cree que es verdad; la *desinformación*, que se refiere a contenido falso y su difusión es de manera deliberada, y la *mala-información*, la cual se basa en la realidad para infligir daño a una persona, organización o país.

Las noticias falsas también han sido incluidas como una categoría en la clasificación de publicaciones con contenido fraudulento: rumores, teorías de conspiración, sátira, desinformación, propaganda, entre otras (Meel and Vishwakarma, 2020). De manera general, se ha considerado que las *fake news* son

*Autor para correspondencia: angelica.espejel2115001@upt.edu.mx
Correo electrónico: angelica.espejel2115001@upt.edu.mx (María Angélica Espejel-Rivera), ricardo.calderon@upt.edu.mx (Ricardo Calderón-Suárez), rosa.ortega@upt.edu.mx (Rosa María Ortega-Mendoza), cesar.camacho@upt.edu.mx (César Joel Camacho-Bello), marquez@upp.edu.mx (Marco Antonio Máquez-Vera)

Historial del manuscrito: recibido el 15/04/2022, última versión-revisada recibida el 21/07/2022, aceptado el 02/08/2022, publicado el 31/08/2022. **DOI:** https://doi.org/10.29057/icbi.v10iEspecial3.9008

noticias no verídicas que tienen la intención de engañar (Zhou and Zafarani, 2020).

La gravedad de las noticias falsas se debe al incremento de su propagación en los últimos años, tanto en redes sociales como fuera de línea. Por ejemplo, durante el año 2020, en *Twitter* la información falsa o engañosa se tuiteó 47 millones de veces, mientras que en *Facebook* se crearon 1,200 millones de interacciones en sitios que difunden *fake news* (Kemp, 2021). También en ese mismo año, se observó que el nivel de participación con contenido engañoso relacionado con temas políticos y del COVID-19 alcanzó niveles récord (Goldstein, 2021). Esta situación ha motivado el estudio del fenómeno de la desinformación desde diversos ámbitos (e.g., Estado, periodístico, derechos humanos, académico y de investigación) para establecer estrategias que garanticen el derecho a la libertad de expresión, así como a la información veraz y oportuna (Mottola, 2020; Santiago et al., 2019).

Desde el campo de la investigación, el estudio sobre las noticias falsas se ha realizado desde diferentes enfoques: (i) *quién las crea*, (ii) *cómo y por qué se generan*, (iii) *cómo se propagan* y (iv) *cómo pueden ser detectadas* (Zhou and Zafarani, 2020). Desde la perspectiva computacional, ha surgido el interés de crear métodos automáticos que permitan su detección para coadyuvar así a contener su propagación. Inicialmente, se desarrollaron modelos basados en el estilo de la escritura de la noticia (Pérez et al., 2017; Zhou et al., 2020). Posteriormente, se realizaron estudios que incluyen el análisis del título de la noticia, la estrategia de publicación y la fiabilidad de la fuente (Nordberg et al., 2020). Recientemente, los algoritmos propuestos combinan técnicas de aprendizaje profundo para aprender representaciones que logren la detección de este tipo de noticias (P et al., 2021). Sin embargo, aún existen retos por resolver debido a su complejidad, diversidad, velocidad de propagación, diferentes modalidades de información (e.g., texto e imágenes) y los costos de la verificación de hechos (Shu et al., 2020).

La mayoría de las investigaciones han utilizado colecciones de noticias escritas en inglés, lo que representa una limitada usabilidad en otros idiomas. Particularmente, en este trabajo se aborda la detección automática de noticias falsas en español con modelos basados en el contenido de la noticia utilizando diferentes representaciones del texto. La metodología combina técnicas de aprendizaje automático y Procesamiento de Lenguaje Natural (PLN).

2. Trabajos relacionados

La principal dificultad en la detección de noticias falsas consiste en encontrar las características que las hacen diferentes de las verdaderas (Xue et al., 2021). En la búsqueda de estos atributos, Horne and Adali (2017) encontraron diferencias significativas, entre ambos tipos de publicaciones, en el título y en el vocabulario que se utiliza para su redacción. Mientras que Mottola (2020) estudió la asociación de los textos fraudulentos con estructuras fijas de expresión y estilo en la escritura (e.g. título escrito en mayúsculas, signos de puntuación exclamativos o puntos suspensivos), así como la presencia de imágenes llamativas, ver Tabla 1.

Tabla 1: Características de las noticias falsas (Mottola, 2020).

Estructurales
Título impactante, llamativo
Uso de mayúsculas
Signos de puntuación (exclamativos, puntos suspensivos)
Incluyen imágenes llamativas
Léxicas
Uso de palabras comunes
Léxico sectorial en noticias relacionadas con conocimientos específicos
Otras
Falta de metadatos y fuentes para verificar

En el caso de noticias falsas en Español, Sued and Rodríguez (2020) analizaron la estructura de un conjunto de noticias fraudulentas que se producen y circulan en Facebook provenientes de Argentina y México. En sus resultados reportaron diferentes estructuras estilísticas: (i) retórica de la hipérbole, que son expresiones enunciativas que exageran un elemento de la realidad para dar mayor fuerza expresiva al mensaje, (ii) sátira, su intención en la fabricación es explícita, (iii) parcialmente fabricadas, donde uno de los elementos puede no ser verificado, (iv) noticias pasadas recontextualizadas, (v) contradicción entre el título y el cuerpo, (vi) imprecisión sobre el contexto espacio temporal, (vii) ausencia de reacciones verbales de actores relevantes y (viii) uso del condicional para referirse al acontecimiento principal. Lo anterior sugiere que la detección de *fake news* requiere de un análisis en diferentes niveles que considere el título, el cuerpo de la noticia y metadatos.

