risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

Recebido/Submission: 20/10/2020
Aceitação/Acceptance: 28/12/2020

Detección de Noticias Falsas en Redes Sociales Basada en Aprendizaje Automático y Profundo: Una Breve Revisión Sistemática

Nathaly Álvarez-Daza[1], Pablo Pico-Valencia[1,2], Juan A. Holgado-Terriza[2]

nathaly.alvarez@pucese.edu.ec; pablo.pico@pucese.edu.ec; jholgado@ugr.es

[1] Pontificia Universidad Católica del Ecuador, Esmeraldas, Ecuador.

[2] Universidad de Granada, Granada, España.

Pages: 632-645

Resumen: Las redes sociales han cambiado la forma en que la sociedad se informa. Redes sociales como Twitter y Facebook cuentan con millones de usuarios quienes muchas veces comparten noticias falsas (*fake news*) sin saberlo. Los contenidos de estas noticias son falsos y no verificados, y llegan a viralizarse, engañando y causando pánico. El objetivo de este estudio es desarrollar una revisión de la literatura en la que se determina cómo el aprendizaje automático y aprendizaje profundo, han apoyado para el desarrollo de clasificadores de noticias falsas en redes sociales. El estudio fue desarrollado a partir de una metodología formal usada en ciencias de la Computación. Los resultados mostraron que los modelos de aprendizaje han sido ampliamente usados para crear sistemas de detección de noticias falsas, predominando la detección en el ámbito político, sociedad, salud y desastres naturales. Se constató que se emplearon mayoritariamente los modelos de aprendizaje automático en comparación con los de aprendizaje profundo, aun así, ambos enfoques demostraron ser eficientes para clasificar noticias falsas, jugando un determinante de los resultados, el set de datos y el método de extracción de características empleado.

Palabras-clave: Aprendizaje automático, aprendizaje profundo, noticias falsas, clasificador.

Detection of Fake News in Social Networks Based on Machine and Deep Learning: A Brief Systematic Literature Review

Abstract: Social networks have changed how society is informed. Social networks such as Twitter and Facebook have millions of users who often share fake news without knowing it. The contents of these news are false and unchecked, and they become viral, deceiving, and causing panic. The objective of this study is to develop a literature review that examines how machine and deep learning have supported the development of social media fake news classifiers. The study was developed from a formal methodology used in computer science. The results showed that learning models have been widely used to create false news detection systems, with

detection predominating in the political field. It was found that machine learning models were mostly used in contrast with deep learning models, however, both approaches demonstrated be efficient to classify fake news, playing a decisive factor of the results, the data set and the feature extraction method used.

Keywords: Machine learning, deep learning, fake news, classifier.

1. Introducción

En la actualidad, las redes sociales se han vuelto una de las tecnologías más populares a nivel mundial. Por su fácil acceso y estar orientadas al intercambio de información, millones de usuarios tienen un perfil en redes sociales como Facebook y Twitter. Además de que las redes sociales permiten a las personas comunicarse con sus amigos y familiares a nivel mundial, les ayudan a mantenerse informados de los sucesos que acontecen en su entorno. A raíz de este uso que la sociedad les ha dado a las redes sociales, han proliferado muchas publicaciones de noticias que no están contrastadas o validadas, originándose así, el concepto de noticias falsas, conocidas también como *fake news* (Dong-Ho Lee, et al., 2019).

Las noticias falsas, son artículos que tienen contenido informativo, que en lugar de informar, desinforman a los cibernautas en redes sociales (Allcott & Gentzkow, 2017). La información publicada y compartida por medio de redes sociales como Twitter y Facebook tiene la particularidad de que llega a muchos países. Esto implica que una noticia falsa se difunde rápidamente e incluso puede llegarse a viralizar, causando pánico en la sociedad. En los últimos años se han difundido cientos de noticias falsas, en distintos ámbitos, entre los que figuran los siguientes: política, deportes, economía, ciencia, entretenimiento, entre otros.

En el ámbito político, se han originado varios casos de noticias falsas en redes sociales que han tenido un impacto mundial. Un caso muy popular fue en las elecciones de Estados Unidos en 2016. En (Allcott & Gentzkow, 2017) se analizaron las publicaciones en las redes sociales en el entorno de las elecciones presidenciales de Estados Unidos. En este panorama, se contabilizaron 115 noticias falsas relacionadas con Donal Trump. Asimismo, en torno a la otra candidata a la presidencia, Hillary Clinton, se compartieron 41 noticias falsas en Facebook. Otro caso relevante, en el ámbito financiero fue el de United Airlines, que, con la subida de un artículo falso a Internet, provocó que el precio de sus acciones cayera. El impacto que tuvo esta noticia falsa fue grave y su efecto persistió por un tiempo antes de volver a recuperarse (Carvalho et al., 2011). Otro caso, también con cierta repercusión, fue la noticia que ha sido compartida en las diferentes plataformas refiriéndose al dióxido de cloro. Según el contenido de la noticia falsa, este químico era la cura contra el covid-19 y a raíz de ello, muchas personas tomaron la decisión de comprarlo y consumirlo, sin saber las consecuencias que este químico puede provocar si se administra sin la supervisión de un médico (Huang & Carley, 2020).

