Muertes de celebridades: Sobre la clasificación automática de noticias falsas no intencionadas

Celebrity deaths: On the automatic classification of unintentional fake news

Fabián Riquelme[1*] https://orcid.org/0000-0003-4491-0148
Eduardo Puraivan[2, 1] https://orcid.org/0000-0003-2134-8922
Diego Rivera[3] https://orcid.org/0000-0002-3628-7053
Magaly Varas[4] https://orcid.org/0000-0002-0016-8028
René Venegas[5] https://orcid.org/0000-0001-5572-651X
Claudia Mellado[6] https://orcid.org/0000-0002-9281-1526
Anaís Berríos[7] https://orcid.org/0009-0009-7709-9559
Kristel Hidalgo[6] https://orcid.org/0009-0002-9188-2290
Ángel Inostroza[1] https://orcid.org/0009-0008-6048-6264
Jean Carlos Órdenes[5] https://orcid.org/0000-0001-9765-8551
Nicolás Riquelme[8] https://orcid.org/0009-0008-2154-8506

Recibido 12 de octubre de 2024, aceptado 27 de noviembre de 2024
Received: October 12, 2024 Accepted: November 27, 2024

RESUMEN

Este estudio aborda el creciente problema del desorden de información, en particular, en el contexto de los anuncios de muertes de celebridades. A medida que la rápida propagación de la desinformación se convierte en un problema social crítico, especialmente a través de los medios de comunicación masiva y las redes sociales, es esencial comprender los mecanismos detrás de este fenómeno. La investigación emplea un enfoque interdisciplinario, que integra perspectivas computacionales, lingüísticas, sociológicas y periodísticas para analizar las características de las noticias falsas no intencionadas. Mediante técnicas de clasificación de aprendizaje automático, se busca diferenciar entre noticias falsas no intencionadas, noticias reales que desacreditan estas afirmaciones falsas y noticias reales. Los hallazgos revelan características lingüísticas significativas que contribuyen al proceso de clasificación. Sin embargo, si bien existen modelos capaces de clasificar ciertos tipos de noticias específicos, ninguno de ellos es capaz de clasificar correctamente todos los tipos de noticias considerados, lo que subraya las complejidades involucradas en la distinción entre información correcta, errónea y desinformación. Este trabajo no solo arroja luz sobre la naturaleza de las noticias falsas no intencionadas, sino que también enfatiza la necesidad de mejorar los procesos de verificación en el periodismo para combatir la propagación viral de información falsa. En última instancia, el estudio exige más investigaciones sobre las implicaciones

[1] Universidad de Valparaíso. Escuela de Ingeniería Informática. Valparaíso, Chile. Email: fabian.riquelme@uv.cl; angel.inostroza@alumnos.uv.cl
[2] Universidad Viña del Mar. Escuela de Ciencias. Viña del Mar, Chile. E-mail: epuraivan@uvm.cl
[3] Pontificia Universidad Católica de Chile. Instituto de Sociología. Santiago, Chile. E-mail: dirivera1@uc.cl
[4] Universidad Viña del Mar. Escuela de Arquitectura, Comunicaciones y Diseño. Viña del Mar, Chile. E-mail: magaly.varas@uvm.cl
[5] Pontificia Universidad Católica de Valparaíso. Instituto de Literatura y Ciencias del Lenguaje. Valparaíso, Chile. E-mail: rene.venegas@pucv.cl; jean.ordenes@pucv.cl
[6] Pontificia Universidad Católica de Valparaíso. Escuela de Periodismo. Valparaíso, Chile. E-mail: claudia.mellado@pucv.cl; kristel.hidalgo.e@mail.pucv.cl
[7] Universidad de Valparaíso. Escuela de Sociología. Valparaíso, Chile. E-mail: anais.berrios@alumnos.uv.cl
[8] Pontificia Universidad Católica de Chile. Departamento de Ciencia de la Computación. Santiago, Chile. E-mail: nariquelme@uc.cl
* Autor de correspondencia: fabian.riquelme@uv.cl

de estos hallazgos para las prácticas de los medios de comunicación y el papel de la tecnología a la hora de abordar los desafíos del desorden de la información en la sociedad contemporánea.

Palabras clave: Noticia falsa, comunicación de masas, aprendizaje automático, procesamiento de lenguaje natural.

ABSTRACT

This study addresses the growing problem of information disorder, particularly in the context of celebrity death announcements. As the rapid spread of disinformation becomes a critical societal issue, particularly through mass media and social media, it is essential to understand the mechanisms behind this phenomenon. The research employs an interdisciplinary approach, integrating computational, linguistic, sociological, and journalistic perspectives to analyze the characteristics of unintentional fake news. Using machine learning classification techniques, we seek to differentiate between unintentional fake news, real news that debunks these false claims, and actual news. The findings reveal significant linguistic features that contribute to the classification process. However, although there are models capable of classifying certain types of news, none can accurately classify all the types considered, highlighting the complexities involved in distinguishing between accurate information, misinformation, and disinformation. This work not only sheds light on the nature of unintentional fake news but also emphasizes the need to improve verification processes in journalism to combat the viral spread of false information. Ultimately, the study calls for further research into the implications of these findings for media practices and the role of technology in addressing the challenges of information disorder in contemporary society.