Debido a la complejidad de esta tarea, Zhou and Zafarani (2020) establecieron cuatro perspectivas para su estudio: *conocimiento*, *estilo*, *propagación* y *credibilidad de la noticia*. Los métodos basados en el conocimiento evalúan la autenticidad del contenido de la noticia, mientras que en el estilo analizan cómo están escritas. Por otro lado, en el modelo de propagación se utiliza la información relacionada con la difusión y el usuario. Por último, el análisis de la credibilidad centra su atención en el perfil de quién crea y quién difunden estas publicaciones. Por su parte, Saquete et al. (2020) propusieron dividir esta tarea en subtareas: i) detección de engaño, postura, controversia o polarización; ii) verificación de hechos; iii) ciberanzuelos y iv) el nivel de credibilidad.

Respecto a los modelos basados en el contenido, Choudhary and Arora (2021) analizaron las características de las *fake news* en diferentes niveles lingüísticos: léxico, sintáctico, semántico o de discurso. Se ha experimentado con diferentes representaciones para estos atributos. Las más comunes se basan en: bolsa de palabras (BoW), n-gramas, TF-IDF (*Term Frecuency-Inverse Document Frequency*), etiquetado POS (*Part of Speech Tagging*), incrustaciones de palabras (*embeddings*) y el uso de diccionarios lingüísticos (Sitaula et al., 2020).

Entre los algoritmos de aprendizaje automático más utilizados se encuentran: Árboles de Decisión (DT), Bosques Aleatorios (RF), Naïve Bayes (NB) y Máquinas de Vectores de Soporte (SVM) (Álvarez et al., 2021; Nordberg et al., 2020; Zhou and Zafarani, 2020). En este contexto, Patwa et al. (2021) usaron una colección de 10,700 publicaciones y artículos de noticias en inglés sobre la COVID-19 para realizar la clasificación de noti-

cias a través de los siguientes algoritmos: Regresión Logística (LR), SVM, DT y K-Vecinos Cercanos (KNN). El clasificador SVM obtuvo el mejor desempeño con un valor $F1$ de 93.2 %.

2.1. Métodos basados en texto con técnicas de aprendizaje profundo

Las técnicas de aprendizaje profundo también han sido utilizadas para detectar automáticamente noticias falsas (Maslej et al., 2019), desde arquitecturas simples hasta las complejas que fusionan diferentes modelos. Entre las más usuales se encuentran las Redes Neuronales Convolucionales (CNN) y Recurrentes (RNN). También ha sobresalido el uso de redes recurrentes especiales, como son GRU (*Gated Recurrent Unit*) y LSTM (*Long Short-Term Memory*). Recientemente, el uso de BERT (*Bidirectional Encoder Representations from Transformers*) ha sido frecuentemente probado en este dominio.

Un estudio comparativo de varias arquitecturas profundas fue realizado por Álvarez et al. (2021). En la Tabla 2 se presentan los mejores resultados reportados por los autores en colecciones de noticias escritas en inglés. En el primer trabajo presentado en la tabla, se utiliza un conjunto de datos disponible en el repositorio *Kaggle*[1], mientras que en los restantes crearon colecciones de noticias a partir de contenido web y blogs informativos. Los resultados fueron estimados a partir de la exactitud (ACC, del inglés *Accuracy*). De forma general, se observa que el desempeño reportado es similar en las diferentes colecciones. Lo anterior muestra la usabilidad de arquitecturas profundas para detectar noticias falsas en el idioma inglés.

Tabla 2: Resultados obtenidos por modelos de aprendizaje profundo descritos en el trabajo de Álvarez et al. (2021).

Autor	Colección	Modelo	ACC
Amine et al. (2019)	Kagle	CNN	0.96
Ahn and Jeong (2019)	creada	BERT	0.96
Liu (2019)	creada	LSTM	0.95
Verma et al. (2019)	creada	GRU	0.91

2.2. Detección de noticias falsas en español

En el año 2020, por primera vez, se llevó a cabo el evento *MEX-A3T: Fake News and Aggressiveness Analysis case study in Mexican Spanish*[2] en colaboración con la iniciativa MexLef como parte del Foro IberLef 2020[3]. El enfoque del foro fue dirigido a resolver tareas en español de México, entre ellas la detección de noticias falsas. De manera general, en las dos ediciones, 2020 y 2021, se han presentado diversas metodologías basadas en *transformers*, redes profundas, así como representaciones tradicionales como BoW, n-gramas y *embeddings*. Sin duda, las colecciones de datos etiquetados del foro se han consolidado como conjuntos de referencia para evaluar diferentes enfoques. En este contexto, *FNDeepML* es otro conjunto de datos etiquetado con noticias en español. Bonet et al. (2021) experimentaron con tales datos para evaluar un modelo que determina la veracidad del contenido y de los elementos esenciales de la estructura periodística: título, subtítulo, introducción, cuerpo y conclusión.

