



**Identificación de hechos y opiniones en redes sociales digitales utilizando
técnicas de aprendizaje automático**

Ruiz Rodríguez, Germán Oswaldo y Tovar Rojas, Jerry Joe

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Trabajo de integración curricular, previo a la obtención del título de Ingeniero en

Tecnologías de la Información

Ing. Jerez Villota Eleana Inés, MSc.

23 de febrero del 2024



Plagiarism and AI Content Detection Report

Ruiz_Tovar_TrabajoIntegracionCurric...

Scan details

Scan time: March 2th, 2024 at 18:16 UTC	Total Pages: 57	Total Words: 14238
--	--------------------	-----------------------

Plagiarism Detection



Types of plagiarism	Words
Identical	0.8% 120
Minor Changes	0.7% 101
Paraphrased	2% 288
Omitted Words	34.1% 4833

AI Content Detection



Text coverage	Words
AI text	0% 0
Human text	100% 9383

[View more](#)

🔍 Plagiarism Results: (22)

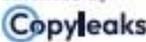
	T-ESPE-058483.pdf	1.6%
	http://repositorio.espe.edu.ec/bitstream/21000/37463/1/t-espe-058483.pdf	
	Admin	
	1 Desarrollo de un prototipo de aplicación móvil para la detección de enfermedades de la piel (como el vitiligo y psoriasis), utilizando...	
<hr/>		
	Procesamiento del Lenguaje Natural Diabolocom	1.2%
	https://es.diabolocom.com/blog/procesamiento-lenguaje-natural-nlp/	
<hr/>		
	Procesamiento del lenguaje natural y extracción de información: Defini...	0.6%
	https://prompt.uno/procesamiento-del-lenguaje-natural/procesamiento-del-lenguaje-natural-y-extraccion-de-...	
	Prompts para IA Inicio Tutoriales Cohere ChatGPT...	
<hr/>		
	Metodología XP Programación Extrema (Metodología ágil) - Diego Calvo	0.6%
	https://www.diegocalvo.es/metodologia-xp-programacion-extrema-metodologia-agil/	
	BIInteligencia de negocio Data Warehouse Minería de datos Cuadro de mando Pentaho Big DataIngesta de datosFlume Sqoop Nifi Sistemas...	

Firma



Ing. Eleana Inés Jerez Villota, MSc.
CC. 1717225039

Certified by



About this report
help.copyleaks.com

copyleaks.com





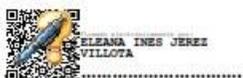
Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Certificación

Certifico que en el trabajo de integración curricular: **“Identificación de hechos y opiniones en redes sociales digitales utilizando técnicas de aprendizaje automático”**, fue realizado por los señores **Ruíz Rodríguez, Germán Oswaldo y Tovar Rojas, Jerry Joe**; el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 18 de marzo de 2024



Ing. Eleana Inés Jerez Villota, MSc.
C.C 1717225039



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Responsabilidad de Autoría

Nosotros, **Ruíz Rodríguez, Germán Oswaldo**, con cédula de ciudadanía N° 1724989338 y **Tovar Rojas, Jerry Joe**, con cédula de ciudadanía N° 1900476837, declaramos que el contenido, ideas y criterios de trabajo de integración curricular: **“Identificación de hechos y opiniones en redes sociales digitales utilizando técnicas de aprendizaje automático”**, es de nuestra autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 18 de marzo de 2024

.....
Ruíz Rodríguez, Germán Oswaldo
C.C 1724989338

.....
Tovar Rojas, Jerry Joe
C.C 1900476837



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Autorización de Publicación

Nosotros, **Ruiz Rodríguez, Germán Oswaldo**, con cédula de ciudadanía N° 1724989338 y **Tovar Rojas, Jerry Joe**, con cédula de ciudadanía N° 1900476837, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: **Título: “Identificación de hechos y opiniones en redes sociales digitales utilizando técnicas de aprendizaje automático”**, en el Repositorio Institucional, cuyo contenido, ideas y criterios son de nuestra responsabilidad.

Sangolquí, 18 de marzo de 2024

.....
Ruiz Rodríguez, Germán Oswaldo
C.C 1724989338

.....
Tovar Rojas, Jerry Joe
C.C 1900476837

Dedicatoria

Dedico este trabajo a mi amada familia y a todas las personas cercanas que han sido pilares fundamentales en mi camino hacia este logro. Su cariño, apoyo, amistad y aliento constante han sido mi mayor fortaleza. Todos han compartido conmigo momentos de alegría y han sido mi fuente de motivación en los momentos difíciles. A todos ustedes, les agradezco de corazón por formar parte esencial de este importante capítulo en mi vida académica. Este logro no sería posible sin su invaluable presencia y respaldo. ¡Gracias por estar siempre a mi lado!

Germán Oswaldo Ruíz Rodríguez

Dedicatoria

Dedico esta tesis a mis padres, Silvia y Polivio. A mi hermana, Coraima. Ya que gracias a su apoyo y acompañamiento me ayudaron a sobresalir en cada momento difícil presentado en mi vida. Han sido mi motor fundamental para continuar adelante y no rendirme en todo este camino. Este trabajo va dedicado a ellos, con todo amor y cariño.

Jerry Joe Tovar Rojas

Agradecimientos

Primero, deseo expresar mi profunda gratitud a Dios por haberme brindado la fuerza y la paciencia necesarias para alcanzar este momento crucial.

A mis amados padres, Cecilia y Oswaldo, pilares fundamentales en mi vida, les dedico este logro con todo mi amor. Su apoyo incondicional y sacrificios han sido la luz que ha guiado mi camino. A mi amada familia, Fausto, Ricardina, Cristina, Israel, Bolivia, Humberto, Michael, Fernanda, Geordy y Christopher, les debo un reconocimiento especial, pues su amor y respaldo han sido mi mayor fortaleza en cada desafío enfrentado.

A Sofía, mi compañera de vida y mi mayor apoyo, le agradezco infinitamente por su amor incondicional y su constante aliento.

A Roberto y Alejandro, mis fieles amigos y compañeros de aventuras, les dedico un espacio en estas líneas por haber compartido conmigo momentos inolvidables y por su inestimable amistad.

A Liliana y Estefanía, mis grandes amigas, a quienes no podría dejar de mencionar con profundo aprecio. Su encuentro en esta travesía universitaria ha sido un regalo invaluable, y en tan poco tiempo se han convertido en personas muy importantes para mí.

Germán Oswaldo Ruíz Rodríguez

Agradecimientos

Al culminar esta etapa educativa, quiero expresar todo mi amor y agradecimiento a mi familia, quienes han sido mi apoyo incondicional, mi motivación constante a lo largo de este camino.

A mis padres y hermana, quienes siempre estuvieron para mí, en las buenas y en las malas, brindándome una educación correcta con valores fuertes e inquebrantables, lo cual me ha permitido entender la importancia del esfuerzo, la perseverancia y el arduo trabajo. Han sido mi ejemplo a seguir en todos los aspectos de la vida, por esta razón es un honor y orgullo devolverles un poquito de lo que me han dado agradeciéndoles todo lo hecho por mí en esta etapa educativa.

También quiero agradecerle a mi novia Cristina De La Cruz, que con sus palabras de aliento y motivación me ha ayudado a estar tranquilo y llevar con calma mi educación estos últimos dos años. Gracias a su alegría, paciencia, comprensión y apoyo incondicional durante esta travesía académica he podido realizar este trabajo de la mejor manera.

A la Universidad de las Fuerzas Armadas ESPE, por brindarme la educación necesaria para mi formación personal y profesional. A la Máster Eleana Jerez, quien muy gentilmente nos ha ayudado con sus consejos y correcciones en este trabajo para culminar mi formación académica.

También quiero extender mi agradecimiento a mis amigos y compañeros, David Díaz, Jean Pierre Toapanta y Marcelo Arias, por su lealtad y compañerismo durante toda mi vida universitaria.

Finalmente, este trabajo de tesis no es solo un esfuerzo personal, sino que, a lo largo de mi vida y carrera universitaria, he tenido el apoyo de grandes personas y seres humanos que me han brindado su guía y orientación para ser un mejor hombre y ser humano. Estoy eternamente agradecido por todo lo que han hecho.

Jerry Joe Tovar Rojas

Índice de Contenidos

Resumen	14
Abstract.....	15
Capítulo I Introducción	16
Planteamiento del problema	18
Objetivos.....	18
Objetivo General.....	18
Objetivos Específicos	18
Justificación.....	19
Alcance.....	20
Capítulo II Revisión de Literatura.....	21
Preguntas de investigación	21
Cadena de búsqueda y palabras clave.....	21
Criterios de Inclusión y Exclusión.....	21
Criterios de inclusión	21
Criterios de exclusión.....	22
Resultados de la Revisión	22
Capítulo III Marco Teórico.....	37
Herramientas, Técnicas y Redes Sociales.....	37
Framework Flask.....	37

Python.....	37
Procesamiento del Lenguaje Natural (NLP)	38
Representaciones de Codificador Bidireccional de Transformadores (BERT).....	40
Metodología XP o Programación Extrema.....	40
Red Social - Facebook	43
Capítulo IV Diseño y Desarrollo	46
Uso de metodología XP.....	46
Desarrollo del sistema	47
Arquitectura de la Solución Propuesta	48
Capítulo V Pruebas Funcionamiento de Software	58
Lighthouse - Evaluación de rendimiento del software	58
Lighthouse - Evaluación de usabilidad del software	59
Lighthouse - Evaluación de seguridad del software	60
Capítulo VI Conclusiones, recomendaciones y trabajos futuros	61
Conclusiones	61
Recomendaciones.....	62
Trabajos futuros	63
Capítulo VII Referencias	64

Índice de Tablas

Tabla 1 Bases de datos electrónicas y artículos primarios	22
Tabla 2 Tipos de contribución.....	25
Tabla 3 Técnicas de Aprendizaje Automático	29
Tabla 4 Redes Sociales & Datasets.....	33

Índice de Figuras

Figura 1 Fases metodología XP	41
Figura 2 Flujograma de la arquitectura de la solución propuesta	49
Figura 3 Arquitectura de la Solución Propuesta	49
Figura 4 Código para recopilación de datos	50
Figura 5 Código para la limpieza	51
Figura 6 Código para analizar el análisis del texto	51
Figura 7 Interfaz gráfica	53
Figura 8 Visualización de resultados	54
Figura 9 Interpretación de resultados	55
Figura 10 Gráfico porcentual de resultados	55
Figura 11 Porcentaje de precisión	56
Figura 12 Evaluación de desempeño - Lighthouse	58
Figura 13 Evaluación de accesibilidad-Lighthouse	59
Figura 14 Evaluación de defensa-Lighthouse	60

Resumen

Las redes sociales digitales como: ResearchGate, una red social de investigadores cuyo enfoque principal es desarrollar ciencia de cualquier disciplina; X (antes conocido como Twitter), utilizada para publicar información en tiempo real acerca de temáticas relacionadas con problemas políticos, económicos, sociales, etc.; Facebook, aquí las personas propagan información mediante texto, fotos o videos en tiempo real acerca de una gran variedad de temáticas; Instagram, una red social utilizada para compartir fotografías y videos; TikTok, el enfoque principal de esta red social es grabar, editar y compartir videos que se publican en tiempo real con varias temáticas relacionadas con tecnología, política, economía y polémica social ; LinkedIn, considerada como una red social digital para profesionales, que muestra un pequeño currículum de las personas; se han convertido en objeto de análisis de muchos investigadores. Por ejemplo, algunos estudios se enfocan en definir modelos de propagación de información, otros en mitigar la información que se propaga identificando a los nodos semilla y algunos en el análisis de la desinformación. Además, en la literatura académica se ha encontrado una clasificación de la información con los términos en inglés: “misinformation”, “disinformation” y “true information”. Sin embargo, no se han encontrado trabajos en los que se identifique tipos de información basados en la argumentación que se propagan intencionalmente en las redes sociales digitales como los hechos y las opiniones utilizando técnicas de aprendizaje automático. La identificación de estos tipos de información, más específica y menos generalizada que la que se ha encontrado, nos permitirá analizar de mejor manera el comportamiento de la información que se propaga en las redes sociales digitales, considerando que en muchas ocasiones puede causar caos y pánico colectivo en la población.

Palabras Clave: Hechos, Opiniones, Identificación, Objetividad, Subjetividad.

Abstract

Digital social networks such as: ResearchGate, a social network of researchers whose main focus is developing science from any discipline; X (formerly known as Twitter), used to publish information in real time about topics related to political, economic, social problems, etc.; Facebook, here people spread information through text, photos or videos in real time about a wide variety of topics; Instagram, a social network used to share photographs and videos; TikTok, the main focus of this social network is to record, edit and share videos that are published in real time with various areas related to technology, politics, economics and social controversy; LinkedIn, considered a digital social network for professionals, which shows a small resume of people; They have become the object of analysis of many researchers. For example, some studies focus on defining information propagation models, others on mitigating the information that is propagated to seed nodes, and some on the analysis of disinformation. Furthermore, in academic literature, a classification of information has been found with the terms in English: "misinformation", "disinformation" and "true information". However, no work has been found that identifies types of information based on argumentation that are intentionally propagated on digital social networks such as facts and opinions using machine learning techniques. The identification of these types of information, more specific and less generalized than what has been found, will allow us to better analyze the behavior of the information that spreads on digital social networks, considering that on many occasions it can cause chaos and panic. collective in the population.

Key words: Facts, Opinions, Identification, Objectivity, Subjectivity.

Capítulo I

Introducción

La identificación de hechos y opiniones es una tarea crítica en el procesamiento del lenguaje natural y juega un papel esencial en diversas aplicaciones, como la minería de opiniones, percepción sobre el contenido de la información y la comprensión automática de texto. En primer lugar, la distinción entre hechos y opiniones se basa en la naturaleza objetiva o subjetiva de la información expresada (Alqahtani, Alsheddi & Alhenaki, 2023). Los hechos son afirmaciones verificables y objetivas que se pueden comprobar mediante evidencia, mientras que las opiniones reflejan la perspectiva subjetiva o el punto de vista del hablante (Albahr & Albahar, 2020).

La identificación de hechos y opiniones también es crucial en la lucha contra la desinformación. Los sistemas automatizados pueden analizar grandes cantidades de información para identificar declaraciones basadas en hechos y opiniones infundadas. La capacidad de discernir entre estos dos tipos de información es esencial para garantizar la calidad y la veracidad de la información que circula en línea.

