



Estrategias basadas en inteligencia artificial para la detección de noticias falsas

Strategies based on artificial intelligence for the detection of fake news

Digmar García Paredes¹

¹ Universidad Nacional de San Martín, Perú.
Email: dgarcia@alumno.unsm.edu.pe

RESUMEN

El artículo examina las estrategias de inteligencia artificial (IA) para combatir las noticias falsas, resaltando el aumento de la desinformación, especialmente durante la pandemia, y su impacto negativo en la toma de decisiones pública. La propagación acelerada de noticias falsas frente a la veracidad subraya la urgencia de métodos efectivos de detección. Mediante la revisión sistemática de literatura, se explora el uso de aprendizaje automático, aprendizaje profundo y procesamiento del lenguaje natural (NLP) para automatizar la identificación de noticias falsas, destacando conjuntos de datos claves como BuzzFeedNews, LIAR y BS Detector, entre otros, esenciales para entrenar algoritmos de detección. El estudio discute varios enfoques y técnicas de IA aplicadas a la detección, incluyendo redes neuronales convolucionales (CNN), LSTM bidireccionales, y la combinación de CNN con LSTM, mostrando mejoras significativas en la precisión y la eficacia. Sin embargo, se señalan las limitaciones de estas técnicas, como la volatilidad de los datos de entrenamiento y la dificultad de adaptar modelos a diferentes contextos de desinformación. La conclusión destaca la IA como una herramienta vital contra las noticias falsas, enfatizando la necesidad de avanzar en la investigación y el desarrollo de tecnologías más sofisticadas para fortalecer la detección de desinformación y proteger la integridad informativa en la sociedad. La lucha contra las noticias falsas es compleja, pero las estrategias basadas en IA muestran un camino prometedor hacia soluciones efectivas.

Palabras Clave: inteligencia artificial; noticias falsas; aprendizaje automático; procesamiento del lenguaje natural; conjunto de datos; redes neuronales.

ABSTRACT

The article examines artificial intelligence (AI) strategies to combat fake news, highlighting the rise in misinformation, especially during the pandemic, and its negative impact on public decision-making. The accelerated spread of fake news in the face of truth underlines the urgency of effective detection methods. Through a systematic literature review, the use of machine learning, deep learning, and natural language processing (NLP) to automate the identification of fake news is explored, highlighting key data sets such as BuzzFeedNews, LIAR, and BS Detector, among others, essential to train detection algorithms.

Cómo citar: Paredes, D. G. (2023). Strategies based on artificial intelligence for the detection of fake news. AWARI; 4, 1-6. DOI: 10.47909/awari.58

Recibido: 21-03-2023 / **Aceptado:** 12-08-2023 / **Publicado:** 19-08-2023

Editor: Adilson Luiz Pinto

Copyright: © 2023 The author(s). This is an open access article distributed under the terms of the CC BY-NC 4.0 license which permits copying and redistributing the material in any medium or format, adapting, transforming, and building upon the material as long as the license terms are followed.

The study discusses various AI approaches and techniques applied to detection, including convolutional neural networks (CNN), bidirectional LSTM, and the combination of CNN with LSTM, showing significant improvements in accuracy and efficiency. However, the limitations of these techniques are pointed out, such as the volatility of the training data and the difficulty of adapting models to different misinformation contexts. The conclusion highlights AI as a vital tool against fake news, emphasizing the need to advance research and develop more sophisticated technologies to strengthen disinformation detection and protect information integrity in society. The fight against fake news is complex, but AI-based strategies show a promising path toward practical solutions.

Keywords: artificial intelligence; fake news; machine learning; natural language processing; dataset; neural networks.

INTRODUCCIÓN

EL TÉRMINO “noticias falsas” se ha empleado ampliamente para referirse a la desinformación, la propaganda, la sátira, el chisme, el clickbait y las noticias de baja calidad (Xu *et al.*, 2022). Estas noticias están dominando la propagación de desinformación, alterando las percepciones y la toma de decisiones de las personas. Según Amer *et al.*, (2022), están diseñadas para influenciar a las personas hacia concepciones sesgadas o falsas. Sin embargo, se reconoce que las noticias falsas y la desinformación son problemas significativos en Internet, que pueden acarrear un costo social considerable.