Hoy en día, ante la falta de colecciones de noticias escritas en español, se han propuesto nuevos enfoques que aprovechan datos etiquetados en otros idiomas. Por ejemplo, Martínez et al. (2021) propusieron cuatro esquemas experimentales basados en una red neuronal recurrente con una unidad LSTM (LSTM-RNN). Uno de sus experimentos consistió en entrenar a la red con instancias en inglés para evaluar su desempeño en la predicción de instancias en español que fueron traducidas al inglés. Por otro lado, en la investigación realizada por De et al. (2021), se presentó un modelo multilingüe basado en un *auto-encoder* para clasificar instancias en idiomas no considerados en el entrenamiento.

En este trabajo de investigación, se estudia el impacto del uso de diversas representaciones creadas para detectar noticias falsas en colecciones de datos en español. Las representaciones corresponden al uso de diferentes características extraídas del contenido de la noticia, así como algunas provenientes del entrenamiento de arquitecturas neuronales.

3. Colecciones de datos

En esta sección se describen los datos usados para entrenar y evaluar los modelos estudiados. También, se presenta un análisis de su contenido para conocer el tipo y cantidad de información presente.

3.1. Estadísticas generales

Para realizar los experimentos, se utilizaron dos colecciones de noticias en español cuyas estadísticas se muestran en la Tabla 3. Los metadatos de ambas colecciones comprenden: clase o categoría (con las etiquetas falsa o verdadera), tópico, encabezado (*título*), texto, fuente y *URL* de la publicación (Posadas et al., 2019).

Tabla 3: Descripción de las colecciones de datos.

Colección		Verdadera	Falsa	Total	
2020	Entrenamiento	338	338	676	971
	Prueba	153	142	295	
2021	Entrenamiento	491	480	971	1543
	Prueba	286	286	572	

La primera colección de noticias se presentó en la competencia *FakeDeS 2020*[4]. Este conjunto de datos se compone de publicaciones realizadas en México en varios sitios web: noticiarios, compañías de medios y foros especializados. La recolección se realizó durante el período enero-julio del año 2018. De acuerdo con los autores, se integra de 971 instancias, de las cuales 676 (70 %) son usadas como particiones de entrenamiento y 295 (30 %) y prueba, respectivamente.

Por otro lado, la segunda colección se utilizó en el evento *FakeDeS 2021*[5]. Una característica distintiva de esta colección

[1] https://www.kaggle.com/competitions/fake-news/data
[2] https://sites.google.com/view/mex-a3t/home
[3] https://sites.google.com/view/iberlef2020/home
[4] https://sites.google.com/view/mex-a3t/data-and-evaluation
[5] https://sites.google.com/view/fakedes/task-data

es que su conjunto de prueba presenta variaciones en la temática y lingüística, con respecto al de entrenamiento, según se observa en la Tabla 4 (Gómez et al., 2021).

Tabla 4: Variación temática en la colección 2021.

	Tópicos	Lugar de publicación
Entrenamiento	ciencia, deporte, economía, educación, entretenimiento, política, salud, seguridad, sociedad	México
Prueba	medio ambiente, ciencia, COVID-19, deporte, internacional, política, sociedad	México, Argentina, Bolivia, Chile, Colombia, Costa Rica, Ecuador, España, Estados Unidos, Francia, Perú, Uruguay, Inglaterra y Venezuela

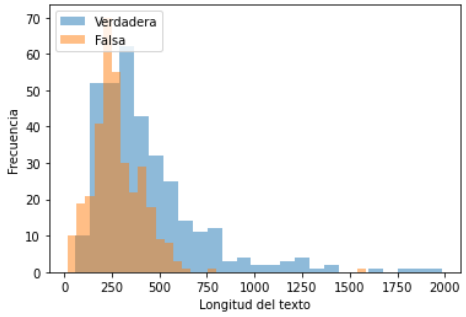
3.2. *Análisis del contenido de las colecciones de datos*

El objetivo de esta sección es estudiar diversas características de cada colección de datos etiquetados para entender el tipo de información que almacenan. Como preprocesamiento, para todos los casos, se realizó conversión a minúsculas y se eliminaron signos de puntuación.