Además, la identificación de hechos y opiniones se relaciona estrechamente con la interpretación contextual del lenguaje. En muchos casos según Alqahtani, Alsheddi y Alhenaki (2023), “la comprensión precisa de una declaración requiere analizar el contexto en el que se presenta, ya que una afirmación puede considerarse un hecho en un contexto y una opinión en otro”.

En última instancia, la mejora continua de las técnicas de identificación de hechos y opiniones es esencial para avanzar en la capacidad de los sistemas de procesamiento del lenguaje natural para comprender de manera más precisa y contextualizada la información expresada en texto, contribuyendo así a una interpretación más fiel y relevante del lenguaje humano.

En el contexto de las redes sociales, la identificación de hechos y opiniones cobra una importancia aún mayor debido a la rápida propagación de información en estas plataformas. En primer

lugar, las redes sociales son canales donde los usuarios comparten una amplia variedad de contenido, que va desde hechos objetivos, como noticias y eventos, hasta opiniones sobre diversos temas (Alqahtani, Alsheddi & Alhenaki,2023). La capacidad de distinguir entre hechos y opiniones se vuelve esencial para los usuarios al evaluar la credibilidad y la veracidad de la información que encuentran en sus feeds (flujo continuo de contenido en redes sociales).

La amplia diversidad de opiniones en las redes sociales resalta la importancia de comprender y respetar la naturaleza subjetiva de muchas expresiones. Aunque algunas afirmaciones se presenten como hechos, es común que los usuarios compartan sus opiniones y perspectivas personales. La identificación precisa de la intencionalidad detrás de un mensaje contribuye a fomentar un diálogo en línea enriquecedor.

Adicionalmente, la identificación de hechos y opiniones desempeña un papel crítico en la gestión de la información en las redes sociales. Con la proliferación masiva de contenido en estas plataformas, que va desde informes objetivos hasta puntos de vista subjetivos, distinguir entre hechos y opiniones se convierte en un factor crucial para los usuarios para percibir más precisamente el contenido de la información publicada por un perfil, página o grupo.

En definitiva, en el entorno dinámico y complejo de las redes sociales, la identificación de hechos y opiniones es esencial para promover un intercambio de información más confiable y fomentar un diálogo respetuoso y constructivo entre los usuarios.

Planteamiento del problema

En un entorno cada vez más interconectado a través de plataformas de redes sociales, se presenta un desafío crítico: la posibilidad de generar caos y pánico mediante la difusión incontrolada de datos "no fiables". La carencia de herramientas efectivas para analizar con precisión el comportamiento de esta información en las redes sociales contribuye a la propagación de desinformación y rumores infundados. En este escenario, surge la imperiosa necesidad de contar con una herramienta que pueda identificar de manera confiable hechos y opiniones en las redes sociales. Este trabajo proponemos una solución que identifica hechos y opiniones en redes sociales con el fin de proporcionar una comprensión más profunda del impacto de la información en línea, con el objetivo de mitigar la desinformación y el caos colectivo en la población.

Objetivos

Objetivo General

Identificar hechos y opiniones en redes sociales digitales utilizando técnicas de aprendizaje automático.

Objetivos Específicos

- Realizar la revisión de literatura de las técnicas de aprendizaje automático utilizadas para identificar hechos y opiniones en las redes sociales digitales.
- Identificar una red social para la recolección de hechos y opiniones
- Delimitar una técnica de aprendizaje automático para identificar hechos y opiniones.
- Ejecución de prueba para la identificación de hechos y opiniones.

Justificación

El análisis de hechos y opiniones en redes sociales, particularmente en plataformas como Facebook, es un campo de estudio de creciente importancia en la era digital. La proliferación de información en estos medios ha generado un entorno en el que la distinción entre hechos y opiniones se ha vuelto cada vez más difusa, lo que conlleva a la propagación de desinformación, noticias falsas y sesgos cognitivos (Rizwan, Naeem, Ahmad, Saleem, El-Affendi, Zaidi & Liaqat, 2022). Esta situación no solo afecta la percepción pública sobre diversos temas, sino que también tiene implicaciones significativas en la toma de decisiones, tanto a nivel individual como colectivo.

En este contexto, el desarrollo de métodos automatizados para analizar y clasificar el contenido de las publicaciones en redes sociales en hechos y opiniones se presenta como una necesidad imperante. La implementación de técnicas de procesamiento de lenguaje natural (NLP) y, el modelo BERT (Bidirectional Encoder Representations from Transformers), ofrece una oportunidad única para abordar este desafío (Imran, Daudpota, Kastrati, & Batra, 2020). Estas tecnologías permiten analizar grandes volúmenes de texto, identificar patrones y extraer características relevantes que facilitan la clasificación precisa del contenido.

La relevancia de este trabajo radica en su contribución al desarrollo de herramientas que permitan una mejor comprensión y gestión de la información en redes sociales. Al proporcionar métodos automatizados para diferenciar entre hechos y opiniones, se fomenta una mayor transparencia y se contribuye a la lucha contra la desinformación. Además, los resultados de esta investigación pueden ser aplicados en diversos campos, como el periodismo, la educación, la política y la salud pública, donde la distinción entre hechos y opiniones es fundamental para una comunicación efectiva y una toma de decisiones informada (Sahu, Ahmad, Jhanjhi, Choudhury & Thakur, 2022).

En resumen, la justificación de este trabajo de tesis reside en la necesidad de abordar los desafíos que plantea la era de la información en las redes sociales, mediante el desarrollo de tecnologías que permitan un análisis eficaz y preciso del contenido. Esto no solo tiene un impacto directo en la calidad de la información disponible para los usuarios, sino que también contribuye al fortalecimiento de la democracia, la educación y la salud pública, al promover una ciudadanía más informada y crítica.

Alcance

El alcance de este trabajo engloba la implementación de herramientas y técnicas de aprendizaje automático para analizar y comprender la información circulante en las redes sociales. El enfoque busca identificar patrones y características que permitan comprender la naturaleza de la información difundida, más que evaluar la veracidad o falsedad de la misma, contribuyendo así a la formación de una perspectiva informada por parte de los usuarios en base a lo que ha sido publicado.

El proyecto comprende la implementación de un sistema prototipo capaz de procesar y clasificar de forma automática el contenido de las redes sociales, ofreciendo información reveladora sobre la naturaleza de las publicaciones. Asimismo, se llevarán a cabo evaluaciones de rendimiento y pruebas para garantizar la eficacia y confiabilidad del sistema desarrollado.

Finalmente, además de analizar la información, se generará gráficos estadísticos a partir de los datos recopilados. Esto con el fin de proporcionar a los usuarios que usen el prototipo una idea general y clara del contenido tratado.

Capítulo II

Revisión de Literatura

Preguntas de investigación

PI1: ¿Cuál es el estado del arte acerca de la identificación de hechos y opiniones?

PI2: ¿Cuáles son los tipos de contribución?

PI3: ¿Cuáles son las técnicas de aprendizaje automático que utilizan los autores en los estudios o trabajos?

PI4: ¿Cuáles son los recursos (datasets) que utilizan en los trabajos para la experimentación?

Cadena de búsqueda y palabras clave

Las palabras claves seleccionadas fueron utilizadas para conformar la cadena de búsqueda ideal, la misma que fue implementada en bases de datos científicas, para encontrar artículos relacionados. La cadena de búsqueda obtenida fue la siguiente:

(Identification OR Detection) AND (Facts OR Opinions OR Objectivity OR Subjectivity) AND (Machine AND Learning) AND (Social AND (Networks OR Media))

Aplicamos la cadena de búsqueda en las bases de datos electrónicas: IEEEExplore y Scopus.

Criterios de Inclusión y Exclusión

A partir de las preguntas de investigación definidas en las secciones anteriores de este trabajo, hemos definido los siguientes criterios de inclusión y exclusión:

Criterios de inclusión

I1. Trabajos publicados desde el año 2020 hasta el 2023.

I2. Trabajos enfocados en el campo de la Informática (Computer Science).

I3. Trabajos escritos en idioma inglés.

Criterios de exclusión

E1. Trabajos duplicados

E2. Trabajos que no reporten en el título o en el resumen acerca de identificación de hechos y opiniones.

E3. Trabajos que no reporten en el título o en el resumen acerca de las técnicas de aprendizaje automático para identificar hechos y opiniones.

Resultados de la Revisión

Luego de aplicar la cadena de búsqueda en las bases de datos electrónicas obtuvimos un total de 178 estudios, eliminamos un total de 9 duplicados, aplicamos los criterios de inclusión y exclusión) y finalmente obtuvimos un total de 69 estudios primarios (ver Tabla 1).

Tabla 1

Bases de datos electrónicas y artículos primarios

Bases digitales	Cantidad de artículos
Scopus	137
IEEE	41
Total	178
Número de estudios duplicados	9
Cumplieron (fit)	69

Nota: Resultados de búsqueda y selección de estudios primarios.

Luego de obtener el conjunto de estudios primarios final, en esta sección responderemos a las preguntas de investigación planteadas:

1. ¿Cuál es el estado del arte acerca de la identificación de hechos y opiniones?

Durante el periodo comprendido entre 2020 y 2023, la investigación en la identificación de hechos y opiniones ha experimentado notables avances, impulsados principalmente por el surgimiento de modelos de lenguaje avanzados, tales como LSTM (Long Short Term Memory), (Imran, Daudpota, Kastrati & Batra, 2020), (Singh, Kumar, Rana & Dwivedi., 2022), (Bilal, Khan, Jna, Musa & Ali, 2023), (Akila & Revathi, 2023), (Fatima, Samad Shaikh, Riaz, Ahmad, El-Affendi, Alyamani, Nabeel, Ali Khan, Yasin & Latif, 2022), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022) y SVM (Support Vector Machine) (Hassan, Alhalangy & Alzahrani, 2023), (Pathan & Prakash, 2022), (Rahman, Tariq, Masood, Subahi, Khalaf & Alotaibi, 2023), (Truong & Tran, 2023), (Angskun, Tipprasert & Angskun, 2022), (Machová, Mach & Adamisín, 2022), (Khalafat, Alqatawna, Al-Sayyed, Eshtay & Kobbaey,, 2021), (Oliveira, Medeiros, & Mattos , 2020), (Rupapara, Rustam, Shahzad, Mehmood, Ashraf & Choi, 2021), (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Studiawan, Sohel & Payne, 2021), (Alqurashi, 2022). Estos modelos han demostrado una capacidad mejorada para comprender el contexto y la semántica, permitiendo una distinción más precisa entre declaraciones de hechos y expresiones de opinión en textos complejos y variados.

Adicionalmente, se ha observado un enfoque creciente en la aplicación de técnicas basadas en evidencia. Estos métodos buscan validar afirmaciones mediante el análisis de patrones y coherencia en el contenido, proporcionando un enfoque más riguroso para la identificación de hechos verificables (Albahr & Albahr, 2020),(Kumar & Taylor, 2023), (Lee, Rustam, Washington, Baraka, Aljedaani & Ashraf, 2022), (Truong & Tran, 2023), (Varlamis, Michail, Glykou & Tsantilas, 2022), (Chen, Zhou, Wang & Li, 2023), (Rahman, Hasan, Billah & Sajuti, 2022), (de Oliveira, Medeiros & Mattos, 2020), (Kaliyar, Goswami & Narang, 2021), (Demilie & Salau, 2022). Paralelamente, la verificación de hechos automatizada ha ganado protagonismo, utilizando algoritmos de aprendizaje automático para evaluar la veracidad de las declaraciones a través del análisis de fuentes y datos adicionales, contribuyendo así a combatir la

propagación de desinformación (Wadhvani, Varshney, Gupta & Kumar, 2023), (Machová, Mach & Adamisín, 2022), (Akila & Revathi, 2023), (Haque, Islam, Islam & Ahsan, 2022), (Rosa, De Silva, Silva, Ayub, Carrillo, Nardelli & Rodríguez, 2020), (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Nijhawan, Attigeri & Ananthakrishna, 2022).

En el campo de la ética, se observa un creciente énfasis transparente en el desarrollo e implementación de tecnologías relacionadas con la inteligencia artificial. Este enfoque busca abordar consideraciones vinculadas a la privacidad y al uso moral de herramientas de investigación y opinión en las redes sociales digitales. El objetivo principal es garantizar una implementación responsable y respetuosa, considerando el potencial impacto social y velando por la protección de la privacidad de los usuarios involucrados en el análisis de contenidos.

En este mismo contexto, la intersección de la inteligencia artificial y la ética adquiere una importancia crucial. La transparencia en los procesos de toma de decisiones de estos modelos y la interpretación de sus resultados son áreas clave que la investigación actual aborda para garantizar la confianza y la aceptación pública de estas tecnologías.

La identificación precisa de hechos y opiniones sigue siendo un desafío, especialmente en el caso de información compleja. El constante desarrollo de las plataformas de redes sociales y la diversidad de métodos de presentación de información plantean desafíos continuos para investigadores y desarrolladores. Por lo tanto, la tecnología de aprendizaje automático debe innovar y adaptarse constantemente para mantenerse al día con la naturaleza dinámica de la información en las redes sociales digitales.

2. ¿Cuáles son los tipos de contribución?

Para proporcionar una visión completa de las contribuciones en este ámbito, se ha llevado a cabo una revisión detallada de los trabajos seleccionados, presentando los hallazgos en una tabla descriptiva (ver Tabla 2). La tabla organiza las contribuciones según diferentes tipos basados en la clasificación de Wieringa's (Wieringa, Maiden, Mead & Rolland, 2006), destacando la diversidad de enfoques adoptados por los investigadores en este campo emergente.