Keywords*: Fake news, mass media, machine learning, natural language processing.*

INTRODUCCIÓN

Una noticia falsa (en inglés, *fake news*) es un mensaje que ha sido creado deliberadamente para desinformar, incluyendo información falsa dentro de un contexto general de veracidad, con el fin de manipular la opinión pública de ciertos grupos sociales. De acuerdo con la UNESCO[1], este término es en sí mismo un oxímoron, ya que una noticia debiese ser por definición una noticia verificable. Las noticias falsas son un ejemplo de desinformación, fenómeno que combina deliberadamente información maliciosa (en inglés, *malinformation*) e información errónea (en inglés, *misinformation*) [2]. Esto se ilustra mejor en la Figura 1. La información maliciosa es aquella creada exclusivamente para difamar o dañar; por ejemplo, los discursos de odio, acosos, y filtración de información sensible y privada. La información errónea se refiere a aquella que posee contenido engañoso o falso, pero no deliberada, debido a malas interpretaciones o traducciones, errores de tipeo u otro tipo de problemas de sus emisores. Todos estos tipos de problemas informativos caben dentro del concepto más general de desorden de la información (en inglés, *information disorder*) [3]. Si bien las noticias falsas han existido desde hace siglos [4], la ubicuidad e instantaneidad de la tecnología actual le permiten una masificación en tiempo real a nivel global, lo que representa un problema crítico para las sociedades a distintos niveles, tales como el político [5], el sanitario [6] o el societal [7].

En este contexto, el objeto de estudio de este trabajo, mucho menos estudiado que las noticias falsas desinformativas, son lo que llamaremos *noticias falsas no intencionadas*, un tipo de información errónea que se produce al adaptar y compartir una noticia basada en una noticia falsa, creyendo que dicha fuente original es veraz. El frecuente uso de Internet y redes sociales como única fuente de información conduce a una pérdida de calidad de contenido informativo, que dificulta a los periodistas verificar adecuadamente los hechos y circunstancias que se publican [8]. Este tipo de fenómeno se da cada vez con mayor frecuencia en los medios de comunicación masivos (en inglés, *mass media*), donde las necesidades económicas y recursos escasos llevan a entornos editoriales acelerados, en que los

Fuente: Basado en [3], [4].

Figura 1. Taxonomía del desorden de la información.

comunicadores no tienen tiempo suficiente para verificar los hechos noticiosos. Lo anterior, puede llevar a publicaciones incorrectas que alimentan la difusión de bulos o engaños, afirmaciones erróneas y rumores no verificados [9]. Las noticias falsas no intencionadas suelen ampliar y complementar el contenido de la noticia falsa original, transformando lo que era una simple publicación de Twitter/X o imagen de Instagram en una aparente nota completa de prensa informativa. Este problema contribuye a la rápida viralización y consolidación de noticias falsas, dificultando aún más su detección [10].

Un caso paradigmático de noticias falsas no intencionadas son los anuncios de muertes de celebridades. Tal es el caso de la noticia de la falsa muerte de Chomsky en junio de 2024, que contrasta con las de otras noticias reales de impacto mediático equivalente, como por ejemplo, la muerte de Stephen Hawking en marzo de 2018.

Al no existir claridad sobre las autorías de este tipo de noticias falsas ni de sus motivos principales, la literatura comparada plantea algunas hipótesis. Desde la psicología política y conductual, estudios remiten al regreso de la idea de la búsqueda de una certeza epistemológica [11], es decir, tener la certeza de que se sabe algo que "nadie más sabe", especialmente útil al pensar en noticias con un carácter conspiranoico, y cómo las y los lectores se involucran con dicha información. Otros trabajos plantean una posibilidad de monetización de contenido, en tiempos en que el consumo de noticias se relaciona de forma directamente proporcional con las métricas de alcance y uso [12], [9], [8], [13]. Se podría sugerir, incluso, que son parte de experimentos sociales, ya que las razones no son tan claras, a diferencia de otras noticias falsas de carácter político o de teorías de conspiración.

Es en su carácter inespecífico en el que encontramos la posibilidad analítica de pensar este problema. En el caso de noticias de muerte de celebridades, no parece haber una finalidad de instalación o manipulación de discursos, pero la intención del creador inicial, sea o no con fines de "experimentación", es igualmente consciente y genera efectos negativos, así que la noticia original sí califica como noticia falsa y, por tanto, como desinformación. En cambio, las noticias de los medios masivos que amplían y amplifican ese contenido falso, originalmente posteado usualmente en redes sociales, califican como información errónea o *misinformation*.

El fenómeno también puede ser abordado desde la urgencia por la primicia periodística [14].

Específicamente, desde la teoría sociológica se ha presentado a la prensa como un campo [15] bajo el cual prima la competencia, provocando una necesidad por mantenerse *ahead of the curve* [16]. Si bien es difícil ligarlo a la generación intencional de contenido falso y desinformación, puede colaborar a entender su rápida y amplia transmisión a otros medios y plataformas de redes sociales.

Actualmente, se distinguen cinco tipos de métodos para la clasificación de noticias falsas [17]. Dos de estas técnicas son manuales. El *fact-checking* o verificación de hechos es una práctica periodística en la que expertos en el tópico de cierta noticia analizan su veracidad, recabando información acerca del hecho noticioso. Es una práctica efectiva de verificación, pero poco eficiente en tiempo y recursos humanos. Debido a su lentitud de verificación, no impide que la noticia falsa se propague y alcance cierto grado de manipulación en el inconsciente colectivo. La otra técnica manual es recurrir al *crowdsourcing*, tarea similar a la anterior, pero realizada por un equipo colaborativo, no necesariamente expertos en el tema. Esta técnica puede ganar en tiempos de respuesta, pero disminuir la calidad de la verificación.