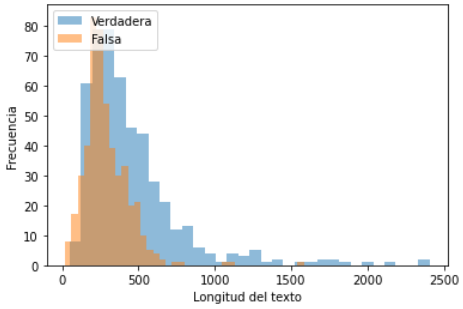
Primero, se exploró la longitud del contenido de la noticia, es decir, el número de *tokens*, en este caso, la cantidad de palabras. Para este análisis, se mantuvieron las palabras vacías (e.g., preposiciones y artículos). Las estadísticas se muestran en la Figura 1. Como se observa, ambas colecciones tienen noticias de longitud variable cuyas distribuciones son similares. La mayoría de los documentos tienen una longitud menor a 1000 palabras. De forma general, el análisis por clase sugiere que las noticias falsas tienden a ser más cortas que las verdaderas.

Segundo, se realizó un análisis de la cantidad de palabras vacías en las colecciones. Como se ha mencionado en la literatura, las palabras vacías no proporcionan un valor semántico, sin embargo, ellas son indispensables en la comunicación natural entre humanos; por lo tanto, su eliminación no es conveniente en todas las tareas de PLN. De ahí que es importante estudiar su aportación en cada tarea. En este contexto, en la Figura 2 se presenta el porcentaje de estos elementos respecto a la longitud el texto para ambas clases. Las distribuciones muestran que son términos altamente frecuentes en el contenido de la noticia, aunque su uso es similar tanto en la clase falsa como en la verdadera.

Tercero, se analizaron diferencias temáticas entre las dos clases de noticias. Específicamente, se realizó una observación de las frecuencias de los términos diferentes a las palabras vacías.

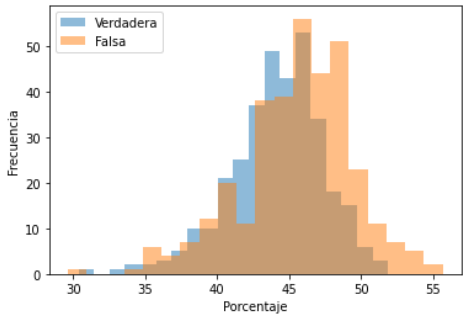


(a) Colección 2020.

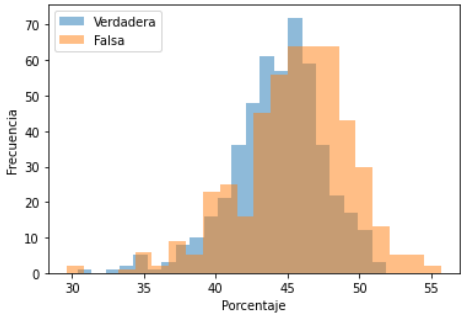


(b) Colección 2021.

Figura 1: Estadísticas de la longitud del texto de la noticia.



(a) Colección 2020.



(b) Colección 2021.

Figura 2: Porcentaje de palabras vacías respecto a la longitud del texto de la noticia.

En las Figuras 3 y 4 se presentan los 40 términos más frecuentes en cada colección de datos. Es posible notar poca diferencia en el uso de términos entre clases de noticias. Más específicamente, aproximadamente el 30 % y 32 % de las palabras son comunes en las dos categorías en las colecciones 2020

y 2021, respectivamente. Estos datos remarcan el desafío de la tarea para discriminar entre los dos tipos de clases de noticias.

(a) Verdaderas. (b) Falsas.

Figura 3: Términos más frecuentes, colección 2020. La frecuencia de las palabras está relacionada con el tamaño de la fuente.

(a) Verdaderas. (b) Falsas.

Figura 4: Términos más frecuentes, colección 2021. La frecuencia de las palabras está relacionada con el tamaño de la fuente.

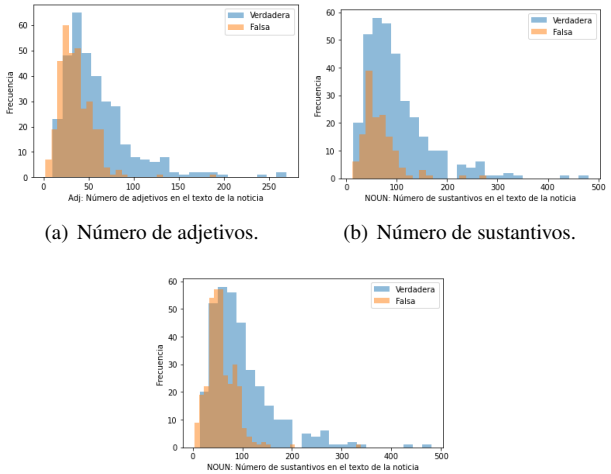
Finalmente, se realizó un estudio de la frecuencia de elementos gramaticales utilizados en el contenido de la noticia, los cuales pueden señalar el estilo de redacción, por ejemplo, el número de sustantivos, verbos y adjetivos. Para identificar la función gramatical de las palabras se empleó un etiquetador POS (*Part of Speech Tagging*)[6]. Los resultados se muestran en las Figuras 5 y 6. Las estadísticas refieren una frecuencia de uso similar en ambas clases de noticias. Por ejemplo, en el corpus 2020, el número de adjetivos utilizados en la clase verdadera representa el 27.7 %, mientras que en la falsa el 24.1 %. Por otro lado, en el corpus 2021, el número de sustantivos en la clase verdadera representa el 40.81 % y en la falsa el 38.48 %. Si bien, no existen diferencias significativas en la frecuencia, es posible encontrar patrones de uso a través de los clasificadores. De ahí que en los experimentos se exploró este tipo de atributos.