El impacto de las noticias falsas ha sido particularmente notable durante tiempos de pandemia, evidenciando un aumento de este fenómeno. En Perú, según Nieves-Cuervo *et al.*, (2021), la vasta cantidad de información generada por la pandemia se ha difundido a través de medios convencionales y digitales, complicando la identificación de fuentes y datos confiables. La difusión intencional de información falsa en Perú está penalizada con hasta cuatro años de prisión, según el artículo 438 del Código Penal peruano (El Peruano, 2021), que sanciona la alteración deliberada de la autenticidad de la información con perjuicio a terceros.

Las noticias falsas se difunden seis veces más rápido que la información verídica (Carrasco-Farré, 2022), lo que resalta el problema de la alta incidencia de noticias falsas en medios digitales con fines de desinformación, ya sea para beneficios políticos o monetarios. El uso de bots y scammers para propagar esta información rápidamente, junto con estrategias que dificultan su detección, son tácticas comunes en este ámbito (Kumar & Singh, 2022).

Existe una variedad de técnicas basadas en conjuntos de datos y modelos que contribuyen a la detección de noticias falsas. Sin embargo, estos recursos no siempre se utilizan adecuadamente, lo que indica una falta de conocimiento documentado sobre estos modelos y conjuntos de datos. Además, las técnicas convencionales de Machine Learning no han demostrado ser eficaces en la detección de noticias falsas, ya que no logran una extracción profunda de información a diferentes niveles (Saleh *et al.*, 2021).

La propagación de noticias falsas en plataformas digitales tiene un impacto directo en eventos reales, como en elecciones presidenciales, y ha jugado un papel significativo en la desinformación durante la pandemia, demostrando el daño real que pueden causar, incluyendo la pérdida de vidas (Megías *et al.*, 2022). Este problema resalta la necesidad de desarrollar soluciones eficaces para la detección automática de noticias falsas, aprovechando las tecnologías de inteligencia artificial.

MATERIAL Y MÉTODO

La búsqueda de soluciones eficientes para reemplazar la verificación manual de noticias falsas ha fomentado el desarrollo de estrategias de detección basadas en inteligencia artificial, utilizando técnicas de aprendizaje automático y profundo para automatizar este proceso. Esto plantea la pregunta: ¿Qué estrategias basadas en inteligencia artificial existen para la detección de noticias falsas? Para responder a esta pregunta, recurrimos al método propuesto por Riesenberg & Justice (2014), que facilita el acceso a fuentes de información relevantes a través de una revisión sistemática, permitiendo una identificación precisa y una síntesis crítica de los artículos relevantes.

Esta investigación se basó en la revisión de literatura para identificar, evaluar y sintetizar el conocimiento generado por investigadores y profesionales (Booth, A., Sutton, A., y Papaioannou, 2021; Randolph, 2019). Se realizó una revisión en las bases de datos de Google Scholar, SpringerLink, IEEE Xplore y ScienceDirect, enfocándose en revistas científicas que tratan sobre inteligencia artificial aplicada a la detección de noticias falsas y en publicaciones accesibles para la investigación. La búsqueda se llevó a cabo utilizando términos como “Artificial Intelligence AND Fake News”, “Deep Learning AND Fake News”, entre otros, enfocándose en los últimos cinco años debido al incremento significativo en la producción de investigaciones relacionadas.

La selección de literatura se hizo con el objetivo de analizar los aportes de la inteligencia artificial en la detección de noticias falsas en las últimas décadas. Documentos seleccionados también fueron proporcionados por servicios del Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC) del Estado Peruano. La gestión de la información se realizó con el gestor bibliográfico Mendeley para la sistematización de la información y la elaboración de citas y referencias.