Otras dos técnicas son automatizadas, ganando en eficiencia al poder entregar resultados casi en tiempo real, pero perdiendo en efectividad. La primera es el uso de técnicas de procesamiento de lenguaje natural o *natural language processing (NLP)*, basada en la aplicación de rasgos lingüísticos y otras técnicas de lingüística computacional para reconocer aquellas características léxicas, sintácticas, gramaticales y semánticas que distinguen una noticia real de una falsa. La segunda refiere a un conjunto de técnicas de aprendizaje automático o *machine learning (ML),* y de aprendizaje profundo o *deep learning (DL)*, que corresponden al uso de inteligencia artificial y estadísticas avanzadas para, mediante un gran conjunto de datos de entrenamiento, enseñarle a un algoritmo a identificar patrones de dichos datos, para así ser capaz de clasificar nuevas noticias, con un cierto grado de exactitud, precisión y confianza. Son muchos los métodos de ML y DL que existen actualmente [9], [18]. Sin embargo, sus principales problemas son, por un lado, la dependencia de contextos (temática, idioma, temporalidad, circunstancias históricas, etc.) al momento de entrenar dichas noticias, esto es, que solo aprenden a clasificar los tipos específicos de noticias con los cuales originalmente se entrenaron, y por otro, la baja interpretabilidad de sus resultados, esto es, la escasa capacidad de comprender por qué una noticia que fue etiquetada como falsa, en realidad lo es, y qué la distingue de una noticia real. Así, el último método es en realidad un conjunto de métodos híbridos; una combinación de los anteriores. Por ejemplo, el uso de rasgos lingüísticos propios de NLP se pueden utilizar para ayudar a dar mayor interpretabilidad a los modelos de ML [19].

Este trabajo tiene por objetivo clasificar automáticamente noticias falsas no intencionadas. Para ello, se aborda el problema de la detección y caracterización automatizada de noticias falsas no intencionadas, empleando un enfoque híbrido basado en modelos de ML y rasgos lingüísticos de NLP. El trabajo se centra, además, en noticias escritas originalmente en español, un idioma de amplio uso en Internet, siendo la tercera lengua más utilizada, aunque bastante menos explorada que el inglés. Su análisis presenta complejidades, dado que cuenta con características lingüísticas propias, como una mayor cantidad de alternativas flexivas y derivativas a nivel morfológico y una mayor flexibilidad a nivel sintáctico [20], lo que implica la indagación de sus rasgos léxicos y gramaticales en profundidad en los textos escritos. Con ello, se evita la pérdida de información sustantiva al solo traducir desde el inglés, solución muy frecuente en este ámbito de investigación. Como caso de estudio, se analizará el fenómeno de las noticias sobre fallecimientos de celebridades, contrastando noticias falsas no intencionadas con las noticias reales que las desmienten, así como con noticias de fallecimientos reales. Como resultados esperados, se pretenden identificar características léxicas, gramaticales y discursivas que permitan distinguir automáticamente noticias reales y falsas no intencionadas.

A continuación, se continúa con el marco de referencia, desde un enfoque interdisciplinario que cubre aristas periodísticas, sociológicas, informáticas y tecnológicas sobre la detección de noticias falsas, desinformación e información errónea. Posteriormente, se detalla la metodología del trabajo, donde se presenta el diseño experimental, para continuar con la configuración experimental basada en un conjunto de datos reales. Luego, se presentan y discuten los resultados obtenidos, y se presentan las principales conclusiones y trabajo futuro del trabajo.

TRABAJO RELACIONADO

Actualmente, es un hecho ampliamente aceptado que el mundo se encuentra dominado por una "infocracia", la cual alude a un régimen de la información controlado mediante algoritmos que analizan los datos de los usuarios humanos [21]. En este contexto, se produce una verdadera "guerra por la información", ya que las distintas voces que interactúan en el espacio público virtual buscan dominar a las demás. Estos mensajes son emitidos por diversos actores desde sus posiciones en los distintos campos sociales [22] y, por tanto, reflejan las relaciones de poder que ejercen. En este contexto, las estrategias que utilizan los actores que componen esos campos de dominio (político, económico, cultural) son de variada naturaleza, empleando como canal los medios de comunicación y, más recientemente, las plataformas digitales que han emergido los últimos 30 años.

Desde hace años que la relación entre la comunicación digital y las noticias falsas es indisociable [17]. Diversos investigadores han intentado comprender esta articulación, refiriéndose con frecuencia a la esfera pública del presente como de "posverdad" [23], [18], [24] fenómeno para el cual se han propuesto actualizaciones y nuevos programas de investigación [25].

Pensar en muertes de personas famosas como noticias falsas es relevante en el sentido de preguntarse por qué estos hechos son posibles y en qué contextos se comparten noticias de este estilo. La sociología introduce para este tipo de procesos el concepto de mercado de atención [26], que busca ayudar a comprender la generación y consumo de cierto tipo de contenido por sobre otro. Hay evidente consenso en que durante la pandemia aumentó el consumo de este tipo de contenidos, fomentado por emociones exacerbadas, desórdenes comunicativos por cámaras de eco [27], filtros burbujas [28] e, incluso, intencionalidad. El enfoque de este trabajo se alinea con el de otros, que permiten pasar desde la preocupación por la amenaza del tipo de contenido hacia la comprensión de por qué en determinados momentos se comparte cierto contenido y no otro [29].