Tabla 2

Tipos de contribución

Tipo de contribución	Estudios
Modelos	(Alqahtani, Alsheddi & Alhenaki, 2023), (Dewani, Memon, Bhatti, Sulaiman, Hamdi, Alshahrani, Alghamdi & Shaikh, 2023)(Hassan, Alhalangy & Alzahrani, 2023), (Kuma & Taylor, 2023), (Mehmood, Farooq, Naseem, Rustam, Villar, Rodríguez & Ashraf, 2022), (Mashraqi & Halawani, 2023), (Alattar & Shaalan, 2021), (Roy & Das, 2023), (Bikakis, Flouris, Patkos & Plexousakis, 2023), (Truică & Apostol, 2023), (Gamal, Alfonse, Jiménez-Zafra & Aref, 2023), (Lee, Rustam, Washington, Barakaz, Aljedaani & Ashraf, 2022), (Fkih, Moulahi & Alabdulatif, 2023), (Wadhvani, Varshney, Gupta & Kumar, 2023), (Rahman, Tariq, Masood, Subahi, Khalaf & Alotaibi, 2023), (Truong & Tran, 2023), (Imran, Daudpota, Kastrati & Batra, 2020), (Angskun, Tipprasert & Angskun, 2022), (Chen, Zhou, Lu & Wang, 2023), (Machová, Mach & Adamišín, 2022), (Languré & Zareei, 2023), (Bilal, Khan, Jan, Musa & Ali, 2023), (M., G., & A., R., 2023), (Khalafat, Alqatawna, Al-Sayyed, Eshtay & Kobbay, 2021)(Zayno & Radhi, 2022), (Vyas, Vyas & Dhiman, 2023), (Akila & Revathi, 2023), (Fatima, Samad

Tipo de contribución	Estudios
Propuestas	<p>Shaikh, Riaz, Ahmad, El-Affendi, Alyamani, Nabeel, Ali Khan, Yasin & Latif, 2022), (Fatima, Samad Shaikh, Riaz, Ahmad, El-Affendi, Alyamani, Nabeel, Ali Khan, Yasin & Latif, 2022), (Haque, Islam, Islam & Ahsan, 2022), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022), (Waheeb, Khan & Shang, 2022), (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Alqurashi, 2022), (Karyukin, Mutanov, Mamykova, Nassimova, Torekul, Sundetova & Negri, 2022)</p> <p>(Baydogan & Alatas, 2021), (Singh, Kuar, Rana & Dwivedi, 2022), (Dashtipour, Gogate, Li, Jiang, Kong & Hussain, 2020), (Keyvanpour, Karimi Zandian & Heidarypanah, 2020), (Rosa, Silva, Silva, Ayub, Carrillo, Nardelli & Rodríguez, 2020), (Studiawan, Sohel & Payne, 2021)</p>
Procesos	<p>(Lighthart, Catal & Tekinerdogan, 2021), (Bikakis, Flouris, Patkos & Plexousakis, 2023), (Elzayady, Mohamed, Badran & Salama, 2022), (Huang, Zavareh & Mustafa, 2023), (Araújo, Pereira & Benevenuto, 2020), (Gozuacik, Sakar & Ozcan, 2021), (Oliveira, Medeiros & Mattos, 2020), (Abuein, Shatnawi & Ghazalat, 2022), (Kaliyar, Goswami & Narang, 2021), (Rupapara, Rustam, Shahzad, Mehmood, Ashraf & Choi, 2021), (Gupta, Goel, Singh, Prasad & Ullah, 2022)</p>
Estudios	<p>(Pathan & Prakash, 2022), (Varlamis, Michail, Glykou & Tsantilas, 2022), (Bilal, Khan, Jan, Musa & Ali, 2023)</p>
Experimentos	<p>(Chrismanto, Sari & Suyanto, 2023), (Dadkhah, Zhang, Weismann, Firouzi & Ghorbani, 2023)</p>
Métodos	<p>(Rahman, Hasan, Billah & Sajuti, 2022), (Sadiq, Aljrees & Ullah, 2023), (Nijhawan, Attigeri & Ananthakrishna, 2022)</p>

Nota: Contribución de estudios primarios.

3. ¿Cuáles son las técnicas de aprendizaje automático que utilizan los autores en los estudios o trabajos?

Para recopilar información, se han utilizado Modelos de Aprendizaje Profundo (Deep Learning). Se observa el uso de modelos como LSTM (Long Short-Term Memory) (Singh, Kuar, Rana & Dwivedi, 2022), (Bilal, Khan, Jan, Musa & Ali, 2023), (Akila & Revathi, 2023), (Fatima, Samad Shaikh, Riaz, Ahmad, El-Affendi, Alyamani, Nabeel, Ali Khan, Yasin & Latif, 2022), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022), para el procesamiento de secuencias, como texto o series temporales, utilizando unidades de memoria especializadas para capturar y recordar patrones a largo plazo, permitiendo la modelización efectiva de dependencias temporales complejas; CNN (Convolutional Neural Network) (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022), (Sadiq, Aljrees & Ullah, 2023), (Alqurashi, 2022), la cual es una técnica diseñada para procesar y analizar datos estructurados, especialmente en imágenes. Utiliza capas de convolución para detectar patrones locales y jerárquicos en los datos, reduciendo la cantidad de parámetros requeridos y permitiendo la identificación de características complejas en imágenes. En cuanto a los Algoritmos de Procesamiento del Lenguaje Natural (NLP), una rama destacada de la inteligencia artificial que se enfoca en la interacción entre las computadoras y el lenguaje humano se busca capacitar a las máquinas para comprender, interpretar y generar texto de manera análoga a los seres humanos. Este campo de investigación ha sido explorado en diversas ocasiones, como se refleja en los trabajos (Anzum, F., & Gavrilova, M. L., 2023), (Baydogan & Alatas, 2021), (Bikakis, Flouris, Patkos & Plexousakis, 2023), (Gamal, Alfonse, Jiménez-Zafra & Aref, 2023), (Zayno & Radhi, 2022), (Gozuacik, Sakar & Ozcan, 2021), (Oliveira, Medeiros & Mattos, 2020), (Rosa, Silva, Silva, Ayub, Carrillo, Nardelli & Rodríguez, 2020), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022) Varios artículos han aplicado técnicas de NLP, como el uso de la técnica Bag of Words (BoW) (Baydogan & Alatas, 2021), (Wadhwani, Varshney, Gupta & Kumar,

2023), que representa documentos como conjuntos no ordenados de palabras, ignorando la gramática y el orden. Cada palabra se considera como una "bolsa" independiente de términos, lo que simplifica la representación y permite analizar la frecuencia de las palabras en un documento para su procesamiento; Term Frequency (TF) (Baydogan & Alatas, 2021), que indica la frecuencia de una palabra específica en un documento. Se calcula dividiendo el número de veces que aparece una palabra en un documento entre el total de palabras en ese documento; Document Vector (Word2Vec) (Baydogan & Alatas, 2021), donde se realiza una representación numérica de un documento. Dicha representación se obtiene promediando los vectores de palabras de dicho documento, capturando así relaciones semánticas para tareas de procesamiento del lenguaje natural y otros enfoques para analizar y comprender el contenido de texto en redes sociales.

Por otra parte, también se han llevado a cabo Enfoques de Clasificación y Detección, en los cuales los autores han implementado métodos como Support Vector Machines (SVM) (Hassan, Alhalangy & Alzahrani, 2023), (Pathan & Prakash, 2022), (Chrismanto, Sari & Suyanto, 2023), (Roy & Das, 2023), (Fkih, Moulahi & Alabdulatif, 2023), (Rahman, Tariq, Masood, Subahi, Khalaf & Alotaibi, 2023), (Truong & Tran, 2023), (Angskun, Tipprasert & Angskun, 2022), (Machová, Mach & Adamišín, 2022), (Akila & Revathi, 2023)(Akila & Revathi, 2023), (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Alqurashi, 2022), donde se usa algoritmos de aprendizaje que encuentran un hiperplano óptimo para separar clases en datos multidimensionales, maximizando la distancia entre ellas, siendo eficaces para la clasificación y regresión, empleando vectores de soporte clave en el proceso; Random Forest (Fkih, Moulahi & Alabdulatif, 2023), (Angskun, Tipprasert & Angskun, 2022), (Zayno & Radhi, 2022), (Akila & Revathi, 2023) (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Alqurashi, 2022), que son múltiples árboles de decisión durante el entrenamiento que combinan sus predicciones para mejorar la precisión y robustez del modelo, reduciendo el riesgo de sobreajuste. Se basa en la idea de "ensamble", donde la combinación de múltiples modelos débiles produce un modelo fuerte y versátil; Decision Trees (Rahman,

Tariq, Masood, Subahi, Khalaf & Alotaibi, 2023), (Angskun, Tipprasert & Angskun, 2022), (Rahman, Hasan, Billah & Sajuti, 2022), (Zayno & Radhi, 2022), donde se toman decisiones a través de una estructura jerárquica de nodos, cada nodo representa una pregunta o condición sobre los datos, y cada rama lleva a una decisión o resultado.

Así mismo, se ha implementado técnicas de clasificación de votación, como Regression Vector Voting Classifier (RVVC) (Rupapara, Rustam, Shahzad, Mehmood, Ashraf & Choi, 2021), que combina múltiples modelos de regresión mediante votación para predecir un valor objetivo, asignando pesos a las predicciones de cada modelo según su desempeño en conjuntos de datos específicos, mejorando así la precisión predictiva del conjunto. También se han aplicado Modelos de Detección de Anomalías basados en aprendizaje profundo y análisis de sentimientos para identificar patrones inusuales en los datos, como en la detección de anomalías en registros del sistema operativo. Algunos estudios han empleado modelos de embedding, como BERT (Bidirectional Encoder Representations from Transformers) (Bilal, Khan, Jan, Musa & Ali, 2023), (Bilal, Khan, Jan, Musa & Ali, 2023), (Alqurashi, 2022), el cual es un modelo de lenguaje preentrenado en aprendizaje profundo, específicamente basado en la arquitectura de transformers. Procesa texto en ambas direcciones (bidireccionalmente), capturando el contexto de las palabras en una oración. Este enfoque mejora la comprensión del significado de las palabras en función del contexto, lo que ha llevado a mejoras significativas en tareas de procesamiento del lenguaje natural como la traducción automática, la respuesta a preguntas y la clasificación de texto; para representar contextualmente el contenido de texto y mejorar la comprensión semántica. (ver Tabla 3).

Tabla 3

Técnicas de Aprendizaje Automático

Técnicas de aprendizaje automático	Estudios
Natural Language Processing (NLP)	(Anzum, F., & Gavrilova, M. L., 2023), (Baydogan & Alatas, 2021),(Bikakis, Flouris,

Técnicas de aprendizaje automático	Estudios
	Patkos & Plexousakis, 2023)(Truică & Apostol, 2023), (Lee, Rustam, Washington, Barakaz, Aljedaani & Ashraf, 2022),(Fkih, Moulahi & Alabdulatif, 2023),(Wadhwani, Varshney, Gupta & Kumar, 2023),(Elzayady, Mohamed, Badran & Salama, 2022),(Huang, Zavareh & Mustafa, 2023),(Imran, Daudpota, Kastrati & Batra, 2020), (Angskun, Tipprasert & Angskun, 2022)
Bag of Words (BoW)	(Baydogan & Alatas, 2021)
Term Frequency (TF)	(Baydogan & Alatas, 2021)
Document Vector (Word2Vec)	(Baydogan & Alatas, 2021)
Decision Tree (DT)	(Baydogan & Alatas, 2021), (Truică & Apostol, 2023), (Gamal, Alfonse, Jiménez-Zafra & Aref, 2023)
Regression Vector Voting Classifier (RVVC)	(Rupapara, Rustam, Shahzad, Mehmood, Ashraf & Choi, 2021)
Support Vector Machines (SVM)	(Hassan, Alhalangy & Alzahrani, 2023), (Pathan & Prakash,2022), (Chrismanto, Sari & Suyanto, 2023), (Roy & Das, 2023), (Alattar & Shaalan, 2021), (Anzum, F., & Gavrilova, M. L., 2023), (Baydogan & Alatas, 2021), (Truică & Apostol, 2023), (Gamal, Alfonse, Jiménez-Zafra & Aref, 2023), (Lee, Rustam, Washington, Barakaz, Aljedaani & Ashraf, 2022), (Fkih, Moulahi & Alabdulatif, 2023), (Elzayady, Mohamed, Badran & Salama, 2022), (Rahman, Tariq, Masood, Subahi, Khalaf & Alotaibi, 2023)
BERT (Bidirectional Encoder Representations from Transformers)	(Bilal, Khan, Jan, Musa & Ali, 2023), (Alqurashi, 2022)
CNN (Convolutional Neural Network)	(Bilal, Khan, Jan, Musa & Ali, 2023), (Alqurashi, 2022),

Técnicas de aprendizaje automático	Estudios
LSTM (Long Short-Term Memory)	(Singh, Kuar, Rana & Dwivedi, 2022), (Bilal, Khan, Jan, Musa & Ali, 2023), (Akila & Revathi, 2023), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022)
Logistic Regression (LR)	(Zayno & Radhi, 2022)
K-Nearest Neighbor (KNN)	(Rahman, Hasan, Billah & Sajuti, 2022), (Zayno & Radhi, 2022)
Random Forest (RF)	(Fkih, Moulahi & Alabdulatif, 2023), (Angskun, Tipprasert & Angskun, 2022), (Zayno & Radhi, 2022), (Akila & Revathi, 2023), (Akila & Revathi, 2023), (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Alqurashi, 2022).
Naive Bayes (NB)	(A., A., & M., A., 2020), (Lighthart, Catal & Tekinerdogan, 2021), (Pathan & Prakash, 2022), (Fkih, Moulahi & Alabdulatif, 2023), (Rahman, Tariq, Masood, Subahi, Khalaf & Alotaibi, 2023), (Truong & Tran, 2023), (Angskun, Tipprasert & Angskun, 2022), (Machová, Mach & Adamišín, 2022), (Rahman, Hasan, Billah & Sajuti, 2022), (Zayno & Radhi, 2022), (Akila & Revathi, 2023), (Akila & Revathi, 2023), (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Alqurashi, 2022).

Nota: Técnicas de aprendizaje en estudios primarios.

La identificación de hechos y opiniones en redes sociales es una tarea desafiante debido a la naturaleza informal y variada del lenguaje utilizado en estas plataformas. Los estudios han aplicado diversas técnicas de aprendizaje automático para abordar este desafío.

Uno de los enfoques comunes es el análisis de sentimiento, que implica la aplicación de algoritmos para determinar la actitud expresada en un mensaje, ya sea positiva, negativa o neutral. Estos

algoritmos pueden entrenarse utilizando conjuntos de datos etiquetados para reconocer patrones específicos en el lenguaje coloquial de las redes sociales.