4. Experimentación

En esta sección se describen los experimentos realizados, tomando como base de conocimiento las colecciones descritas en la Sección 3. Los modelos fueron entrenados y evaluados con las particiones definidas en los conjuntos de datos como entrenamiento y prueba, respectivamente.

4.1. *Configuración Experimental*

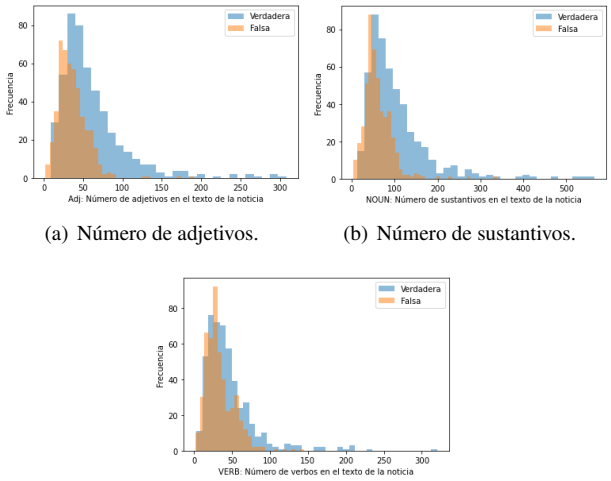
En esta sección se detalla la configuración tomada como base para realizar la experimentación y el análisis de resultados.



(a) Número de adjetivos. (b) Número de sustantivos.

(c) Número de verbos.

Figura 5: Análisis de elementos POS en la colección 2020.



(a) Número de adjetivos. (b) Número de sustantivos.

(c) Número de verbos.

Figura 6: Análisis de elementos POS en la colección 2021.

4.1.1. *Preprocesamiento*

Consiste en la preparación del texto de la noticia para la etapa de clasificación. En este trabajo se estudia la combinación de diferentes técnicas tradicionales de preprocesamiento como la conversión del texto a minúsculas, la eliminación de palabras vacías y signos de puntuación. La configuración de preprocesamiento utilizada es señalada en cada experimento.

4.1.2. *Algoritmos de clasificación*

En los experimentos se evaluaron los siguientes clasificadores: SVM, LR y RF. También se aplicaron algunas técnicas de aprendizaje profundo: CNN, GRU y GRU con una capa de atención. El desempeño se reporta mediante las siguientes medidas:

- Exactitud (ACC), la cual representa la fracción de todas las instancias que se clasificaron correctamente.

[6]https://www.nltk.org

- $F1$, la cual combina las medidas de precisión y sensibilidad (*recall*) en un único valor.

4.1.3. *Modelos de referencia*

Para la construcción de los modelos de referencia (i.e. *baseline*) se consideró la configuración descrita en los trabajos de Posadas et al. (2019) y Gómez et al. (2021). Se basan en la tradicional bolsa de palabras con preprocesamiento que incluye conversión a minúsculas y eliminación de signos de puntuación. La Tabla 5 reporta el desempeño de tales modelos.

Tabla 5: Desempeño de los modelos de referencia.

Colección		SVM	LR	RF
2020	ACC	0.7627	0.7763	0.7797
	F1	0.7626	0.7763	0.7795
2021	ACC	0.7273	0.7168	0.7080
	F1	0.7271	0.7164	0.7051

4.2. *Análisis del impacto del preprocesamiento*

El objetivo de este experimento es analizar el impacto generado en el desempeño de los modelos cuando se aplican diferentes niveles de preprocesamiento. Específicamente, se explora: conversión a minúsculas (minus), uso de signos de puntuación (!,¿.") y palabras vacías (pvac). A partir de los resultados presentados en la Tabla 6, se observa que el desempeño mejora cuando se mantienen las palabras vacías, independientemente del uso de los otros dos tipos de procesamiento. Estos resultados sugieren que el estilo de redacción, el cual también está definido por el uso de palabras vacías, es un elemento importante para discriminar noticias falsas de verdaderas.

Tabla 6: Resultados con diferentes tipos de preprocesamientos. Se reporta el desempeño obtenido por el clasificador SVM.

	Preprocesamiento				
	minus	!,¿."	pvac	ACC	F1
Baseline	✓		✓	**0.7627**	**0.7626**
2020		✓	✓	0.7627	0.7626
		✓		0.7153	0.7152
			✓	0.7627	0.7626
				0.7288	0.7281
	✓	✓	✓	0.7627	0.7627
	✓	✓		0.7322	0.7318
	✓			0.7288	0.7281
Baseline	✓		✓	**0.7273**	**0.7271**
2021		✓	✓	0.7255	0.7254
		✓		0.6853	0.6825
			✓	0.7273	0.7271
				0.7220	0.7219
	✓	✓	✓	0.7255	0.7254
	✓	✓		0.6906	0.6875
	✓			0.7220	0.7219