Desde el punto de vista computacional y tecnológico, las mayores capacidades de cómputo y de recolección de datos, sumada a los últimos avances en inteligencia artificial, han permitido avanzar hacia soluciones parciales en contextos específicos para la detección y clasificación de diversos tipos de desórdenes de información. Sin embargo, han sido estas mismas tecnologías las que han permitido una mayor sofisticación y efectividad en la creación de noticias falsas e información maliciosa [30]. En el caso particular de la información errónea o *misinformation*, los métodos de aprendizaje automático se han aplicado en múltiples contextos, tales como crisis sociales [31] o sanitarias [32]. Uno de los principales retos actuales es transitar hacia métodos híbridos que utilicen técnicas de inteligencia artificial explicables [3]. En particular, la combinación de aprendizaje automático y procesamiento de lenguaje natural parecen una vía apropiada para la búsqueda de mayor interpretabilidad de los resultados de análisis [19].

METODOLOGÍA

Para este trabajo se usará una metodología híbrida, utilizando técnicas de clasificación y análisis cuantitativos, propios del aprendizaje automático y la lingüística computacional, así como técnicas cualitativas de análisis del discurso. Lo anterior permitirá un enfoque no solo descriptivo sino también explicativo, soportado por una mayor interpretabilidad de los resultados de análisis. Concretamente, más que el solo clasificar una noticia como real o falsa, un problema complejo muy dependiente del contexto y de la calidad de los datos, se busca también comprender las posibles diferencias lingüísticas y discursivas entre noticias reales y falsas no intencionadas.

Este trabajo se propone clasificar noticias de muertes de celebridades, dentro de las cuales abundan noticias falsas no intencionadas. El procedimiento general se describe en la Figura 2 y se ejemplifica en la Figura 3, siendo ambos detallados a continuación.

Recolección de datos

Las técnicas de aprendizaje automático supervisado requieren de etapas de entrenamiento, prueba (o testing) y validación. Para las etapas de entrenamiento y prueba se utilizó la base de datos ya existente CLNews, empleada para la clasificación de rumores falsos en español [3]. CLNews contiene 300 árboles de propagación de contenido en Twitter (hoy X), relacionados con *trending topics* de eventos sociales, políticos y deportivos de Chile. Estos árboles,

Fuente: Elaboración propia.

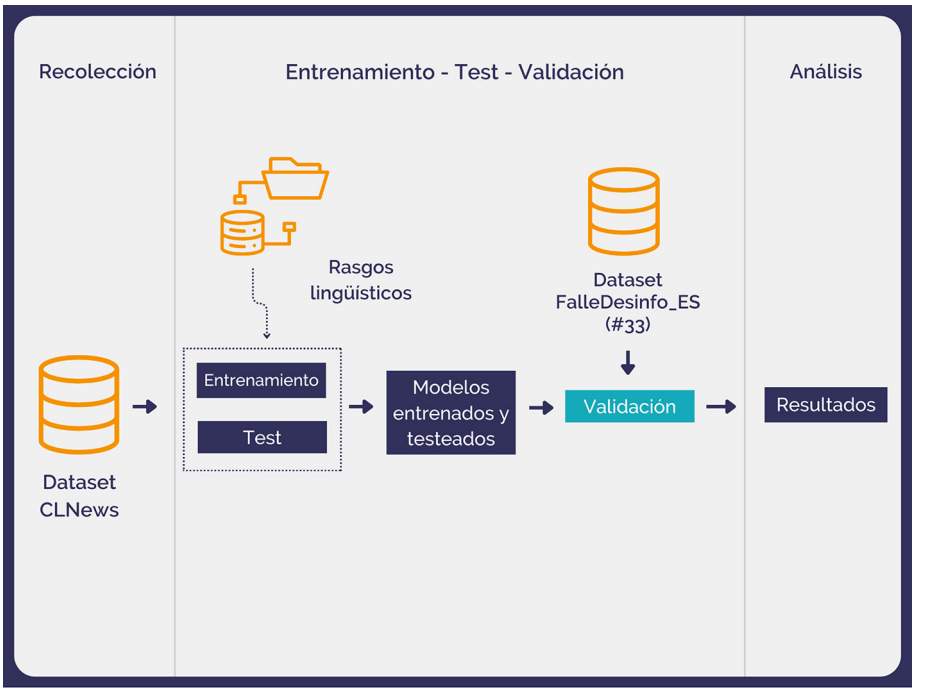

Figura 2. Proceso de manipulación de los datos.

Fuente: Elaboración propia.