Naive Bayes (NB) (Ligthart, Catal & Tekinerdogan, 2021), (Pathan & Prakash, 2022), (Fkih, Moulahi & Alabdulatif, 2023), (Rahman, Tariq, Masood, Subahi, Khalaf & Alotaibi, 2023), (Truong & Tran, 2023), (Angskun, Tipprasert & Angskun, 2022), (Machová, Mach & Adamišín, 2022), (Rahman, Hasan, Billah & Sajuti, 2022), (Zayno & Radhi, 2022), (Akila & Revathi, 2023)(Akila & Revathi, 2023), (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Alqurashi, 2022) también se utiliza ampliamente, donde los modelos de aprendizaje automático son entrenados para distinguir entre hechos y opiniones en base a características específicas del texto. Estos modelos pueden beneficiarse del uso de técnicas avanzadas como las redes neuronales profundas, especialmente cuando se trata de analizar grandes volúmenes de datos sociales no estructurados.

El Natural Language Processing (NLP) (Anzum & Gavrilova, 2023), (Baydogan & Alatas, 2021), (Bikakis, Flouris, Patkos & Plexousakis, 2023), (Gamal, Alfonse, Jiménez-Zafra & Aref, 2023), (Zayno & Radhi, 2022), (Gozuacik, Sakar & Ozcan, 2021), (Oliveira, Medeiros & Mattos, 2020), (Rosa, Silva, Silva, Ayub, Carrillo, Nardelli & Rodríguez, 2020), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022)(Renjith, Abraham, Jyothi, Chandran & Thomson, 2022), desempeña un papel crucial, ya que permite a los investigadores extraer características lingüísticas relevantes para identificar la subjetividad en los mensajes. Además, los modelos preentrenados en tareas de NLP, como BERT (Bidirectional Encoder Representations from Transformers) (Bilal, Khan, Jan, Musa & Ali, 2023), (Bilal, Khan, Jan, Musa & Ali, 2023), (Alqurashi, 2022); han demostrado ser efectivos para capturar el contexto y la semántica en el lenguaje social.

El uso de metadatos y contextos sociales también se ha explorado para mejorar la identificación de opiniones, ya que la información sobre el autor, la red de conexiones y el contexto temporal puede proporcionar pistas valiosas sobre la intención detrás de un mensaje.

La identificación de hechos y opiniones en redes sociales implica una combinación de técnicas de análisis de sentimiento, clasificación de texto, procesamiento de lenguaje natural y consideraciones contextualizadas. Estos enfoques se adaptan a la naturaleza dinámica y compleja de las interacciones en línea, permitiendo a los investigadores comprender mejor la diversidad de expresiones en las plataformas sociales.

4. ¿Cuáles son los recursos (datasets) que utilizan en los trabajos para la experimentación?

En la investigación y propuestas revisadas, se observa que las técnicas de aprendizaje automático para la identificación de hechos y opiniones en redes sociales digitales han sido aplicadas principalmente en plataformas como Facebook (ver Tabla 4). Estos estudios abordan diversas temáticas, incluyendo la detección de discurso de odio, análisis de sentimientos, detección de eventos, y clasificación de noticias en estos entornos. Además, se mencionan plataformas como WhatsApp, Telegram, Instagram, Amazon, entre otras. Estos resultados sugieren que la aplicación de técnicas de aprendizaje automático en la identificación de hechos y opiniones se ha extendido a diversas redes sociales digitales, destacando la importancia de estas herramientas en la comprensión y análisis de contenido en línea.

Tabla 4

Redes Sociales & Datasets

Redes Sociales / Datasets	Estudios
Twitter	(Alqahtani, Alsheddi & Alhenaki, 2023), (Mehmood, Farooq, Naseem, Rustam, Villar, Rodríguez & Ashraf, 2022), (Mashraqi & Halawani, 2023), (Alattar &

Redes Sociales / Datasets	Estudios
	Shaalán, 2021), (Alattar & Shaalan, 2021), (Anzum & Gavrilova, 2023), (Lee, Rustam, Washington, Barakaz, Aljedaani & Ashraf, 2022), (Wadhvani, Varshney, Gupta & Kumar, 2023), (Truong & Tran, 2023), (Huang, Zavareh & Mustafa, 2023), (Imran, Daudpota, Kastrati & Batra, 2020), (Angskun, Tipprasert & Angskun, 2022), (Singh, Kuar, Rana & Dwivedi, 2022), (Bilal, Khan, Jan, Musa & Ali, 2023), (Bilal, Khan, Jan, Musa & Ali, 2023), (Fatima, Samad Shaikh, Riaz, Ahmad, El-Affendi, Alyamani, Nabeel, Ali Khan, Yasin & Latif, 2022), (Haque, Islam, Islam & Ahsan, 2022), (Dadkhah, Zhang, Weismann, Firouzi & Ghorbani, 2023), (Sadiq, Aljrees & Ullah, 2023), (Waheeb, Khan & Shang, 2022), (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Alqurashi, 2022), (Nijhawan, Attigeri & Ananthakrishna, 2022), (Zhang, Li, Ying, Li, Tang, & Lu, 2020)
Facebook	(Bilal, Khan, Jan, Musa & Ali, 2023), (Bilal, Khan, Jan, Musa & Ali, 2023), (Abuein, Shatnawi & Ghazalat, 2022)
Amazon	(Huang, Zavareh & Mustafa, 2023)
Telegram	(Waheeb, Khan & Shang, 2022)
Instagram	(Waheeb, Khan & Shang, 2022)
English Content Dataset	(Araújo, Pereira & Benevenuto, 2020)
YouTube	(Sadiq, Aljrees & Ullah, 2023)
Whatsapp	(Roy & Das, 2023)

Nota: Redes sociales en estudios primarios

La identificación de hechos y opiniones en redes sociales se ha convertido en un campo clave para aplicar técnicas de aprendizaje automático, dado el vasto volumen de datos generados por usuarios en plataformas populares. (X) antes conocido como Twitter (Alqahtani, Alsheddi & Alhenaki, 2023), (Mehmood, Farooq, Naseem, Rustam, Villar, Rodríguez & Ashraf, 2022), (Mashraqi & Halawani, 2023), (Alattar & Shaalan, 2021), (Alattar & Shaalan, 2021), (Anzum & Gavrilova, 2023), (Lee, Rustam, Washington, Barakaz, Aljedaani & Ashraf, 2022), (Wadhvani, Varshney, Gupta & Kumar, 2023), (Truong &

Tran, 2023), (Huang, Zavareh & Mustafa, 2023), (Imran, Daudpota, Kastrati & Batra, 2020)(Imran, Daudpota, Kastrati & Batra, 2020), (Angskun, Tipprasert & Angskun, 2022), (Singh, Kuar, Rana & Dwivedi, 2022), (Singh, Kuar, Rana & Dwivedi, 2022), (Bilal, Khan, Jan, Musa & Ali, 2023), (Fatima, Samad Shaikh, Riaz, Ahmad, El-Affendi, Alyamani, Nabeel, Ali Khan, Yasin & Latif, 2022), (Haque, Islam, Islam & Ahsan, 2022), (Dadkhah, Zhang, Weismann, Firouzi & Ghorbani, 2023), (Sadiq, Aljrees & Ullah, 2023), (Waheeb, Khan & Shang, 2022), (Alqurashi, 2022), (Nijhawan, Attigeri & Ananthakrishna, 2022), (Zhang, Li,, Ying, Li, Tang & Lu, 2020), conocido por sus breves tweets, ha sido objeto de análisis de sentimiento y clasificación de texto para comprender las opiniones sobre eventos actuales y temas específicos. La brevedad de los mensajes desafía a los modelos a capturar la sutileza de las expresiones en un espacio limitado.

Facebook (Bilal, Khan, Jan, Musa & Ali, 2023), (Bilal, Khan, Jan, Musa & Ali, 2023), (Abuein, Shatnawi & Ghazalat, 2022), una plataforma multifacética que abarca diferentes tipos de contenido, como texto, imágenes y videos, se presta para técnicas de clasificación de texto y análisis de sentimiento. Aquí, la identificación de opiniones se extiende a diversas formas de contenido multimedia, lo que implica desafíos adicionales en el análisis visual de imágenes.

Instagram (Waheeb, Khan & Shang, 2022), centrado en compartir imágenes y videos, a menudo involucra la identificación de opiniones mediante el análisis de comentarios y descripciones de imágenes. El análisis visual juega un papel crucial para interpretar las imágenes compartidas y comprender el contexto emocional detrás de ellas.

YouTube (Sadiq, Aljrees & Ullah, 2023), con su enfoque en videos, presenta oportunidades y desafíos para la identificación de opiniones. El análisis de sentimiento se aplica a los comentarios, mientras que el análisis de texto puede extenderse a las descripciones de videos. La combinación de datos visuales y textuales amplía la complejidad de la tarea.

La identificación de hechos y opiniones en redes sociales implica la aplicación de diversas técnicas de aprendizaje automático adaptadas a las características específicas de cada plataforma. La combinación de análisis de sentimiento, clasificación de texto y, en algunos casos, análisis visual, permite a los investigadores y empresas comprender las complejidades de las expresiones de los usuarios en entornos sociales digitales.

Capítulo III

Marco Teórico

Herramientas, Técnicas y Redes Sociales

Framework Flask

Flask es un micro framework para el desarrollo web Python. Un marco, en los términos más simples, es una biblioteca o colección de bibliotecas que tiene como objetivo resolver una parte de un problema genérico en lugar de uno específico completo. Al crear aplicaciones web, hay algunos problemas que siempre deberán resolverse, como el enrutamiento desde las URL a los recursos, la inserción de datos dinámicos en HTML y la interacción con un usuario final (Dwyer,2016).

Flask es un framework web simple y flexible que facilita el desarrollo de aplicaciones web en Python sin imponer restricciones innecesarias. Es ideal para proyectos pequeños y medianos donde la simplicidad y la rapidez de desarrollo son prioritarias.

Python

Es un lenguaje de programación versátil, accesible y poderoso que ha dejado una marca significativa en diversas áreas del desarrollo de software. Su sintaxis clara, su filosofía de diseño y su sólida comunidad hacen que Python sea una elección popular tanto para principiantes como para profesionales experimentados (¿Qué es Python? - Explicación del lenguaje Python - AWS, n.d.).

Como lenguaje de programación, se compone de varias librerías, entre ellas:

- **Torch (PyTorch):** Es una librería de Python utilizada para el aprendizaje profundo. Te permite crear y entrenar redes neuronales, que son modelos matemáticos que aprenden de los datos.

- **Transformers:** Esta librería proporciona modelos de procesamiento de lenguaje natural (NLP) preentrenados, como BERT. Te ayuda a analizar y comprender el texto de manera avanzada.
- **Pandas:** Es una librería para manipulación y análisis de datos. Te permite trabajar con datos en forma de tablas, similar a Excel, pero con muchas más funcionalidades.
- **Scikit-learn:** Es una librería para aprendizaje automático. Te ofrece herramientas para crear modelos predictivos, como clasificadores o regresiones.
- **Selenium:** Es una herramienta para automatizar navegadores web. Te permite simular la interacción de un usuario con una página web, como hacer clic en botones o extraer información.
- **Matplotlib:** Es una librería para crear visualizaciones de datos en Python, como gráficos de barras o líneas.
- **Time:** Es un módulo de Python que permite trabajar con el tiempo. Por ejemplo, se puede usar para pausar la ejecución del programa durante unos segundos, obtener la hora actual o medir cuánto tiempo tarda en ejecutarse una parte del código.
- **re:** Es un módulo que proporciona operaciones con expresiones regulares en Python. Las expresiones regulares son una herramienta poderosa para trabajar con cadenas de texto, permitiéndote buscar patrones específicos, reemplazar texto o dividir cadenas de texto de formas complejas.

Procesamiento del Lenguaje Natural (NLP)

De acuerdo a la revisión de literatura, creemos que la mejor opción en técnicas de procesamiento de texto para identificar hechos y opiniones y que ha dado buenos resultados es Natural Language Processing (NLP) (Baydogan & Alatas, 2021), (Bikakis, Flouris, Patkos & Plexousakis, 2023), (Gamal, Alfonse, Jiménez-Zafra & Aref, 2023), (Zayno & Radhi, 2022), (Gozuacik, Sakar & Ozcan, 2021),

(Oliveira, Medeiros & Mattos, 2020), (Rosa, Silva, Silva, Ayub, Carrillo, Nardelli & Rodríguez, 2020), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022.).

La elección de utilizar Natural Language Processing (NLP) para la identificación de hechos y opiniones sería justificada por varias razones fundamentales. En primer lugar, NLP se especializa en la comprensión y procesamiento del lenguaje natural, lo cual es esencial cuando la tarea implica analizar y clasificar expresiones verbales. Al abordar la identificación de opiniones, Natural Language Processing (NLP) es capaz de capturar la complejidad inherente en el lenguaje humano, permitiendo diferenciar entre hechos objetivos y evaluaciones subjetivas (Zayno & Radhi, 2022), (Gozuacik, Sakar & Ozcan, 2021).

Además, si los datos de entrada consisten en texto, como comentarios, reseñas o publicaciones en redes sociales, Natural Language Processing (NLP) se presenta como la opción más lógica. La capacidad de extraer características significativas del texto facilita el aprendizaje de patrones asociados con hechos y opiniones, proporcionando una perspectiva profunda sobre la naturaleza de la información textual.

Otra ventaja significativa radica en la capacidad de Natural Language Processing (NLP) para manejar la complejidad del lenguaje, especialmente en contextos donde se requiere distinguir entre hechos y opiniones en situaciones más elaboradas. Técnicas avanzadas, como el uso de modelos de lenguaje preentrenados como BERT (Bidirectional Encoder Representations from Transformers) (Bilal, Khan, Jan, Musa & Ali, 2023), (Bilal, Khan, Jan, Musa & Ali, 2023), (Alqurashi, 2022), permiten capturar semántica y contexto, brindando mayor precisión en la clasificación.

Finalmente, la interpretabilidad de los resultados también juega un papel crucial. Natural Language Processing (NLP) ofrece la ventaja de ser interpretable, permitiendo a los usuarios entender

cómo se toman las decisiones al analizar las características específicas del lenguaje que contribuyen a la identificación de hechos u opiniones. Esta transparencia es valiosa, especialmente en aplicaciones donde la comprensión del razonamiento detrás de las predicciones es esencial. En conjunto, estas razones respaldan la elección de Natural Language Processing (NLP) como una herramienta efectiva para la identificación de hechos y opiniones en entornos basados en texto.

