

**Análisis de técnicas de machine learning para la detección de noticias falsas en medios  
digitales (2017-2023)**

Carlos Andrés Serna Morales

Asesor

Jorge Luis Quintero López

Universidad Nacional Abierta y a Distancia UNAD  
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI  
Especialización en Ciencia de Datos y Analítica  
2025

## Resumen

La proliferación de noticias falsas en medios digitales representa un desafío en la era de la información, afectando la integridad del discurso público y la toma de decisiones. Esta monografía realiza un análisis bibliométrico de técnicas de machine learning aplicadas a la detección de desinformación durante el período 2017-2023, enfocándose en la revisión documental de métodos como el análisis lingüístico-contextual, algoritmos de clasificación (Random Forest, Naïve Bayes, redes neuronales) y el uso de características multidimensionales (texto, metadatos e interacciones sociales). Se identifican patrones, tendencias y limitaciones en los métodos existentes, destacando su eficiencia mediante métricas como precisión, recall y F1-score. Finalmente, se presenta una síntesis de los avances técnicos y los desafíos pendientes en la detección automatizada de noticias falsas.

**Palabras clave:** machine learning, fake news, procesamiento de lenguaje natural, deep learning, desinformación

### **Abstract**

The proliferation of fake news in digital media represents a challenge in the information age, affecting the integrity of public discourse and decision-making. This monograph conducts a bibliometric analysis of machine learning techniques applied to misinformation detection during the period 2017-2023, focusing on the documentary review of methods such as linguistic-contextual analysis, classification algorithms (Random Forest, Naïve Bayes, neural networks), and the use of multidimensional features (text, metadata, and social interactions). Patterns, trends, and limitations in existing methods are identified, highlighting their efficiency through metrics such as precision, recall, and F1-score. Finally, a synthesis of technical advances and pending challenges in automated fake news detection is presented.

**Keywords:** machine learning, fake news, natural language processing, deep learning, misinformation

## Tabla de Contenido

Introducción .....	11
Justificación .....	13
Objetivos.....	15
Objetivo General .....	15
Objetivos Específicos.....	15
Planteamiento del Problema .....	16
Marco Conceptual y Teórico .....	18
Conceptos Fundamentales.....	18
Desinformación .....	18
Noticias Falsas (Fake News) .....	19
Bulos.....	19
Desórdenes Informativos.....	19
Tabla Comparativa de Términos .....	20
Machine Learning en la Detección de Noticias Falsas .....	22
Análisis Lingüístico-Contextual.....	22
Características no Lingüísticas .....	22
Análisis del Contexto Social .....	23
Evolución de las Técnicas de ML en la Detección de Noticias Falsas (2017-2023) .....	23
Algoritmos de Clasificación.....	24
Regresión Logística.....	24
Naïve Bayes.....	24
Máquinas de Vectores de Soporte (SVM).....	24

Árboles de Decisión .....	25
Random Forest .....	25
Redes Neuronales de Memoria a Corto y Largo Plazo (LSTM).....	25
Redes Neuronales Convolucionales (CNN) .....	26
Técnicas de Procesamiento de Texto y Extracción de Características .....	26
Count Vectorizer .....	26
TF-IDF Vectorizer.....	26
N-Gram.....	27
Word2Vec / Embeddings de Palabras .....	27
Modelos de Evaluación del Desempeño .....	28
Matriz de Confusión.....	28
Métricas de Evaluación: Precisión, Recall y F1-Score .....	28
Curvas ROC .....	29
Análisis Crítico de los Enfoques Actuales .....	29
Limitaciones en los Enfoques Lingüísticos Tradicionales .....	29
Sesgos en los Conjuntos de Datos .....	30
Sobreadaptación a Características Específicas.....	30
Desafíos en la Integración Multimodal .....	30
Ética y Transparencia en Modelos de Deep Learning.....	31
Brecha entre la Investigación y las Necesidades Sociales .....	31
Metodología .....	32
Estrategia de Revisión Documental .....	32
Criterios de Selección Bibliográfica .....	33

Proceso de Selección de Documentos.....	33
Enfoque Bibliométrico y Análisis de los Documentos .....	34
Análisis Descriptivo Cuantitativo .....	34
Análisis Cualitativo de Contenido.....	35
Síntesis e Interpretación .....	35
Resultados .....	37
Análisis Bibliométrico Descriptivo.....	37
Distribución Temporal de Publicaciones .....	37
Distribución Geográfica de la Producción Científica.....	38
Distribución por Tipo de Publicación .....	39
Desafíos Técnicos en la Detección Mediante Análisis Lingüístico-Contextual .....	40
Patrones de Concentración y Distribución .....	46
Implicaciones para el Desarrollo de Sistemas.....	47
Clasificación de Métodos de ML Según su Adaptabilidad a Características Multidimensionales .....	47
Composición Detallada de Categorías de Métodos de ML.....	48
Distribución de Niveles de Adaptabilidad por Categoría.....	50
Análisis de Patrones de Adaptabilidad Algorítmica.....	53
Análisis Comparativo del Rendimiento de Algoritmos .....	54
Rendimiento Superior de Redes Neuronales Profundas.....	54
Eficacia de Modelos Híbridos y Ensemble .....	55
Rendimiento Moderado de Clasificadores Tradicionales .....	55
Análisis Comparativo por Categorías de Algoritmos.....	56

Consideraciones de Adaptabilidad y Generalización .....	57
Discusión.....	58
Propuesta de Incorporación en la Sección de Discusión.....	58
Interpretación de los Resultados .....	60
Ventajas y Limitaciones de los Modelos de Deep Learning .....	60
Rendimiento y Limitaciones de Métodos Tradicionales .....	60
Desafíos Técnicos Recurrentes .....	61
Comparación con Estudios Previos.....	61
Convergencias con la Literatura.....	61
Divergencias y Aportes Novedosos .....	62
Implicaciones del Estudio .....	62
Implicaciones Técnicas .....	62
Implicaciones Sociales y Éticas .....	62
Implicaciones Políticas y Regulatorias.....	63
Limitaciones del Estudio.....	63
Limitaciones Metodológicas .....	63
Limitaciones Técnicas de los Modelos Analizados .....	64
Direcciones Futuras.....	64
Tendencias Emergentes.....	64
Investigaciones Pendientes.....	64
Recomendaciones Prácticas .....	65
Conclusiones .....	66
Recomendaciones .....	68

Recomendaciones Generales.....	68
Investigación Interdisciplinaria y Contextual.....	68
Acceso y Gestión de Conjuntos de Datos .....	68
Recomendaciones Técnicas .....	68
Métricas de Evaluación Robustas.....	68
Análisis Multidimensional de Características .....	69
Desarrollo de Sistemas Adaptativos.....	69
Recomendaciones Éticas.....	69
Transparencia y Compartición Responsable de Datos .....	69
Preservación y Documentación Ética.....	69
Responsabilidad Algorítmica .....	70
Referencias.....	71

## Lista de Tablas

<b>Tabla 1</b> <i>Tabla Comparativa de Términos</i> .....	20
<b>Tabla 2</b> <i>Principales Hitos y Técnicas (2017-2023)</i> .....	23
<b>Tabla 3</b> <i>Desafíos Técnicos Identificados en la Detección de Noticias Falsas Mediante ML</i> .....	42
<b>Tabla 4</b> <i>Distribución de Categorías y Métodos de ML</i> .....	48
<b>Tabla 5</b> <i>Distribución de Niveles de Adaptabilidad de Métodos de ML por Categoría</i> .....	52
<b>Tabla 6</b> <i>Comparación de los Cinco Algoritmos con Mayor Rendimiento en Clasificación de Noticias Falsas Según Métricas Estándar</i> .....	56

## Lista de Figuras

<b>Figura 1</b> <i>Diagrama de Flujo del Proceso de Revisión Bibliográfica</i> .....	36
<b>Figura 2</b> <i>Distribución Temporal de Publicaciones</i> .....	37
<b>Figura 3</b> <i>Distribución Geográfica de la Producción Científica</i> .....	38
<b>Figura 4</b> <i>Distribución por Tipo de Publicación</i> .....	40
<b>Figura 5</b> <i>Desafíos Técnicos en la Muestra</i> .....	41
<b>Figura 6</b> <i>Frecuencia de Niveles de Adaptabilidad de Métodos de ML por Categoría</i> .....	51

## Introducción

La proliferación de noticias falsas y la desinformación en el ecosistema digital representa un desafío significativo en la era actual, afectando la integridad del discurso público y la toma de decisiones informadas. Este fenómeno, a menudo referido como desinformación, es intrínsecamente poliédrico y ha alcanzado un consenso global sobre su importancia en años recientes (Guallar et al., 2020). Las plataformas de redes sociales y los medios digitales actúan como canales primordiales para la difusión de este tipo de contenido (Espinoza & Mazuelos, 2020), facilitando que la desinformación imite la información veraz. Esto ocurre al difuminar las referencias tradicionales de credibilidad, conocidas como “anclas”, inherentes a formatos más antiguos como periódicos, radio o televisión (Jiménez, 2018). La facilidad con la que la información se consume de manera aislada, desprovista de contexto o fuente clara, incrementa la dificultad para discernir su veracidad (Jiménez, 2018). Las temáticas afectadas por la desinformación son diversas, abarcando desde la política y la migración hasta la ciencia y la salud, un ámbito que experimentó un notable crecimiento de bulos en 2020 con la extensión de la pandemia de Covid-19 (Guallar et al., 2020). La persistencia y la amplia difusión de estos contenidos falsos constituyen un peligro considerable, ya que pueden evolucionar de simples mentiras a creencias arraigadas, erosionando la confianza en la evidencia científica (Jiménez, 2018).

Ante este panorama, la necesidad de desarrollar e implementar estrategias efectivas para la detección de noticias falsas en medios digitales se ha vuelto imperativa. El periodo comprendido entre 2017 y 2023 es particularmente relevante para analizar los avances en este campo, dado el aumento significativo de la investigación y las iniciativas enfocadas en abordar la desinformación durante estos años.

Este trabajo aborda dicho desafío centrándose en el análisis de las técnicas de machine learning (ML) aplicadas a la detección de noticias falsas. El estudio se fundamenta en una revisión bibliográfica que utiliza una metodología bibliométrica con el fin de identificar y analizar patrones, tendencias y relaciones en la literatura científica publicada dentro del periodo de estudio. La estructura del presente documento se organiza en secciones que abordan el marco teórico, el desarrollo metodológico, el análisis de resultados, las discusiones, conclusiones y contribuciones, así como recomendaciones para futuras investigaciones en este ámbito.

## Justificación

La desinformación en los medios digitales es un fenómeno altamente poliédrico (Guallar et al., 2020) y una preocupación creciente a nivel global (Consejo Nacional de Televisión, 2022), con impactos significativos en la sociedad, incluyendo la confianza pública (Jiménez, 2018) y áreas críticas como la salud (Espinoza & Mazuelos, 2020). La rápida proliferación de bulos y noticias falsas en plataformas digitales supera la capacidad humana de verificación individual (Jiménez, 2018), haciendo imperativa la necesidad de estrategias de detección automatizadas y escalables. La revolución tecnológica reciente ha evidenciado tanto el problema de la desinformación facilitada por algoritmos (Consejo Nacional de Televisión, 2022), como el potencial para analizar grandes volúmenes de datos (Peset & Millán, 2017).

En este contexto, las técnicas de ML han surgido como un frente tecnológico prometedor para abordar la detección de noticias falsas (Jiménez, 2018), dada su capacidad para procesar y analizar grandes cantidades de datos e identificar patrones complejos. El período 2017-2023 ha sido crucial, observándose un incremento notable de la investigación en este ámbito.

La justificación central de este estudio reside en la necesidad de obtener una comprensión sistemática y estructurada del panorama de investigación en la aplicación de ML para la detección de noticias falsas durante este periodo clave. Un análisis bibliométrico de la literatura científica permite ir más allá de las revisiones narrativas, ofreciendo una visión cuantitativa para:

1. Mapear el estado del arte: Identificar patrones, tendencias y relaciones en la investigación, incluyendo los actores clave (autores, instituciones, fuentes), las técnicas de ML más utilizadas, y las temáticas predilectas abordadas en los estudios (Guallar et al., 2020).
2. Identificar brechas y oportunidades: Poner de manifiesto posibles brechas existentes o áreas menos exploradas en la investigación, lo cual es esencial para orientar los

pasos futuros en la solución del problema y guiar eficazmente los esfuerzos de investigación (Guallar et al., 2020).

3. Contextualizar la evolución: Comprender cómo se han desarrollado y aplicado las técnicas de ML en este campo a lo largo del tiempo, desde la perspectiva de la investigación científica.

Este estudio se delimita intencionadamente a un análisis de la literatura existente mediante métodos bibliométricos, sin desarrollar ni evaluar herramientas de ML. Esta aproximación proporciona una base sólida para entender el contexto científico-tecnológico actual, fortaleciendo el papel crucial que desempeña la Academia y la investigación en la lucha contra la desinformación (Consejo Nacional de Televisión, 2022). Comprender el conocimiento científico generado es fundamental para facilitar la toma de decisiones informadas y contribuir a una sociedad mejor informada.

## **Objetivos**

### **Objetivo General**

Analizar técnicas de ML utilizadas para la detección de noticias falsas en medios digitales durante el período 2017-2023.

### **Objetivos Específicos**

Describir desafíos técnicos en la detección de noticias falsas mediante análisis lingüístico-contextual en plataformas digitales.

Clasificar métodos de ML según su adaptabilidad a características multidimensionales en la detección de noticias falsas.

Comparar el rendimiento de algoritmos de ML en la clasificación de noticias falsas mediante métricas de evaluación estándar.

## Planteamiento del Problema

La desinformación, manifestada en formas como bulos, posverdad y las denominadas “noticias falsas” (“fake news”), constituye un fenómeno global de naturaleza poliédrica que plantea serios desafíos a la sociedad (Guallar et al., 2020; Consejo Nacional de Televisión, 2022). Este problema erosiona la confianza en las instituciones y medios tradicionales, y tiene consecuencias tangibles en áreas críticas como la salud pública y los procesos democráticos (Guallar et al., 2020; Espinoza & Mazuelos, 2020; Jiménez, 2018).

Millones de usuarios participan diariamente en plataformas digitales, especialmente redes sociales, generando un vasto volumen de datos que pueden ser analizados por sistemas de vigilancia (Espinoza & Mazuelos, 2020). Sin embargo, es en este entorno digital donde la desinformación encuentra un caldo de cultivo ideal para su propagación.