RESULTADOS DE LA REVISIÓN

La proliferación de noticias falsas puede tener un impacto significativo en la percepción del lector, llevando a decisiones erróneas con posibles consecuencias financieras y de salud (Shu *et al.*, 2017). Es vital encontrar soluciones alternativas a la verificación manual de noticias falsas, y una propuesta es aplicar sistemas de detección basados en inteligencia artificial (Umer *et al.*, 2020). Los métodos de aprendizaje automático han demostrado ser efectivos en este ámbito, creando modelos precisos, eficaces y prácticos que permiten una detección casi en tiempo real de fenómenos indeseables.

El Procesamiento del Lenguaje Natural (NLP) se utiliza para determinar patrones lingüísticos que diferencian noticias falsas de reales, empleando algoritmos y técnicas de aprendizaje automático (Shahbazi & Byun, 2021). Este enfoque es crucial para abordar la difusión masiva de información errónea de manera escalable y efectiva, complementando la labor de los verificadores de hechos humanos (Khan *et al.*, 2021).

Conjuntos de datos para la detección de noticias falsas

Los conjuntos de datos desempeñan un papel fundamental en el entrenamiento de algoritmos para reconocer patrones de desinformación. Se destacan varios conjuntos de datos utilizados para la detección de noticias falsas, como

BuzzFeedNews: El conjunto de datos contiene noticias sobre las elecciones estadounidenses de 2016 publicadas en Facebook por 9 agencias de noticias. Este conjunto de datos contiene 1.262 artículos que en su mayoría son ciertos, 212 que son una mezcla de verdadero y falso, y 87 que son falsos, después de limpiar el conjunto de datos de cuerpos de noticias o títulos faltantes (Shrestha & Spezzano, 2021).

LIAR: El conjunto de datos se obtiene de PolitiFact 24 utilizando una API (Wang, 2017) y consta de 12.836 artículos de varios campos. La clasificación se ha realizado en categorías como “pantalones de fuego”, “falso”, “apenas cierto”, “verdadero a medias”, “mayormente cierto” y “verdadero”.

BS Detector: el conjunto de datos se genera mediante la función de autoanotación del kit de herramientas que rastrea todos los enlaces en una página web determinada y determina las fuentes no confiables comparándolas con una lista de dominios proporcionada de antemano (Oshikawa *et al.*, 2020).

Fake News Challenge (FNC): El conjunto de datos consiste en titulares o contenido de artículos de noticias y tiene como objetivo etiquetar la postura asociada con él en categorías como “de acuerdo”, “en desacuerdo”, “discutido” y “no relacionado”. Mientras que el conjunto de trenes consta de 49 972 registros, el conjunto de prueba consta de 25 413 registros. El dominio del conjunto de datos incluye varios temas como política, salud, medio ambiente, estilo de vida, etc. (Qawasmeh *et al.*, 2019).

CRED BANK: El conjunto de datos consta de 60 millones de tweets de octubre de 2015 durante un período de 96 días. Todos los tweets han sido segregados y atribuidos a más de 1000 eventos y la credibilidad de cada evento verificada por asociados de Amazon-Mechanical-Turk

BuzzFace: el conjunto de datos es una extensión de BuzzFeed con 2263 artículos junto con 1,6 millones de comentarios (Santia & Williams, 2018).

FacebookHoax: el conjunto de datos consta de 15 500 publicaciones repartidas en 32 páginas relacionadas con teorías científicas y de conspiración con más de 2 millones de Me gusta (Tacchini *et al.*, 2017).

FakeNewsNet: Es un conjunto de datos multidimensional que contiene varias categorías de información, como contexto de noticias, contexto social, información espacial y temporal. Los datos han sido tomados de repositorios como PolitiFact y GossipCop (Shu *et al.*, 2020).

Detección de noticias falsas mediante aprendizaje automático y profundo

Diversos estudios han abordado la detección de noticias falsas a través de técnicas avanzadas de aprendizaje automático y profundo. Por ejemplo, Umer *et al.*, (2020) utilizaron una red neuronal híbrida para analizar la postura de un artículo respecto a su titular, logrando mejorar la precisión y la puntuación F1. Otros estudios, como el de Senhadji & Ahmed (2022), han comparado el rendimiento de clasificadores como Naive Bayes y LSTM, encontrando ventajas significativas en ciertos enfoques sobre otros.