Representaciones de Codificador Bidireccional de Transformadores (BERT)

De acuerdo a la revisión de literatura, creemos que la mejor opción en técnicas de aprendizaje automático es BERT, que significa "Bidirectional Encoder Representations from Transformers", es una técnica revolucionaria en el campo del procesamiento del lenguaje natural (PLN). Desarrollado por investigadores de Google, BERT ha cambiado la forma en que las máquinas entienden el lenguaje humano.

Los modelos de lenguaje previamente entrenados proporcionan un contexto para las palabras, que previamente han aprendido la aparición y las representaciones de palabras a partir de datos de entrenamiento no anotados. Las representaciones de codificador bidireccional de transformadores (BERT) son un modelo de lenguaje previamente entrenado que está diseñado para considerar el contexto de una palabra tanto del lado izquierdo como del derecho simultáneamente (Devlin et al., 2019)

BERT ha sido un gran avance en la comprensión del lenguaje por parte de las máquinas. Ha abierto nuevas posibilidades para aplicaciones que van desde los asistentes virtuales hasta los motores de búsqueda, mejorando la forma en que interactuamos con la tecnología en nuestro día a día.

Metodología XP o Programación Extrema

Según (Kent Beck, 1999), Extreme Programming (XP) fue diseñada para abordar las necesidades específicas del desarrollo de software realizado por equipos pequeños frente a requisitos vagos y cambiantes. Esta nueva metodología liviana desafía muchos principios convencionales, incluida la

suposición de larga data de que el costo de cambiar una pieza de software necesariamente aumenta dramáticamente con el transcurso del tiempo.

La Programación Extrema es una metodología ágil que enfatiza la comunicación, la adaptabilidad y la entrega continua de software funcional. A través de prácticas como la retroalimentación constante, la programación en pareja y la atención a la calidad del código, busca mejorar la satisfacción del cliente y la eficiencia del equipo de desarrollo.

Figura 1

Fases metodología XP



Nota: En la figura se muestra las fases de la metodología XP. Obtenida de <https://www.sinnaps.com/blog-gestion-proyectos/metodologia-xp>

Aquí se presentan las actividades clave que se llevaron a cabo respecto a la programación XP del proyecto:

Fase 1: Planificación

Se definen los requisitos del proyecto a través de historias de usuario y se planifican iteraciones cortas para abordar estos requisitos de manera incremental.

Fase 2: Diseño

Se promueve un diseño simple y se realizan mejoras iterativas. Se utiliza la refactorización para mejorar la estructura del código sin cambiar su comportamiento.

Fase 3: Codificación

Se fomenta la programación en parejas para aumentar la calidad del código y la productividad. Se escribe el código de manera que pase todas las pruebas automatizadas y se integra continuamente con el código existente.

Fase 4: Pruebas

Las pruebas automatizadas son fundamentales en XP. Se escriben pruebas antes del código (Test-Driven Development) para garantizar que el software cumpla con los requisitos y se mantenga la calidad a lo largo del tiempo.

Fase 5: Lanzamiento

Se realizan lanzamientos frecuentes y en pequeñas iteraciones para obtener retroalimentación temprana de los usuarios y adaptarse rápidamente a sus necesidades.

Planificación: En la etapa de planificación, se determinó el tipo de información a analizar, reflejados en la variedad de contenido en Facebook, y se ajustaron las técnicas de recolección y procesamiento de datos en consecuencia.

Diseño: El diseño flexible del código ha facilitado cambios y mejoras constantes. Esta adaptabilidad fue clave para incorporar BERT, una técnica avanzada de aprendizaje automático, para clasificar textos en hechos y opiniones.

Codificación: Se utilizó la programación en parejas para crear un código sólido y eficaz para la extracción y análisis de datos. Esta técnica fue muy adecuada para manejar la complejidad y los cambios frecuentes en los datos de Facebook, garantizando así un código fácil de mantener y de buena calidad

Pruebas: Se utilizó el Desarrollo Guiado por Pruebas para verificar que la herramienta cumpliera con los requisitos y mantuviera una buena calidad en la clasificación. Las pruebas automatizadas se integraron en el proceso de desarrollo, garantizando que los cambios en el código no afectaran la funcionalidad existente.

Lanzamientos: La fase de lanzamiento de la metodología XP en nuestro proyecto se ha materializado en la distribución de pasos específicos, detallando comandos para la instalación y configuración del proyecto en un entorno diferente. **Integración y Visualización de Resultados:** Con Flask, se proporcionó una interfaz web para mostrar los resultados del análisis de BERT. Las visualizaciones y estadísticas ofrecen una interpretación visual clara de los datos procesados, mejorando la comprensión del usuario sobre los resultados obtenidos.

Red Social - Facebook

Observamos que la mayoría de los trabajos utilizan (X) antes conocido como Twitter (Alqahtani, Alsheddi & Alhenaki, 2023), (Mehmood, Farooq, Naseem, Rustam, Villar, Rodríguez & Ashraf, 2022), (Mashraqi & Halawani, 2023), (Alattar & Shaalan, 2021), (Alattar & Shaalan, 2021), (Lee, Rustam, Washington, Barakaz, Aljedaani & Ashraf, 2022), (Wadhwani, Varshney, Gupta & Kumar, 2023), (Truong & Tran, 2023), (Huang, Zavareh & Mustafa, 2023), (Imran, Daudpota, Kastrati & Batra, 2020), (Imran, Daudpota, Kastrati & Batra, 2020), (Angskun, Tipprasert & Angskun, 2022), (Singh, Kuar, Rana & Dwivedi, 2022), (Singh, Kuar, Rana & Dwivedi, 2022), (Bilal, Khan, Jan, Musa & Ali, 2023), (Fatima, Samad Shaikh, Riaz, Ahmad, El-Affendi, Alyamani, Nabeel, Ali Khan, Yasin & Latif, 2022), (Haque, Islam, Islam & Ahsan,

2022), (Dadkhah, Zhang, Weismann, Firouzi & Ghorbani, 2023), (Sadiq, Aljrees & Ullah, 2023), (Waheeb, Khan & Shang, 2022), (Salsabila, Tyas, Romadhona & Purwitasari, 2023), (Alqurashi, 2022), (Nijhawan, Attigeri & Ananthakrishna, 2022) y queremos aportar una herramienta para la red social Facebook (Bilal, Khan, Jan, Musa & Ali, 2023).

La elección de Facebook como plataforma para la identificación de hechos y opiniones se justificaría por varias razones estratégicas y prácticas. En primer lugar, Facebook es una de las redes sociales más grandes y diversificadas, con miles de millones de usuarios que comparten una variedad de contenidos, desde publicaciones de texto hasta imágenes y videos. Esto proporciona un conjunto de datos rico y complejo que es esencial para abordar la tarea de identificación de hechos y opiniones (Huang, Zavareh & Mustafa, 2023),

Además, la naturaleza interactiva de Facebook facilita la observación de discusiones y conversaciones en tiempo real, lo que puede ser valioso para comprender cómo se forman y evolucionan las opiniones en un entorno social dinámico. La plataforma también refleja la diversidad de expresiones lingüísticas, desde mensajes informales hasta discusiones más estructuradas, lo que desafía positivamente la identificación de opiniones y la aplicación de técnicas avanzadas de Natural Language Processing (NLP) (Anzum & Gavrilova, 2023), (Baydogan & Alatas, 2021), (Bikakis, Flouris, Patkos & Plexousakis, 2023), (Gamal, Alfonse, Jiménez-Zafra & Aref, 2023), (Zayno & Radhi, 2022), (Gozuacik, Sakar & Ozcan, 2021), (Oliveira, Medeiros & Mattos, 2020), (Rosa, Silva, Silva, Ayub, Carrillo, Nardelli & Rodríguez, 2020), (Renjith, Abraham, Jyothi, Chandran & Thomson, 2022)(Renjith, Abraham, Jyothi, Chandran & Thomson, 2022).

La demografía diversa de los usuarios en Facebook también es un factor clave. Si la tarea de identificación de hechos y opiniones está vinculada a objetivos específicos de investigación o marketing,

la representación de una amplia gama de usuarios en la plataforma puede ser beneficiosa para obtener una perspectiva más holística.

La elección de Facebook para la identificación de hechos y opiniones se basa en la riqueza de sus datos, la diversidad de expresiones, la interactividad de la plataforma y la representación demográfica, factores que contribuyen a la creación de un entorno de investigación valioso y relevante para comprender las opiniones en el contexto social digital.

Capítulo IV

Diseño y Desarrollo

Uso de metodología XP

La metodología de programación extrema (XP) es un enfoque ágil para el desarrollo de software que enfatiza la retroalimentación continua, la simplicidad, la comunicación y el coraje. A continuación, se describe cómo se aplicaron las fases de XP a este proyecto:

Planificación: En esta etapa, se identificaron los requisitos del proyecto, como la extracción de publicaciones de Facebook y el análisis de estas utilizando BERT para determinar si son hechos u opiniones.

Diseño: Se diseña la arquitectura del proyecto de forma simple y directa. En este caso, se decide utilizar Flask para el servidor web y Python para el procesamiento de datos y el análisis con BERT.

Codificación: Se comienza a escribir el código en iteraciones cortas. Se implementaría la extracción de publicaciones, el análisis con BERT y la visualización de resultados en la página web.

Pruebas: En XP, las pruebas son fundamentales y se escriben antes del código (Test Driven Development). Se crearon pruebas para cada funcionalidad, como la extracción de datos, el procesamiento y el análisis con BERT. Estas pruebas ayudan a garantizar que el código cumple con los requisitos y funciona correctamente.

Integración: Se integrarían continuamente las nuevas funcionalidades y se realizarían pruebas de integración para asegurar que todas las partes del sistema funcionan juntas.

Lanzamiento: Una vez que el sistema es estable y cumple con los requisitos, se prepara para el lanzamiento. Esto podría incluir la configuración del entorno de producción y la implementación del sistema.

Mantenimiento: Después del lanzamiento, el proyecto entra en una fase de mantenimiento, donde se harían ajustes y mejoras basadas en la retroalimentación de los usuarios. Se planifican nuevas iteraciones para implementar estas mejoras.

Cada una de estas fases se realizaría de forma iterativa y con un enfoque en la simplicidad y la retroalimentación continua, que son principios clave de la programación extrema.

Desarrollo del sistema

El Sistema de Identificación de hechos y opiniones en redes sociales digitales utilizando técnicas de aprendizaje automático está desarrollado en el lenguaje de programación Python, es una aplicación informática diseñada para analizar hechos y opiniones a partir del análisis de textos. Este sistema utiliza técnicas aprendizaje automático, específicamente NLP (Procesamiento de Lenguaje Natural), para extraer características relevantes de los textos, como la frecuencia de palabras, la presencia de palabras clave y la estructura gramatical; adicional se utiliza el modelo BERT (Representaciones de Codificadores Bidireccionales de Transformadores). Este modelo es un lenguaje preentrenado por Google, que utiliza la arquitectura de Transformer para entender el contexto y la relación entre las palabras en un texto.

El proceso que se está desarrollando en el aplicativo es:

Recopilación y Preparación de Datos: Se recopilaron textos de publicaciones de la red social Facebook mediante el envío de credenciales de usuario a través del BackEnd accediendo a las publicaciones y comentarios de interés. Estas se almacenan de manera temporal para su posterior procesamiento.

Carga y Preprocesamiento de Datos: El código, carga un conjunto de textos de publicaciones de Facebook. Las publicaciones se convierten en tensores para poder ser procesados por el modelo. Además, se realiza la normalización para convertir el texto a un formato estándar, como pasar todo a minúsculas.

Entrenamiento del modelo: El modelo empieza a aprender utilizando el conjunto de entrenamiento y se ratifica en el conjunto de pruebas durante 3 épocas. En cada época, el modelo BERT se entrenará durante tres pasadas completas sobre el conjunto de datos de entrenamiento.

Guardado del Modelo Entrenado: Una vez completado el entrenamiento, guardar el modelo entrenado para su posterior uso en la inferencia o la evaluación en el conjunto de prueba.

Evaluación del Modelo: Finalmente, evaluar el modelo entrenado en un conjunto de prueba para medir su rendimiento en la clasificación de hechos y opiniones, utilizando métricas como la precisión.

Despliegue y Uso

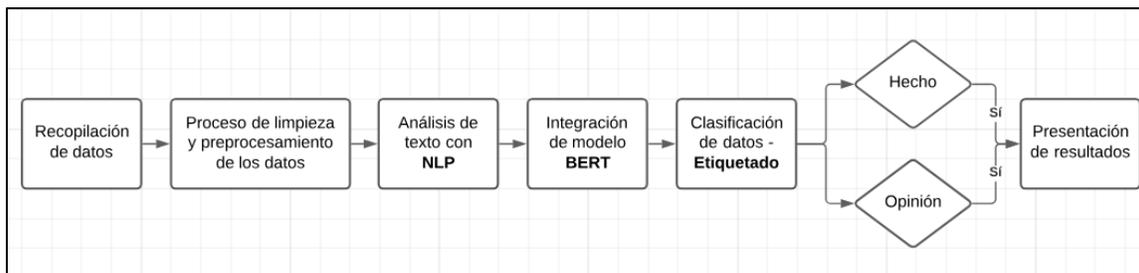
El modelo entrenado puede implementarse en un entorno de producción para su uso en la clasificación automatizada de hechos y opiniones mediante la técnica de procesamiento lenguaje natural (NLP) y el modelo BERT.

Arquitectura de la Solución Propuesta

Diseñar una arquitectura para la identificación de hechos y opiniones en la red social Facebook mediante Natural Language Processing (NLP) implica una combinación de técnicas específicas para el análisis del texto presente en publicaciones. El flujograma y la arquitectura por considerar serían los siguientes:

Figura 2

Flujograma de la arquitectura de la solución propuesta

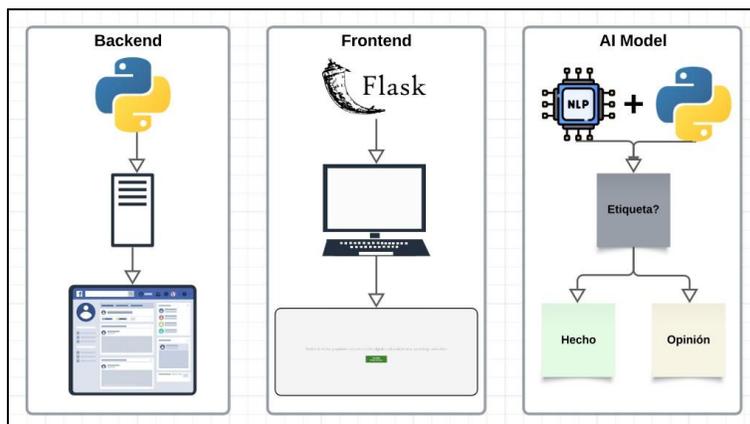


Nota. La figura representa el flujo utilizado para el desarrollo del proyecto.