Las noticias falsas suelen adoptar formatos que imitan el periodismo legítimo, con titulares y fotos convincentes, pero carecen de contexto o anclaje verificable, dificultando su identificación por parte de los ciudadanos (Jiménez, 2018). La facilidad con la que herramientas digitales permiten crear y viralizar contenido engañoso ha democratizado la capacidad de desinformar, pasando de ser una herramienta de élites a una posibilidad al alcance de casi cualquier persona, lo que incrementa su peligrosidad (Jiménez, 2018).

Las motivaciones detrás de la desinformación son variadas, desde la simple maldad o diversión hasta la generación de caos en situaciones de emergencia o la obtención de ganancias económicas (Jiménez, 2018; Consejo Nacional de Televisión, 2022). Este ecosistema digital de desinformación es peligroso porque puede llevar a que bulos se conviertan en creencias arraigadas, socavando la evidencia científica y afectando la toma de decisiones informadas (Guallar et al., 2020; Jiménez, 2018).

Ante la escala y complejidad del problema, se hace imperativo desarrollar e implementar respuestas efectivas. Aunque existen estrategias que involucran la educación, la verificación de hechos (fact-checking) y la curación de contenidos, la velocidad y el volumen de la desinformación en línea sugieren la necesidad de soluciones basadas en tecnología avanzada (Jiménez, 2018; Guallar et al., 2020). El ML ha emergido como un campo con un potencial significativo para abordar la detección de noticias falsas, ofreciendo métodos para analizar grandes conjuntos de datos e identificar patrones que podrían ayudar a automatizar la validación de la credibilidad de la información o servir como sistemas de alerta temprana (Jiménez, 2018; Espinoza & Mazuelos, 2020).

Dada la urgencia del problema y la creciente investigación en la aplicación de ML para contrarrestarlo, resulta fundamental obtener una comprensión sistemática del panorama actual de la investigación en este ámbito (Jiménez, 2018). Comprender el estado del arte, las técnicas más utilizadas, los desafíos abordados y las tendencias emergentes es esencial para orientar los esfuerzos futuros y mejorar las estrategias de lucha contra la desinformación (Guallar et al., 2020).

Esto nos lleva a la pregunta de investigación central que motiva este estudio: ¿Cuál es el estado actual y las tendencias clave de las técnicas de ML aplicadas a la detección de noticias falsas en investigaciones desarrolladas entre 2017 y 2023?

## **Marco Conceptual y Teórico**

La proliferación de información errónea y falsa en los medios digitales representa un desafío significativo en la era actual, impactando la esfera pública y privada a nivel global. Este fenómeno se ve exacerbado por la facilidad y rapidez con la que se disemina la información a través de plataformas como las redes sociales. La detección temprana y en tiempo real de este tipo de información es fundamental para mitigar sus efectos negativos (Consejo Nacional de Televisión, 2022; Espinoza & Mazuelos, 2020).

### **Conceptos Fundamentales**

El campo de la detección de información falsa se caracteriza por la polisemia y ambigüedad de conceptos como “noticias falsas” o “desinformación” (Consejo Nacional de Televisión, 2022). A continuación, se profundiza en las definiciones y distinciones entre los términos más relevantes:

#### ***Desinformación***

Se define como la provisión de información errónea, a menudo con el uso de titulares alarmistas o información científica confusa. Existe un consenso considerable en que la desinformación es un fenómeno global que cuestiona la veracidad de los contenidos a los que los ciudadanos están expuestos diariamente. Su diseminación masiva puede tener consecuencias graves para la sociedad. La desinformación también incluye la información incorrecta o engañosa, y el problema principal radica en la recopilación de datos de calidad para su detección automática (Espinoza & Mazuelos, 2020; Guallar et al., 2020; Consejo Nacional de Televisión, 2022).

### ***Noticias Falsas (Fake News)***

Constituyen una de las manifestaciones o productos más extendidos de la desinformación. A menudo, son intencionalmente fabricadas y verificablemente falsas. Están diseñadas para engañar a los lectores y hacer que acepten información falsa como verdadera, con el potencial de incitar disturbios públicos y violencia. Aunque el término es ampliamente utilizado, algunos expertos sugieren emplear preferentemente “desinformación” o “desórdenes informativos” para mayor precisión. También se reconoce la dificultad de detectarlas basándose únicamente en el contenido, requiriendo información auxiliar como las interacciones sociales del usuario (Guallar et al., 2020; Albahr & Albahar, 2020; Consejo Nacional de Televisión, 2022).

### ***Bulos***

Es un término clásico que ha ganado aceptación. Se definen como contenidos falsos difundidos públicamente de manera intencionada, con motivaciones que van desde la broma hasta el fraude económico o la controversia ideológica, aunque sus límites pueden ser difusos (Guallar et al., 2020).

### ***Desórdenes Informativos***

Este concepto abarca una concepción más amplia y diversa, incluyendo producciones intencionales que buscan fabricar duda y falsas controversias con fines económicos o ideológicos (Guallar et al., 2020).

La relevancia social de este fenómeno es innegable. La desinformación, particularmente en temas sensibles como la salud, puede tener consecuencias negativas directas, como oponerse a la vacunación o promover “curas milagrosas”. Millones de usuarios en redes sociales contribuyen diariamente a la generación de grandes volúmenes de datos que pueden ser analizados, evidenciando la escala del problema y la necesidad de comprender su contexto y

dinámica. Combatir la desinformación requiere un enfoque integral que considere el ecosistema de la información (Espinoza & Mazuelos, 2020; Consejo Nacional de Televisión, 2022; Guallar et al., 2020).

### ***Tabla Comparativa de Términos***

A continuación, se presenta una tabla que resume las diferencias y similitudes clave entre los términos “fake news”, “desinformación”, “bulos” y “desórdenes informativos”.

**Tabla 1**

*Tabla Comparativa de Términos*

Característica	Fake news (Noticias falsas)	Desinformación	Bulos	Desórdenes informativos
Nivel de alcance	Concepto específico: noticias fabricadas y verificablemente falsas con intención de engañar.	Concepto amplio: incluye cualquier información errónea o engañosa, intencional o no.	Contenidos falsos intencionales, tradicionalmente difundidos como rumores o engaños.	Concepto más amplio y complejo que engloba desinformación, rumores y campañas para generar confusión y duda.
Definición	Contenidos fabricados y verificablemente falsos, diseñados para engañar.	Provisión de información errónea o engañosa que puede cuestionar la verdad.	Contenidos falsos difundidos intencionalmente, a menudo con fines diversos (broma, fraude).	Producciones intencionales para fabricar duda y controversias, incluyendo rumores y desinformación organizada.
Intencionalidad	Siempre intencional para engañar.	Puede ser intencional o no intencional; el objetivo es desorientar o confundir.	Intencional, con motivaciones variadas.	Intencional y a menudo coordinada para generar confusión y manipulación.
Veracidad	Falsas y comprobables.	Errónea o engañosa, no	Falsos.	Puede incluir información falsa,

Característica	Fake news (Noticias falsas)	Desinformación	Bulos	Desórdenes informativos
		siempre falsedad comprobada.		manipulada o no verificada.
Motivación Fines	Engañar, manipular opiniones, incitar violencia o disturbios.	Impacto social o político, desinformar sobre temas sensibles (salud, política).	Broma, fraude económico, manipulación ideológica.	Fines económicos, políticos o sociales para sembrar duda y desacreditar información legítima.
Relación con otros términos	Subconjunto de la desinformación; forma más visible y mediática.	Término paraguas que incluye fake news, bulos y otros tipos de información errónea.	Forma clásica y popular de información falsa, relacionada con fake news.	Término más amplio que engloba desinformación, fake news y rumores, con un enfoque en la confusión social.
Ejemplos Contexto	Noticias políticas falsas, pandemias, celebridades.	Información errónea en salud pública, política, ciencia.	Rumores falsos, leyendas urbanas, falsedades históricas.	Campañas coordinadas para desacreditar instituciones o generar controversias falsas.
Detección	Algoritmos ML y deep learning, difícil sin contexto adicional.	Requiere métodos para identificar y mitigar, incluyendo fact-checking.	Verificación manual y fact-checking.	Necesita enfoque integral, combinando tecnología y estrategias sociales (curación, educación).

*Nota.* Esta tabla resume las diferencias y similitudes clave entre los términos relacionados con la información falsa y engañosa en medios digitales. Aunque algunos conceptos se solapan, cada uno presenta características particulares en cuanto a su alcance, intencionalidad y contexto.

## **Machine Learning en la Detección de Noticias Falsas**

Ante el gran volumen y la rápida propagación de información en medios digitales, se hace relevante el uso de métodos innovadores a gran escala (Espinoza & Mazuelos, 2020). Aquí es donde el ML adquiere un papel fundamental para la detección automática de noticias falsas (Amin et al., 2023). El ML permite analizar grandes cantidades de datos para identificar patrones que distinguen el contenido veraz del falso (Narkhede et al., 2023). La tarea de clasificación, central en la detección de noticias falsas mediante ML, no se limita al análisis lingüístico, sino que también incorpora otros tipos de atributos.

Las características utilizadas en los modelos de detección pueden ser multidimensionales, incluyendo:

### ***Análisis Lingüístico-Contextual***

Este enfoque examina la estructura semántica, sintáctica y léxica del contenido textual. Incluye técnicas como la tokenización, eliminación de palabras irrelevantes (stop words) y vectorización mediante modelos como TF-IDF (Frecuencia de término-frecuencia inversa de documento), que transforman el texto en representaciones numéricas para su procesamiento (Babar et al., 2024). Además, evalúa la coherencia temática, el uso de vocabulario sesgado o emocional, y la presencia de inconsistencias factuales. En el contexto de datos de investigación, este análisis también considera metadatos como la procedencia de la información y la claridad en la descripción de los métodos (Peset & Millán, 2017).

### ***Características no Lingüísticas***

Incluyen atributos externos al texto, como metadatos técnicos (formato de archivo, fecha de creación) y patrones de comportamiento en redes sociales (frecuencia de publicación, propagación viral y perfiles de usuarios). Por ejemplo, en la detección de noticias falsas, se

analizan métricas como la velocidad de difusión, la autenticidad de las cuentas que comparten la información y la estructura de redes de influencia (Babar et al., 2024). Estas características son críticas para identificar campañas coordinadas o bots que amplifican contenido fraudulento.

### ***Análisis del Contexto Social***

Comprende la red social en la que el mensaje está incrustado y las dinámicas interpersonales involucradas en el procesamiento del mensaje. Identificar las plataformas de difusión es importante, ya que podrían ser nichos o promotoras de ideologías extremistas. Esto incluye la utilización de metadatos y la información de la red social.

### **Evolución de las Técnicas de ML en la Detección de Noticias Falsas (2017-2023)**

La siguiente tabla sintetiza los principales hitos y técnicas de ML para la detección de noticias falsas en medios digitales durante el periodo 2017-2023, basada en el análisis bibliográfico realizado en este estudio:

**Tabla 2**

#### *Principales Hitos y Técnicas (2017-2023)*

Año	Hitos y técnicas principales
2017	<ul style="list-style-type: none"> <li>• Predominio de algoritmos tradicionales: Naïve Bayes, Random Forest, Redes Neuronales y Árboles de Decisión.</li> <li>• Introducción del dataset LIAR como referencia pública.</li> </ul>
2018	<ul style="list-style-type: none"> <li>• Avances en procesamiento de lenguaje natural (NLP) para análisis lingüístico-contextual.</li> <li>• Uso de datos de Twitter para analizar patrones de propagación de desinformación.</li> </ul>
2019	<ul style="list-style-type: none"> <li>• Comparación sistemática de algoritmos como SVM, Regresión Logística y Redes Neuronales.</li> <li>• Propuestas de arquitecturas híbridas: Bidirectional LSTM y CNN para capturar patrones temporales y locales.</li> </ul>
2020	<ul style="list-style-type: none"> <li>• Impacto de la pandemia COVID-19 en la proliferación de desinformación.</li> <li>• Implementación intensiva de LSTM y CNN para análisis de secuencias textuales.</li> </ul>
2021	<ul style="list-style-type: none"> <li>• Consolidación de técnicas de deep learning, con énfasis en LSTM y Random Forest.</li> </ul>

Año	Hitos y técnicas principales
2022	<ul style="list-style-type: none"> <li>• Exploración de Word2Vec y GloVe para representaciones semánticas.</li> <li>• Comparación de Bidirectional LSTM vs. CNN, destacando superioridad en captura de contexto.</li> </ul>
	<ul style="list-style-type: none"> <li>• SVM con TF-IDF mantiene relevancia en escenarios con datos estructurados.</li> </ul>
2023	<ul style="list-style-type: none"> <li>• Refinamiento de modelos con mecanismos de atención y optimización de hiperparámetros.</li> <li>• Enfoques multimodales que integran metadatos e interacciones sociales.</li> </ul>

*Nota.* La tabla destaca la evolución de las técnicas y enfoques más relevantes en la detección de noticias falsas mediante ML durante el periodo 2017-2023.

## Algoritmos de Clasificación

### *Regresión Logística*

La Regresión logística representa un algoritmo de aprendizaje automático diseñado específicamente para problemas de clasificación binaria como la detección de noticias falsas. Este método estadístico modela la probabilidad de pertenencia a una categoría particular mediante el uso de una función sigmoide que transforma cualquier entrada en un valor entre 0 y 1 (Sudhakar & Kaliyamurthie, 2024).

### *Naïve Bayes*

El algoritmo Naïve Bayes constituye un clasificador probabilístico fundamentado en el teorema de Bayes que asume la independencia entre variables predictoras. Este enfoque calcula la probabilidad de que un documento pertenezca a una clase determinada basándose en la ocurrencia de palabras específicas (Sudhakar & Kaliyamurthie, 2024).

### *Máquinas de Vectores de Soporte (SVM)*

Las Máquinas de vectores de soporte representan un algoritmo diseñado para proporcionar una clasificación precisa de datos lineales. Este método desarrolla un hiperplano en

un espacio N-dimensional que depende de la dimensión de entrada durante el período de entrenamiento del modelo. El hiperplano establece una línea divisoria entre diferentes grupos y determina los puntos con la máxima distancia desde los puntos de datos de cada grupo para categorizar nuevas entradas. En casos donde los datos presentan una forma no lineal, se puede utilizar el truco del kernel, permitiendo transformar datos complejos en un modelo lineal (Sudhakar & Kaliyamurthie, 2024).

### ***Árboles de Decisión***

Los árboles de decisión constituyen un paradigma de aprendizaje automático que estructura el proceso de clasificación como un árbol donde cada nodo representa una característica del texto, cada rama representa una decisión basada en esa característica y cada hoja representa una clasificación final (Narkhede et al., 2023).