Estos avances subrayan la importancia de la inteligencia artificial en la lucha contra la desinformación, demostrando que es posible desarrollar sistemas eficientes para la detección de noticias falsas. La integración de diferentes métodos y técnicas, incluyendo NLP y aprendizaje profundo, ofrece perspectivas prometedoras para abordar este desafío global.

Limitaciones de las técnicas sujetas a la resolución de esta problemática

Es importante reconocer las limitaciones de las técnicas actuales para resolver esta problemática. Una limitante significativa es la volatilidad de los datos de entrenamiento sobre noticias falsas en comparación con otras áreas de investigación. Por ejemplo, un modelo entrenado con datos de noticias falsas difundidas durante elecciones presidenciales podría no ser efectivo para detectar noticias falsas sobre una

pandemia. Además, la libre y fácil propagación de información a través de plataformas digitales complica la identificación automática de noticias falsas.

CONCLUSIONES

Este artículo presenta una revisión sistemática sobre los avances en la detección de noticias falsas basada en inteligencia artificial, ofreciendo una visión integral y actualizada del tema. A través de esta revisión, se evidencia el apoyo de diferentes algoritmos en la detección de noticias falsas, destacando la importancia de la inteligencia artificial y la necesidad de continuar desarrollando tecnologías y enfoques innovadores en este campo. La tendencia creciente de noticias falsas y las medidas insuficientes para controlarlas subrayan la necesidad urgente de implementar soluciones eficaces y confiables para mitigar su impacto negativo en la sociedad.

Conflicto de interés

La autora declara que no hay conflicto de interés.

Declaración de consentimiento de datos

Los datos utilizados en el estudio están presentes en el artículo. ●

REFERENCIAS

- BAHAD, P., SAXENA, P., & KAMAL, R. (2019). Fake News Detection using Bi-directional LSTM-Recurrent Neural Network. *Procedia Computer Science*, 165(2019), 74-82. <https://doi.org/10.1016/j.procs.2020.01.072>
- BALAKRISHNAN, V., NG, K. S., & RAHIM, H. A. (2021). To share or not to share – The underlying motives of sharing fake news amidst the COVID-19 pandemic in Malaysia. *Technology in Society*, 66, 101676. <https://doi.org/10.1016/j.techsoc.2021.101676>
- BATTA, M. (2020). Machine Learning Algorithms – A Review. *International Journal of Science and Research (IJ)*, 9(1), 381-386. <https://doi.org/10.21275/ART20203995>
- BOOTH, A., SUTTON, A., Y PAPAIOANNOU, D. (2021). Systematic Approaches to a Successful