Esta arquitectura proporciona una estructura general para la identificación de hechos y opiniones en la red social Facebook mediante Natural Language Processing (NLP) y el modelo BERT.

Figura 3

Arquitectura de la Solución Propuesta



Nota. La figura representa la arquitectura utilizada para el desarrollo del proyecto.

BackEnd: Desarrollar y mostrar un aplicativo de clasificación de hechos y opiniones de las redes sociales digitales como Facebook utilizando el Procesamiento del Lenguaje Natural (NLP) y el modelo de Representaciones de Codificador Bidireccional de Transformers (BERT). El propósito es que este modelo pueda analizar publicaciones y predecir a qué etiqueta pertenece el texto. Las etiquetas en este contexto son “Hecho” y “Opinión”, se llevan a cabo las siguientes etapas:

Recopilación de Datos:

Para acceder a los datos de Facebook, el sistema utiliza un método de autenticación directa mediante el envío de credenciales, lo que implica la navegación por la interfaz de usuario utilizando la herramienta Selenium, para acceder a las publicaciones.

Figura 4

Código para recopilación de datos

```
def obtener_publicaciones(driver, cantidad_a_encontrar, max_desplazamientos=50):
    texto_de_publicaciones = []
    contador_desplazamiento = 0

    while len(texto_de_publicaciones) < cantidad_a_encontrar and contador_desplazamiento < max_desplazamientos:
        altura_actual = driver.execute_script("return document.body.scrollHeight")
        driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
        WebDriverWait(driver, 30).until(lambda d: d.execute_script("return document.body.scrollHeight") > altura_actual)
        time.sleep(2) # Ajustar basado en la carga de la página

        div_elements = driver.find_elements(By.XPATH, '//div[@data-ad-preview="message"]')
        for div_element in div_elements:
            texto_de_la_publicacion = div_element.text
            if texto_de_la_publicacion and texto_de_la_publicacion not in texto_de_publicaciones:
                texto_de_publicaciones.append(texto_de_la_publicacion)

        contador_desplazamiento += 1

    return texto_de_publicaciones

def main():
    # Carga el modelo y el tokenizer
    tokenizer = BertTokenizer.from_pretrained(TOKENIZER_PATH)
    model = BertForSequenceClassification.from_pretrained(MODEL_PATH)
    model.eval() # Asegura que el modelo esté en modo evaluación

    options = webdriver.ChromeOptions()
    options.add_argument('--disable-extensions')
    options.add_argument('--headless')
    options.add_argument('--disable-gpu')
    options.add_argument('--no-sandbox')
    driver = webdriver.Chrome(options=options)

    try:
        driver.get('https://www.facebook.com/')
        WebDriverWait(driver, 30).until(lambda d: d.find_element(By.NAME, 'email')).send_keys('correo electrónico') # Inscribe el correo electrónico
        driver.find_element(By.NAME, 'pass').send_keys('contraseña') # Inscribe la contraseña
        driver.find_element(By.NAME, 'login').click()
        WebDriverWait(driver, 30).until(lambda d: d.url_contains('https://www.facebook.com/'))
        time.sleep(3)

        enlace = "enlace" # Ajusta el enlace de facebook
        driver.get(enlace)

        texto_publicaciones = obtener_publicaciones(driver, 10) # Ajusta la cantidad deseada
```

Nota. La figura representa el código para obtener datos de Facebook.

Proceso de limpieza y Preprocesamiento del texto:

Se implementan funciones de limpieza en el BackEnd para eliminar elementos irrelevantes como, menciones, hashtags y emojis. También se corrigen errores ortográficos y se normaliza el texto (por ejemplo, convirtiendo todo a minúsculas). Adicional, se aplican técnicas preprocesamiento como la tokenización, la eliminación de palabras vacías y la lematización para transformar el texto en un formato adecuado para el análisis.

Figura 5*Código para la limpieza*

```
def limpiar_texto(texto):
    texto_limpio = re.sub(r'^a-zA-ZáéíóúÁÉÍÓÚñ\S', '', texto.lower())
    texto_limpio = re.sub(r'http\S+', '', texto_limpio)
    texto_limpio = re.sub(r'\s+', '', texto_limpio)
    texto_limpio = re.sub(r'ver más', '', texto_limpio).strip() # Eliminar la frase "ver más"
    texto_limpio = re.sub(r'Ver más', '', texto_limpio).strip() # Eliminar la frase "ver más"

    return texto_limpio
```

Nota. La figura representa el código para la limpieza del texto obtenido de Facebook.

Preparación para el modelo BERT

Los datos preprocesados se estructuran de acuerdo con los requisitos del modelo BERT, incluyendo con la conversión a tokens BERT y la creación de máscaras de atención si es necesario.

Análisis y clasificación

Los datos preparados se envían al modelo BERT para su análisis y clasificación como Hechos u Opiniones. Los resultados se almacenan y se envían para su posterior visualización.

Figura 6*Código para analizar el análisis del texto*

```
def analizar_con_bert(texto, tokenizer, model):
    inputs = tokenizer(texto, return_tensors="pt", padding=True, truncation=True, max_length=1024)
    with torch.no_grad():
        outputs = model(**inputs)

    logits = outputs.logits
    probabilities = torch.softmax(logits, dim=1).tolist()[0]
    prediction = 'Fact' if torch.argmax(logits) == 0 else 'Opinion'
    return prediction, probabilities
```

Nota. La figura representa el código para el análisis del texto obtenido de Facebook.

Manejo de resultados

Los resultados del análisis se gestionan en el BackEnd aplicando el análisis de texto con NLP y el modelo BERT, y posteriormente ser enviados al FrontEnd en donde se visualiza la clasificación de las

publicaciones en hechos y opiniones, incluye gráfico de barras para indicar la proporción entre las etiquetas y a su vez proporciona información valiosa sobre las tendencias y patrones en las publicaciones de Facebook.

En resumen, el manejo de resultados en el análisis de hechos y opiniones implica visualizar y analizar los datos de manera efectiva, así como comunicar los hallazgos y utilizarlos para informar decisiones y acciones. Adicional a ello, este proceso es fundamental para asegurar que los resultados del análisis sean útiles y tengan un impacto positivo.

FrontEnd: En esta interfaz permite al usuario interactuar con el sistema de análisis de hechos y opiniones. Se centra en recoger las entradas del usuario, como publicaciones de redes sociales, y mostrar los resultados del análisis realizado por el BackEnd. El código incluye tecnologías web como HTML y CSS, y con ello el framework FLASK para una estructura y diseño más eficiente. El FrontEnd se comunica con el BackEnd mediante el uso de solicitudes HTTP (como GET o POST) para enviar y recibir datos, asegurando una experiencia de usuario fluida y dinámica.

Interfaz de Usuario (UI): Está diseñada para proporcionar una experiencia sencilla e intuitiva que permite al usuario analizar publicaciones de Facebook y comprender los resultados del análisis, e incluye la siguiente característica clave:

- Un botón claramente etiquetado que el usuario puede presionar para comenzar su análisis, siendo el punto de partida para todo el proceso.

Figura 7*Interfaz gráfica*

Nota. La figura representa la interfaz inicial del aplicativo web. Fuente Propia.

Visualización de Resultados: Después de ejecutar el sistema para su análisis, se muestra una versión limpia del texto en la interfaz. Esta versión ha sido procesada para eliminar elementos innecesarios como enlaces, menciones y emojis, permitiendo al usuario ver claramente el texto que se está analizando.

Luego del análisis, se muestra la clasificación del texto como "Hecho" u "Opinión". Esta clasificación se presenta de manera destacada para que el usuario pueda identificar rápidamente el resultado.

Finalmente, se muestra la probabilidad o nivel de confianza que el sistema tiene en su predicción. Por ejemplo, aparece como "Hecho (90% de confianza)" u "Opinión (60% de confianza)", proporcionando una medida cuantitativa de la certeza del resultado.

Figura 8

Visualización de resultados

Resultados:

#	Texto Original	Contenido	Predicción	Probabilidades
1	I had to do something quick and easy in order to start promoting this book. Soooo, I made this cover on Canva. This will not be the final cover but I hope it's an attention-grabber. Your thoughts?	I had to do something quick and easy in order to start promoting this book soooo I made this cover on canva this will not be the final cover but I hope its an attentiongrabber your thoughts	Fact	[0.5, 0.4]
2	Hey all! I'm currently writing a fantasy novel and I could do with some advance if possible? I'm writing a chapter that goes back in history explaining an ancient battle that took place which shaped the land today. I have described the battle and the events after as much as I can with out spoiling the story. The chapter is only around 3000 words. Can I just cut back to the current time line without ending the chapter? Or would it be best to go back and try to extend the chap... Ver más	hey all im currently writing a fantasy novel and I could do with some advance if possible im writing a chapter that goes back in history explaining an ancient battle that took place which shaped the land today I have described the battle and the events after as much as I can with out spoiling the story the chapter is only around words can I just cut back to the current time line without ending the chapter or would it be best to go back and try to extend the chap	Opinion	[0.0, 0.9]
3	I published my 432-page fire service memoir through Amazon KDP. I have had a bit of limited success selling through Amazon, but do well in selling author copies. My book was written by a fireman (me), about firemen and for firemen. Obviously, my target audience is those persons involved in the Emergency Services. Therefore, there is a market for my book. Because I sell primarily author copies, I have ordered from Amazon at least ten times in the past few months. Now, comes th... Ver más	i published my page fire service memoir through amazon kdp i have had a bit of limited success selling through amazon but do well in selling author copies my book was written by a fireman me about firemen and for firemen obviously my target audience is those persons involved in the emergency services therefore there is a market for my book because i sell primarily author copies I have ordered from amazon at least ten times in the past few months now comes th	Opinion	[0.3, 0.6]
4	Hello all, this is the book cover for my story made by my friend who's only 14. I really want to give her a chance to show her skills! The story is based on a man who kills people one after the other and the case goes on for years, the police think they are near to getting him but he's always 10 steps ahead! At last they know he snags his victims head with a joyo but he replaced the regular rope with one of his own. Your thoughts are most welcome (negative and positive is ok... Ver más	hello all this is the book cover for my story made by my friend whos only i really want to give her a chance to show her skills the story is based on a man who kills people one after the other and the case goes on for years the police think they are near to getting him but hes always 10 steps ahead at last they know he snags his victims head with a joyo but he replaced the regular rope with one of his own your thoughts are most welcome negative and positive is ok	Opinion	[0.4, 0.5]
5	Ok folks, I really need inspiration, and I mean like, major inspiration. I've been running dry on ideas since november last year, and it sucks please help	ok folks i really need inspiration and i mean like major inspiration ive been running dry on ideas since november last year and it sucks please help	Opinion	[0.0, 0.9]

Nota. La figura representa la visualización de resultados del aplicativo web. Fuente Propia.

Este proyecto está diseñado para facilitar el análisis de publicaciones de Facebook con un solo clic en un botón, permitiendo al usuario visualizar el texto limpio, recibir una predicción de clasificación con su probabilidad asociada y comprender los resultados de manera clara y eficiente.

Interpretación de Resultados: Una vez que se muestra el resultado de la predicción, se analiza cómo el modelo ha clasificado el texto. Si las probabilidades son altas y la etiqueta predicha es razonable con lo que esperaba para ese texto, entonces el modelo está funcionando adecuadamente. Si las predicciones no son precisas, podría ser necesario ajustar y de ser el caso mejorar el modelo.

Figura 9

Interpretación de resultados

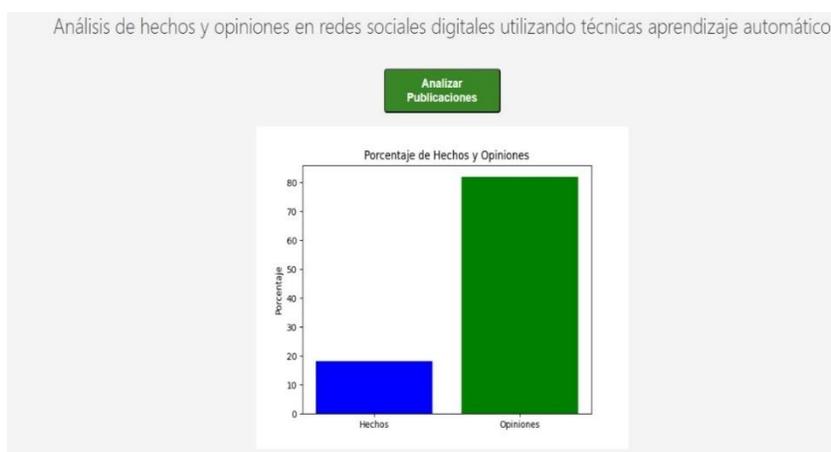
Contenido	Predicción	Probabilidades
hi everybody i want to write to enhance my writing skills that is why i entered this group i also write often but i am a new writer i can learn how to write when reading these posts i hope i can connect with someone to assist me in writing	Opinion	[0.2, 0.7]
i would like to write a bio about a little known woman in late s and early s england adventurous mountaineer photographer author of several books i learned about her when i found a few of her photo albums documenting her climbs travels and people including a shoot she did with arthur conan doyle in a viking costume i have lots of research material on her but what other research do you recommend im starting to develop a timeline for her activities over a fe	Opinion	[0.0, 0.9]
as you know selfpromotion is not allowed in this community however because we know so many of you are excited about what youre working on every monday well start a thread to allow you to share whatever youd like your latest blog post link to your website or author facebook page a bit of your latest workinprogress upcoming writing contests etc lets motivate one another by sharing and supporting each others work please keep selfpromo this this thread only	Fact	[0.7, 0.2]

Nota. La figura representa la visualización del contenido analizado, la predicción y la probabilidad obtenida Fuente Propia.

En este caso, se analizan múltiples publicaciones, se muestra mediante un gráfico estadístico que resumen los resultados generales, como el porcentaje de publicaciones clasificadas como Hechos versus Opiniones.

Figura 10

Gráfico porcentual de resultados



Nota. La figura representa el gráfico porcentual en base a los Hechos y Opiniones analizados.

