**DETECCIÓN DE FAKE NEWS**

**USANDO MODELOS DE NLP**

Pasos a seguir:

1. Análisis del Estado del Arte del Problema: Se llevará a cabo un exhaustivo estudio del estado actual del problema de detección de noticias falsas, Deep Learning y Procesamiento del Lenguaje Natural (PLN).
2. Análisis de Conjuntos de Datos Disponibles: Se analizarán los conjuntos de datos más populares y recientes en la literatura, seleccionando dos para su resolución. Uno de estos conjuntos será en inglés y el otro en español.

* **Fact checking dataset[1]**:
  + Contiene 221 declaraciones obtenidas de Politifact y Channel.
* Mide la veracidad en una escala de cinco opciones.
* **EMERGENT[2]**:
  + Incluye 2.595 artículos de sitios web de rumores como Snopes y Twitter.
* Se utiliza para la clasificación de postura, con afirmaciones y documentos a favor o en contra.
* **LIAR[3]**:
* Contiene 12.836 declaraciones cortas recogidas de Politifact.
* Cada declaración se mide en una escala de seis opciones de veracidad y también incluye información sobre la temática, partido político, contexto y orador.
* **Fever[4]**:
* Contiene 185.445 afirmaciones generadas a partir de Wikipedia.
* Cada declaración está etiquetada como soportada, refutada o sin suficiente información, además de señalar las frases de Wikipedia utilizadas como evidencia.
* **FAKESNEWSNET[5]:**
* Consiste en la cabecera y el cuerpo del texto de artículos de noticias falsas obtenidos en BuzzFeed y Politifact.
* Recolecta información de Twitter sobre el compromiso social de estos artículos.
* **BUZZFEEDNEWS[6]**:
* Contiene 2.282 publicaciones de 9 agencias de Facebook, verificadas por 5 periodistas de BuzzFeed.
* Los artículos están recogidos de periódicos de distintas ideologías políticas.
* **SOME-LIKE-IT-HOAX[7]:** 
  + Contiene 15.500 publicaciones de 32 páginas de Facebook, incluyendo páginas de conspiración y organizaciones científicas.
* **PHEME[8]**:
  + Es un conjunto de datos de Twitter que contiene 330 hilos sobre 9 eventos de actualidad, etiquetados como verdadero o falso.
* **CREDBANK[9]**:
  + Contiene 60 millones de tweets agrupados en 1.049 eventos, clasificados en una escala de cinco de veracidad por 30 anotadores humanos.
* **FNC-1 dataset[10]:**
* Contiene 49.972 artículos de EMERGENT sobre política, sociedad y tecnología.
* Está preparado para detectar si la noticia es falsa o no relacionando el título con el cuerpo de la noticia.
* **Spanish Fake News Corpus [11]:**
* Contiene 971 noticias en español sobre 9 dominios distintos, clasificadas en verdadero o falso.
* Una versión más reciente incluye 572 noticias adicionales, con énfasis en temas actuales como el COVID-19.
* **CONSTRAINT@AAAI 2021[12]:**
* Consiste en 10.700 publicaciones en inglés, con 5.100 noticias falsas y 5.600 verdaderas.
* Las noticias verdaderas se recogen de Twitter, mientras que las falsas provienen de varias fuentes, incluyendo Twitter, Facebook, Whatsapp y sitios de fact-checking como Politifact, NewsChecker y Boomlive.

1. Desarrollo de un Sistema de Deep Learning para Resolver el Problema: Se desarrollará un sistema de Deep Learning para abordar el problema, utilizando únicamente el texto de la noticia (título y cuerpo) sin ningún tipo de información adicional.

\*finetunning

1. Implementación de una Interfaz Web Básica para el Problema: Se creará e implementará una interfaz web para demostrar de manera interactiva el funcionamiento del sistema desarrollado.

# REFERENCIAS

[1] Andreas Vlachos and Sebastian Riedel. Fact checking: Task definition and dataset construction. In Proceedings of the ACL 2014 workshop on language technologies and computational social science, pages 18–22,2014.

[2] William Ferreira y Andreas Vlachos. "Emergent: a novel data-set for stance classification." En Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies, ACL, 2016.

[3] William YangWang. ”liar, liar pants on fire”: A new benchmark datasetfor fake news detection. arXiv preprint arXiv:1705.00648, 2017.

[4] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, y Arpit Mittal. "Fever: a large-scale dataset for fact extraction and verification." arXiv preprint arXiv:1803.05355, 2018.

[5] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, y Huan Liu. "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, 2020.

[6] Benjamin Horne y Sibel Adali. "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news." En Proceedings of the International AAAI Conference on Web and Social Media, volumen 11, páginas 759–766, 2017..

[7] Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, y Luca de Alfaro. "Some like it hoax: Automated fake news detection in social networks." arXiv preprint arXiv:1704.07506, 2017

[8] Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, y Peter Tolmie. "Analysing how people orient to and spread rumours in social media by looking at conversational threads." PloS One, 11(3):e0150989, 2016.

[9] Tanushree Mitra y Eric Gilbert. "Credbank: A large-scale social media corpus with associated credibility annotations." En Proceedings of the International AAAI Conference on Web and Social Media, volumen 9, páginas 258–267, 2015.

[10] Benjamin Riedel, Isabelle Augenstein, Georgios P Spithourakis, y Sebastian Riedel. "A simple but tough-to-beat baseline for the fake news challenge stance detection task." arXiv preprint arXiv:1707.03264, 2017.

[11] Juan-Pablo Posadas-Durán, Helena Gómez-Adorno, Grigori Sidorov, y Jesús Jaime Moreno Escobar. "Detection of fake news in a new corpus for the Spanish language." \*Journal of Intelligent & Fuzzy Systems\*, 36(5):4869–4876, 2019.

[12] Parth Patwa, Mohit Bhardwaj, Vineeth Guptha, Gitanjali Kumari, Shivam Sharma, Srinivas PYKL, Amitava Das, Asif Ekbal, Shad Akhtar, y Tanmoy Chakraborty. "Overview of constraint 2021 shared tasks: Detecting English COVID-19 fake news and Hindi hostile posts." En Proceedings of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT). Springer, 2021.