- Literature Review. *Reino Unido: The University of Sheffield*. <https://us.sagepub.com/en-us/nam/systematic-approaches-to-a-successful-literature-review/book244586>
- BUZEA, M. C., TRAUSAN-MATU, S., & REBEDEA, T. (2022). Automatic Fake News Detection for Romanian Online News. *Information (Switzerland)*, 13(3), 1-13. <https://doi.org/10.3390/info13030151>
- CHOI, H., & KO, Y. (2021). Using Adversarial Learning and Biterm Topic Model for an Effective Fake News Video Detection System on Heterogeneous Topics and Short Texts. *IEEE Access*, 9, 164846-164853. <https://doi.org/10.1109/ACCESS.2021.3122978>
- DONG, X., VICTOR, U., & QIAN, L. (2020). Two-Path Deep Semisupervised Learning for Timely Fake News Detection. *IEEE Transactions on Computational Social Systems*, 7(6), 1386-1398. <https://doi.org/10.1109/TCSS.2020.3027639>
- KHAN, T., MICHALAS, A., & AKHUNZADA, A. (2021). Fake news outbreak 2021: Can we stop the viral spread? *Journal of Network and Computer Applications*, 190(April), 103112. <https://doi.org/10.1016/j.jnca.2021.103112>
- KISHWAR, A., & ZAFAR, A. (2023). Fake news detection on Pakistani news using machine learning and deep learning. *Expert Systems with Applications*, 211(August 2022). <https://doi.org/10.1016/j.eswa.2022.118558>
- KOZIK, R., KULA, S., CHORAŚ, M., & WOŹNIAK, M. (2022). Technical solution to counter potential crime: Text analysis to detect fake news and disinformation. *Journal of Computational Science*, 60(October 2021), 101576. <https://doi.org/10.1016/j.jocs.2022.101576>
- LI, D., GUO, H., WANG, Z., & ZHENG, Z. (2021). Unsupervised Fake News Detection Based on Autoencoder. *IEEE Access*, 9, 29356-29365. <https://doi.org/10.1109/ACCESS.2021.3058809>
- OSHIKAWA, R., QIAN, J., & WANG, W. Y. (2020). A survey on natural language processing for fake news detection. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, 6086-6093.
- 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. <https://doi.org/10.1109/SNAMS.2019.8931873>
- RANDOLPH, J. (2019). A Guide to Writing the Dissertation Literature Review. *Practical Assessment, Research, and Evaluation*, 14(1). <https://doi.org/https://doi.org/10.7275/b0az-8t74>
- RIESENBERG, L. A., & JUSTICE, E. M. (2014). Revisión sistemática de la bibliografía (parte 1). *Nursing (Ed. Española)*, 31(6), 61-64. <https://doi.org/10.1016/j.nursi.2014.12.019>
- SALEH, H., ALHARBI, A., & ALSAMHI, S. H. (2021). OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection. *IEEE Access*, 9, 129471-129489. <https://doi.org/10.1109/ACCESS.2021.3112806>
- SANTIA, G. C., & WILLIAMS, J. R. (2018). Buzz face: A news veracity dataset with facebook user commentary and egos. *12th International AAAI Conference on Web and Social Media, ICWSM 2018, Icwsm*, 531-540.
- SENHADJI, S., & AHMED, R. A. S. (2022). Fake news detection using naïve Bayes and long short term memory algorithms. *IAES International Journal of Artificial Intelligence*, 11(2), 746-752. <https://doi.org/10.11591/ijai.v11.i2.pp746-752>
- SHAHBAZI, Z., & BYUN, Y. C. (2021). Fake Media Detection Based on Natural Language Processing and Blockchain Approaches. *IEEE Access*, 9, 128442-128453. <https://doi.org/10.1109/ACCESS.2021.3112607>
- SHRESTHA, A., & SPEZZANO, F. (2021). *Textual Characteristics of News Title and Body to Detect Fake News: A Reproducibility Study* (pp. 120-133). https://doi.org/10.1007/978-3-030-72240-1_9
- SHU, K., MAHUESWARAN, D., WANG, S., LEE, D., & LIU, H. (2020). FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data*, 8(3), 171-188. <https://doi.org/10.1089/big.2020.0062>
- SHU, K., SLIVA, A., WANG, S., TANG, J., & LIU, H. (2017). Fake News Detection on Social Media. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36. <https://doi.org/10.1145/3137597.3137600>
- SONG, C., SHU, K., & WU, B. (2021). Temporally evolving graph neural network for fake news detection. *Information Processing &*

- Management*, 58(6), 102712. <https://doi.org/10.1016/j.ipm.2021.102712>
- TACCHINI, E., BALLARIN, G., DELLA VEDOVA, M. L., MORET, S., & DE ALFARO, L. (2017). Some like it Hoax: Automated fake news detection in social networks. *CEUR Workshop Proceedings*, 1960. <https://developers.facebook.com/docs/graph-api>.
- UMER, M., IMTIAZ, Z., ULLAH, S., MEHMOOD, A., CHOI, G. S., & ON, B.-W. (2020). Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM). *IEEE Access*, 8, 156695-156706. <https://doi.org/10.1109/ACCESS.2020.3019735>
- WANG, W. Y. (2017). "Liar, liar pants on fire": A new benchmark dataset for fake news detection. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2, 422-426. <https://doi.org/10.18653/v1/P17-2067>
- YANG, Y., ZHENG, L., ZHANG, J., CUI, Q., LI, Z., & YU, P. S. (2018). *TI-CNN: Convolutional Neural Networks for Fake News Detection*. <http://arxiv.org/abs/1806.00749>