4.3. *Selección de las características más discriminativas*

Comúnmente, en conjuntos de datos similares a los que se usan en esta investigación, las representaciones basadas en frecuencia consisten de vectores con alta dimensionalidad. Sin embargo, un gran número de atributos no supone una mayor probabilidad de éxito (Spasova, 2017). La selección de características es una técnica utilizada para reducir la dimensionalidad en los datos. Este proceso se puede realizar a través de varias técnicas *filter*, *wrapper* o métodos híbridos (Maseda, 2019). Particularmente, el objetivo de este experimento es observar si el desempeño de los métodos de clasificación mejora reduciendo el conjunto de atributos. Para ello, se aplicó la técnica *filter* chi-cuadrada. Esta técnica ha sido aplicada con éxito en problemas de NLP (Meesad et al., 2011). Este experimento se realizó con la representación BoW expuesta en la Sección 4.1.3.

En las Figuras 7 y 8, se muestra el desempeño obtenido por los clasificadores según la variación del número de características relevantes seleccionadas. A partir de los resultados, es posible observar que cada uno de los algoritmos de clasificación presentan un comportamiento diferente para cada subconjunto de características.

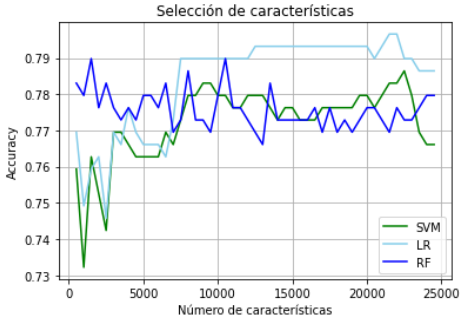


Figura 7: Resultados usando selección de características, colección 2020.

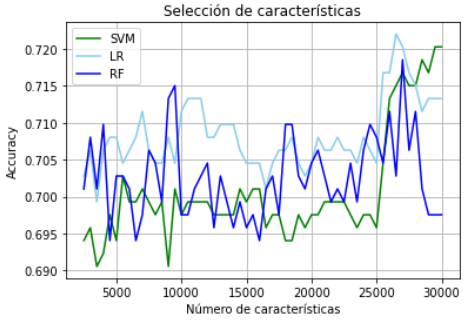


Figura 8: Resultados usando selección de características, colección 2021.

4.4. *Representaciones tradicionales*

En esta sección se evalúa el desempeño de representaciones tradicionales (RepTrad) para tareas de clasificación de texto (Sebastiani, 2002). Particularmente, se evalúa TF-IDF, bolsa de atributos POS y la representación formada por la concatenación de BoW con atributos POS. Estas representaciones son detalladas enseguida. TF-IDF permite expresar la importancia de un término en una noticia que forma parte de la colección. Por otro lado, para estudiar el impacto de los atributos estilísticos en la tarea sobre la detección las noticias falsas, se utilizó una bolsa de atributos POS. La finalidad es identificar patrones estilísticos capturados por el uso de clases de palabras, como verbos, adjetivos, sustantivos. Finalmente, en el interés de combinar atributos estilísticos y temáticos, la tercera representación consiste en la concatenación de las representaciones BoW y POS.

En la Tabla 7 se presentan los resultados de los algoritmos de clasificación con las diferentes representaciones propuestas y con las mejores 10000 características, seleccionadas con chi-cuadrada. Se incluye el *baseline* presentado en la Sección 4.1.3.

Tabla 7: Resultados con representaciones tradicionales.

	RepTrad		SVM	LR	RF
2020	***Baseline***	ACC	0.7627	0.7763	0.7797
		F1	0.7626	0.7763	0.7795
	BoW	ACC	0.7797	**0.7898**	0.7729
		F1	0.7796	**0.7898**	0.7726
	TFIDF	ACC	0.7559	0.7492	0.7898
		F1	0.7547	0.7485	0.7847
	POS	ACC	0.7017	0.7119	0.7254
		F1	0.7015	0.7119	0.7253
	BoW-POS	ACC	0.7593	0.7695	0.7729
		F1	0.7593	0.7694	0.7725
2021	***Baseline***	ACC	0.7273	0.7168	0.7080
		F1	0.7271	0.7164	0.7051
	BoW	ACC	0.6976	0.7045	0.6976
		F1	0.6964	0.7024	0.6947
	TFIDF	ACC	0.6573	0.6399	**0.7290**
		F1	0.6374	0.6260	**0.7283**
	POS	ACC	0.6906	0.6766	0.6713
		F1	0.6894	0.6746	0.6660
	BoW-POS	ACC	0.6958	0.6713	0.7115
		F1	0.6951	0.6674	0.7078

En general, se observa que los resultados de clasificación dependen de la configuración del modelo implementado, como la representación textual y la técnica de aprendizaje automático. Además, estos modelos solo consideran la ocurrencia de las palabras, sin aprovechar información de su contexto. Las arquitecturas expuestas en la siguiente sección consideran la relación o asociación entre términos.