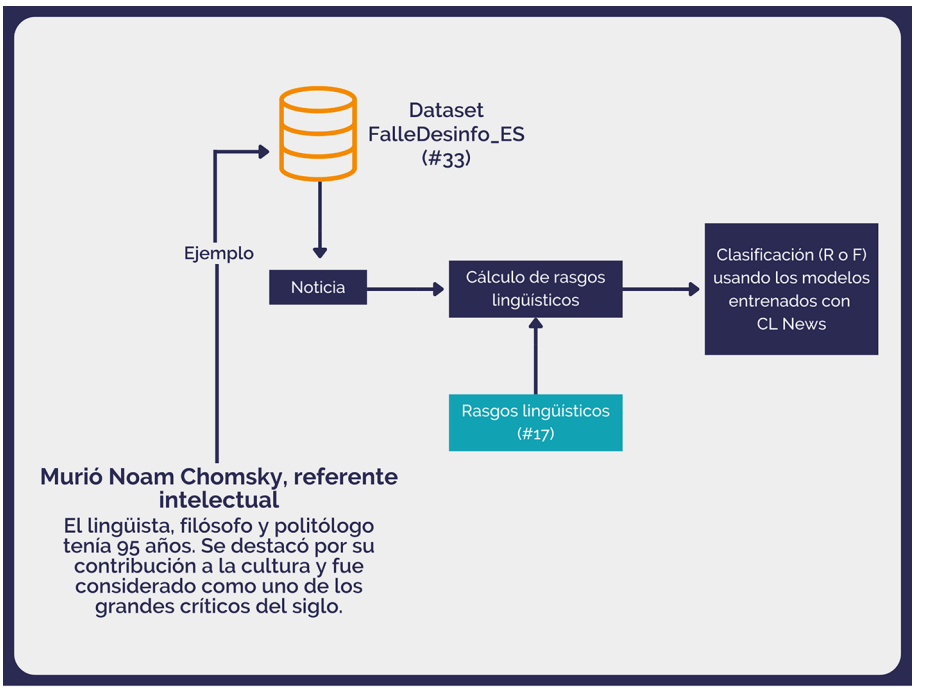

Figura 3. Ejemplo de proceso de manipulación de los datos.

recolectados entre junio de 2019 y enero de 2020, fueron etiquetados manualmente, quedando en 87 rumores verdaderos, 53 falsos, 49 no verificables, y 111 no rumores. Nótese que tras estas etapas de entrenamiento y prueba, se obtuvo un modelo listo para ser validado con nuevas noticias.

Para la etapa de validación, el conjunto de datos o *dataset* se construyó manualmente a partir de una serie de noticias de muertes de celebridades, escritas exclusivamente en español, sin traducciones, y publicadas en medios de habla hispana. Se consideraron tres tipos de noticias, provenientes de medios de prensa en línea, incluyendo título, bajada, y cuerpo de las mismas:

– T1: Noticias falsas no intencionadas.
– T2: Noticias reales, que desmienten la información errónea de T1.
– T3: Noticias reales.

El grupo T2 se plantea como una novedad metodológica, ya que se trata de una noticia real, pero con características distintas a las que tendría una noticia real sobre la muerte de una persona. El hecho noticioso, en estos casos, no es el fallecimiento de una celebridad, sino el que se haya desmentido el fallecimiento de esta.

Cada grupo contenía la misma cantidad de noticias. Para una misma celebridad, las noticias de T1 y T2 deberán tener la misma fecha de publicación, para evitar que noticias de días posteriores puedan ampliar el contenido agregando contextos adicionales que generen ruido adicional a las comparaciones. Cabe destacar que el conjunto de datos de validación no necesita ser grande, ya que esto no representa una limitación teórica ni práctica para el propósito descrito. El conjunto de datos de validación utilizado, llamado FalleDesinfo_ES, se detalla más adelante en la sección de configuración experimental.

Análisis de datos

Para la clasificación de noticias se utilizaron cinco modelos clásicos de aprendizaje automático supervisado, basados en rasgos lingüísticos [19] para el idioma español: *support vector machine* (SVM), de kernel lineal y con *radial basis function* (RBF); *random forest* (RF), XGBoost, y *logistic regression* (LR) [33]. Como ya se mencionó anteriormente, estos modelos se entrenaron y probaron con la base de datos CLNews, y se validaron con el nuevo conjunto de datos FalleDesinfo_ES.

Los rasgos lingüísticos son métricas implementadas algorítmicamente y que permiten cuantificar características específicas de un texto, tales como: estadísticas relacionadas a sus palabras (a todas ellas o a ciertos tipos específicos, como adjetivos, adverbios, sustantivos, etc.), oraciones o párrafos; conteos de términos pertenecientes a ciertos diccionarios específicos, entre otros. En este trabajo se consideran diversas métricas ya existentes, basadas en trabajos previos [19] e implementadas específicamente para el idioma español, identificadas a través de la plataforma de análisis cuantitativo de textos en español PACTE [45].

Para comparar las noticias, se aplicaron los clasificadores y rasgos lingüísticos sobre todas las noticias de cada tipo (T1, T2 y T3). Luego, se compararon los resultados de las clasificaciones: dentro de cada tipo, entre tipos y entre clasificadores.

Los clasificadores permiten etiquetar cada noticia del conjunto de datos como "real" o "falsa". El rendimiento de estos clasificadores se evalúa mediante las métricas tradicionales de *accuracy* (exactitud), *precision* (precisión), *recall* (exhaustividad) y *F1-score*, indicadas en ecuaciones (1)-(4):

$$accuracy = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

$$precision = TP / (TP + FP) \quad (2)$$

$$recall = TP / (TP + FN) \quad (3)$$

$$F1 = 2\ precision / (precision + recall) \quad (4)$$

donde *TP* corresponde a los verdaderos positivos (*true positives*), *TN* a los verdaderos negativos (*true negatives*), *FP* a los falsos positivos (*false positives*), y *FN* a los falsos negativos (*false negatives*).

Estas métricas permiten saber el grado de confiabilidad con que el clasificador realizó su etiquetado. Cabe destacar que el F1-score se concibe como la media armónica entre precisión, el valor que considera los ejemplos correctos identificados (TP) y los ejemplos identificados que corresponden a otra clase (FP); y exhaustividad, el valor que considera los ejemplos correctos identificados (TP) y aquellos que no se

lograron identificar para la clase respectiva. Ello permite combinar en un solo valor ambas medidas, lo que resulta práctico para comparar el rendimiento combinado de la precisión y la exhaustividad entre varios resultados de clasificación.