Los casos descritos anteriormente son solo algunos ejemplos de noticia falsas de los últimos años. Esto ha llamado la atención de los investigadores para crear herramientas orientadas a detectar noticias falsas en redes sociales y en Internet en general. En este sentido, se ha empleado la Inteligencia Artificial (IA), y específicamente las técnicas de aprendizaje automático. Puntualmente, para crear sistemas predictivos que sean

capaces de clasificar noticias falsas en las redes sociales se han planteado métodos de minería de texto. Estos sistemas emplean como un componente principal, un modelo de aprendizaje basado en datos, el cual es entrenado a partir de tweets, posts of cometido web descargados de las redes sociales, micro blogs o webs. Sin embargo, existen muchos modelos y no se conoce de manera formal cuáles son las características de los clasificadores especializados y publicados en la literatura. Algoritmos de aprendizaje automático, conocidos también como aprendizaje de máquina (i.e., árboles de decisión, bosques aleatorios, Naïve Bayes, máquina de vectores de soporte), han sido aplicados con éxito (Gilda, 2018; Lakshmanarao et al., 2019). Asimismo, modelos de aprendizaje profundo (i.e., redes convolucionales, con memoria, recurrentes) han sido empleados de manera independiente o fusionados con redes neuronales profundas (Kresnakova et al., 2019). Es así, como se han originado muchos modelos para clasificar noticias falsas en redes sociales.

El objetivo de este estudio es investigar cuáles han sido los principales modelos de IA que se han propuesto en la literatura para detectar noticias falsas en redes sociales. Para ello, se plantea el desarrollo de una revisión de la literatura basada en una metodología formal, como es la metodología de (Kitchenham & Charters, 2007), ampliamente usada para el desarrollo de este tipo de estudios en las ciencias computacionales. Los resultados de este estudio permitirán describir experiencias de investigadores y nivel de exactitud alcanzado por los sistemas de detección automática de noticias falsas implementados a partir de minería de textos y modelos de aprendizaje automático.

El documento está organizado de la siguiente manera. La Sección 2, describe la metodología empleada para llevar a cabo el proceso de revisión sistemática de la literatura, enfatizando en la estrategia de búsqueda diseñada. La Sección 3 presenta los resultados, esto es, se responde de manera detallada a cada una de las preguntas de investigación (PIs) formuladas en base al análisis de los estudios primarios recuperados. También, se discuten brevemente dichos resultados. Finalmente, en la Sección 4 se describen las conclusiones.

2. Metodología

Una revisión sistemática de la literatura consiste en identificar, evaluar e interpretar los estudios más relevantes de un tema en particular. Para llevar a cabo el proceso de revisión sistemática planteado en este estudio, se empleó la metodología propuesta por (Kitchenham & Charters, 2007), la cual es una metodología formal para la ejecución de trabajos de esta naturaleza en las ciencias computacionales.

2.1. Preguntas de investigación

Las noticias falsas han despertado un gran interés entre los investigadores especialistas en IA. Es así como surgió la siguiente pregunta, ¿cómo los modelos de aprendizaje de la IA han coadyuvado en la creación de clasificadores para la detección automática de noticias falsas en redes sociales, el principal medio tecnológico usada por los internautas para informarse? Para responder esta pregunta, estratégicamente se plantearon cuatro PIs específicas.

- **PI1:** ¿Cuáles han sido las principales propuestas basadas en técnicas de aprendizaje automático para detectar noticias falsas en redes sociales?
- **PI2:** ¿Cuál es el alcance que han tenido los modelos de aprendizaje, en términos de exactitud, para detectar noticias falsas en medios sociales?
- **PI3:** ¿Qué herramientas de software y datos se han usado para crear modelos predictivos de detección de noticias falsas en redes sociales?
- **PI4:** ¿En qué ámbitos se ha detectado noticias falsas en redes sociales?

2.2. Estrategia de búsqueda

La búsqueda realizada en este estudio consideró trabajos disponibles en la literatura que han abordado la detección automática de noticias falsas en redes sociales desde distintas perspectivas, esto es, teórica y, experimental o práctica. Para buscar y recuperar los estudios primarios a partir de los cuales se pudo responder las PIs planteadas, la estrategia de búsqueda se dividió en tres pasos: definición de las fuentes de información (FI), formulación de la ecuación de búsqueda, y selección de los estudios primarios recuperados en las FI.

Definición de las FI. El primer paso de la estrategia de búsqueda fue identificar y definir qué FIs se emplearían para realizar la revisión sistemática. Para este caso en particular, se seleccionaron dos bibliotecas digitales (*ACM, IEEE Xplore*) y dos bases de datos documentales (*Scopus y Web of Science*). Ambos tipos de FIs son especializadas en el campo de la Computación e indexan trabajos de impacto.

Cadena de búsqueda. Para realizar la búsqueda de los estudios primarios, a partir de los cuales se pudiera responder a las PIs formuladas, se definió una cadena de búsqueda la cual se detalla como sigue: ((“machine” OR “deep”) AND (“learning”) AND (“fake new*”)). Dicha cadena de búsqueda se adaptó para ser aplicada en cada una de las FIs seleccionadas.