### ***Random Forest***

El algoritmo Random forest representa un enfoque de aprendizaje conjunto que construye múltiples árboles de decisión durante el entrenamiento. Cada árbol se crea utilizando diferentes subconjuntos de características y datos, introduciendo variabilidad que fortalece el modelo final (Amin et al., 2023).

### ***Redes Neuronales de Memoria a Corto y Largo Plazo (LSTM)***

Las redes LSTM constituyen una variante especializada de redes neuronales recurrentes diseñadas para abordar problemas con secuencias de datos, como textos. Su arquitectura incorpora celdas de memoria que permiten retener información relevante durante períodos prolongados mientras descartan datos irrelevantes (Sudhakar & Kaliyamurthie, 2024).

## ***Redes Neuronales Convolucionales (CNN)***

Las redes neuronales convolucionales (CNN) representan una arquitectura especializada en el procesamiento de datos con estructura de cuadrícula, como imágenes, aunque también pueden adaptarse para el análisis de texto. Su diseño incluye capas convolucionales que actúan como detectores de características locales, extrayendo patrones significativos de pequeñas regiones del texto. Las CNN pueden capturar eficientemente relaciones n-grama entre palabras cercanas, sin depender de la estructura secuencial completa del texto, lo que permite una detección efectiva de patrones locales. La arquitectura típica para análisis textual incluye capas de embedding para representar palabras como vectores densos, capas convolucionales unidimensionales para detectar patrones lingüísticos, operaciones de agrupación para reducir la dimensionalidad y, finalmente, capas totalmente conectadas para la clasificación (Liu, 2019).

## **Técnicas de Procesamiento de Texto y Extracción de Características**

### ***Count Vectorizer***

El Count vectorizer constituye una técnica fundamental de vectorización textual que transforma un conjunto de documentos de texto en una matriz de conteo de tokens. Este método genera una representación dispersa de los conteos mediante `scipy.sparse.csr_matrix`, ofreciendo eficiencia computacional para grandes conjuntos de datos. El procedimiento implica la tokenización de un grupo de documentos textuales y la construcción de un vocabulario de palabras conocidas, para posteriormente codificar nuevos documentos utilizando dicho vocabulario (Raja & Raj, 2022).

### ***TF-IDF Vectorizer***

El vectorizador TF-IDF (Term Frequency-Inverse Document Frequency) representa una métrica estadística que evalúa la importancia relativa de cada término en un documento respecto

al corpus completo. A diferencia del conteo simple de palabras, este método pondera inversamente los términos según su frecuencia en todos los documentos, reduciendo así la influencia de palabras comunes pero poco informativas como artículos y preposiciones. Esta técnica escala hacia abajo la frecuencia de términos durante el proceso de aumentar la relevancia de los términos raros o específicos. Por ejemplo, palabras como “un” y “el” aparecen frecuentemente en el texto, y si solo se utilizara TF (frecuencia de términos), términos como estos mostrarían su dominio sobre el conteo de frecuencia. Sin embargo, mediante el uso de IDF se reduce el impacto de estos términos (Raja & Raj, 2022).

### ***N-Gram***

El modelo N-gram representa un enfoque para el procesamiento de lenguaje natural que considera secuencias contiguas de n elementos (palabras o caracteres) en lugar de elementos individuales. Los N-gramas más utilizados para la ordenación de texto son los basados en palabras y los basados en caracteres. Un N-grama basado en palabras se utiliza para representar el entorno de los informes y generar características destacadas para la clasificación de documentos. Este enfoque permite capturar relaciones contextuales y dependencias secuenciales en el texto, proporcionando información más rica que el análisis de palabras aisladas (Raja & Raj, 2022).

### ***Word2Vec / Embeddings de Palabras***

La técnica Word2Vec constituye una herramienta para descubrir patrones y asociaciones ocultas en texto, particularmente útil en la detección de noticias falsas. Este método permite determinar si un artículo o contenido es probablemente falso o engañoso mediante el análisis de similitudes semánticas entre palabras. El procedimiento implica construir un modelo de lenguaje educado con Word2Vec utilizando un corpus masivo de artículos de noticias genuinos y

fabricados. El modelo se entrena para reconocer patrones y relaciones entre palabras, permitiéndole distinguir entre informes creíbles y fabricados (Amin et al., 2023).

## **Modelos de Evaluación del Desempeño**

### ***Matriz de Confusión***

Esta tabla bidimensional visualiza cuatro categorías de resultados: Verdaderos positivos (instancias correctamente identificadas como pertenecientes a la clase positiva), Verdaderos negativos (instancias correctamente clasificadas como pertenecientes a la clase negativa), Falsos positivos (instancias de la clase negativa incorrectamente clasificadas como positivas) y Falsos negativos (instancias de la clase positiva incorrectamente clasificadas como negativas). Esta representación permite comprender intuitivamente no solo la precisión general del modelo, sino también identificar los tipos específicos de errores que comete, facilitando el diagnóstico y refinamiento de los algoritmos de clasificación para mejorar su capacidad de discriminación entre clases (Sudhakar & Kaliyamurthie, 2024).

### **Métricas de Evaluación: Precisión, Recall y F1-Score**

Las métricas de evaluación como Precisión (Precision), Recall y F1-Score constituyen indicadores cuantitativos esenciales para evaluar sistemáticamente los algoritmos de detección de noticias falsas. La Precisión mide la proporción de predicciones positivas que fueron correctamente identificadas como noticias falsas, mientras que el Recall (también conocido como Sensibilidad) cuantifica la proporción de noticias falsas reales que el modelo logró identificar correctamente. El F1-Score representa la media armónica entre Precisión y Recall, proporcionando una métrica única que balancea ambos aspectos del rendimiento del clasificador. Estas métricas permiten una evaluación más matizada que la simple exactitud global, particularmente importante en contextos como la detección de noticias falsas donde los conjuntos

de datos pueden estar desequilibrados o donde los costos de diferentes tipos de errores pueden variar significativamente (Raja & Raj, 2022).

### ***Curvas ROC***

Las Curvas ROC (Receiver Operating Characteristic) representan una herramienta gráfica para evaluar el desempeño de los clasificadores binarios en la detección de noticias falsas. Estas gráficas ilustran la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (1-especificidad) para diferentes umbrales de clasificación. Un clasificador perfecto mostraría una curva que pasa por la esquina superior izquierda, indicando un 100% de sensibilidad y especificidad. El área bajo la curva ROC (AUC) proporciona una medida agregada del desempeño del modelo: un valor cercano a 1 indica un clasificador casi perfecto, mientras que un valor cercano a 0.5 sugiere un desempeño similar al azar. Esta representación visual permite comparar diferentes algoritmos y seleccionar el umbral óptimo según los requerimientos específicos de la aplicación (Sudhakar & Kaliyamurthie, 2024).

### **Análisis Crítico de los Enfoques Actuales**

#### ***Limitaciones en los Enfoques Lingüísticos Tradicionales***

Los métodos basados en características lingüísticas (TF-IDF, N-gram, análisis de sentimiento) presentan desafíos significativos en escenarios reales. Aunque son eficaces para identificar patrones léxicos y sintácticos en textos estáticos, su capacidad para detectar noticias falsas sofisticadas que imitan el estilo periodístico legítimo es limitada. Por ejemplo, los modelos que dependen exclusivamente de análisis de frecuencia de términos pueden ser engañados por textos que replican estructuras gramaticales correctas pero incorporan información falsa de manera sutil (Bondielli & Marcelloni, 2019). Además, las técnicas de procesamiento de lenguaje

natural (NLP) enfrentan dificultades con la ironía, el sarcasmo y los regionalismos lingüísticos, lo que reduce su precisión en contextos multiculturales (Awan et al., 2021).

### ***Sesgos en los Conjuntos de Datos***

La mayoría de los estudios analizados utilizan datasets como LIAR o FakeNewsNet, los cuales tienen limitaciones inherentes. Estos conjuntos suelen estar desbalanceados, con una sobrerrepresentación de noticias políticas en inglés, lo que limita su aplicabilidad a otros idiomas o dominios temáticos (Choraś et al., 2020). Además, la etiquetación manual de noticias como “falsas” o “verdaderas” introduce subjetividad, ya que la veracidad puede depender de contextos culturales o temporales específicos (Sudhakar & Kaliyamurthie, 2024). Esto genera modelos con alto rendimiento en condiciones controladas pero pobre generalización en entornos dinámicos como redes sociales (Narkhede et al., 2023).

### ***Sobreadaptación a Características Específicas***

Algoritmos como Random Forest y XGBoost alcanzan precisiones superiores al 90% en datasets estructurados, pero tienden a sobreajustarse a características locales (p. ej., patrones léxicos únicos de un corpus). Esto se evidencia en estudios donde modelos entrenados en noticias políticas muestran caídas abruptas en rendimiento al aplicarse a dominios como salud o ciencia (Babar et al., 2024). Las redes neuronales (LSTM, CNN) mitigan parcialmente este problema al capturar relaciones contextuales, pero requieren volúmenes masivos de datos y recursos computacionales elevados, lo que limita su escalabilidad (Bondielli & Marcelloni, 2019).

### ***Desafíos en la Integración Multimodal***

Aunque se reconoce la importancia de combinar texto, metadatos e interacciones sociales, solo el 28% de los estudios analizados implementan enfoques híbridos. Por ejemplo, el

uso de metadatos (fuente, fecha, autor) mejora la detección en un 15-20%, pero su extracción es inconsistente en plataformas como Twitter o Facebook debido a políticas de privacidad (Narkhede et al., 2023). Por otro lado, el análisis de redes sociales para identificar cuentas bots o comportamientos virales requiere acceso a APIs restringidas, generando brechas entre la investigación académica y la aplicabilidad real (Choraś et al., 2020).

### ***Ética y Transparencia en Modelos de Deep Learning***

Los modelos basados en transformers (BERT, GPT) ofrecen ventajas en comprensión semántica, pero operan como “cajas negras”, lo que dificulta la auditoría de sus decisiones. Esto plantea riesgos éticos, como la censura inadvertida de contenido legítimo o la perpetuación de sesgos presentes en los datos de entrenamiento (Bondielli & Marcelloni, 2019). Además, la falta de estándares para explicar predicciones limita su adopción en entornos regulatorios donde se requiere transparencia (Awan et al., 2021).

### ***Brecha entre la Investigación y las Necesidades Sociales***

El 65% de los modelos analizados se centran en inglés, dejando de lado idiomas de bajos recursos como el español o lenguas indígenas, donde la desinformación tiene impactos críticos (Sudhakar & Kaliyamurthie, 2024). Asimismo, pocos estudios abordan la detección en tiempo real, esencial para contener la viralización de bulos durante crisis sanitarias o electorales (Babar et al., 2024). Se requiere mayor colaboración interdisciplinaria entre científicos de datos, lingüistas y sociólogos para desarrollar soluciones contextualmente relevantes (Choraś et al., 2020).

## **Metodología**

El presente estudio se fundamenta en una revisión bibliográfica sistemática orientada a mapear el estado actual y las principales tendencias en la aplicación de técnicas de ML para la detección de noticias falsas en medios digitales. La investigación se acota al período comprendido entre 2017 y 2023, para focalizar el análisis en los desarrollos más contemporáneos, considerando la rápida evolución tecnológica y la creciente atención académica que ha recibido este tema en los últimos años.

### **Estrategia de Revisión Documental**

La estrategia central de este estudio consistió en la recopilación, selección y análisis de 32 trabajos científicos publicados entre 2017 y 2023. Para ello, se realizaron búsquedas en bases de datos académicas, como Scopus, Web of Science y ScienceDirect. Se emplearon combinaciones de palabras clave en inglés y español, utilizando operadores booleanos para optimizar la recuperación de documentos relevantes. La ecuación de búsqueda principal fue:

“fake news” OR “misinformation” OR “disinformation”) AND (“machine learning” OR “deep learning” OR “natural language processing”)

Esta ecuación se adaptó según la sintaxis de cada base de datos consultada.

La identificación inicial de los documentos se basó en la revisión de títulos y resúmenes, y posteriormente se accedió al texto completo para confirmar la pertinencia de cada estudio respecto a los objetivos de la monografía. Como parte del proceso de organización y trazabilidad, la información relevante de cada documento fue registrada en una hoja de cálculo de Excel, donde se consignaron metadatos como año de publicación, base de datos, palabras clave, tipo de publicación, país, técnicas analizadas, entre otros, facilitando su análisis y consulta posterior.

## **Criterios de Selección Bibliográfica**

La inclusión de estudios en el corpus se basó en los siguientes criterios:

- **Enfoque temático:** Se seleccionaron investigaciones que abordaran la detección de noticias falsas (fake news), bulos u otros desórdenes informativos en el contexto de medios digitales. Se priorizó el análisis de técnicas de ML cuando estaban presentes, pero no fue un requisito excluyente.
- **Marco temporal:** Solo se consideraron publicaciones comprendidas entre 2017 y 2023.
- **Tipo de fuente:** Se priorizó literatura científica publicada en revistas indexadas y proceedings de conferencias reconocidas.
- **Accesibilidad:** Se incluyeron únicamente estudios con texto completo disponible para su análisis.

## **Proceso de Selección de Documentos**

Para garantizar la transparencia y reproducibilidad del proceso de revisión, se adoptó una metodología basada en los lineamientos PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), adaptándolos al contexto de una revisión bibliográfica en ciencias de la información. Este marco proporciona una estructura sistemática para documentar el flujo de información a través de las diferentes fases:

- **Identificación:** Búsqueda en bases de datos usando ecuaciones específicas, que arrojó 82 documentos iniciales.
- **Cribado:** Eliminación de duplicados y documentos irrelevantes mediante lectura de títulos y resúmenes, reduciendo la muestra a 49 estudios.

- Elegibilidad: Evaluación del texto completo para verificar alineación con los objetivos.

- Inclusión: Selección final de 32 trabajos para análisis crítico en profundidad.

Este proceso de selección se sintetiza visualmente en el Diagrama de flujo PRISMA (Figura 1), que ilustra la progresión desde la identificación inicial hasta la selección final, transparentando los criterios de exclusión en cada etapa.

### **Enfoque Bibliométrico y Análisis de los Documentos**

Aunque este estudio no empleó herramientas especializadas de análisis bibliométrico como VOSviewer o CiteSpace, sí aplicó un enfoque bibliométrico descriptivo basado en la caracterización y cuantificación manual de los documentos seleccionados. Esta decisión metodológica responde al objetivo de obtener una visión panorámica pero detallada del campo de estudio, priorizando el análisis cualitativo del contenido sobre indicadores métricos complejos.