En cuanto a los rasgos lingüísticos, estos arrojan valores numéricos. Como parte del análisis de los clasificadores, identificar los rasgos lingüísticos más relevantes nos ayuda a identificar las características más significativas para las predicciones del modelo. Para un modelo SVM lineal y la regresión logística, estos valores se basan en los coeficientes del modelo, así valores de mayor magnitud significan mayor relevancia. En un Random Forest se mide cuánta información aporta cada característica al reducir la incertidumbre en los árboles. XGBoost evalúa la importancia según la frecuencia con que cada característica se usa para dividir los datos en los árboles del modelo. Para el SVM con kernel RBF, se utiliza la permutación de importancia, que mide cuánto cae la precisión del modelo cuando se alteran los valores de una característica. Estas métricas nos permiten identificar las características clave que influyen en los resultados del modelo.

Debido a la naturaleza de las noticias, se espera encontrar diferencias significativas entre las noticias de los grupos T1 y T2. En cambio, las diferencias entre los grupos T1 y T3 puede que sean menos evidentes.

CONFIGURACIÓN EXPERIMENTAL

A continuación, se presentará la aplicación del diseño experimental visto en la sección anterior. Se detallan el conjunto de datos, el rendimiento de los clasificadores automáticos previamente testeados, y una descripción de los rasgos lingüísticos que resultaron significativos para este trabajo y que tienen implicancias en la siguiente sección de Resultados.

Conjunto de datos

Para el conjunto de datos de validación de los clasificadores, se recolectaron manualmente 11 noticias de cada tipo, totalizando 33 noticias en total. Este conjunto de datos, denominado FalleDesinfo_ES, ha sido publicado para su libre disposición [34]. El grupo T1 está conformado por noticias relacionadas con el supuesto fallecimiento del lingüista e intelectual Noam Chomsky, difundidas el 18 de junio de 2024 y que amplificaban una noticia falsa inicialmente difundida por plataformas de redes sociales (Twitter/X y Facebook). El grupo T2 contiene noticias publicadas ese mismo día y que desmienten lo anterior. Más aún, algunos medios de difusión de T1 y T2 se repiten. Finalmente, el grupo T3 contiene noticias reales sobre el fallecimiento del físico Stephen Hawking, el 14 de marzo de 2018.

La Figura 4 muestra el interés de búsqueda mundial normalizado despertado en Internet por las figuras de Stephen Hawking y Noam Chomsky de acuerdo con Google Trends, entre una semana antes y después de las noticias de sus decesos. El decrecimiento menos abrupto para el caso de Chomsky se debe a las noticias de los días siguientes que desmintieron su fallecimiento.

La Figura 5 muestra la geolocalización de las mismas búsquedas durante el mismo lapso de tiempo. Las búsquedas normalizadas de Hawking las lidera Honduras, Paraguay, Reino Unido, Australia y Panamá.

Fuente: Google Trends.

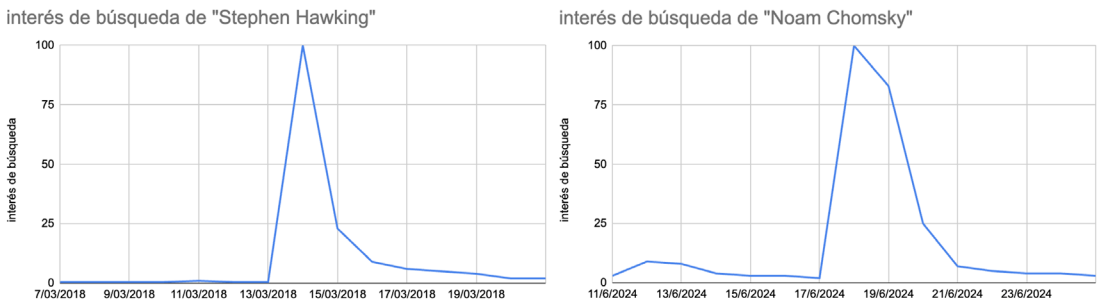

Figura 4. Interés de búsquedas de Stephen Hawking (izquierda) y Noam Chomsky (derecha) en Internet, incluyendo la fecha de la noticia de sus fallecimientos.

Fuente: Google Trends.

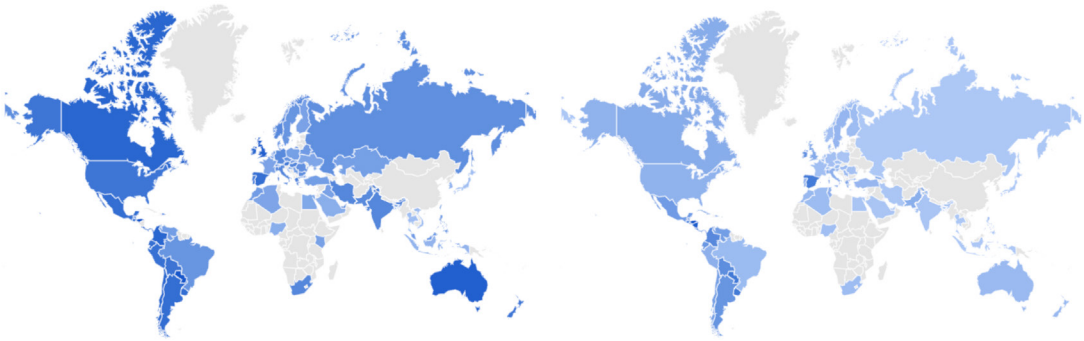

Figura 5. Geolocalización de búsquedas de Stephen Hawking (izquierda) y Noam Chomsky (derecha) en Internet durante el mismo lapso de tiempo anterior.