Criterios de inclusión y exclusión. Los criterios de inclusión definidos para este estudio estuvieron enfocados en tomar en cuenta solo aquellas propuestas que analizaran y emplearan mecanismos de IA para detectar noticias falsas en redes sociales. De manera complementaria, se definieron cuatro criterios de exclusión orientados a descartar estudios que no aportaran a la investigación. Estos criterios de exclusión fueron los siguientes: documentos repetidos, documentos escritos en un idioma diferente al inglés, documentos inaccesibles, y documentos publicados antes del 2015. Los artículos recuperados con la cadena de búsqueda que no cumplieron estos criterios no fueron analizados. Los resultados obtenidos, luego de aplicar la estrategia de búsqueda, se muestra en la Figura 1.

La selección de los estudios se realizó en dos etapas; (i) lectura del título, resumen, y las conclusiones de cada artículo, y (ii) lectura completa de los artículos. Estas etapas se aplicaron a mediados de agosto de 2020. De los 106 artículos recuperados, 55 fueron no repetidos. Estos 55 estudios se redujeron a 40 tras leer el título y resumen en detalle y descartar 15 estudios más. En consecuencia, quedaron por analizar de forma exhaustiva

40 estudios. Sin embargo, éstos se redujeron a 36 ya que 4 eran libros y uno era un análisis teórico.

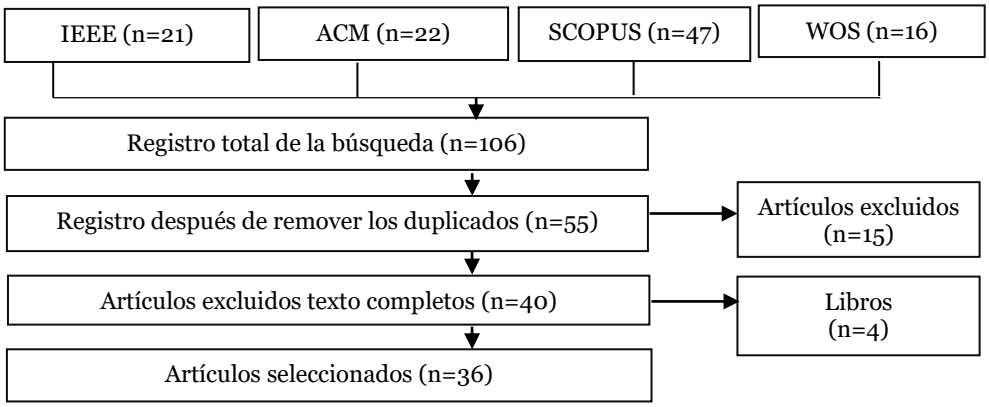

Figura 1 – Proceso de selección de estudios primarios

3. Resultados y discusión

En base a la información y resultados descritos en los 36 artículos sujetos a estudio, en esta sección, se presentan los resultados y se responden a cada una de las PIs formuladas en esta investigación. Dichos resultados se resumen como sigue:

PI1: ¿Cuáles han sido las principales propuestas basadas en técnicas de aprendizaje automático para detectar noticias falsas en redes sociales? Para clasificar noticias falsas en redes sociales se han empleado tanto los algoritmos de aprendizaje automático y aprendizaje profundo. En el primer bloque de la Tabla 1, se listan 12 estudios en los que se ha propuesto solo el uso de al menos uno de los algoritmos de aprendizaje automático para entrenar clasificadores automáticos de noticias falsas en redes sociales. En el bloque 2 de la misma tabla, se observa también que otros 13 estudios se limitaron a aplicar exclusivamente modelos de aprendizaje profundo (redes neuronales) para entrenar clasificadores de noticias falsas. Es importante señalar, que como muestra el tercer bloque de la Tabla 1, se han desarrollado 10 propuestas en las que también se han realizado pruebas tanto con modelos de aprendizaje automático como modelos de aprendizaje profundo. Algunas de estas propuestas consistieron en comprobar cuál de los enfoques proveía mayor exactitud (i.e., S23, S33). Otras, menos frecuentes, desarrollaron modelos híbridos, principalmente de aprendizaje profundo (i.e., S2, S26), para mejorar el nivel de exactitud que alcanzaban al aplicar los algoritmos de aprendizaje automático y profundo de manera independiente. Uno de los 36 artículos no figura en la tabla ya que fue una propuesta solo de análisis.