4.5. Técnicas de aprendizaje profundo

El objetivo de esta sección es evaluar algunas arquitecturas profundas para la detección de noticias falsas. Los modelos neuronales estudiados son capaces de seleccionar características para crear representaciones vectoriales. Como entrada reciben incrustaciones de palabras comúnmente conocidas como *embeddings*. Particularmente, se usaron *embeddings* preentrenados en español: *Word2Vec*[7] y *FastText*[8]. De forma específica, se estudian las siguientes técnicas de aprendizaje profundo:

- **CNN multicanal**: Se compone de 5 canales, los cuales representan el tamaño de filtro usado en cada capa Convolucional (tamaño 1 a 5). Cada capa contiene 750 filtros con una capa posterior *Global MaxPooling*. Se usó la función de activación *relu* y *Dropout*=0.2. Los canales son concatenados a través de una capa densa con 10 neuronas, la cual es conectada a la capa de salida.
- **Modelo GRU**: Se conforma por una capa GRU bidireccional con 512 unidades recurrentes y *Dropout* = 0.2. Como capas de clasificación se utilizaron una capa densa con 32 neuronas conectada a la capa de salida.
- **Modelo GRU con capa de atención (GRUatt)**: Contiene la siguiente secuencia: GRU bidireccional con 768 unidades recurrentes, una capa de atención con 768 unidades, *Dropout*= 0.2 y una capa densa con 32 neuronas, la cual es conectada a la capa de salida.

Dado que es un problema de clasificación binario, la capa de salida de los tres modelos se diseñó con una neurona y función de activación sigmoide. Se utilizó el optimizador Adam. Las arquitecturas fueron entrenadas usando *Early Stopping* con paciencia (*patience*) de 10. El número de neuronas en las capas se estableció de acuerdo con la optimización de hiperparámetros realizada a través de una búsqueda aleatoria (i.e., *Random Search*).

En la Tabla 8 se presentan los resultados obtenidos con los tres modelos propuestos. De forma general, se observa que el desempeño está relacionado con el tipo de *embeddings* y la arquitectura diseñada. Por ejemplo, las arquitecturas CNN y GRUatt presentan un mejor desempeño con *embeddings* de *Word2Vec* en ambas colecciones. También se observa que el modelo profundo GRUatt con embeddings de Word2Vec obtiene el mejor resultado en ambas colecciones, lo que indica que la capa de atención agregada a la GRU enriquece el desempeño.

Tabla 8: Resultados con técnicas de aprendizaje profundo.

Corpus	*Embeddings*		CNN	GRU	GRUatt
2020	*Word2Vec*	ACC	0.7559	0.7220	**0.8000**
		F1	0.7551	0.7185	**0.7987**
	FastText	ACC	0.7390	0.7627	0.7051
		F1	0.7384	0.7627	0.6984
2021	*Word2Vec*	ACC	0.6731	0.6853	**0.7045**
		F1	0.6709	0.6838	**0.7009**
	FastText	ACC	0.6119	0.6573	0.6031
		F1	0.6105	0.6567	0.5545

5. Conclusiones y trabajo futuro

En este trabajo se estudian diversas representaciones textuales para la detección automática de noticias falsas en español. Los modelos presentados en este trabajo sólo consideran el contenido de la noticia sin usar metadatos. Los resultados muestran que el desempeño de los modelos, depende en gran medida de las características usadas en la representación del texto, así como de los algoritmos de aprendizaje automático aplicados. En los experimentos con la colección 2021 se observa un desempeño menor al que se obtiene con el corpus 2020, lo que sugiere que los modelos evaluados tienen una dificultad para detectar las variaciones temáticas y lingüísticas que se presentan en este conjunto de noticias. Por otro lado, se observa que los modelos tradicionales y profundos obtuvieron resultados comparables;

[7] https://www.kaggle.com/datasets/rtatman/pretrained-word-vectors-for-spanish
[8] https://fasttext.cc/docs/en/crawl-vectors.html

sin embargo, en la colección 2020 el mejor desempeño se obtuvo con un modelo profundo que consiste de una GRU con capa de atención. En este contexto, también se observa que los resultados dependen en gran medida de la configuración de la arquitectura.

Cabe señalar que los modelos estudiados, no superan el desempeño mostrado en el estado del arte. Lo anterior, propicia el interés de profundizar en el estudio representaciones que consideren otras características léxicas y de estilo. También, es importante estudiar la relación existente entre los metadatos de la noticia.

Referencias

Ahn, Y. C. and Jeong, C. S. (2019). Natural language contents evaluation system for detecting fake news using deep learning. In *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 289–292.

Álvarez, N., Pico, P., and Holgado, J. (2021). Detección de noticias falsas en redes sociales basada en aprendizaje automático y profundo: Una breve revisión sistemática. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, 41:632–645.

Amine, B. M., Drif, A., and Giordano, S. (2019). Merging deep learning model for fake news detection. In *2019 International Conference on Advanced Electrical Engineering (ICAEE)*, pages 1–4.

Bonet, A., Piad, A., Saquete, E., Martínez, P., and García, M. (2021). Exploiting discourse structure of traditional digital media to enhance automatic fake news detection. *Expert Systems with Applications*, 169:114340.