Para Chomsky, los diez primeros países son todos hispanoamericanos, liderando la lista Honduras, Nicaragua, El Salvador, Paraguay y Uruguay.

Cabe mencionar que todos los países presentes en la lista han enfrentado o estado expuestos a campañas de desinformación, especialmente en contextos de elecciones y crisis sociales, e incluso, en algunos casos, han sido llevadas a cabo (y financiadas) por agencias de gobierno [35], [36]. Esto con los objetivos de influir la opinión pública, conseguir mayor apoyo de la ciudadanía y/o dañar la reputación de adversarios políticos. Es en este ecosistema digital que resaltan los casos de Honduras [37] y Nicaragua [38], en que las estrategias de desinformación están sistematizadas e institucionalizadas, y se utilizan principalmente las técnicas de cybertroop [36] y farm troll [39]. Así es como en esta coyuntura se podría conjeturar que la constante exposición de los hondureños a fake news, haya generado cierto escepticismo que los lleva a cuestionar la credibilidad de determinadas fuentes (redes sociales, principalmente) y a verificar la información en otros medios, utilizando para ello el buscador Google. Serían necesarios más estudios focalizados para poder comprender mejor este ranking.

Los medios de las noticias recolectadas en el conjunto de datos FalleDesinfo_ES para cada grupo fueron los siguientes:

– T1: Continental, Debate, e-consulta, El Cronista, Página12, La Tercera (2), Línea Directa, LT10 (fuente original: Cadena 3), Radio UChile, RDN (fuente original: ABC).
– T2: API, Cadena 3, Debate, El Comercio de Perú, El Diario de la República, El Periodista, La Tercera, MSN, Portafolio, T13, Wired.
– T3: As, Clarín, CNN Español, DW, El Mundo, Euro News, El Heraldo, Infobae, La Tercera, Público, T13.

Algunas noticias de T1 debieron extraerse del historial de sitios web almacenados en Internet Archive (archive.org), mientras que llama la atención que otras siguen disponibles en línea hasta el momento de la escritura de este artículo.

Clasificadores automáticos

Como ya se mencionó anteriormente, los cinco clasificadores utilizados fueron entrenados previamente con la base de datos CLNews de rumores falsos en español [40]. Se utilizó un 90% de los datos para entrenamiento y un 10% para testing, obteniendo para estos últimos los rendimientos ilustrados en la Tabla 1. De entre todos, destaca el modelo SVM lineal, quedando RF y XGBoost con rendimientos bastante más discretos que los demás. Esta información es relevante al momento de interpretar los resultados obtenidos en la siguiente sección, para los nuevos datos de validación.

Rasgos lingüísticos

Los rasgos lingüísticos considerados pueden agruparse en cinco categorías: variables de superficie, variables de categorías gramaticales, variables

Tabla 1. Rendimiento de clasificadores testeados previamente.

Modelo	Accuracy	Precisión	Recall	F1-Score
SVM (lineal)	0,857	0,883	0,857	0,846
SVM (RBF)	0,714	0,802	0,714	0,645
RF	0,714	0,802	0,714	0,645
XGBoost	0,643	0,629	0,643	0,632
LR	0,643	0,629	0,643	0,632

léxicas con función discursiva, léxico de polaridad y léxico de emociones y sentimientos.

Las variables de superficie consideran las unidades lingüísticas que se pueden reconocer en la superficie del texto, como son las oraciones, párrafos, palabras y caracteres. Cada una de estas unidades pueden medirse en relación con la cantidad total, máxima, mínima y mediana. Para la primera categoría se consideran cinco rasgos: cantidad total de oraciones, cantidad mínima y máxima de palabras, cantidad máxima de caracteres y cantidad máxima de caracteres sin espacios.

Las variables de categorías gramaticales refieren a la clasificación de palabras del texto. Entre ellas puede mencionarse a verbos, sustantivos, adjetivos, adverbios, entre otras. Al igual que la variable anterior, pueden medirse con relación a la cantidad total, máxima, mínima, media y mediana. A estas se agrega la desviación estándar, la idf (*inverse document frequency*) y la tf-idf (*term frequency–inverse document frequency*). Para la segunda categoría se consideran ocho rasgos: cantidad máxima, mínima, media y mediana de preposiciones, cantidad máxima de auxiliares y tf-idf, desviación estándar de numerales e idf de sustantivos propios.

Las variables léxicas con función discursiva se enfocan en aquellas palabras que cumplen una función discursiva dentro del texto como marcadores o modalizadores. Se miden mediante la cuantificación total. Para esta tercera categoría se analizan dos rasgos: marcadores discursivos textuales contraargumentativos y conectores de opinión generalizadores.

El léxico de polaridad refiere a aquellas palabras asociadas a una polaridad emocional. Serán de polaridad negativa cuando se asocian a una emoción negativa y de polaridad positiva cuando la emoción sea positiva. La clasificación de polaridad se realiza mediante diccionarios que presentan diversos grados. Para el rasgo de esta categoría, se utiliza el diccionario creado por Finn Årup Nielsen (Affin). En la cuarta categoría se destaca el rasgo muy positivo Affin.