Grupo	#	Pais	Año	Fuente	AM	AP	Referencia
GRUPO 1	S8	Canadá	2020	Scopus	x		(Ibrishimova & Li, 2020)
	S10	India	2019	Scopus	x		(Kaliyar et al., 2019)
	S14	Pakistán	2019	Scopus	x		(Kareem & Awan, 2019)
	S18	Malaysia	2019	Scopus	x		(Mokhtar et al., 2019)
	S19	India	2019	Scopus	x		(Jain et al., 2019)
	S20	India	2019	Scopus	x		(A. Kumar et al., 2019)
	S22	India	2019	Scopus	x		(Lakshmanarao et al., 2019)
	S27	Brasil	2019	Scopus	x		(Reis et al., 2019)
	S44	India	2017	Scopus	x		(Gilda, 2018)
	S45	Alemania	2016	Scopus	x		(Masood & Aker, 2018)
	S46	Colombia	2018	Scopus	x		(Rolong, et al., 2018)
	S47	Canadá	2017	Scopus	x		(Ahmed et al., 2017)
GRUPO 2	*S1*	España	2019	Scopus		x	(Kong et al., 2020)
	S9	Singapur	2019	Scopus		x	(Liu, 2019)
	S11	UK	2019	Scopus		x	(Han & Mehta, 2019)
	S12	Eslovaquia	2019	Scopus		x	(Kresnakova et al., 2019)
	S13	Argelia	2019	Scopus		x	(Amine et al., 2019)
	S15	Jordania	2019	ACM		x	(Abedalla et al., 2019)
	S17	Jordania	2019	Scopus		x	(Qawasmeh et al., 2019)
	S21	India	2019	Scopus		x	(Verma et al., 2019)
	S24	Portugal	2018	Scopus		x	(Borges et al., 2019)
	S25	Corea	2019	Scopus		x	(Ye-Chan Ahn, 2019)
	S26	India	2019	Scopus		x	(Barua et al., 2019)
	S32	Egipto	2018	Scopus		x	(Sherry Girgis, Eslam Amer, 2018)
	S34	Corea	2019	Scopus		x	(Dong-Ho Lee, et al., 2019)
GRUPO 3	S2	India	2019	Scopus	x	x	(S. Kumar et al., 2020)
	S5	EE. UU.	2020	ACM	x	x	(Singh et al., 2020)
	S16	Cyprus	2019	ACM	x	x	(Katsaros et al., 2019)
	S23	India	2019	Scopus	x	x	(Hiramath, Chaitra K, 2020)
	S28	Bangladesh	2019	Scopus	x	x	(Abdullah-All-Tanvir et al., 2019)
	S30	India	2019	Scopus	x	x	(Poddar et al., 2019)
	S33	Tailandia	2018	Scopus	x	x	(Supanya Aphinwongsophon, 2018)
	S35	India	2019	Scopus	x	x	(Tanik Saikh, et al., 2019)
	S39	India	2019	Scopus	x	x	(Arvinder Pal Singh Bali, et al., 2019)
	S42	Romania	2019	Scopus	x	x	(Adrian M. P. Brasoveanu, 2019)

Tabla 1 – Propuestas orientadas a detectar noticias falsas usando aprendizaje automático

Los datos muestran que el uso de la IA aplicada para la detección de noticias falsas se ha venido trabajando ampliamente en los últimos años; mayoritariamente, las propuestas analizadas se publicaron en 2019 (26 estudios), muchas de las cuales fueron desarrolladas por investigadores de la India (12 estudios). Esto se debe a que en ese país en 2019 se está trabajando en una fuerte campaña para palear las noticias falsas. Por tanto, la academia está jugando un papel importante para resolver este problema, que, en términos de política, ha causado desestabilización.

PI2: ¿Cuál es el alcance que han tenido los modelos de aprendizaje, en términos de exactitud, para detectar noticias falsas en medios sociales?

Modelos de aprendizaje automático para detectar noticias falsas. Los resultados obtenidos mostraron que muchos de los algoritmos de clasificación empleados en aprendizaje automático han sido experimentados para entrenar modelos de detección de noticias falsas. En la Tabla 2, es posible ver que mayoritariamente se aplicaron los algoritmos relacionados con las máquinas de vectores de soporte (16 estudios), seguido de la regresión logística (9 estudios) y a la par los algoritmos basados en el Teorema de Bayes, bosques aleatorios y árboles de decisión con 7, 7 y 6 estudios respectivamente. Otros algoritmos aplicados, aunque con menor frecuencia fue el algoritmo de vecinos más cercanos y algoritmos menos populares como el de potenciación del gradiente y descenso estocástico de gradiente. Sin embargo, los algoritmos que alcanzaron mejor exactitud (*accuracy*) para detectar noticias falsas fueron: las máquinas de vectores de soporte con 99,90%, regresión logística con 91,60% y el algoritmo de potenciación de gradiente con el 91,05%.

#	NB	DT	RF	KNN	SVM	LR	GB	AUC	AB	SGD
S5	21,00	*	34,00	*	34,00	*	*	*	*	*
S8	*	*	*	*	*	40,00	*	*	*	*
S10	*	*	*	*	*	*	86,00	*	*	*
S14	68,00	64,00	65,00	70,00	68,00	69,00	*	*	*	*
S16	*	47,00	55,00	*	58,00	58,00	*	*	*	*
S18	*	*	*	*	*	77,00	*	*	*	*
S19	*	*	*	*	93,60	*	*	*	*	*
S20	*	*	*	*	59,00	*	*	*	*	*
S22	*	**82,70**	**90,70**	**79,20**	75,50	*	*	*	*	*
S23	89,00	*	77,00	*	79,00	75,00	*	*	*	*
S27	*	*	*	*	*	*	*	88,00	*	*
S28	**89,06**	*	*	*	89,34	69,47	*	*	*	*
S30	86,30	82,5	*	*	89,10	**91,60**	*	*	*	*
S33	76,08	*	*	*	**99,90**	*	*	*	*	*
S35	*	*	*	*	56,04	*	*	*	*	*
S39	*	*	86,63	72,54	63,55	*	**91,05**	**89,25**	**89,25**	*
S42	*	26,20	*	*	28,40	26,00	*	*	*	*

#	NB	DT	RF	KNN	SVM	LR	GB	AUC	AB	SGD
S44	*	67,60	64,80	*	73,60		65,77	*	*	**65,70**
S45	*	*	*	*	*	*	*	*	*	*
S46	88,10	*	*	*	*	*	*	*	*	*
S47	*	*	*	*	92,00	*	*	*	*	*
S2	*	*	*	*	58,68	57,58	*	*	*	*
Σ	7	6	7	3	16	9	3	2	1	1

Tabla 2 – Uso de los modelos de aprendizaje automático en los estudios analizados. NB=Naïve Bayes, DT=árbol de decisión, RF=bosque aleatorio, KNN=vecinos más cercanos, SVM=vectores de soporte, LR=Regresión logística, GB=potenciación de gradiente, AUC-ROC=área bajo la curva, AB=potenciación adaptativa, SGD=descenso estocástico de gradiente.