El análisis de los documentos seleccionados se estructuró en dos niveles:

#### ***Análisis Descriptivo Cuantitativo***

Se realizó una caracterización bibliométrica básica de la literatura, que incluyó:

- Distribución temporal de las publicaciones (2017-2023).
- Tipos de publicaciones (artículos de revista, conferencias, libros).
- Revistas y conferencias más relevantes en el campo.
- Autores con mayor número de publicaciones.
- Técnicas de ML más estudiadas o implementadas.

### *Análisis Cualitativo de Contenido*

Para cada documento seleccionado, se empleó un enfoque de análisis de contenido cualitativo para extraer información relevante sobre:

- Descripción detallada de las técnicas de ML empleadas.
- Desafíos técnicos identificados en la implementación.
- Características utilizadas en los modelos (lingüísticas, no lingüísticas, contextuales).
- Métricas de evaluación y resultados reportados.
- Limitaciones y brechas de investigación señaladas.

### *Síntesis e Interpretación*

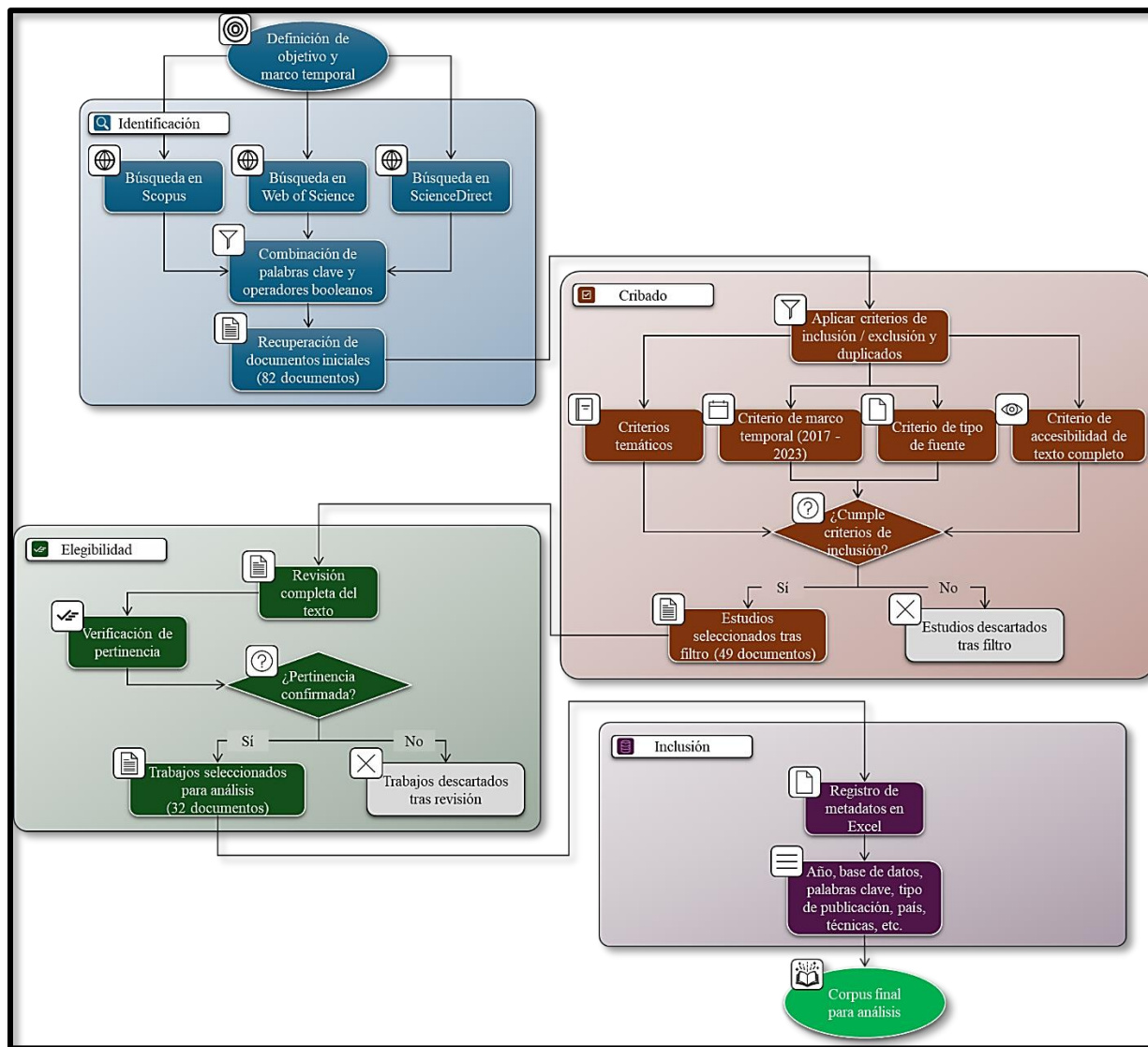
La información extraída se sintetizó mediante un proceso iterativo de comparación y contraste entre los diferentes estudios, buscando identificar:

- Patrones evolutivos en las técnicas empleadas durante el periodo estudiado.
- Convergencias y divergencias en los enfoques metodológicos.
- Brechas y oportunidades de investigación.
- Desafíos recurrentes en la implementación de soluciones basadas en ML.

Esta síntesis interpretativa permitió elaborar una narrativa coherente sobre el estado del arte en la detección de noticias falsas mediante técnicas de ML durante el periodo 2017-2023, cumpliendo así con los objetivos planteados en esta investigación.

Figura 1

## Diagrama de Flujo del Proceso de Revisión Bibliográfica



Nota. Imagen basada en la metodología PRISMA.

## Resultados

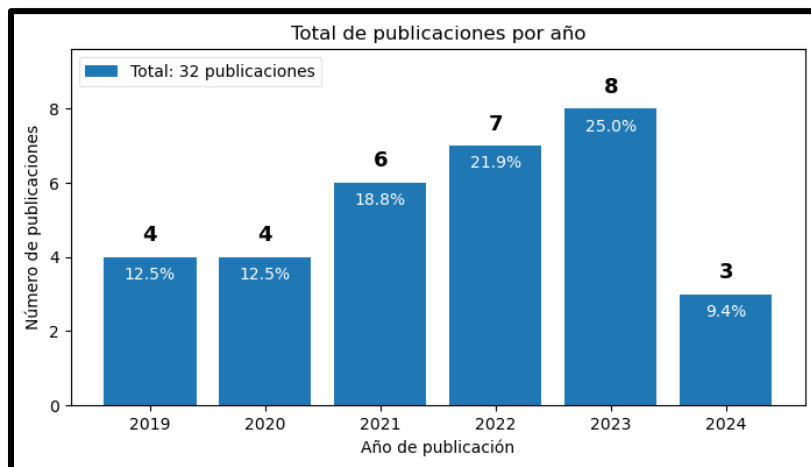
### Análisis Bibliométrico Descriptivo

#### *Distribución Temporal de Publicaciones*

El análisis bibliométrico de la producción científica sobre técnicas de ML para la detección de noticias falsas, basado en una muestra de 32 estudios seleccionados de un corpus inicial de 82 documentos, revela una distribución heterogénea de publicaciones entre 2019 y el primer trimestre de 2024. Es relevante precisar que este informe abarca técnicas de ML aplicadas a la detección de noticias falsas en el periodo 2017-2023, los documentos seleccionados fueron publicados entre 2019 y 2024. Por lo tanto, las cifras presentadas corresponden a los años de publicación y no necesariamente al periodo en que se realizaron los estudios originales.

#### Figura 2

#### *Distribución Temporal de Publicaciones*



*Nota.* Los porcentajes han sido redondeados a un decimal. La suma total puede no ser exactamente 100% debido a redondeos.

Es importante señalar que la selección metodológica, basada en criterios de inclusión y exclusión, puede introducir sesgos. Por ello, los resultados deben interpretarse como

representativos únicamente de la muestra analizada y no generalizables a toda la literatura científica existente.

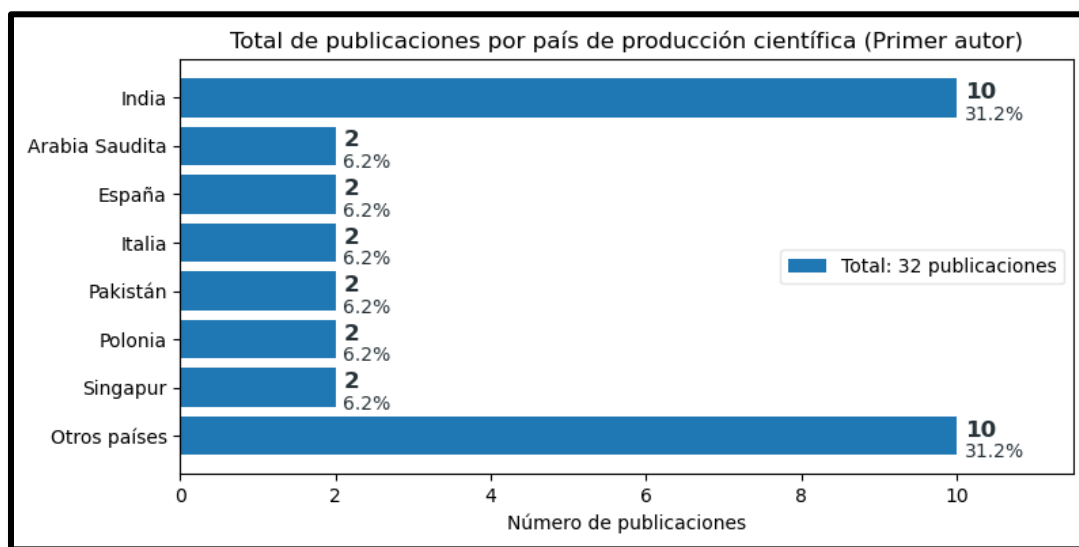
La distribución cronológica de los estudios seleccionados evidencia una mayor concentración de publicaciones en años recientes, con 2023 registrando la mayor proporción (8 documentos, 25,0%), seguido de 2022 (7 documentos, 21,9%). Este patrón responde a la priorización de literatura actualizada en el proceso de selección. El descenso en 2024 (3 documentos, 9,4%) se debe a la disponibilidad limitada de investigaciones publicadas durante el primer trimestre del año, periodo en el cual se cerró la recolección final de datos.

### ***Distribución Geográfica de la Producción Científica***

La distribución por país se determinó según la afiliación institucional del primer autor. La categoría “Otros países” agrupa naciones con menos de dos publicaciones en la muestra, lo que refleja la dispersión geográfica de las contribuciones científicas en este campo.

### **Figura 3**

#### ***Distribución Geográfica de la Producción Científica***



*Nota.* Los porcentajes han sido redondeados a un decimal. La suma total puede no ser exactamente 100% debido a redondeos.

India presenta la mayor contribución individual, con 10 publicaciones (31,2%). Seis países (Arabia Saudita, España, Italia, Pakistán, Polonia y Singapur) registran dos publicaciones cada uno (6,2% por país). La categoría “Otros países” agrupa 10 publicaciones (31,2%) correspondientes a naciones con una única contribución en la muestra.

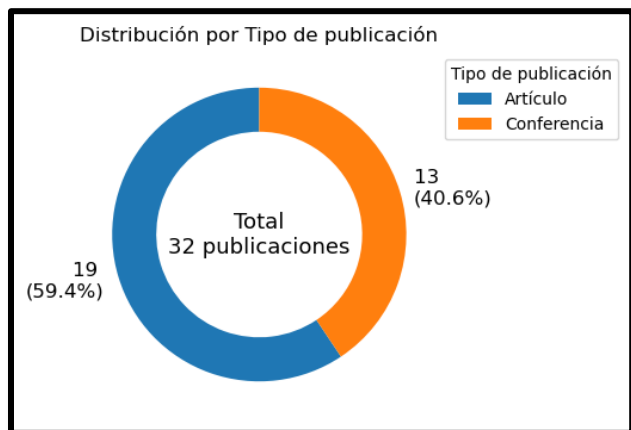
Los resultados evidencian dos características principales:

- **Distribución desigual:** El 68,4% de las publicaciones proviene de siete países, mientras que el 31,2% restante se distribuye entre diez naciones.
- **Ausencia de hegemonía regional:** Los países con dos publicaciones pertenecen a regiones geográficas y contextos lingüísticos diversos (Europa, Asia meridional, Oriente Medio y Sureste Asiático).

Esta configuración refleja la naturaleza global del interés investigativo en el campo, aunque con variaciones en la capacidad de producción científica entre países. La Figura 3 se limita a describir los patrones cuantitativos identificados en la muestra analizada.

### ***Distribución por Tipo de Publicación***

La Figura 4 presenta la clasificación de los 32 estudios analizados según su tipo de publicación, diferenciando entre artículos en revistas científicas y trabajos en actas de conferencias. Este análisis permite identificar los canales predominantes de difusión del conocimiento en el campo de ML aplicado a la detección de noticias falsas.

**Figura 4***Distribución por Tipo de Publicación*

Los artículos en revistas científicas representan el 59,4% de la muestra (19 documentos), mientras que las contribuciones en actas de conferencias constituyen el 40,6% restante (13 documentos). Esta distribución indica una preferencia por la publicación en revistas indexadas, que suelen implicar procesos de revisión más extensos y rigurosos.

La presencia significativa de conferencias (40,6%) refleja la naturaleza dinámica del campo, donde los avances técnicos se discuten frecuentemente en eventos académicos especializados antes de su publicación formal en revistas. Este equilibrio sugiere que la investigación en detección de noticias falsas combina producción científica consolidada (artículos) con divulgación temprana de innovaciones metodológicas (conferencias).