Choudhary, A. and Arora, A. (2021). Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications*, 169:114171.

De, A., Bandyopadhyay, D., Gain, B., and Ekbal, A. (2021). A transformer-based approach to multilingual fake news detection in low-resource languages. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 21(1).

Goldstein, A. (2021). Socialmedia engagement with deceptive sites reached record highs in 2020. Technical report, The German Marshal Fund of the United States, https://www.gmfus.org/news/social-media-engagement-deceptive-sites-reached-record-highs-2020.

Gómez, H., Posadas, J., Bel, G., and Porto, C. (2021). Overview of fakedes at iberlef 2021: Fake news detection in spanish shared task. *Procesamiento del lenguaje natural*, 67:223–231.

Horne, B. and Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the international AAAI conference on web and social media*, pages 759–766.

Ireton, C. and Posetti, J. (2018). Journalism, fake news and disinformation: handbook for journalism education and training. Technical report, UNESCO, https://unesdoc.unesco.org/ark:/48223/pf0000265552.

Kemp, S. (2021). Digital 2021: global overview report. Technical report, we are social, https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/.

Liu, H. (2019). A location independent machine learning approach for early fake news detection. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 4740–4746.

Martínez, K., Ávarez, A., and Arias, J. (2021). Fake news detection in spanish using deep learning techniques. *arXiv preprint arXiv:2110.06461*.

Maseda, M. (2019). Reducción de la dimensionalidad mediante métodos de selección de características en microarrays de adn. Master's thesis, Universitat Oberta de Catalanya.

Maslej, V., Sarnovsky, M., and Butka, P. (2019). Deep learning methods for fake news detection. In *2019 IEEE 19th International Symposium on Computational Intelligence and Informatics and 7th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics (CINTI-MACRo)*, pages 000143–000148. IEEE.

Meel, P. and Vishwakarma, D. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*.

Meesad, P., Boonrawd, P., and Nuipian, V. (2011). A chi-square-test for word importance differentiation in text classification. volume 6, pages 110–114, Singapore. IACSIT Press.

Mottola, S. (2020). Las fake news como fenómeno social. anáisis lingüístico y poder persuasivo de bulos en italiano y español. *Discurso y Sociedad*, 14(3):683–706.

Nordberg, P., Kärestada, J., and Nohlberg, M. (2020). Automatic detection of fake news. *CEUR Workshop Proceedings*, 2789(23):168–179.

P, D., Chakraborty, T., Long, C., and Kumar, S. (2021). *Data Science for Fake News*, volume 42. Springer.

Patwa, P., Sharma, S., Pykl, S., and Guptha, V. (2021). Fighting an infodemic: Covid-19 fake news dataset. *CONSTRAINT-2021*.

Pérez, V.and Kleinberg, B., Lefevre, A., and Mihalcea, R. (2017). Automatic detection of fake news. *arXiv*.

Posadas, J., Gómez, H., Sidorov, G., and Escobar, J. (2019). Detection of fake news in a new corpus for the spanish language. *Journal of Intelligent and Fuzzy Systems*, 36(5):4868–4876.

Santiago, R., Adame, C., and Palacios, C. (2019). Reporte sobre las campañas de desinformación, "noticias falsasƶ su impacto en el derecho de la libertad de expresión. *Comisión Nacional de Derechos Humanos México*, pages 10–12.

Saquete, E., Tomás, D., Moreda, P., Martínez, P., and Palomar, M. (2020). Fighting post-truth using natural language processing: A review and open challenges. *Expert Systems with Applications*, 141(112943).

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Comput. Surv.*, 34(1):1?47.

Shu, K., Wang, S., Lee, D., and Liu, H. (2020). *Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements*, pages 1–19. Springer International Publishing, Cham.

Sitaula, N., Mohan, C., Grygiel, J., Zhou, X., and Zafarani, R. (2020). Credibility-based fake news detection. In *Disinformation, Misinformation, and Fake News in Social Media*, pages 163–182. Springer.

Spasova, R. (2017). *Desarrollo y evaluación de métodos de selección de características para la predicción de eventos adversos en pacientes polimedicados*. PhD thesis, Universidad Pública de Navarra.

Sued, G. and Rodríguez, M. (2020). Noticias falsas en facebook: narrativas, circulación y verificación. los casos de argentina y méxico. *Estudios sobre el Mensaje Periodístico*, 26(3):1229–1242.

Verma, A., Mittal, V., and Dawn, S. (2019). Find: Fake information and news detections using deep learning. In *2019 Twelfth International Conference on Contemporary Computing (IC3)*, pages 1–7.

Xue, J., Wang, Y., Tian, Y., Li, Y., Shi, L., and Wei, L. (2021). Detecting fake news by exploring the consistency of multimodal data. *Information Processing and Management*, 58(5):102610.

Zhou, X., Jain, A., Phoha, V., and Zafarani, R. (2020). Fake news early detection: A theory-driven model. *Digital Threats: Research and Practice*, 1(2):1–25.

Zhou, X. and Zafarani, R. (2020). A survey of fake news. *ACM Computing Surveys*, 53(5):1–40.