Modelos de aprendizaje profundo para detectar noticias falsas. Los algoritmos de aprendizaje automático no han sido los únicos empleados para intentar crear sistemas de IA para detectar noticias falsas en redes sociales. Modelos muy populares y eficientes basados en redes neuronales también han sido usados para intentar conseguir un clasificador más efectivo que aquellos basados en aprendizaje automático. En el segundo grupo de la Tabla 1 se listan 13 propuestas en los que se aplicaron redes neuronales para detectar noticias falsas. Adicionalmente, 12 estudios más, listados en el tercer grupo de la misma tabla, aplicaron estos modelos de aprendizaje profundo, en conjunto con modelos de aprendizaje automático. Como en el enfoque anteriormente descrito, el principal motivo de uso de ambos modelos, fue principalmente para comparar y determinar cuál de los dos enfoques de aprendizaje de la IA se adaptaba mejor a los sets de datos de noticias falsas empleados para entrenar los modelos.

Aunque dentro del aprendizaje profundo existe una variedad de modelos orientados a procesar la información usando redes neuronales artificiales, la investigación permitió identificar que solo tres de los modelos de redes neuronales existentes han sido principalmente aplicados para crear clasificadores de noticias falsas en redes sociales. Entre estos modelos de redes neuronales, resumidos en la Tabla 3, figuran los siguientes: las redes con memoria (10 estudios), redes convolucionales (7 estudios) y las redes recurrentes (5 estudios). Los resultados en términos de exactitud (*accuracy*) muestran que el modelo de red neuronal que mejor se adaptó fue una red neuronal genérica de la cual no se dio detalles con 99,90% (S33), luego, las redes convolucionales con el 96,00% (S13) y finalmente, el modelo de memoria a largo plazo con el 95,30% (S9). Sin embargo, como se evidenció en el estudio S2, la creación de modelos híbridos, han permitido alcanzar valores de precisión (accuracy) más competentes que los obtenidos por modelos que operan de forma independiente, por ejemplo, de 73.29% de precisión obtenido a partir de una red convolucional (CNN) y el 80,62% de precisión alcanzado por una red con memoria (LSTM), un modelo denominado CNN + LSTM bidireccional, alcanzó una precisión de 88.78%. Esto abre un abanico de posibilidades para seguir explorando modelos híbridos de aprendizaje profundo (i.e., S17, S26)

#	CNN	LSTM	GRU	BERT	MLP	GDL	Modelo genérico	Vanilla
S1	93,70	91,00	*	*	*	*	*	*
S9	92,40	**95,30**	*	*	*	*	*	*
S11	*	*	*	*	*	**92,70**	*	*
S12	97,5	91,80	*	*	**89,8**	*	*	*
S13	**96,00**	*	*	*	*	*	*	*
S15	*	72,20	*	*	*	*	*	*
S17	87,47		*	*	*	*	*	*
S21	*	94,30	**91,90**	*	*	*	*	*
S24	*	*	83,38	*	*	*	*	*
S25	*	*	*	96,25	*	*	*	*
S26	*	88,61		*	*	*	*	*
S32	*	21,66	21,70	*	*	*	*	21,50
S34	52,80	41,70	*	*	*	*	*	*
S2	73,29	80,62	*	*	*	*	*	*
S5	*	*	*	72,10	*	*	*	*
S16	60,00	*	*	*	58,00	*	*	*
S23	*	*	*	*	*	*	91,00	*
S28	*	76,00	74,00	*	*	*	*	*
S30	*	*	*	*	*	*	49,90	*
S33	*	*	*	*	*	*	**99,90**	*
S35	*	*	*	*	72,00	*	*	*
S39	*	*	*	*	88,36	*	*	*
S42	29,00	32,40	54,90	*	*	*	*	*
Σ	9	12	6	2	4	1	3	1

Tabla 3 – Uso de los modelos de aprendizaje profundo en los estudios analizados. CNN=redes convolucionales, LSTM=memoria de largo plazo, RNN=unidades recurrentes, BERT=representación del codificador bidireccional de los transformadores, MLP=perceptrón multicapas, GDL=aprendizaje profundo geométrico.

