

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

Johan Alejandro Zambrano Palma

19 de octubre de 2023

Resumen

Las redes sociales han transformado la forma en que la sociedad se informa, pero también han dado lugar a la propagación de noticias falsas sin verificar. Este estudio analiza cómo el aprendizaje automático y el aprendizaje profundo se utilizan para desarrollar clasificadores de noticias falsas en las redes sociales. Los resultados muestran que estos modelos son ampliamente empleados, especialmente en ámbitos como la política, sociedad, salud y desastres naturales. Se observa que el aprendizaje automático se utiliza más que el aprendizaje profundo, pero ambos enfoques son efectivos, con la calidad de los datos y el método de extracción de características siendo factores clave en su rendimiento.

1. Introducción

Las redes sociales son ampliamente utilizadas en todo el mundo, lo que ha dado lugar a la proliferación de noticias falsas o "fake news." [1] Estas noticias desinforman a las personas y a menudo se viralizan, causando pánico en la sociedad. Se han observado casos significativos de noticias falsas en ámbitos como la política, las finanzas y la salud.

Para abordar este problema, los investigadores han utilizado técnicas de inteligencia artificial, como el aprendizaje automático, para desarrollar sistemas de detección de noticias falsas en las redes sociales. Estos sistemas se basan en la minería de texto y emplean algoritmos de aprendizaje automático, como árboles de decisión y máquinas de soporte vectorial, así como modelos de aprendizaje profundo, como redes convolucionales.

El objetivo de este estudio es realizar una revisión de la literatura sobre los modelos de inteligencia artificial utilizados para detectar noticias falsas en redes sociales. La metodología empleada se basa en una revisión sistemática, y los resultados proporcionarán información sobre la eficacia de estos sistemas.

El documento se divide en varias secciones, incluyendo una descripción de la metodología utilizada, los resultados detallados y las conclusiones obtenidas a partir del análisis de los estudios primarios recuperados.

2. Metodología

Una revisión sistemática de la literatura implica la identificación y evaluación de los estudios más relevantes sobre un tema específico. En este estudio, se utilizó la metodología propuesta por Kitchenham y Charters en 2007, una metodología formal ampliamente reconocida en ciencias computacionales, para llevar a cabo este proceso de revisión sistemática.

2.1. Preguntas de investigación

Las noticias falsas han atraído la atención de investigadores especializados en inteligencia artificial (IA). El estudio se enfoca en responder la pregunta sobre cómo los modelos de aprendizaje de IA han contribuido a la creación de clasificadores para detectar automáticamente noticias falsas en las redes sociales, un medio clave para la información. El estudio plantea cuatro preguntas específicas (PIs) para abordar esta cuestión.

- 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

En este estudio, se buscó investigar la detección automática de noticias falsas en redes sociales a través de una revisión sistemática de la literatura. Se llevaron a cabo tres pasos en la estrategia de búsqueda: definición de las fuentes de información (bibliotecas digitales y bases de datos documentales), formulación de una cadena de búsqueda específica y selección de los estudios primarios recuperados. Los criterios de inclusión se centraron en trabajos que utilizaran inteligencia artificial para detectar noticias falsas en redes sociales, mientras que se establecieron criterios de exclusión para descartar documentos repetidos, escritos en idiomas distintos al inglés, inaccesibles o publicados antes de 2015.

En total, se recuperaron 106 artículos, de los cuales 55 eran únicos. Después de analizar en detalle títulos y resúmenes, se redujeron a 40 estudios para una revisión exhaustiva. Sin embargo, 4 de ellos eran libros y uno era un análisis teórico, dejando un total de 36 estudios para su revisión completa.

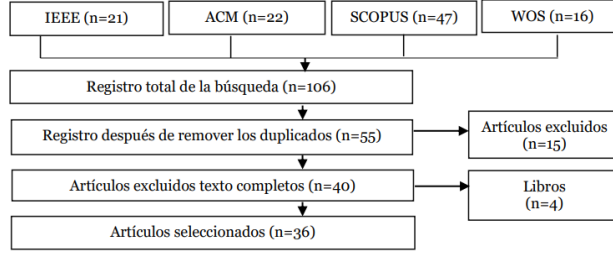


Figura 1: Proceso de selección de estudios primarios

3. Resultados y discusión

En esta sección, se presentan los resultados basados en los 36 artículos estudiados y se responden a las preguntas de investigación formuladas en este estudio. Los resultados se resumen a continuación:

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 utilizado tanto algoritmos de aprendizaje automático como de aprendizaje profundo. De los 36 estudios analizados, 12 se centraron en el uso exclusivo de algoritmos de aprendizaje automático, 13 se enfocaron exclusivamente en modelos de aprendizaje profundo (redes neuronales), y 10 estudios exploraron tanto modelos de aprendizaje automático como de aprendizaje profundo. Algunos de estos últimos compararon la efectividad de ambos enfoques, mientras que otros desarrollaron modelos híbridos, mayoritariamente basados en aprendizaje profundo, para mejorar la precisión en la detección de noticias falsas.

Los datos indican que el uso de la inteligencia artificial para abordar el problema de las noticias falsas ha crecido en los últimos años, con la mayoría de las propuestas publicadas en 2019. Se destaca que muchos de estos estudios provienen de investigadores en la India, donde se ha impulsado una campaña para combatir las noticias falsas, y la academia desempeña un papel importante en la resolución de este problema que ha tenido repercusiones políticas significativas.