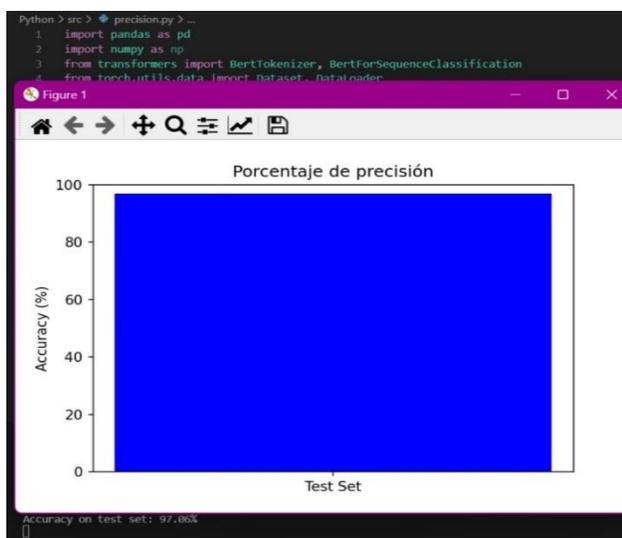
Precisión: Es una métrica clave para evaluar la capacidad del modelo BERT en clasificar correctamente las publicaciones de Facebook como hechos u opiniones. Se calcula como el número de predicciones correctas dividido por el número total de predicciones realizadas por el modelo.

La precisión es importante en este proyecto porque:

- **Confianza en el Modelo:** Una alta precisión significa que el modelo es confiable y que los usuarios pueden confiar en los resultados del análisis para tomar decisiones informadas basadas en los datos de las redes sociales.
- **Calidad del Análisis:** La precisión refleja la calidad del análisis de NLP y el modelo BERT en comprender y clasificar correctamente el contenido de las publicaciones.
- **Mejora Continua:** Monitorear la precisión del modelo ayuda a identificar áreas de mejora y a ajustar el modelo o los datos de entrenamiento para aumentar su rendimiento.

Figura 11

Porcentaje de precisión



Nota. La figura presenta un 97.06% de precisión siendo exitoso el rendimiento del modelo.

En resumen, la precisión dentro de este proyecto es una medida fundamental para evaluar el rendimiento del modelo en la clasificación de hechos y opiniones en publicaciones de Facebook, y es crucial para garantizar la confiabilidad y la calidad del análisis.

Capítulo V

Pruebas Funcionamiento de Software

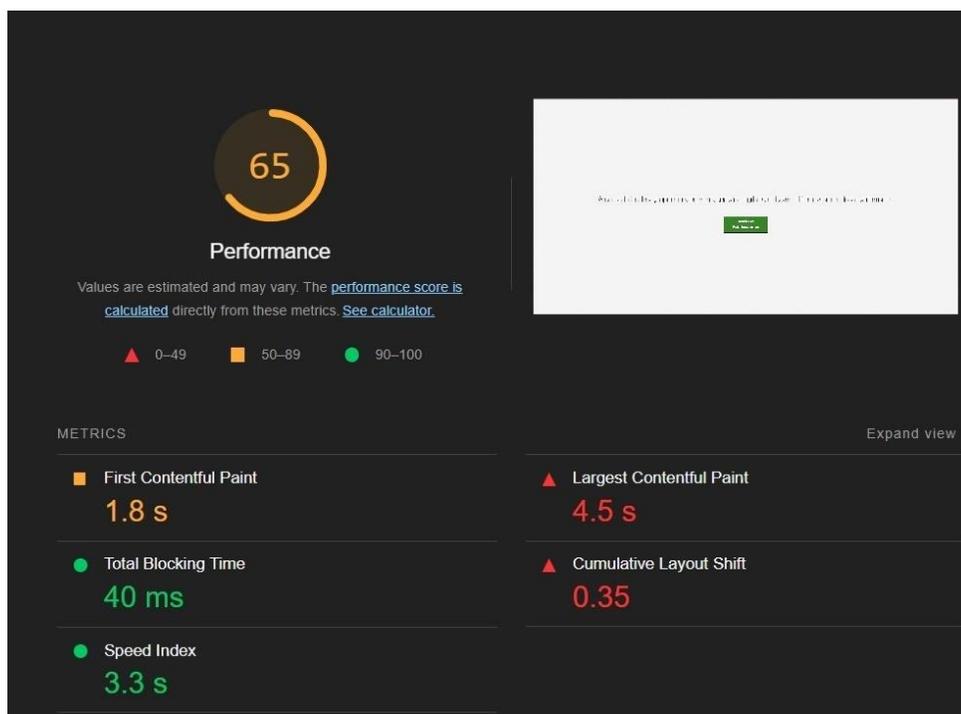
Lighthouse de Google permite evaluar pruebas de rendimiento, de usabilidad y de seguridad por lo cual se ha tomado en cuenta esta herramienta que permite tener un análisis profundo.

Lighthouse - Evaluación de rendimiento del software

Como se puede ver en la figura 12, las pruebas de rendimiento mostraron un resultado del 65%, lo cual está por encima de lo normal. Esto significa que hay margen para mejorar, especialmente en áreas como la velocidad de carga y el uso de recursos adicionales.

Figura 12

Evaluación de desempeño - Lighthouse



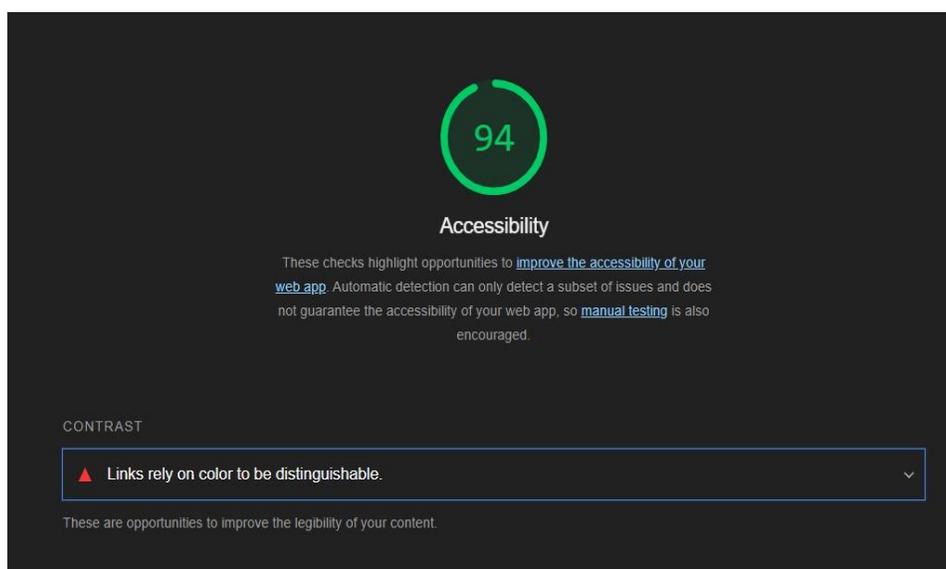
Nota. La figura representa un 65% de asertividad con relación al rendimiento.

Lighthouse - Evaluación de usabilidad del software

En la figura 13, se muestran los resultados de accesibilidad, los cuales arrojaron un puntaje del 94%. Se identificaron que los enlaces dependen del color para ser distinguibles para una mejor identificación. A pesar de este resultado, es poco probable que esto cause problemas en la mayoría de las interacciones del usuario.

Figura 13

Evaluación de accesibilidad-Lighthouse



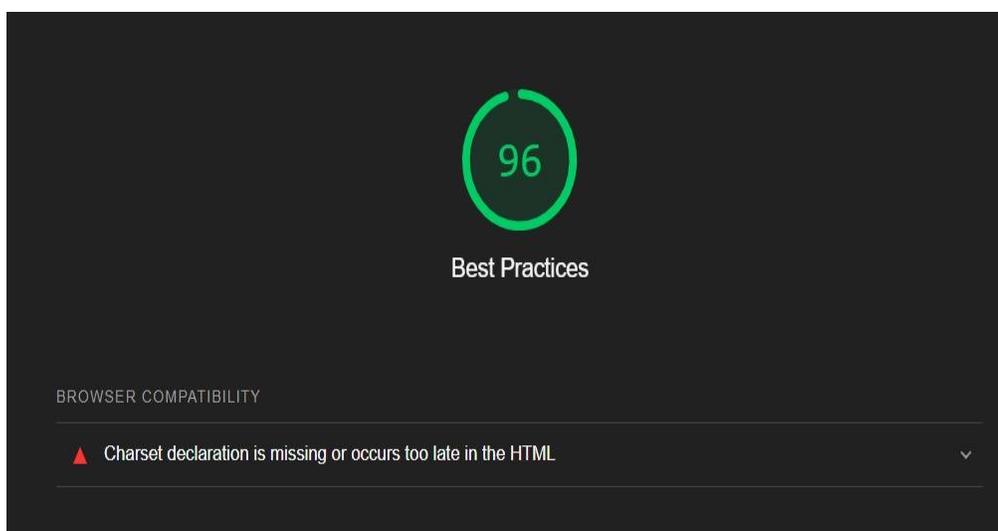
Nota. La figura muestra los resultados de la herramienta Lighthouse de Google que presenta resultados positivos.

Lighthouse - Evaluación de seguridad del software

Por último, en la prueba de seguridad, el software obtuvo un 96% de aprobación, lo que significa que está libre de problemas de seguridad. Esto incluye políticas de seguridad adecuadas, conexiones seguras y medidas para prevenir la fuga de información. Estos aspectos se visualizan a detalle en la figura 14, en la cual indica que falta la declaración de caracteres o aparece demasiado tarde en el HTML, esto hace referencia a la codificación que se está utilizando, como UTF-8, y generalmente se coloca dentro de la etiqueta <meta> en la sección <head> del HTML.

Figura 14

Evaluación de defensa-Lighthouse



Nota. La figura muestra que el software es seguro con un porcentaje del 96%.

Capítulo VI

Conclusiones, recomendaciones y trabajos futuros

Conclusiones

- Se realizó la revisión de literatura y se encontraron estudios con técnicas de aprendizaje automático tales como, Naive Bayes y Convolutional Neural Networks, entre otras; que se han implementado en diferentes redes sociales digitales, siendo Facebook la segunda red social con más trabajos relacionados.
- La capacidad de discernir entre hechos y opiniones es fundamental para combatir la desinformación en línea, asegurando que la información verificable y objetiva se distinga de las opiniones subjetivas y potencialmente sesgadas.
- El uso de BERT, un modelo de procesamiento del lenguaje natural es fundamental para mejorar la identificación de hechos y opiniones en textos. Este modelo avanzado puede ayudar analizar y comprender el contenido de manera más precisa, lo que es especialmente útil en el contexto de las redes sociales digitales.
- Se realizaron análisis de rendimiento utilizando la herramienta Lighthouse de Google, obteniendo una puntuación de 65 sobre 100. La exploración de la usabilidad produjo un resultado de 94 sobre 100, mientras que la evaluación de la seguridad reveló un puntaje de 96 sobre 100. En el contexto de las pruebas unitarias, se ejecutaron satisfactoriamente sin errores.

Recomendaciones

- Para realizar una revisión exhaustiva de literatura, utilizar otras bases de datos académicas como Google Scholar o Web of Science, en donde se aborda la clasificación de texto en identificación de hechos u opiniones mediante técnicas de aprendizaje automático en redes sociales digitales.
- Explorar técnicas avanzadas de preprocesamiento de texto para mejorar la calidad de los datos antes de alimentarlos al modelo BERT. Esto puede incluir el uso de técnicas de normalización de texto más sofisticadas, la identificación y el manejo de jergas o abreviaturas específicas de las redes sociales, y la implementación de métodos para manejar el sarcasmo o la ironía, que pueden ser desafiantes para la clasificación de hechos y opiniones.
- Además de la precisión, utilizar otras métricas como la exhaustividad, el puntaje F1 y la matriz de confusión para evaluar el modelo de manera más completa. Esto proporcionará una visión más detallada de su rendimiento y ayudará a identificar áreas específicas de mejora.
- Continuar mejorando la interfaz de usuario en el FrontEnd para proporcionar una experiencia más intuitiva y enriquecedora. Esto puede incluir la implementación de características interactivas adicionales, como la visualización en tiempo real de la clasificación y la probabilidad, y la mejora de la accesibilidad y el diseño responsive.

Trabajos futuros

En el futuro, sería factible expandir y enriquecer este estudio explorando nuevas formas de entender y analizar las interacciones en las redes sociales. Además, profundizar en la comprensión de cómo las personas expresan sus pensamientos y sentimientos en línea, y cómo esto impacta en la percepción de la realidad. La colaboración con educadores y formadores de opinión podría ser clave para crear conciencia sobre la importancia de una comunicación responsable y crítica en línea. Finalmente, explorar maneras de promover un uso más saludable y positivo de las redes sociales, donde el respeto por la diversidad de opiniones y la búsqueda de la verdad sean valores fundamentales.