Es importante señalar que los porcentajes de exactitud alcanzados por los algoritmos de aprendizaje automático y aprendizaje profundo es directamente dependiente de los datos y set de datos empleados para entrenar los modelos; y dependiente también del método de extracción de características aplicados sobre los datos. En la mayoría de los estudios se aplicaron métodos como: de frecuencia de término (i.e., S14), frecuencia de término - frecuencia de documento inversa (TF-IDF) (i.e., S9, S28, S44, S30), vectores globales (*GloVe*) (i.e. S9, S16) y el método *CountVectorizer* (i.e., S9, S16)

PI3: ¿Qué herramientas de software y datos se han usado para crear modelos predictivos de detección de noticias falsas en redes sociales? En lo relacionado a las herramientas de desarrollo empleadas para crear clasificadores de noticias falsas en redes sociales se empleó mayoritariamente la herramienta de programación Python (18 estudios). Python es actualmente la herramienta más popular y usada para el desarrollo de sistemas de IA, y específicamente para el desarrollo de los sistemas de minería de texto, que integran modelos de clasificación automática, se está empleando este lenguaje de programación. También se ha evidenciado el uso de las librerías de aprendizaje automático de Python como: *sci-kit learn* (6 estudios; S2, S5, S14, S16, S44, S46), Keras (4 estudios; S9, S12, S26, S42) y TensorFlow (4 estudios; S12, S13, S21, S26). De manera complementaria, también se ha empleado la librería para procesamiento de lenguaje natural, NLTK, *Natural Language Toolkit,* por sus siglas en inglés (9 estudios; S9, S15, S21, S14, S16, S39, S44, S46, S47). En otros estudios se evidenció también el uso de la herramienta de procesamiento de lenguaje natural de Google (4 estudios; S5, S8, S26, S27). Asimismo, hubo estudios que no detallaban aspectos de las herramientas de desarrollo como es el caso de los estudios S8, S10, S11, S18, S30, S33, S35, S24 y S20 haciendo que no se puedan replicar los experimentos en algunos casos.

En lo que respecta a los datos usados por las propuestas analizadas, éstas emplearon principalmente set de datos accesibles de forma pública en repositorios como: Kaggle (9 estudios; S12, S13, S8, S10, S20, S22, S30, S39, S46), GitHub (1 estudio; S17) y PolitiFact (4 estudios; S11, S16, S2, S32, S46). Se evidenció también que algunos estudios emplearon en menor proporción otras fuentes de datos como: Twitter (S28, S33), T*he New York Times* y *Washington Post* (S1), Reuter.com o *BuzzFeed* (S27, S46, S11, S27).

Es importante resaltar que en general los modelos aplicaron minería de texto sobre set de datos previamente especificados. De esta manera, los investigadores optimizaron el tiempo requerido para realizar la recolección y preprocesamiento de los datos que son el insumo para aplicar el proceso de minería de textos y llevar a cabo el proceso de entrenamiento del clasificador aplicando los algoritmos de aprendizaje supervisado, aprendizaje en el que se conocen las etiquetas de salida en el set de datos de entrenamiento. Asimismo, parte de los datos que conformaban el set de datos fueron empleados para llevar a cabo la fase de pruebas de los modelos. No obstante, en algunas propuestas, los investigadores crearon sus propios sets de datos para desarrollar sus modelos predictivos de clasificación de noticias falsas. Usaron como fuente, información de webs y blogs de noticieros. En resumen, los sets de datos empleados fueron los siguientes: NYT (S5, S9), LIAR (S16, S42, S32), News (S23), Fake News (S13, S8), Fake News Net (S2), Fake Real News (S46) y FNC-1 (S15, S17, S24, S35, S45). Otros menos populares también fueron usados o creados (S26, S19, S28, S20).

PI4: ¿En qué ámbitos se ha detectado noticias falsas en redes sociales? Cuatro de los ámbitos en los que se han enfocado las propuestas estudiadas, y orientadas a detectar noticias falsas, fueron las siguientes: política (26 estudios; S5, S8, S10, S11, S14, S16, S20, S27, S30, S39, S42, S44, S45, S47, S1, S2, S9, S12, S13, S15, S17, S24, S25, S26, S32, S34), negocios y economía (3 estudios; S5, S25, S34), sociedad, deporte y cultura (4 estudios; S5, S14, S5, S3), ciencia, tecnología y salud (3 estudios; S5, S14, S5), entretenimiento (2 estudios; S14, S34) y desastres naturales (S33). Existieron también, 7 estudios en los que no se especificaba el ámbito de las noticias falsas analizadas (S18,

S19, S22, S23, S28, S46, S21). Se evidenció que la mayor cantidad de noticias falsas están en el ámbito político, esto es, información sobre gobernantes y sus acciones. Es por ello, que a nivel de nación, los gobiernos tienen interés en palear las noticias falsas, muchas de las cuales desestabilizan sus gobiernos causando caos entre la población.

4. Conclusiones

Los resultados mostraron que los modelos de aprendizaje de IA han sido ampliamente empleados para la creación de sistemas de detección automática de noticias falsas. Con altos y bajos porcentajes de exactitud (accuracy), 22 de los clasificadores analizados integraron algoritmos basados en: árboles de decisión, teorema de Bayes, regresión logística, vecinos más cercanos, bosques aleatorios y máquinas de vectores de soporte. Este último tuvo la mejor medida de precisión, 99,9%. Por otro lado, 23 estudios restantes emplearon redes neuronales (RN). Las redes convolucionales, redes con memoria a corto plazo y las redes recurrentes, fueron las más usadas, alcanzando las redes convolucionales una precisión del 97% y un modelo genérico de red neuronal un 99,90%. Esto permite concluir que los clasificadores que aplicaron aprendizaje automático fueron más exactos para clasificar noticias falsas a partir de tweets. Sin embargo, muchos modelos están optando por combinar tanto el aprendizaje automático como el aprendizaje profundo para optimizar el proceso en distintas etapas y conseguir así una mejor medida de exactitud a la hora de detectar noticias falsas.