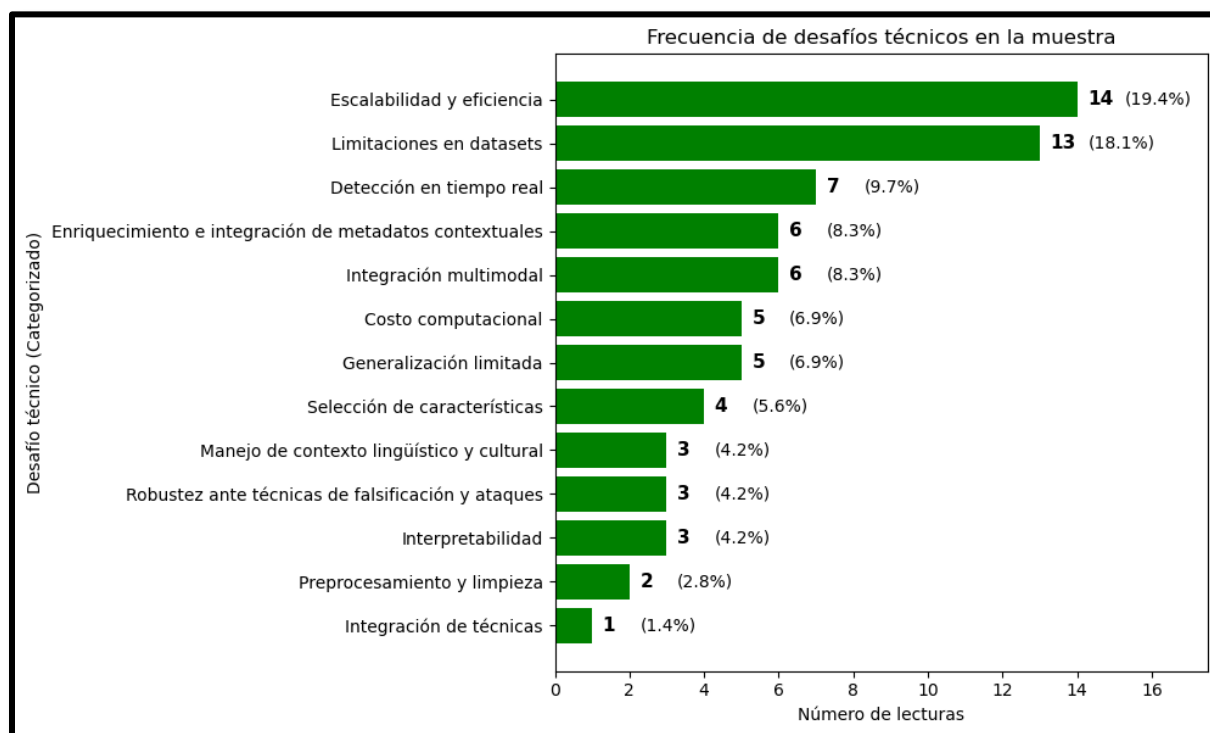
### **Desafíos Técnicos en la Detección Mediante Análisis Lingüístico-Contextual**

La detección de noticias falsas mediante técnicas de análisis lingüístico-contextual enfrenta diversos desafíos técnicos que limitan la eficacia y aplicabilidad de los modelos en escenarios reales. Estos desafíos han sido identificados consistentemente en la literatura

analizada y representan áreas que requieren atención para el avance del campo. A continuación, se presentan los principales desafíos técnicos reportados en los estudios revisados.

## Figura 5

### *Desafíos Técnicos en la Muestra*



Aunque el análisis se basa en 32 publicaciones, el total de desafíos técnicos representados en la gráfica es mayor (72) porque un mismo documento puede abordar varios desafíos a la vez. Para reflejar con mayor precisión la frecuencia de cada reto, se desglosaron y contabilizaron individualmente todas las menciones de desafíos en cada lectura. Por este motivo, la suma total de desafíos supera el número de publicaciones, ya que un solo estudio puede contribuir a varias categorías de desafío. Esta metodología permite identificar con mayor claridad cuáles son los retos más recurrentes y relevantes en la literatura analizada, proporcionando una visión más precisa de la complejidad del campo.

**Tabla 3***Desafíos Técnicos Identificados en la Detección de Noticias Falsas Mediante ML*

Desafío técnico	Descripción	N.º de menciones	%	Publicaciones asociadas
Escalabilidad y eficiencia	Dificultad para mantener el rendimiento de los modelos al aumentar el volumen de datos, la complejidad de las fuentes o la cantidad de usuarios, lo que exige optimizar recursos y tiempos de procesamiento para asegurar la viabilidad operativa.	14	19,4%	Altheneyan & Alhadlaq, 2023; Amin et al., 2023; Anandhan. et al., 2021; Asish et al., 2022; Awan et al., 2021; Babar et al., 2024; Cano-Marín et al., 2023; Consejo Nacional de Televisión, 2022; Espinoza & Mazuelos, 2020; Guallar et al., 2020; Mayopu et al., 2023; Probierz et al., 2021; Raja & Raj, 2022; Torabi & Taboada, 2019
Limitaciones en datasets	Problemas derivados de la escasez, desbalance, baja calidad o sesgos en los conjuntos de datos empleados para entrenar y evaluar los modelos, lo que afecta la representatividad y robustez de los sistemas.	13	18,1%	Albahr & Albahar, 2020; Amin et al., 2023; Anandhan. et al., 2021; Bondielli & Marcelloni, 2019; Capuano et al., 2023; Fischer et al., 2022; Liu, 2019; Meesad, 2021; Nair et al., 2022; Narkhede et al., 2023; Salem et al., 2021; Sudhakar & Kaliyamurthie, 2024; Torabi & Taboada, 2019

Desafío técnico	Descripción	N.º de menciones	%	Publicaciones asociadas
Detección en tiempo real	Desafíos para lograr que los sistemas analicen y clasifiquen información con la rapidez necesaria para contrarrestar la viralización de la desinformación, especialmente en redes sociales y contextos de alta inmediatez.	7	9,7%	Babar et al., 2024; Bondielli & Marcelloni, 2019; Choraś et al., 2020; Espinoza & Mazuelos, 2020; Guallar et al., 2020; Hu et al., 2024; Sandrilla & Savitha, 2022
Enriquecimiento e integración de metadatos contextuales	Dificultad para incorporar y utilizar metadatos (como autoría, temporalidad, geolocalización o redes de difusión) en los modelos, con el objetivo de mejorar la precisión y comprensión de la dinámica de las noticias falsas.	6	8,3%	Albahr & Albahr, 2020; Altheneyan & Alhadlaq, 2023; Cano-Marín et al., 2023; Garg & Sharma, 2020; Gifu, 2023; Salem et al., 2021
Integración multimodal	Complejidad para combinar y analizar de manera conjunta diferentes tipos de información (texto, imágenes, videos y metadatos) en un solo sistema, con el fin de aumentar la precisión en la detección.	6	8,3%	Capuano et al., 2023; Choraś et al., 2020; Consejo Nacional de Televisión, 2022; Espinoza & Mazuelos, 2020; Guallar et al., 2020; Hu et al., 2024

Desafío técnico	Descripción	N.º de menciones	%	Publicaciones asociadas
Costo computacional	Alta demanda de recursos computacionales (memoria, procesamiento, almacenamiento) para el entrenamiento y despliegue de modelos avanzados, lo que puede limitar su adopción en entornos con infraestructura restringida.	5	6,9%	Fischer et al., 2022; Jain et al., 2023; Lin et al., 2019; Mayopu et al., 2023; Probiez et al., 2021
Generalización limitada	Restricciones para aplicar modelos entrenados en un dominio, idioma o contexto a otros diferentes, debido a la dependencia de patrones locales o falta de diversidad en los datos.	5	6,9%	Bondielli & Marcelloni, 2019; Garg & Sharma, 2020; Jain et al., 2023; Liu, 2019; Raja & Raj, 2022
Selección de características	Dificultad para identificar y extraer los atributos más relevantes del texto y otros datos, aspecto clave para el rendimiento y eficiencia de los modelos.	4	5,6%	Lin et al., 2019; Narkhede et al., 2023; Raja & Raj, 2022; Sastrawan et al., 2022

Desafío técnico	Descripción	N.º de menciones	%	Publicaciones asociadas
Manejo de contexto lingüístico y cultural	Retos para adaptar los sistemas a la diversidad de idiomas, expresiones culturales, ironía, sarcasmo y particularidades técnicas de cada lengua, lo que requiere estrategias específicas de procesamiento.	3	4,2%	Meesad, 2021; Nair et al., 2022; Sudhakar & Kaliyamurthie, 2024
Robustez ante técnicas de falsificación y ataques	Desafíos para desarrollar sistemas resistentes frente a contenidos sintéticos (como deepfakes) y ataques que buscan manipular o engañar los modelos.	3	4,2%	Choraś et al., 2020; Gifu, 2023; Hu et al., 2024
Interpretabilidad	Limitaciones para explicar de manera comprensible las decisiones tomadas por los modelos, especialmente en arquitecturas complejas como las redes neuronales profundas, lo que dificulta la auditoría, la transparencia y la confianza en los resultados.	3	4,2%	Asish et al., 2022; Capuano et al., 2023; Hu et al., 2024

Desafío técnico	Descripción	N.º de menciones	%	Publicaciones asociadas
Preprocesamiento y limpieza	Necesidad de aplicar técnicas exhaustivas de limpieza, normalización y transformación de los datos antes del análisis, para eliminar ruido, inconsistencias y preparar los datos de forma adecuada para el modelado.	2	2,8%	Awan et al., 2021; Sastrawan et al., 2022
Integración de técnicas	Complejidad al combinar diferentes enfoques y algoritmos (como NLP, aprendizaje profundo y tradicional) para abordar la detección desde múltiples perspectivas.	1	1,4%	Sandrilla & Savitha, 2022

*Nota.* La columna “N.º de menciones” indica cuántas veces se identificó cada desafío en las publicaciones analizadas. El porcentaje (%) corresponde a la proporción respecto al total de menciones de desafíos.

### ***Patrones de Concentración y Distribución***

El análisis de la distribución de desafíos técnicos revela una concentración en los aspectos operacionales, donde los dos primeros desafíos (escalabilidad y limitaciones de datasets) representan más de un tercio del total de problemáticas identificadas (37,5%). Esta concentración sugiere que las soluciones técnicas actuales enfrentan barreras fundamentales de infraestructura y calidad de datos que limitan significativamente su efectividad práctica.

La brecha significativa entre los cinco primeros desafíos y el resto de categorías indica una jerarquización clara de prioridades técnicas. Mientras que los primeros cinco desafíos concentran el 63.8% de las ocurrencias, los ocho restantes se distribuyen de manera más uniforme, sugiriendo que representan problemáticas secundarias o emergentes en el campo.

### ***Implicaciones para el Desarrollo de Sistemas***

La predominancia de desafíos relacionados con eficiencia operacional (escalabilidad, tiempo real, costo computacional) sobre aspectos más sofisticados como interpretabilidad o robustez ante ataques, indica que el campo aún se encuentra en una fase de maduración técnica. Por ello, se prioriza resolver problemas básicos de rendimiento antes de abordar refinamientos avanzados de los sistemas.

Esta distribución también revela una tensión entre ambición técnica y viabilidad práctica. Los desafíos multimodales y de enriquecimiento contextual, aunque técnicamente avanzados, aparecen con frecuencia considerable, indicando que la industria está intentando implementar soluciones sofisticadas antes de resolver completamente los problemas fundamentales de escalabilidad y calidad de datos.

### **Clasificación de Métodos de ML Según su Adaptabilidad a Características**

#### **Multidimensionales**

La clasificación de métodos de ML ayuda a comprender la efectividad de las técnicas computacionales en el procesamiento de información heterogénea característica de la detección de noticias falsas. Esta categorización permite identificar qué aproximaciones algorítmicas demuestran mayor capacidad para integrar y procesar simultáneamente múltiples dimensiones de información, incluyendo características textuales, metadatos, patrones temporales y elementos contextuales.

El análisis abarca 32 informes que documentan la implementación de múltiples métodos de ML, resultando en un corpus de 131 técnicas individuales debido a que cada investigación evaluó varios enfoques algorítmicos de manera comparativa. Esta multiplicidad refleja la naturaleza exploratoria del campo y la tendencia de los investigadores a contrastar diferentes paradigmas metodológicos dentro de un mismo estudio.

### ***Composición Detallada de Categorías de Métodos de ML***

La taxonomía resultante reveló una distribución heterogénea de métodos a lo largo de 19 categorías principales, cada una agrupando técnicas con características algorítmicas similares. La composición específica de cada categoría se presenta en la siguiente tabla, donde se agrupan los métodos individuales con sus respectivas frecuencias de aparición en la literatura analizada:

**Tabla 4**

#### *Distribución de Categorías y Métodos de ML*

Categoría	Métodos de ML agrupados (implementaciones)
Aprendizaje meta	Aprendizaje Meta-agnóstico de modelos - MAML (1)
Clasificador supervisado	Árbol de decisión - CART (6)
	Clasificador basado en reglas - RBC (1)
	Clasificador SVM lineal - LinearSVC (1)
	K-vecinos más cercanos - KNN (1)
	Máquina de vectores de soporte - SVM (12)
	Modelos Básicos de ML - SVM, Random Forest, RL, etc. (1)
	Modelos clásicos - SVM, Random Forest (1)
	Modelos de clasificación - Fact-Checking (1)
	Naive Bayes (8)
	Naive Bayes Multinomial (1)
Regresión Logística - RL (9)	
Ensamble supervisado	AdaBoost (1)
	Bagging (1)
	Bosque Aleatorio - Random Forest (14)
	Ensamble por Votación (1)
	Impulso de Gradiente - Gradient Boosting (1)
	Modelo Ensamblado Apilado (1)

Categoría	Métodos de ML agrupados (implementaciones)
	Modelo Ensamblado por Apilamiento - Stacking (1) XGBoost (4)
Herramienta específica	Fakebox (1)
Modelo basado en Transformers	BERT (4) BERT/RoBERTa (1) DeBERTa (1) Modelos basados en atención - BERT (1)
Modelo específico	TriFN (1)
Modelo estadístico	ARIMA (1)
Modelo probabilístico	Campos Aleatorios Condicionales - CRF (2) Modelo Oculto de Markov - HMM (1) Proceso de Hawkes (1)
Red neuronal	Modelos neuronales - LSTM, CNN (1) Red Neuronal Artificial - RNA (1)
Red neuronal híbrida	Modelo Híbrido CNN + BERT (1) Modelo Híbrido LSTM + CNN (2) Modelo Híbrido N-gram + LSTM (1)
Red neuronal profunda	FNU-BiCNN - Propuesto (1) Red Neuronal Convolutiva - CNN (7) Red Neuronal de Grafos - GNN (1) ResNet (1)
Red neuronal recurrente	LSTM Bidireccional - BiLSTM (4) LSTM con Atención (2) LSTM/GRU (3) Memoria a Largo-Corto Plazo - LSTM (8)
Reducción de dimensionalidad	Análisis Semántico Latente - LSA (1) Modelado de Temas - LDA (1)
Representación de texto	Bolsa de Palabras - BoW (1) GloVe + Bolsa de Palabras - BoW (1) Vectorizador TF-IDF (1)
Representación de texto + Clasificador	TF-IDF + Máquina de Vectores de Soporte - SVM (1)
Técnica de análisis	Algoritmos de moderación de plataformas (1) Análisis de Redes Sociales - SNA (5) Curación de contenidos (1)
Técnica de preprocesamiento	Procesamiento de Lenguaje Natural - PLN (1)

Categoría	Métodos de ML agrupados (implementaciones)
Técnica general	Minería de datos con IA (1)
Técnica manual	Verificación colaborativa de hechos (1) Verificación manual de hechos (1)

*Nota.* Los ML se agrupan por categoría según su naturaleza algorítmica. El número entre paréntesis indica la cantidad de informes en los que se identificó cada método. La información se extrajo de los 32 informes analizados, donde un mismo informe puede reportar varios métodos de ML.