Capítulo VII

Referencias

- Abuein, Q., Shatnawi, M., & Ghazalat, L. (2022). Detection of Americans? Behavior toward Islam on Facebook. doi:10.5614/itbj.ict.res.appl.2022.16.3.7
- Akila, R., & Revathi, S. (2023). Fine Grained Analysis of Intention for Social Media Reviews. doi:10.58346/JISIS.2023.I2.003
- Alattar, F., & Shaalan, K. (2021). A Survey on Opinion Reason Mining and Interpreting Sentiment Variations. doi:10.1109/ACCESS.2021.3063921
- Albahr, A., & Albahr, M. (2020). An empirical comparison of fake news detection using different machine learning algorithms. *International Journal of Advanced Computer Science and Applications(IJACSA)*. doi:10.14569/IJACSA.2020.0110917
- Alqahtani, A., Alhenaki, L., & Alsheddi, A. (2023). Text-based Sarcasm Detection on Social Networks: A Systematic Review. *International Journal of Advanced Computer Science and Applications(IJACSA)*. doi:10.14569/IJACSA.2023.0140336
- Alqurashi, T. (2021). Stance Analysis of Distance Education in the Kingdom of Saudi Arabia during the COVID-19 Pandemic Using Arabic Twitter Data. doi:10.3390/s22031006
- Amazon Web Services, Inc. (n.d.). *¿Qué es Python? - Explicación del lenguaje Python.* . (s.f). Obtenido de *¿Qué es Python? - Explicación del lenguaje Python.* : <https://aws.amazon.com/es/what-is/python/>
- Angskun, J., Tipprasert, S., & Angskun, T. (2020). Big data analytics on social networks for real-time depression detection. *Journal of Big Data*. doi:10.1186/s40537-022-00622-2

- Anzum, F., & Gavrilova, M. (2023). Emotion Detection From Micro-Blogs Using Novel Input Representation. doi:10.1109/ACCESS.2023.3248506
- Ashir, A. (2021). A Generalized Method for Sentiment Analysis across Different Sources. *Applied Computational Intelligence and Soft Computing*. doi:10.1155/2021/2529984
- Bang, T. H., & Cuong, T. V. (2023). A framework for fake news detection based on the wisdom of crowds and the ensemble learning model. *Computer Science and Information Systems*. doi:10.1109/ACCESS.2021.3083638
- Baydogan, C., & Alatas, B. (2021). Metaheuristic Ant Lion and Moth Flame Optimization-Based Novel Approach for Automatic Detection of Hate Speech in Online Social Networks. *Social networking (online);Blogs;Measurement;Machine learning algorithms;Search problems;Feature extraction;Deep learning*. doi:10.1109/ACCESS.2021.3102277
- Bikakis, A., Plexousakis, D., Patkos, T., & Plexousakis, D. (2023). Sketching the vision of the Web of Debates. *Frontiers in Artificial Intelligence*. doi:10.3389/frai.2023.1124045
- Bilal, M., Khan, A., Jan, S., Musa, S., & Ali, S. (2023). Roman Urdu Hate Speech Detection Using Transformer-Based Model for Cyber Security Applications. doi:10.3390/s23083909
- Canive, T. (27 de May de 2020). *Metodología XP o programación extrema: ¿Qué es y cómo aplicarla?* *Sinnaps*. Obtenido de <https://www.sinnaps.com/blog-gestion-proyectos/metodologia-xp#caracteristicas-de-la-metodologia-xp>
- Chen, J., Zhou, G., Lu, J., Wang, S., & Li, S. (2023). Trust-Aware Evidence Reasoning and Spatiotemporal Feature Aggregation for Explainable Fake News Detection. doi:10.3390/app13095703

- Chrismanto, A. R., Sari, A. K., & Suyanto, Y. (2023). Enhancing Spam Comment Detection on Social Media With Emoji Feature and Post-Comment Pairs Approach Using Ensemble Methods of Machine Learning. doi:10.1109/ACCESS.2023.3299853
- Dadkhah, S., Zhang, X., & Weismann, A. G. (2023). The Largest Social Media Ground-Truth Dataset for Real/Fake Content: TruthSeeker. doi:10.1109/TCSS.2023.3322303
- Dashtipour, K., Gogate, M., Li, J., Jiang, F., Kong, B., & Hussain, A. (2020). A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. *Neurocomputing*. doi:10.1016/j.neucom.2019.10.009
- De León Languré, A., & Zareei, M. (2023). Breaking Barriers in Sentiment Analysis and Text Emotion Detection: Toward a Unified Assessment Framework. doi:10.1109/ACCESS.2023.3331323
- de Oliveira, N., Medeiros, D., & Mattos, D. (2020). A Sensitive Stylistic Approach to Identify Fake News on Social Networking. doi:10.1109/LSP.2020.3008087
- Demilie, W. B., & Salau, A. O. (2022). Detection of fake news and hate speech for Ethiopian languages: a systematic review of the approaches. *Journal of Big Data*.
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Dewani, A., Memon, M. A., Bhatti, S., Sulaiman, A., Hamdi, M., Alshahrani, H., . . . Shaikh, A. (2023). Detection of Cyberbullying Patterns in Low Resource Colloquial Roman Urdu Microtext using Natural Language Processing, Machine Learning, and Ensemble Techniques. doi:10.3390/app13042062

- Dhelim, S., Chen, L., Das, S., Ning, H., Nugent, C., Leavey, G., . . . Bantry-White, E. (2023). Detecting Mental Distresses Using Social Behavior Analysis in the Context of COVID-19: A Survey. doi:10.1145/3589784
- Dwyer, G. (2016). *Flask By Example*. Packt Publishing Ltd.
- Elzayady, H., M. M., K. B., & G. S. (2022). Detecting Arabic textual threats in social media using artificial intelligence: An overview. *Indonesian Journal of Electrical Engineering and Computer Science*. doi:10.11591/ijeecs.v25.i3.pp1712-1722
- Fahad, A., Alhalangy, A., & Hassan, A. (2023). Fake Accounts Identification in Mobile Communication Networks Based on Machine Learning. *International Journal of Interactive Mobile Technologies*. doi:10.3991/ijim.v17i04.37645
- Fkih, F., Moulahi, T., & Alabdulatif, A. (2023). Machine Learning Model for Offensive Speech Detection in Online. *WSEAS Transactions on Information Science and Applications*. doi:10.37394/23209.2023.20.2
- Gamal , D., Alfonse, M., Jiménez-Zafra, S. M., & Aref, M. (2023). Intelligent Multi-Lingual Cyber-Hate Detection in Online Social Networks: Taxonomy, Approaches, Datasets, and Open Challenges. doi:10.3390/bdcc7020058
- Gidwani, M., & Rao, A. (2023). Comparative analysis of rumor detection in social networks using various classifiers. doi:10.15622/ia.22.4.3
- Gozuacik, N., Saka, O., & Ozcan, S. (2020). Social media-based opinion retrieval for product analysis using multi-task deep neural networks. doi:10.1016/j.eswa.2021.115388

- Haddaoui, B., Chiheb, R., Faizi, R., & Afia, A. (2023). Sentiment analysis in SemEval: a review of sentiment identification approaches. *International Journal of Electrical and Computer Engineering (IJECE)*. doi:10.11591/ijece.v13i3.pp3322-3338
- Haque, R., Islam, N., Islam, M., & Ahsan, M. M. (2022). A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning. doi:10.3390/technologies10030057
- Huang, H., Zavareh, A. A., & Mustafa, M. B. (2023). Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions. *Electronic commerce; Machine learning; Deep learning; Sentiment analysis; Databases; Social networking (online); Business; Natural language processing*. doi:10.1109/ACCESS.2023.3307308
- Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets.
- Kaliyar, R. K., Goswami, A., & Narang, P. (2021). EchoFakeD: improving fake news detection in social media with an efficient deep neural network. doi:10.1007/s00521-020-05611-1
- Karyukin, V., Mutanov, G., Mamykova, Z., Nassimova, G., Torekul, S., Sundetova, Z., & Negri, M. (2022). On the development of an information system for monitoring user opinion and its role for the public. *Journal of Big Data*.
- Kent, B. (October). de 1999). *Extreme Programming Explained. O'Reilly Media*. Obtenido de <https://www.oreilly.com/library/view/extreme-programming-explained/0201616416/>
- Keyvanpou, M., Zandian, Z. K., & Heidarypana, M. (2020). OMLML: a helpful opinion mining method based on lexicon and machine learning in social networks. doi:10.1007/s13278-019-0622-6
- Khalafa, M., Alqatawna, J., Al-Sayyed, R., Eshtay, M., & Kobbaey, T. (2021). Violence Detection over Online Social Networks: An Arabic Sentiment Analysis Approach. doi:10.3991/ijim.v15i14.23029

Krommyda, M., Rigos, A., Bouklas, K., & Amditis, A. (2021). An Experimental Analysis of Data Annotation Methodologies for Emotion Detection in Short Text Posted on Social Media.

doi:10.3390/informatics8010019

Kumar, A., & Taylor, J. (2023). Feature importance in the age of explainable AI: Case study of detecting fake news & misinformation via a multi-modal framework. doi:10.1016/j.ejor.2023.10.003

Kusal, S., Patil, S., Kotecha, K., Aluvalu, R., & Varadarajan, V. (2021). AI Based Emotion Detection for Textual Big Data: Techniques and Contribution. doi:10.3390/bdcc5030043

Lee, E., Rustam, F., Washington, P. B., Barakaz, F., Aljedaani, W., & Ashraf, I. (2022). Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets Using Stacked Ensemble GCR-NN Model. doi:10.1109/ACCESS.2022.3144266

Ligthart, A., Catal, C., & Tekinerdogan, B. (2022). Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification. *Applied Soft Computing*.

doi:10.1016/j.asoc.2020.107023

Lopes Ros, R., Jordane De Silva, M., Henrique Silva, D., Ayub, M. S., Carrillo, D., Nardelli, P., & Zegarra Rodríguez, D. (2020). Event Detection System Based on User Behavior Changes in Online Social Networks: Case of the COVID-19 Pandemic. doi:10.1109/ACCESS.2020.3020391

Machová, K., Mach, M., & Adamišín, K. (2022). Machine Learning and Lexicon Approach to Texts

Processing in the Detection of Degrees of Toxicity in Online Discussions. doi:10.3390/s22176468

Mashraq, A., & Halawan, H. (2023). Dragonfly Optimization with Deep Learning Enabled Sentiment Analysis for Arabic Tweets. *Computer Systems Science and Engineering*.

doi:10.32604/csse.2023.031246

- Meedin, N., Caldera, M., Perera, S., & Perera, I. (2022). A Novel Annotation Scheme to Generate Hate Speech Corpus through Crowdsourcing and Active Learning. *International Journal of Advanced Computer Science and Applications(IJACSA)*. doi:10.14569/IJACSA.2022.0131146
- Mehmood, A., Farooq, M. S., Naseem, A., Rustam, F., Villar, M. G., Rodríguez, C. L., & Ashraf, I. (2022). Threatening URDU Language Detection from Tweets Using Machine Learning. doi:10.3390/app122010342
- Nijhawan, T., Attigeri, G., & Ananthakrishna, T. (2022). Stress detection using natural language processing and machine learning over social interactions. *Journal of Big Data*.
- NLTK :: Natural Language Toolkit. (n.d.). (s.f.).* Obtenido de <https://www.nltk.org/>
- Pathan, A., & Prakash, C. (2022). Cross-Domain Aspect Detection and Categorization using Machine Learning for Aspect-based Opinion Mining. *International Journal of Information Management Data Insights*. doi:10.1016/j.jjime.2022.100099
- Prakash, J. S., Kumar, A., Rana, N., & Dwivedi, Y. (2020). Attention-Based LSTM Network for Rumor Veracity Estimation of Tweets. *Information Systems Frontiers*. doi:10.1007/s10796-020-10040-5
- Rahman, H., Tariq, J., Masood, A., Subahi, A., Khalaf, O. I., & Alotaibi, Y. (2023). Multi-Tier Sentiment Analysis of Social Media Text Using Supervised Machine Learning. *Computers, Materials and Continua*. doi:10.32604/cmc.2023.033190
- Rahman, M., Hasan, M., Billah, M. M., & Sajuti, R. J. (2022). Political Fake News Detection from Different News Source on Social Media using Machine Learning Techniques. doi:10.53799/ajse.v21i1.383
- Renjith, S., Abraham, A., Jyothi, S., Chandran, L., & Thomson, J. (2022). An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. *Journal of King Saud University - Computer and Information Sciences*. doi:10.1016/j.jksuci.2021.11.010

- Roy, B., & Das, S. (2023). Perceptible sentiment analysis of students' WhatsApp group chats in valence, arousal, and dominance space. *Journal of King Saud University - Computer and Information Sciences*. doi:10.1007/s13278-022-01016-1
- Rubia , F., Shaik, N. S., Riaz, A., Ahmad, S., El-Affendi, M., Alyamani, K., . . . Latif, R. A. (2022). Applications of Intelligent Systems on Health and the Environment. doi:10.1155/2022/6561622
- Rupapara, V., Rustam, F., & Shahzad, H. F. (2021). Impact of SMOTE on Imbalanced Text Features for Toxic Comments Classification Using RVVC Model. doi:10.1109/ACCESS.2021.3083638
- Sadiq, S., Aljrees, T., & Ullah, S. (2023). Deepfake Detection on Social Media: Leveraging Deep Learning and FastText Embeddings for Identifying Machine-Generated Tweets. doi:10.1109/ACCESS.2023.3308515
- Salsabila, S., Romadhona, Y., Permataning Tyas, S. M., & Purwitasari, D. (2023). Aspect-based Sentiment and Correlation-based Emotion Detection on Tweets for Understanding Public Opinion of Covid-19. doi:10.20473/jisebi.9.1.84-94
- Studiawan, H., Sohel, F., & Payne, C. (2020). Anomaly Detection in Operating System Logs with Deep Learning-Based Sentiment Analysis. doi:10.1109/TDSC.2020.3037903
- Tesfagergish, S. G., Kapočiūtė-Dzikienė, J., & Damaševičius, R. (2022). Zero-Shot Emotion Detection for Semi-Supervised Sentiment Analysis Using Sentence Transformers and Ensemble Learning. doi:10.3390/app12178662
- Truică, C.-O., & Apostol, E.-S. (2023). It's All in the Embedding! Fake News Detection Using Document Embeddings. doi:10.3390/math11030508
- Varlamis, I., Michail, D., Glykou, F., & Tsantilas, P. (2022). A Survey on the Use of Graph Convolutional Networks for Combating Fake News. doi:10.3390/fi14030070

- Vyas, P., Vyas, G., & Dhiman, G. (2023). RUemo—The Classification Framework for Russia-Ukraine War-Related Societal Emotions on Twitter through Machine Learning. doi:10.3390/a16020069
- Wadhvani, G. K., Varshney, P. K., Gupta, A., & Kumar, S. (2023). Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models Using Twitter Data on Russia–Ukraine War. *SN Computer Science*. doi:10.1007/s42979-023-01790-5
- Waheeb, S. A., Khan, N. A., & Shang, X. (2022). Topic Modeling and Sentiment Analysis of Online Education in the COVID-19 Era Using Social Networks Based Datasets. doi:10.3390/electronics11050715
- Wieringa, R., Maiden, N., Mead, N., & Rolland, C. (2005). Requirements engineering paper classification and evaluation criteria: a proposal and a discussion.
- Zayno, M., & Radhi, A. M. (2022). Data Mining Methods for Extracting Rumors Using Social Analysis Tools. doi:10.24996/ijcs.2022.63.8.36
- Zhang, X., Li, W., Ying, H., Li, F., & Tang, S. (2020). Emotion Detection in Online Social Networks: A Multilabel Learning Approach. doi:10.1109/JIOT.2020.3004376