A pesar de que se han desarrollado una alta gama de modelos predictivos orientados a clasificar noticias falsas a partir de tweets, es aún un reto integrar estos modelos en la red social Twitter para advertir a los usuarios previo a compartir contenido. Los modelos propuestos solo han identificado características principales de los tweets que incorporan contenido de noticias falsas por lo que existe una brecha entre la detección de noticias falsas compatibles no solo con la red social Twitter; sino que sea compatible con cualquier otra de las redes sociales ampliamente usadas como lo es Facebook o blogs disponibles en Internet.

La difusión de noticias falsas a través de las redes sociales o en la web general causa impacto negativo a la sociedad de general. Aunque se ha identificado que mayoritariamente, las noticias falsas que mayormente se difunden son las del ámbito político, económico y respecto a desastres naturales, es importante ahondar esfuerzos para que aplicando la IA, y el aprendizaje automático, sea posible también identificar noticias falsas en otros ámbitos, y que no sean solo en formato texto; sino que cubra el espectro de tipos de noticias falsas descritas en (Manzoor et al., 2019), tales como: basadas en imágenes, en usuario, en conocimiento, en estilo y basadas en postura.

Referencias

Abdullah-All-Tanvir, Mahir, E. M., Akhter, S., & Huq, M. R. (2019). Detecting Fake News using Machine Learning and Deep Learning Algorithms. *2019 7th International Conference on Smart Computing and Communications, ICSCC 2019*, 1–5.

Abedalla, A., Al-Sadi, A., & Abdullah, M. (2019). A closer look at fake news detection: A deep learning perspective. *ACM International Conference Proceeding Series*, 24–28.

Adrian M. P. Brasoveanu, y R. A. (2019). Semantic Fake News Detection: A Machine Learning Perspective. *Springer Nature Switzerland*, *2*(June), 283–296.

Ahmed, H., Traore, I., & Saad, S. (2017). Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. *First International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, *10618*, 169–181.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236.

Amine, B. M., Drif, A., & Giordano, S. (2019). Merging deep learning model for fake news detection. *2019 International Conference on Advanced Electrical Engineering, ICAEE 2019*, 5–8.

Arvinder Pal Singh Bali, Mexson Fernandes, S. C., & Goel, and M. (2019). *Comparative Performance of Machine Learning Algorithms for Fake News Detection* (Vol. 1046, Issue July). Springer Singapore.

Barua, R., Maity, R., Minj, D., Barua, T., & Layek, A. K. (2019). F-NAD: An Application for Fake News Article Detection using Machine Learning Techniques. *2019 IEEE Bombay Section Signature Conference, IBSSC 2019*, *2019Januar*, 0–5.

Borges, L., Martins, B., & Calado, P. (2019). Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *Journal of Data and Information Quality*, *11*(3).

Carvalho, C., Klagge, N., & Moench, E. (2011). The persistent effects of a false news shock. *Journal of Empirical Finance*, *18*(4), 597–615.

Dong-Ho Lee, Yu-Ri Kim, Hyeong-Jun Kim, Seung-Myun Park, Y.-J. Y. (2019). Fake News Detection Using Deel Learning. *Journal of Information Processing Systems*, *15*(5), 1119–1130.

Gerardo Ernesto Rolong Agudelo1, Octavio José Salcedo Parra1, 2, & Velandia, and J. B. (2018). Raising a Model for Fake News Detection Using Machine Learning in Python. *Challenges and Opportunities in the Digital Era*, *11195*(October), 325–336.

Gilda, S. (2018). Evaluating machine learning algorithms for fake news detection. *IEEE Student Conference on Research and Development: Inspiring Technology for Humanity, SCOReD 2017 - Proceedings*, *2018-Janua*, 110–115.

Han, W., & Mehta, V. (2019). Fake news detection in social networks using machine learning and deep learning: Performance evaluation. *Proceedings - IEEE International Conference on Industrial Internet Cloud, ICII 2019*, *Icii*, 375–380.

Hiramath, Chaitra K, P. G. . D. (2020). Fake News Detection Using Deep Learning Techniques. *ISCAIE 2020 - IEEE 10th Symposium on Computer Applications and Industrial Electronics*, *Ml*, 102–107.

Huang, B., & Carley, K. M. (2020). *Disinformation and Misinformation on Twitter during the Novel Coronavirus Outbreak*. http://arxiv.org/abs/2006.04278

Ibrishimova, M. D., & Li, K. F. (2020). A machine learning approach to fake news detection using knowledge verification and natural language processing. *Advances in Intelligent Systems and Computing*, *1035*, 223–234.

Jain, A., Shakya, A., Khatter, H., & Gupta, A. K. (2019). A smart System for Fake News Detection Using Machine Learning. *IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques, ICICT 2019*, 2–5.

Kaliyar, R. K., Goswami, A., & Narang, P. (2019). Multiclass Fake News Detection using Ensemble Machine Learning. *Proceedings of the 2019 IEEE 9th International Conference on Advanced Computing, IACC 2019*, 103–107.

Kareem, I., & Awan, S. M. (2019). Pakistani Media Fake News Classification using Machine Learning Classifiers. *3rd International Conference on Innovative Computing, ICIC 2019*, *Icic*, 1–6.