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

Se han empleado tanto algoritmos de aprendizaje automático como de aprendizaje profundo para detectar noticias falsas. Entre los algoritmos de aprendizaje automático, los más utilizados incluyen máquinas de vectores de soporte, regresión logística, algoritmos basados en el Teorema de Bayes, bosques aleatorios y árboles de decisión. Los algoritmos de máquinas de vectores de soporte obtuvieron la mayor precisión con un 99,90 %, seguidos de la regresión logística con un 91,60 %.

Grupo	#	País	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)
GRUPO 2	S46	Colombia	2018	Scopus	x		(Rolong, et al., 2018)
	S47	Canadá	2017	Scopus	x		(Ahmed et al., 2017)
	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)
GRUPO 3	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)
	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)
GRUPO 3	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)

Figura 2: Propuestas orientadas a detectar noticias falsas usando aprendizaje automático

En el ámbito del aprendizaje profundo, se aplicaron redes neuronales, incluyendo redes con memoria, redes convolucionales y redes recurrentes. La red neuronal genérica obtuvo la mejor precisión con un 99,90 %, seguida de las redes convolucionales con un 96,00 % y las redes de memoria a largo plazo con un 95,30 %. Además, se observó que los modelos híbridos que combinan diferentes enfoques de aprendizaje profundo pueden lograr una precisión aún mayor, como un modelo que alcanzó un 88,78 % de precisión al combinar redes convolucionales y redes de memoria.

La precisión de estos algoritmos tanto de aprendizaje automático como de aprendizaje profundo depende de los datos utilizados para entrenar los modelos y del método de extracción de características aplicado, que a menudo incluye técnicas como la frecuencia de término, TF-IDF, vectores globales (GloVe) y CountVectorizer.

PI3: ¿Qué herramientas de software y datos se han usado para

crear modelos predictivos de detección de noticias falsas en redes sociales?

En cuanto a las herramientas de desarrollo utilizadas para crear clasificadores de noticias falsas en redes sociales, Python fue la elección principal en la mayoría de los estudios (18 en total). Python es ampliamente utilizado en la inteligencia artificial y en la minería de texto, lo que lo convierte en una herramienta popular. Se emplearon bibliotecas de aprendizaje automático de Python, como sci-kit learn, Keras y TensorFlow, además de la librería para procesamiento de lenguaje natural, NLTK.

En cuanto a los datos utilizados, los estudios aprovecharon conjuntos de datos públicos disponibles en repositorios como Kaggle, GitHub y PolitiFact. Algunos también recurrieron a fuentes adicionales, como Twitter, The New York Times, Washington Post, Reuter.com y BuzzFeed. En general, se aplicó la minería de texto a conjuntos de datos predefinidos, lo que agilizó el proceso de entrenamiento de clasificadores. Algunos investigadores crearon sus propios conjuntos de datos utilizando información de sitios web y blogs de noticias.

Los conjuntos de datos utilizados en los estudios incluyeron nombres como NYT, LIAR, News, Fake News, Fake News Net, Fake Real News y FNC-1, entre otros. En resumen, se emplearon diversas fuentes de datos y conjuntos de datos preexistentes, así como algunos creados específicamente para estos fines.

PI4: ¿En qué ámbitos se ha detectado noticias falsas en redes sociales?

Los estudios examinados se enfocaron en detectar noticias falsas en varios ámbitos, siendo la política el más destacado con 26 estudios. Otros ámbitos incluyeron negocios y economía (3 estudios), sociedad, deportes y cultura (4 estudios), ciencia, tecnología y salud (3 estudios), entretenimiento (2 estudios) y desastres naturales (1 estudio). En 7 estudios no se especificó el ámbito de las noticias falsas analizadas. La política fue el ámbito más destacado, ya que las noticias falsas en este campo pueden tener un impacto significativo en la estabilidad de los gobiernos y la sociedad.

4. Conclusiones

Los modelos de aprendizaje de inteligencia artificial se han utilizado ampliamente para crear sistemas de detección automática de noticias falsas. Algunos clasificadores se basaron en algoritmos de aprendizaje automático como árboles de decisión, Teorema de Bayes, regresión logística, vecinos más cercanos, bosques aleatorios y máquinas de vectores de soporte, con la última logrando la mejor precisión del 99.9%. Otros 23 estudios utilizaron redes neuronales, incluyendo redes convolucionales, redes con memoria a corto plazo y redes recurrentes, con redes convolucionales alcanzando una precisión del 97% y un modelo genérico de red neuronal obteniendo un 99.90% de precisión.

A pesar de los avances en la detección de noticias falsas en tweets, sigue siendo un desafío integrar estos modelos en redes sociales como Twitter para advertir a los usuarios antes de compartir contenido. Los modelos actuales se

centran en identificar las principales características de los tweets que contienen noticias falsas, y existe una brecha en la detección de noticias falsas en otras redes sociales como Facebook y en otros formatos más allá del texto.

A pesar de que las noticias falsas en los ámbitos político, económico y de desastres naturales son las más comunes, es importante ampliar los esfuerzos para detectar noticias falsas en otros contextos y formatos, como imágenes, usuarios, conocimiento, estilo y postura, utilizando técnicas de inteligencia artificial y aprendizaje automático.

Referencias

- [1] Hyeong-Jun Kim Seung-Myun Park Y.-J. Y Dong-Ho Lee, Yu-Ri Kim. Fake news detection using deep learning. *Journal of Information Processing Systems*, 15(5):1119–1130, 2019.