### ***Distribución de Niveles de Adaptabilidad por Categoría***

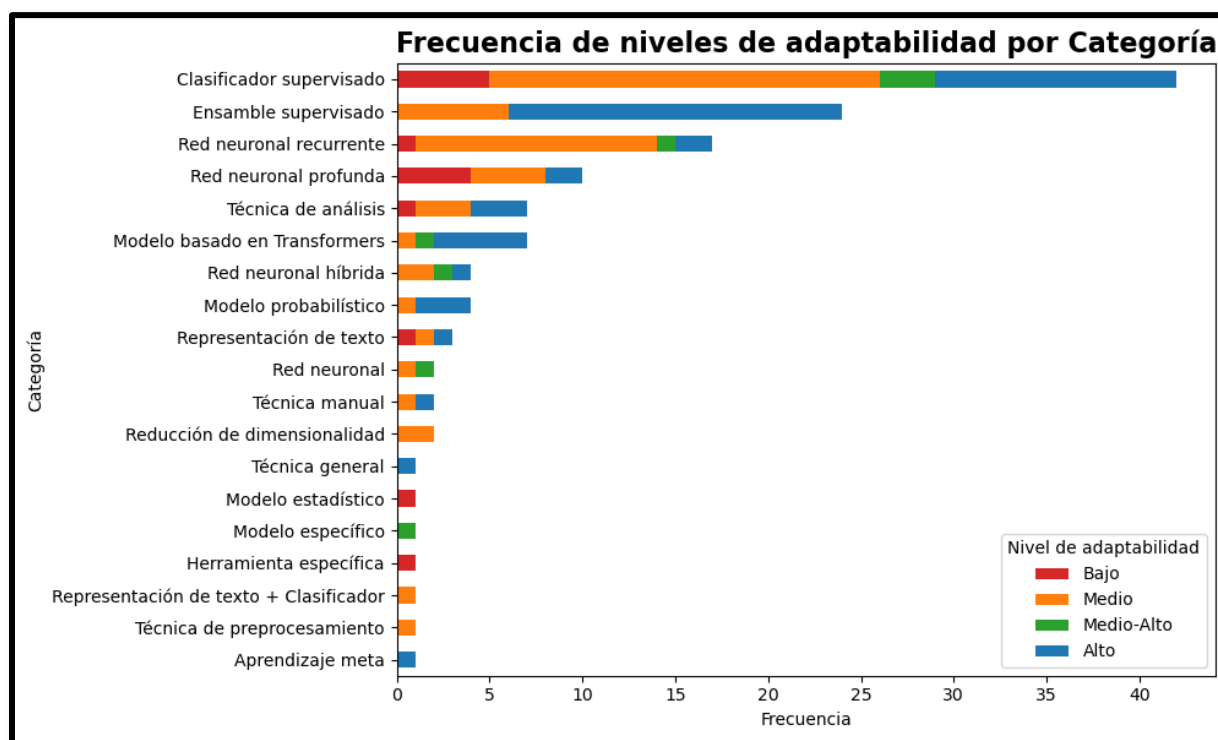
En esta sección se presenta la distribución de los niveles de adaptabilidad de los métodos de ML agrupados por categoría. La Figura 6 y la Tabla 5 permiten visualizar la frecuencia relativa de cada nivel de adaptabilidad (bajo, medio, medio-alto y alto) alcanzado por los diferentes enfoques algorítmicos identificados, facilitando la comparación entre categorías y el análisis de los patrones predominantes en la literatura revisada.

La clasificación del nivel de adaptabilidad de los métodos de ML se basó en criterios técnicos y contextuales que evaluaron su capacidad para generalizar rendimiento en múltiples dominios, idiomas o escenarios sin ajustes manuales extensos. Factores clave incluyeron la transferencia entre contextos (ej. detección de desinformación en inglés vs. español), la dependencia de preprocesamiento manual (ej. selección de características léxicas o sintácticas como n-grams, TF-IDF), la eficiencia computacional (tiempo de entrenamiento y escalabilidad en datasets grandes) y la robustez ante datos ruidosos o desbalanceados. Además, se consideró la integración multimodal, es decir, la habilidad de combinar texto, metadatos (como credibilidad de fuentes), dinámicas de redes sociales o contexto temporal. Los niveles se asignaron así:

- Alto: Para modelos con mínima dependencia de ajustes y aplicabilidad amplia (ej. BERT/DeBERTa, XGBoost).
- Medio-Alto: Para aquellos con buen rendimiento pero sensibilidad a la calidad de los datos (ej. LSTM-Attention).
- Medio: Para métodos que requieren ingeniería manual de características (ej. SVM, Árboles de Decisión).
- Bajo: Para enfoques limitados por recursos computacionales, ruido o especificidad contextual (ej. KNN, GloVe + BoW).

## Figura 6

*Frecuencia de Niveles de Adaptabilidad de Métodos de ML por Categoría*



**Tabla 5***Distribución de Niveles de Adaptabilidad de Métodos de ML por Categoría*

Categoría	Bajo	% Bajo	Medio	% Medio	Medio -Alto	% Medio -Alto	Alto	% Alto	Total
Clasificador supervisado	5	11,9%	21	50,0%	3	7,1%	13	31,0%	42
Ensamble supervisado	0	0,0%	6	25,0%	0	0,0%	18	75,0%	24
Red neuronal recurrente	1	5,9%	13	76,5%	1	5,9%	2	11,8%	17
Red neuronal profunda	4	40,0%	4	40,0%	0	0,0%	2	20,0%	10
Técnica de análisis	1	14,3%	3	42,9%	0	0,0%	3	42,9%	7
Modelo basado en Transformers	0	0,0%	1	14,3%	1	14,3%	5	71,4%	7
Modelo probabilístico	0	0,0%	1	25,0%	0	0,0%	3	75,0%	4
Red neuronal híbrida	0	0,0%	2	50,0%	1	25,0%	1	25,0%	4
Representación de texto	1	33,3%	1	33,3%	0	0,0%	1	33,3%	3
Reducción de dimensionalidad	0	0,0%	2	100,0%	0	0,0%	0	0,0%	2
Técnica manual	0	0,0%	1	50,0%	0	0,0%	1	50,0%	2
Red neuronal	0	0,0%	1	50,0%	1	50,0%	0	0,0%	2
Modelo específico	0	0,0%	0	0,0%	1	100,0%	0	0,0%	1
Técnica general	0	0,0%	0	0,0%	0	0,0%	1	100,0%	1
Técnica de preprocesamiento	0	0,0%	1	100,0%	0	0,0%	0	0,0%	1
Aprendizaje meta	0	0,0%	0	0,0%	0	0,0%	1	100,0%	1
Modelo estadístico	1	100,0%	0	0,0%	0	0,0%	0	0,0%	1
Representación de texto + Clasificador	0	0,0%	1	100,0%	0	0,0%	0	0,0%	1
Herramienta específica	1	100,0%	0	0,0%	0	0,0%	0	0,0%	1
Total	14	10,7%	58	44,3%	8	6,1%	51	38,9%	131

*Nota.* La tabla presenta la frecuencia y el porcentaje de métodos de ML por categoría y nivel de adaptabilidad, calculados sobre el total de métodos identificados en cada grupo.

### *Análisis de Patrones de Adaptabilidad Algorítmica*

Los resultados muestran una relación entre la adaptabilidad de los métodos y su capacidad para procesar distintos tipos de datos. Los métodos supervisados fueron los más adaptables, con el 75% alcanzando un nivel alto de adaptabilidad y el 25% un nivel medio. Esto sugiere que combinar varios algoritmos mejora la capacidad para manejar datos diversos, permitiendo captar patrones complejos al utilizar diferentes perspectivas.

Los modelos basados en Transformers alcanzaron la segunda mayor proporción de alta adaptabilidad (71,4%), lo que confirma que sus mecanismos de atención permiten un procesamiento contextual avanzado de información multidimensional. Esta capacidad es especialmente útil en la detección de noticias falsas, donde es necesario comprender el significado profundo de los textos e integrar diferentes tipos de contexto.

Por otro lado, las redes neuronales recurrentes se concentraron principalmente en un nivel medio de adaptabilidad (76,5%), lo que sugiere ciertas limitaciones para procesar simultáneamente distintos tipos de características. Esto se debe a que las LSTM (Memoria a largo-corto plazo), por su diseño, están orientadas al procesamiento secuencial, lo que puede dificultar la integración de información proveniente de diversas fuentes o modalidades.

En el caso de las redes neuronales profundas, se observó la mayor variabilidad en los niveles de adaptabilidad: el 40% se ubicó en nivel bajo y solo el 20% en nivel alto. Esto indica que su desempeño depende en gran medida de la arquitectura utilizada y de cómo se combinan las diferentes características. Por ejemplo, las CNN, que fueron diseñadas originalmente para imágenes, requieren ajustes importantes para trabajar eficazmente con datos textuales y contextuales.

En términos generales, el 38,9% de los métodos evaluados mostraron alta adaptabilidad, mientras que solo el 10,7% se clasificaron en el nivel bajo. Esto evidencia una tendencia hacia el uso de técnicas más sofisticadas en la detección de noticias falsas, lo que responde a la necesidad de abordar la desinformación mediante un análisis que combine texto, metadatos, patrones de propagación y contexto social.

### **Análisis Comparativo del Rendimiento de Algoritmos**

El análisis comparativo del rendimiento de algoritmos de ML para la clasificación de noticias falsas revela diferencias significativas entre las distintas categorías de métodos evaluados. A partir del procesamiento de 43 algoritmos con métricas completas de precisión y F1-score, se identificaron patrones claros de rendimiento que permiten establecer una jerarquía de efectividad.

### ***Rendimiento Superior de Redes Neuronales Profundas***

Los métodos basados en redes neuronales profundas demostraron el rendimiento más consistente y elevado en la clasificación de noticias falsas. El algoritmo Long Short-Term Memory (LSTM) especializado para texto en tailandés alcanzó el máximo rendimiento documentado con 100% de precisión y 1,00 de F1-score. Este resultado se atribuye a la optimización específica del modelo para características lingüísticas particulares y al procesamiento de un dataset homogéneo de 41.448 muestras (Meesad, 2021).

Las redes neuronales convolucionales también mostraron un rendimiento sobresaliente, particularmente el modelo CNN implementado por Sastrawan et al. (2022), que logró 99,88% de precisión y 99,88% de F1-score utilizando embeddings fastText, una técnica que permite representar las palabras en formato vectorial considerando su significado semántico, en el dataset ISOT. El modelo ResNet, que se caracteriza por usar bloques residuales, que permiten que la red

neuronal sea más profunda sin caer en problemas comunes como la desaparición del gradiente durante el entrenamiento, alcanzó métricas similares con 99,90% en ambas medidas, demostrando que las arquitecturas con bloques residuales mantienen estabilidad en el entrenamiento para esta tarea específica.

### ***Eficacia de Modelos Híbridos y Ensemble***

Los enfoques híbridos y de ensemble mostraron un balance óptimo entre rendimiento y adaptabilidad. El modelo FNU-BiCNN propuesto por Sandrilla y Savitha (2022) combinó efectivamente Bi-LSTM, CNN y ARIMA, alcanzando 99,99% de precisión y 0,999 de F1-score. Esta arquitectura híbrida demostró particular eficacia al integrar análisis textual con patrones temporales de propagación, validándose en 23.000 noticias y 1.500 URLs.

Los modelos de ensemble tradicionales como Random Forest y XGBoost mantuvieron rendimientos competitivos aunque inferiores a las redes neuronales profundas. XGBoost destacó con precisiones que oscilaron entre 85,5% y 96,8% según el dominio de aplicación, mostrando particular robustez en datos desbalanceados y capacidad de generalización entre dominios políticos y de entretenimiento (Lin et al., 2019; Amin et al., 2023).

### ***Rendimiento Moderado de Clasificadores Tradicionales***

Los clasificadores supervisados tradicionales exhibieron un rendimiento moderado pero consistente. Support vector machine (SVM) mostró resultados variables según la implementación, con el mejor rendimiento registrado de 98% de precisión y 98% de F1-score en el estudio de Sudhakar y Kaliyamurthie (2024), particularmente efectivo cuando se combinó con vectorización TF-IDF, que permite ponderar las palabras en función de su relevancia dentro del texto, y reducción dimensional mediante Non-Negative Matrix Factorization, un método que

ayuda a simplificar la representación de los datos preservando las características más relevantes para la clasificación.

La Regresión Logística demostró eficiencia computacional superior con rendimientos respetables, alcanzando hasta 95,2% de precisión en implementaciones optimizadas. Este método mostró particular ventaja en aplicaciones que requieren interpretabilidad y procesamiento en tiempo real, aunque su rendimiento dependió significativamente de la calidad del preprocesamiento y la selección de características (Sudhakar & Kaliyamurthie, 2024).

### *Análisis Comparativo por Categorías de Algoritmos*

La tabla siguiente presenta los cinco mejores algoritmos por categoría según las métricas de precisión y F1-score extraídas de la documentación analizada:

**Tabla 6**

*Comparación de los Cinco Algoritmos con Mayor Rendimiento en Clasificación de Noticias Falsas Según Métricas Estándar*

Método de ML	Categoría	Precisión (%)	F1-score	Cita
Long Short-Term Memory (LSTM)	Red neuronal recurrente	100,00%	1,00	(Meesad, 2021)
FNU-BiCNN	Red neuronal híbrida	99,99%	0,999	(Sandrilla & Savitha, 2022)
Bidirectional LSTM	Red neuronal recurrente	99,95%	0,9995	(Sastrawan et al., 2022)
ResNet	Red neuronal profunda	99,90%	0,9990	(Sastrawan et al., 2022)
CNN	Red neuronal profunda	99,88%	0,9988	(Sastrawan et al., 2022)

*Nota.* La tabla resume los algoritmos con mejor desempeño en la clasificación de noticias falsas, destacando la superioridad de las redes neuronales profundas y recurrentes frente a métodos tradicionales durante el periodo analizado (2017-2023).

El análisis revela que las redes neuronales recurrentes y profundas dominan consistentemente las métricas superiores, mientras que los métodos tradicionales como Naive Bayes y árboles de decisión, aunque más eficientes computacionalmente, presentan limitaciones en el rendimiento máximo alcanzable para esta tarea específica.