Katsaros, D., Stavropoulos, G., & Bridge, I. (2019). *Which machine learning paradigm for fake news detection ?* 383–387.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering*.

Kong, S. H., Tan, L. M., Gan, K. H., & Samsudin, N. H. (2020). Fake News Detection using Deep Learning. *ISCAIE 2020 - IEEE 10th Symposium on Computer Applications and Industrial Electronics*, *Ml*, 102–107.

Kresnakova, V. M., Sarnovsky, M., & Butka, P. (2019). Deep learning methods for Fake News detection. *IEEE Joint 19th International Symposium on Computational Intelligence and Informatics and 7th International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics, CINTI-MACRo 2019 - Proceedings*, 143–148.

Kumar, A., Singh, S., & Kaur, G. (2019). Fake news detection of Indian and United States election data using machine learning algorithm. *International Journal of Innovative Technology and Exploring Engineering*, *8*(11), 1559–1563.

Kumar, S., Asthana, R., Upadhyay, S., Upreti, N., & Akbar, M. (2020). Fake news detection using deep learning models: A novel approach. *Transactions on Emerging Telecommunications Technologies*, *31*(2), 1–23.

Lakshmanarao, A., Swathi, Y., & Srinivasa Ravi Kiran, T. (2019). An effecient fake news detection system using machine learning. *International Journal of Innovative Technology and Exploring Engineering*, *8*(10), 3125–3129.

Liu, H. (2019). A Location Independent Machine Learning Approach for Early Fake News Detection. *2019 IEEE International Conference on Big Data (Big Data)*, 4740–4746.

Manzoor, S. I., Singla, J., & Nikita. (2019). Fake news detection using machine learning approaches: A systematic review. *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019*, *Icoei*, 230–234.

Masood, R., & Aker, A. (2018). The fake news challenge: Stance detection using traditional machine learning approaches. *IC3K 2018 - Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, *3*(Kmis), 128–135.

Mokhtar, M. S., Jusoh, Y. Y., Admodisastro, N., Pa, N., & Amruddin, A. Y. (2019). Fakebuster: Fake news detection system using logistic regression technique in machine learning. *International Journal of Engineering and Advanced Technology*, *9*(1), 2407–2410.

Poddar, K., Amali, G. B. D., & Umadevi, K. S. (2019). Comparison of Various Machine Learning Models for Accurate Detection of Fake News. *2019 Innovations in Power and Advanced Computing Technologies, i-PACT 2019*, 1–5.

Qawasmeh, E., Tawalbeh, M., & Abdullah, M. (2019). Automatic Identification of Fake News Using Deep Learning. *2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS 2019*, 383–388.

Reis, J. C. S., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Explainable machine learning for fake news detection. *WebSci 2019 - Proceedings of the 11th ACM Conference on Web Science*, 17–26.

Sherry Girgis, Eslam Amer, M. G. (2018). Deep Learning Algorithms for Detecting Fake News in Online Text. *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, 93–97.

Singh, R., Chun, S. A., & Atluri, V. (2020). Developing Machine Learning Models to Automate News Classification. *The 21st Annual International Conference on Digital Government Research*, 354–355.

Supanya Aphinwongsophon, P. C. (2018). *Detecting Fake nes with Machine Learning Method*. 528–531.

Tanik Saikh, Amit Anand, AsifEkbal, and P. B. (2019). A Novel Approach Towards Fake News Detection: Deep Learning Augmented with Textual Entailment Features. In *Nldb 2019* (Vol. 1, Issue Table 1). Springer International Publishing.

Verma, A., Mittal, V., & Dawn, S. (2019). FIND: Fake Information and News Detections using Deep Learning. *2019 12th International Conference on Contemporary Computing, IC3 2019*, 1–7.

Ye-Chan Ahn, C.-S. J. (2019). Natural Language Contents Evaluation System for Detecting Fake News using Deep Learning. *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, *February*, 1–9.

Critérios Editoriais

A RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação) é um periódico científico, que foca a investigação e a aplicação prática inovadora no domínio dos sistemas e tecnologias de informação.

O Conselho Editorial da RISTI incentiva potenciais autores a submeterem artigos originais e inovadores para avaliação pelo Conselho Científico.

A submissão de artigos para publicação na RISTI deve realizar-se de acordo com as chamadas de artigos e as instruções e normas disponibilizadas no sítio Web da revista (http://www.risti.xyz).

Todos os artigos submetidos são avaliados por um conjunto de membros do Conselho Científico, não inferior a três elementos.

Em cada número da revista são publicados entre cinco a oito dos melhores artigos submetidos.

Criterios Editoriales

La RISTI (Revista Ibérica de Sistemas y Tecnologías de la Información) es un periódico científico, centrado en la investigación y en la aplicación práctica innovadora en el dominio de los sistemas y tecnologías de la información.

El Consejo Editorial de la RISTI incentiva autores potenciales a enviar sus artículos originales e innovadores para evaluación por el Consejo Científico.

Lo envío de artículos para publicación en la RISTI debe hacerse de conformidad con las llamadas de los artículos y las instrucciones y normas establecidas en el sitio Web de la revista (http://www.risti.xyz).

Todos los trabajos enviados son evaluados por un número de miembros del Consejo Científico de no menos de tres elementos.

En cada número de la revista se publican cinco a ocho de los mejores artículos enviados.