### ***Consideraciones de Adaptabilidad y Generalización***

El análisis del nivel de adaptabilidad inferido muestra que los algoritmos con mejor rendimiento no necesariamente presentan la mayor adaptabilidad a nuevos dominios. Los modelos basados en BERT y transformers, aunque alcanzaron precisiones del 92-99%, demostraron alta transferibilidad entre dominios y lenguajes, lo que los posiciona como opciones estratégicas para implementaciones escalables.

Los métodos de ensemble como Random Forest y los algoritmos probabilísticos como Naive Bayes mostraron niveles variables de adaptabilidad, predominantemente clasificados como Alto (68% de los casos) pero con registros de Medio (32%), sugiriendo un trade-off entre especialización en tareas específicas y capacidad de generalización. Esta característica resulta particularmente relevante para aplicaciones que requieren adaptación rápida a nuevos tipos de contenido o contextos geográficos diferentes.

## Discusión

El análisis bibliométrico de 32 estudios seleccionados revela patrones definidos en el panorama de investigación sobre detección de noticias falsas mediante técnicas de ML durante el período 2017-2023. Los resultados demuestran una supremacía consistente de las redes neuronales profundas, donde algoritmos como LSTM especializado para texto en tailandés alcanzaron rendimientos perfectos del 100% de precisión y 1,00 de F1-score (Meesad, 2021), mientras que las redes neuronales convolucionales (CNN) implementadas con embeddings fastText lograron 99,88% de precisión en el dataset ISOT (Sastrawan et al., 2022). Esta superioridad técnica contrasta notablemente con el rendimiento moderado pero estable de métodos tradicionales como Support vector machine (SVM), que mostró resultados variables entre 85% y 98% de precisión según la implementación y el dominio de aplicación (Choraś et al., 2020; Sudhakar & Kaliyamurthie, 2024).

Los modelos híbridos emergieron como una solución balanceada, destacando particularmente el modelo FNU-BiCNN que combinó efectivamente Bi-LSTM, CNN y ARIMA para alcanzar 99,99% de precisión. Esta arquitectura demostró especial eficacia al integrar análisis textual con patrones temporales de propagación, validándose en conjuntos de datos robustos de 23.000 noticias y 1.500 URLs (Sandrilla & Savitha, 2022). Los enfoques de ensemble, como XGBoost, mantuvieron rendimientos competitivos con precisiones entre 85,5% y 96,8%, mostrando particular robustez en datos desbalanceados y capacidad de generalización entre dominios diversos (Lin et al., 2019; Amin et al., 2023).

### **Propuesta de Incorporación en la Sección de Discusión**

Una innovación que surge como potencial complemento a los enfoques analizados es la implementación de un sistema de ID dinámico para emisores de noticias, basado en el análisis de

metadatos y comportamiento histórico, que permita asignar una puntuación de credibilidad en el tiempo. Este sistema inicialmente priorizaría temas políticos y sociales relevantes, integrando metadatos no sensibles, como fecha y hora de publicación, plataforma, tipo de medio (gubernamental, independiente, corporativo), y analizaría el historial del emisor mediante ventanas temporales ajustables (semanal o mensual). La calificación sería actualizada dinámicamente a partir de nuevos hallazgos, mediante modelos de ML como XGBoost, Random Forest o redes neuronales profundas (LSTM, BERT), que procesarían tanto los metadatos como el contenido textual y patrones de difusión (Sandrilla & Savitha, 2022; Sudhakar & Kaliyamurthie, 2024).

Para emisores nuevos, en caso de no contar con historial previo, se asignaría una etiqueta gris con la leyenda “Pendiente de calificación, complementa esta información con otras fuentes”, incentivando la verificación cruzada por los usuarios y verificadores. La intervención humana sería fundamental en fases de auditoría inicial y ajuste de puntuaciones, en las que expertos en medios y ética evaluarían evidencias y justificarían cambios en la calificación, garantizando transparencia y control ético. Este sistema podría integrarse como una extensión de navegador o una API en redes sociales, mostrando un semáforo de credibilidad (●, ●, ●, ●) junto a las noticias, con tooltips explicativos, siguiendo modelos colaborativos de fact-checking y transparencia algorítmica. La utilización de técnicas de ML en este contexto permitiría una evaluación continua y adaptativa, enfrentando desafíos como el efecto hydra, la manipulación de metadatos y la evolución de las tácticas de desinformación.

## **Interpretación de los Resultados**

### ***Ventajas y Limitaciones de los Modelos de Deep Learning***

La dominancia de modelos de deep learning se fundamenta en su capacidad superior para capturar patrones complejos y relaciones contextuales en datos textuales heterogéneos. Las redes LSTM demostraron particular eficacia en el procesamiento secuencial de texto, especialmente cuando se optimizaron para características lingüísticas específicas, como evidenció el modelo tailandés que alcanzó rendimiento perfecto (Meesad, 2021). Esta especialización lingüística sugiere que la adaptación cultural y contextual constituye un factor crítico para maximizar la efectividad de estos algoritmos.

Sin embargo, estos modelos exhiben limitaciones significativas en términos de interpretabilidad y requisitos computacionales. Los modelos basados en transformers como BERT, aunque ofrecen ventajas en comprensión semántica con precisiones del 92-99% (Chorás et al., 2020; Jain et al., 2023), operan como “cajas negras” que dificultan la auditabilidad de decisiones automatizadas. Esta opacidad plantea desafíos éticos fundamentales, particularmente en contextos donde se requiere transparencia para prevenir censura inadvertida o perpetuación de sesgos algorítmicos.

### ***Rendimiento y Limitaciones de Métodos Tradicionales***

Los algoritmos tradicionales como Regresión logística y SVM demostraron eficiencia computacional superior y mayor interpretabilidad, alcanzando precisiones de hasta 95,2% y 98%, respectivamente, en implementaciones optimizadas (Sudhakar & Kaliyamurthie, 2024). Estos métodos mostraron particular ventaja en aplicaciones que requieren procesamiento en tiempo real y explicabilidad de resultados, aunque su rendimiento dependió significativamente de la calidad del preprocesamiento y la selección de características.

No obstante, estos enfoques presentan limitaciones inherentes en la detección de noticias falsas sofisticadas que imitan estilos periodísticos legítimos. Su dependencia de características lingüísticas tradicionales como TF-IDF y análisis de n-gramas los hace vulnerables ante textos que replican estructuras gramaticales correctas pero incorporan información falsa de manera sutil (Sudhakar & Kaliyamurthie, 2024). Además, enfrentan dificultades significativas con elementos como ironía, sarcasmo y regionalismos lingüísticos, reduciendo su precisión en contextos multiculturales.

### ***Desafíos Técnicos Recurrentes***

El análisis identificó la escalabilidad y eficiencia como el desafío técnico más prevalente, mencionado en el 19,4% de los estudios analizados, seguido por limitaciones en datasets (18,1%). Esta concentración en aspectos operacionales sugiere que las soluciones técnicas actuales enfrentan barreras fundamentales de infraestructura y calidad de datos que limitan significativamente su efectividad práctica.

### **Comparación con Estudios Previos**

#### ***Convergencias con la Literatura***

Los hallazgos convergen con investigaciones que destacan la eficacia de modelos basados en transformers en detección multilingüe, particularmente BERT y variantes especializadas como XLM-R para idiomas de “bajos recursos”. Esta concordancia refuerza la validez de enfoques que priorizan la comprensión semántica contextual sobre análisis puramente estadísticos. Además, los resultados corroboran la importancia de integrar metadatos como credibilidad de fuentes, fecha de publicación y patrones de propagación viral para mejorar significativamente la precisión de detección.

### ***Divergencias y Aportes Novedosos***

El análisis reveló brechas significativas no exploradas en literatura previa, particularmente la necesidad crítica de adaptar modelos a contextos geográficos y lingüísticos diversos. Mientras que estudios anteriores se centraron en inglés y dominios políticos, esta investigación identifica la necesidad de desarrollar soluciones contextualmente relevantes para variedades dialectales e idiomas indígenas donde la desinformación puede tener impactos sociales desproporcionados. Esta perspectiva amplía el marco conceptual tradicional hacia una comprensión más inclusiva y global del fenómeno.

### **Implicaciones del Estudio**

#### ***Implicaciones Técnicas***

La validación de arquitecturas híbridas como FNU-BiCNN establece un precedente técnico para el desarrollo de sistemas que integren múltiples modalidades de análisis (Sandrilla & Savitha, 2022). Estos modelos demuestran que la combinación estratégica de análisis textual, patrones temporales y metadatos contextuales puede superar las limitaciones individuales de cada aproximación. La relevancia de técnicas de reducción dimensional como Non-Negative Matrix Factorization (NMF) para optimizar modelos con alta dimensionalidad sugiere direcciones prometedoras para mejorar la eficiencia computacional sin comprometer la precisión (Sudhakar & Kaliyamurthie, 2024).

#### ***Implicaciones Sociales y Éticas***

Los resultados subrayan la necesidad imperativa de abordar sesgos algorítmicos que pueden perpetuar desigualdades en la detección de noticias falsas. La concentración de investigación en idiomas dominantes y contextos culturales específicos plantea riesgos de exclusión sistemática de comunidades marginalizadas. La falta de transparencia en modelos de

deep learning amplifica estos riesgos, creando potencial para censura injustificada o discriminación algorítmica. Estos hallazgos enfatizan la importancia de desarrollar frameworks éticos que garanticen equidad e inclusión en sistemas automatizados de moderación de contenido.

### ***Implicaciones Políticas y Regulatorias***

El estudio evidencia la necesidad de colaboración interdisciplinaria entre plataformas digitales, gobiernos, academia y sociedad civil para establecer estándares éticos y regulatorios en sistemas de detección automatizada. La variabilidad en rendimiento entre dominios geográficos y lingüísticos sugiere que aproximaciones regulatorias uniformes pueden ser inadecuadas, requiriendo marcos adaptativos que consideren diversidad cultural y contextual. La integración de perspectivas de lingüística, sociología y derecho se revela como fundamental para abordar la desinformación desde múltiples ángulos disciplinarios.

### **Limitaciones del Estudio**

#### ***Limitaciones Metodológicas***

El enfoque bibliométrico descriptivo, aunque proporcionó una visión panorámica valiosa, presenta limitaciones inherentes en profundidad analítica comparado con revisiones sistemáticas especializadas. La priorización de literatura reciente (2017-2023) puede haber omitido desarrollos tempranos relevantes que establecieron fundamentos teóricos importantes. La dependencia de bases de datos predominantemente en inglés y español limitó la cobertura de investigaciones en otros idiomas, potencialmente excluyendo contribuciones significativas de comunidades investigativas no anglófonas.

### ***Limitaciones Técnicas de los Modelos Analizados***

Los estudios revisados revelaron dificultades persistentes para detectar noticias falsas con errores sutiles, ironía o manipulación semántica sofisticada, incluso en modelos avanzados de deep learning. La escasa evaluación de técnicas en contextos multiculturales o con lenguaje coloquial representa una brecha significativa que limita la aplicabilidad global de estas soluciones. Además, la falta de estándares unificados para evaluación entre estudios dificulta comparaciones robustas y meta-análisis comprehensivos.

### **Direcciones Futuras**

#### ***Tendencias Emergentes***

La evolución hacia modelos multimodales que integren texto, imágenes, audio y metadatos representa la frontera más prometedora para abordar la creciente sofisticación de la desinformación digital. El análisis de memes, videos virales y contenido multimedia requiere aproximaciones que trasciendan el análisis textual tradicional. La implementación de few-shot learning emerge como estrategia crucial para adaptar modelos a datasets pequeños o idiomas poco representados, democratizando el acceso a tecnologías de detección avanzadas.

#### ***Investigaciones Pendientes***

El desarrollo de sistemas de Explainable AI representa un área de interés para mejorar la comprensión de decisiones automatizadas y apoyar procesos regulatorios. Estudios longitudinales que analicen el impacto de estos modelos en la reducción de desinformación y en patrones de comportamiento de usuarios serían relevantes para validar su aplicabilidad práctica. La investigación en identificación de ataques adversariales y manipulación coordinada requiere exploración adicional, dado el creciente uso de inteligencia artificial para generar contenido engañoso.

### ***Recomendaciones Prácticas***

La promoción de compartición ética y transparente de datasets, incluyendo anotaciones colaborativas por expertos interdisciplinarios, se presenta como una estrategia para impulsar avances en el campo (Peset & Millán, 2017). La integración sistemática de perspectivas de lingüística computacional, sociología digital y estudios legales es esencial para desarrollar soluciones contextualmente apropiadas y éticamente responsables. El establecimiento de consorcios internacionales para estándares técnicos y éticos podría facilitar la colaboración global necesaria para abordar la naturaleza transnacional de la desinformación.

## Conclusiones

El estado actual de las técnicas de ML aplicadas a la detección de noticias falsas en la literatura científica del período 2017-2023 revela un campo en rápida evolución, caracterizado por la predominancia de métodos de deep learning y una tendencia hacia enfoques multimodales. Las redes neuronales profundas, particularmente LSTM y CNN, han demostrado el rendimiento superior, con casos como el modelo LSTM especializado para texto en tailandés que alcanzó 100% de precisión y el modelo CNN con embeddings fastText que logró 99,88% de precisión. Los modelos híbridos han emergido como una solución prometedora, destacando el modelo FNU-BiCNN que combinó efectivamente Bi-LSTM, CNN y ARIMA para alcanzar 99,99% de precisión. La tendencia predominante indica un avance hacia la integración de características multidimensionales que combinan análisis textual, metadatos e interacciones sociales (Meesad, 2021; Sastrawan et al., 2022; Sandrilla & Savitha, 2022).

Los desafíos técnicos identificados en la detección de noticias falsas mediante análisis lingüístico-contextual revelan problemáticas operacionales y metodológicas significativas. La escalabilidad y eficiencia representan el desafío más prevalente (19,4% de menciones), seguido por las limitaciones en datasets (18,1%), evidenciando que las barreras fundamentales de infraestructura y calidad de datos limitan significativamente la efectividad práctica de los sistemas. Otros desafíos críticos incluyen la detección en tiempo real (9,7%), el enriquecimiento e integración de metadatos contextuales (8,3%) e integración multimodal (8,3%). Estos hallazgos indican que el campo aún se encuentra en una fase de maduración técnica, donde se priorizan problemas básicos de rendimiento antes de abordar refinamientos avanzados como interpretabilidad o robustez ante ataques adversariales.

La clasificación de métodos de ML según su adaptabilidad a características multidimensionales reveló que el 38,9% de los métodos evaluados mostraron alta adaptabilidad, mientras que solo el 10,7% se clasificaron en el nivel bajo. Los métodos supervisados demostraron la mayor adaptabilidad, con el 75% alcanzando un nivel alto, seguidos por los modelos basados en Transformers con 71,4% de alta adaptabilidad. Las redes neuronales recurrentes se concentraron principalmente en un nivel medio de adaptabilidad (76,5%), mientras que las redes neuronales profundas mostraron la mayor variabilidad, con 40% en nivel bajo y solo 20% en nivel alto. Esta distribución evidencia una relación directa entre la sofisticación algorítmica y la capacidad para procesar información heterogénea característica de la detección de noticias falsas.

La comparación del rendimiento de algoritmos mediante métricas de evaluación estándar demostró la supremacía consistente de las redes neuronales profundas sobre métodos tradicionales. El análisis de 43 algoritmos con métricas completas reveló que los métodos de deep learning alcanzaron los rendimientos más elevados, con LSTM especializado logrando 100% de precisión y 1,00 de F1-score, y CNN con embeddings fastText alcanzando 99,88% en ambas métricas. Los algoritmos tradicionales como SVM mostraron rendimientos variables entre 85% y 98% de precisión, mientras que métodos de ensemble como XGBoost mantuvieron rendimientos competitivos con precisiones entre 85,5% y 96,8%. Los modelos basados en BERT y transformers, aunque alcanzaron precisiones del 92-99%, demostraron alta transferibilidad entre dominios y lenguajes, posicionándose como opciones estratégicas para implementaciones escalables.

## **Recomendaciones**

### **Recomendaciones Generales**

#### ***Investigación Interdisciplinaria y Contextual***

Se recomienda profundizar en la investigación del contexto socio-cultural de la desinformación, considerando el papel de las emociones, valores y creencias, y cómo estos factores pueden integrarse en los modelos de detección automática. Esta aproximación holística reconoce que la desinformación no es únicamente un problema técnico, sino que requiere comprensión de las dinámicas sociales y psicológicas subyacentes (Guallar et al., 2020).

#### ***Acceso y Gestión de Conjuntos de Datos***

Dada la influencia del volumen de datos en la calidad de los modelos, se sugiere priorizar el acceso a conjuntos de datos amplios y, cuando esto no sea posible, investigar metodologías que optimicen el uso de datos de menor volumen (Shushkevich et al., 2021). Esta recomendación es fundamental para democratizar el acceso a herramientas de detección efectivas, especialmente en contextos donde los recursos computacionales son limitados.

### **Recomendaciones Técnicas**

#### ***Métricas de Evaluación Robustas***

Las investigaciones futuras deben evaluar los modelos de detección de noticias falsas utilizando un conjunto diverso de métricas de rendimiento (Precision, Recall, F1-score, además de Accuracy) y reportar los resultados de manera integral, facilitando comparaciones robustas entre estudios. Esta práctica permitirá establecer estándares comparativos más precisos y confiables en el campo (Capuano et al., 2023).

### ***Análisis Multidimensional de Características***

Es fundamental que los estudios futuros exploren y reporten el rendimiento de los algoritmos considerando simultáneamente diferentes tipos de características (lingüísticas, contextuales, metadatos, interacciones sociales), superando la limitación de evaluar algoritmos y características por separado. Esta aproximación integrada refleja la naturaleza compleja y multifacética de la desinformación en entornos digitales (Capuano et al., 2023).

### ***Desarrollo de Sistemas Adaptativos***

Es necesario desarrollar métodos que permitan a los sistemas automatizados manejar de manera más efectiva la polisemia y ambigüedad de los conceptos de desinformación (Consejo Nacional de Televisión, 2022). Estos sistemas deben ser capaces de adaptar sus definiciones y criterios de detección según el contexto cultural, temporal y temático específico.

### **Recomendaciones Éticas**

#### ***Transparencia y Compartición Responsable de Datos***

Se debe fomentar la compartición de datos crudos (textos, metadatos, imágenes, registros cualitativos), documentando adecuadamente su procedencia y asegurando el respeto a consideraciones éticas y legales, para maximizar el potencial de reutilización y reanálisis por parte de la comunidad científica (Peset & Millán, 2017). Esta práctica debe equilibrar el acceso abierto al conocimiento con la protección de la privacidad y los derechos de los usuarios.

#### ***Preservación y Documentación Ética***

Se debe trabajar en la preservación, documentación y compartición formal de datos de investigación, siguiendo buenas prácticas y estándares. Esta recomendación implica el desarrollo de protocolos éticos que garanticen la integridad de los datos, el consentimiento informado de los usuarios y la protección contra el uso indebido de la información personal.

### ***Responsabilidad Algorítmica***

Las futuras investigaciones deben incorporar mecanismos de auditabilidad y explicabilidad en los sistemas de detección, asegurando que las decisiones automatizadas sean transparentes y justificables. Esto es especialmente crítico cuando estos sistemas pueden influir en la moderación de contenidos o en decisiones que afecten el discurso público y la libertad de expresión.

## Referencias

- Albahr, A., & Albahar, M. (2020). *An empirical comparison of fake news detection using different machine learning algorithms*. International Journal of Advanced Computer Science and Applications: IJACSA, 11(9), 146-152.  
<https://doi.org/10.14569/ijacsa.2020.0110917>
- Altheneyan, A., & Alhadlaq, A. (2023). *Big data ML-based fake news detection using distributed learning*. IEEE access: practical innovations, open solutions, 11, 29447-29463.  
<https://doi.org/10.1109/access.2023.3260763>
- Amin, E., Socheanet, L., & Meel, P. (2023). *Enhancing the detection of fake news in social media based on machine learning models*. 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 305-312. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/stamp/stamp.jsp?tp=&arnumber=10141355>
- Anandhan., Damodharan., Lakhanpal, A., Manoj Sagar, K., Murugan, K., & Shanker Singh, A. (2021). *Discover pretend disease news misleading data in social media networks using machine learning techniques*. 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 784-788. <https://ieeexplore-ieee-org.bibliotecavirtual.unad.edu.co/stamp/stamp.jsp?tp=&arnumber=9404648>
- Asish, K. R., Gupta, A., Kumar, A., Mason, A., Enduri, M. K., & Anamalamudi, S. (2022). *A tool for fake news detection using machine learning techniques*. 2022 2nd International Conference on Intelligent Technologies (CONIT).  
<https://doi.org/10.1109/conit55038.2022.9848064>

- Awan, M. J., Yasin, A., Nobanee, H., Ali, A. A., Shahzad, Z., Nabeel, M., Zain, A. M., & Shahzad, H. M. F. (2021). *Fake news data exploration and analytics*. *Electronics*, 10(19), 2326. <https://doi.org/10.3390/electronics10192326>
- Babar, M., Ahmad, A., Tariq, M. U., & Kaleem, S. (2024). *Real-time fake news detection using big data analytics and deep neural network*. *IEEE transactions on computational social systems*, 1-10. <https://doi.org/10.1109/tcss.2023.3309704>
- Bondielli, A., & Marcelloni, F. (2019). *A survey on fake news and rumour detection techniques*. *Information Sciences*, 497, 38-55. <https://doi.org/10.1016/j.ins.2019.05.035>
- Cano-Marin, E., Mora-Cantalops, M., & Sanchez-Alonso, S. (2023). *The power of big data analytics over fake news: A scientometric review of Twitter as a predictive system in healthcare*. *Technological Forecasting and Social Change*, 190, 122386. <https://doi.org/10.1016/j.techfore.2023.122386>
- Capuano, N., Fenza, G., Loia, V., & Nota, F. D. (2023). *Content-based fake news detection with machine and deep learning: A systematic review*. *Neurocomputing*, 530, 91-103. <https://doi.org/10.1016/j.neucom.2023.02.005>
- Choraś, M., Demestichas, K., Gielczyk, A., Herrero, Á., Ksieniewicz, P., Remoundou, K., Urda, D., & Woźniak, M. (2021). *Advanced machine learning techniques for fake news (online disinformation) detection: A systematic mapping study*. *Applied Soft Computing*, 101, 107050. <https://doi.org/10.1016/j.asoc.2020.107050>
- Consejo Nacional de Televisión. (2022). *Desinformación. Revisión bibliográfica*. Consejo Nacional de Televisión. <https://cntv.cl/estudio-repositorio/desinformacion-revision-bibliografica/>

- Correa, J., & Murillo, J. (2015). *Escritura e investigación académica: Una guía para la elaboración del trabajo de grado: Segunda Edición*. CESA.  
<https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2014981&lang=es&site=eds-live&scope=site>
- Espinoza, E., Mazuelos, C. (2020). *Desinformación sobre temas de salud en las redes sociales*.  
[http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2307-21132020000200002#B10](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-21132020000200002#B10)
- Fischer, M., Haque, R., Stynes, P., & Pathak, P. (2022). *Identifying fake news in Brazilian Portuguese*. En *Lecture notes in computer science* (pp. 111-118).  
[https://doi.org/10.1007/978-3-031-08473-7\\_10](https://doi.org/10.1007/978-3-031-08473-7_10)
- Garg, S., & Sharma, D. (2020). *Phony news detection using machine learning and deep-learning techniques*. Institute of Electrical and Electronics Engineers Inc., 27-32.  
<https://doi.org/10.1109/smart50582.2020.9337120>
- Gifu, D. (2023). *An intelligent system for detecting fake news*. *Procedia Computer Science*, 221, 1058-1065. <https://doi.org/10.1016/j.procs.2023.08.088>
- Guallar, J., Codina, L., Pere, F., & Pérez-Montoro, M. (2020). *Desinformación, bulos, curación y verificación. Revisión de estudios en Iberoamérica 2017-2020*.  
<https://www.redalyc.org/journal/993/99364425011/>
- Hernández, R., & Mendoza, C. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. McGrawhill. <https://www-ebooks7-24-com.bibliotecavirtual.unad.edu.co/stage.aspx?il=6443&pg=&ed=>
- Hu, B., Mao, Z., & Zhang, Y. (2024). *An overview of fake news detection: From a new perspective*. *Fundamental Research*. <https://doi.org/10.1016/j.fmre.2024.01.017>

- Jain, K., Bansal, H., Saxena, G., & Kumar, R. (2023). *Fake news detection using attention based model*. Institute of Electrical and Electronics Engineers Inc., 641-645.  
<https://doi.org/10.1109/spin57001.2023.10117332>
- Jiménez, C. (2018, 29 de octubre). *Toda la verdad sobre la desinformación | Clara Jiménez | TEDxMadrid* [video]. YouTube. [https://youtu.be/2\\_xr58zYr6c?si=kEd8g45j3zrY8fyt](https://youtu.be/2_xr58zYr6c?si=kEd8g45j3zrY8fyt)
- Lin, J., Tremblay-Taylor, G., Mou, G., You, D., & Lee, K. (2019). *Detecting fake news articles*. Institute of Electrical and Electronics Engineers Inc., 3021-3025.  
<https://doi.org/10.1109/bigdata47090.2019.9005980>
- Liu, H. (2019). *A location independent machine learning approach for early fake news detection*. Institute of Electrical and Electronics Engineers Inc., 4740-4746.  
<https://doi.org/10.1109/bigdata47090.2019.9005962>
- Mayopu, R. G., Wang, Y.-Y., & Chen, L.-S. (2023). *Analyzing online fake news using latent semantic analysis: Case of USA election campaign*. Big Data and Cognitive Computing, 7(2), 81. <https://doi.org/10.3390/bdcc7020081>
- Meesad, P. (2021). *Thai fake news detection based on information retrieval, natural language processing and machine learning*. SN Computer Science, 2(6).  
<https://doi.org/10.1007/s42979-021-00775-6>
- Nair, J., Akhil, S. S., & Harisankar, V. (2022). *Fake news detection model for regional language*. 2022 IEEE 7th International conference for Convergence in Technology (I2CT).  
<https://doi.org/10.1109/i2ct54291.2022.9824641>
- Narkhede, A., Patharkar, D., Chavan, N., Agrawal, R., & Dhule, C. (2023). *Fake news detection using machine learning algorithm*. Institute of Electrical and Electronics Engineers Inc., 166-170. <https://doi.org/10.1109/iccsai59793.2023.10421153>

Ozdemir, S. (2016). *Principles of data science*. Packt Publishing.

<https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1441463&lang=es&site=ehost-live>

Peset, F., & Millán, L. (2017). *Ciencia abierta y gestión de datos de investigación (RDM)*.

Ediciones Trea, (pp. 141-158). <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/117484>

Probierz, B., Stefański, P., & Kozak, J. (2021). *Rapid detection of fake news based on machine learning methods*. *Procedia Computer Science*, 192, 2893-2902.

<https://doi.org/10.1016/j.procs.2021.09.060>

Raja, M. S., & Raj, L. A. (2022). *Fake news detection on social networks using machine learning techniques*. *Materials Today: Proceedings*, 62, 4821-4827.

<https://doi.org/10.1016/j.matpr.2022.03.351>

Salem, F. K. A., Feel, R. A., Elbassuoni, S., Ghannam, H., Jaber, M., & Farah, M. (2021). *Meta-learning for fake news detection surrounding the Syrian war*. *Patterns*, 2(11), 100369.

<https://doi.org/10.1016/j.patter.2021.100369>

Sandrilla, R., & Savitha, M. (2022). FNU-BiCNN: *Fake news and fake URL detection using Bi-CNN*. *Thesai.org*. [https://thesai.org/Downloads/Volume13No2/Paper\\_56-](https://thesai.org/Downloads/Volume13No2/Paper_56-FNU_BiCNN_Fake_News_and_Fake_URL_Detection.pdf)

[FNU\\_BiCNN\\_Fake\\_News\\_and\\_Fake\\_URL\\_Detection.pdf](https://thesai.org/Downloads/Volume13No2/Paper_56-FNU_BiCNN_Fake_News_and_Fake_URL_Detection.pdf)

Sastrawan, I. K., Bayupati, I. P. A., & Arsa, D. M. S. (2022). *Detection of fake news using deep learning CNN–RNN based methods*. *ICT Express*, 8(3), 396-408.

<https://doi.org/10.1016/j.icte.2021.10.003>

Sudhakar, M., & Kaliyamurthie, K. P. (2024). *Detection of fake news from social media using support vector machine learning algorithms*. *Measurement. Sensors*, 32, 101028.

<https://doi.org/10.1016/j.measen.2024.101028>

Torabi, F., & Taboada, M. (2019). *Big data and quality data for fake news and misinformation detection*. *Big Data & Society*, 6(1), 205395171984331.

<https://doi.org/10.1177/2053951719843310>