Finalmente, el léxico de emociones y sentimientos consiste en aquellas palabras asociadas a emociones y sentimientos en un texto. Al igual que la categoría previa, las palabras se clasifican mediante diccionarios. Para el rasgo analizado sobre la emoción de enojo, se usa el diccionario NRC. Este fue desarrollado por Saif Mohammad y Peter Turney y establece ocho emociones básicas: enojo, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto para clasificar las palabras del texto.

RESULTADOS Y DISCUSIÓN

Características lingüísticas de noticias de muertes
Los rasgos lingüísticos relevantes para cada modelo de clasificación se ilustran en la Tabla 2. En las columnas de los modelos, un 1 significa que la métrica está entre los diez más relevantes para dicho modelo, y un 0, que no lo está. La columna "Total" representa el número de modelos para los cuales el rasgo lingüístico fue relevante, considerando todo el conjunto de datos. Los rasgos se listan por orden de importancia de acuerdo a este valor. A continuación, las últimas tres columnas muestran el número de modelos para los cuales el rasgo lingüístico es relevante, segmentando por cada tipo de noticias: T1, T2, y T3.

Tal como se observa, los cuatro rasgos lingüísticos más relevantes en la clasificación de noticias, considerando el conjunto de datos de validación completo, están relacionados con la extensión de las oraciones en términos de caracteres (R2-R3), el uso de adposiciones (R1) y de sustantivos propios (R4). Luego, se observan más rasgos relacionados con la extensión de las

Tabla 2. Rasgos lingüísticos relevantes por cada modelo de clasificación.

ID	Rasgo lingüístico	Modelo					Total	T1	T2	T3
		SVM (lineal)	SVM (RBF)	RF	XG Boost	LR				
R1	**promedio de adposiciones por oración**	**1**	**1**	**1**	**1**	**1**	**5**	**3**	**4**	**1**
R2	máximo núm. de caracteres por oración	1	1	1	1	1	5	1	3	1
R3	máximo núm. de caracteres sin espacios por oración	1	1	1	1	1	5	3	2	2
R4	idf de sustantivos propios en toda la noticia	1	1	1	1	1	5	2	4	4
R5	mediana de adposiciones por oración	1	1	1	0	1	4	3	2	4
R6	tf-idf de auxiliares en toda la noticia	1	0	1	1	1	4	4	1	5
R7	máximo núm. de palabras por oración	1	0	1	1	1	4	1	3	1
R8	frecuencia de conectores de opinión generalizadores	1	1	0	0	1	3	3	2	5
R9	mínimo núm. de palabras por oración	0	1	1	1	0	3	3	5	4
R10	máximo núm. de adposiciones por oración	0	1	0	1	0	2	3	2	0
R11	mínimo núm. de adposiciones por oración	0	1	0	0	1	2	2	3	0
R12	medida de sentimiento enojo según léxico NRC	0	0	1	1	0	2	3	1	2
R13	frecuencia de marcadores contraargumentativos	1	0	0	0	1	2	4	2	4
R14	núm. total de oraciones en toda la noticia	0	0	1	1	0	2	4	4	4
R15	máximo num. de auxiliares por oración	0	1	0	0	0	1	4	5	3
R16	desviación estándar de núm. numerales por oración	1	0	0	0	0	1	4	4	5
R17	medida de sentimiento muy positivo según léxico Affin	0	0	0	0	0	0	3	3	5

oraciones (R7, R9) y del texto (R14), así como con el uso de adposiciones (R5, R10-R11), y se agregan otros rasgos relevantes sobre la frecuencia de auxiliares (R6), de conectores de opinión generalizadores (R8), y de marcadores contraargumentativos (R13). Para dos de los modelos resultan también relevantes los léxicos que reflejan enojo (R12).

Si se analizan los tipos de noticias por separado, entonces emerge el rasgo R17 de léxicos de sentimientos muy positivos como relevantes para cada uno de ellos, y aumentan las coincidencias de relevancia también para los rasgos R15 y R16, que consideran los auxiliares y numerales, respectivamente. Para las noticias reales de T3, se destaca una menor relevancia de las adposiciones, y una mayor relevancia en la frecuencia de auxiliares, conectores de opinión generalizadores, numerales y sentimientos muy positivos.

Este análisis permite identificar diferencias lingüísticas entre las noticias de T1 y T2, pero sobre todo entre ellas con T3. Por lo tanto, como se verá a continuación, permiten hallar diferenciaciones entre T1 y T3 que resultan poco claras si se consideran sólo modelos de clasificación automática.

Clasificación automática de noticias de muertes

Los resultados de los clasificadores se presentan en la Tabla 3. En verde se destaca cuando se trata de una noticia bien clasificada de acuerdo al tipo de noticia, y en rojo cuando no está bien clasificada. Las columnas de "Unificación" refieren a los resultados de clasificación conjunta de todos los clasificadores. La columna "Correctitud" marca "Sí" en caso de que la mayoría de clasificadores hayan etiquetado correctamente, respecto al tipo de noticia. Este resultado suele ser útil para la toma de decisiones, ya que mayores coincidencias en los resultados de los modelos se traducen en una mayor confianza para los analistas.

De acuerdo a lo esperado, los resultados varían bastante dependiendo del grupo de las noticias. Para el grupo T1 (noticias falsas no intencionadas sobre Chomsky), los modelos SVM lineal y LR resultan ser los mejores (91% de éxito de clasificación), mientras que RF solo acertó en tres etiquetados (18% de éxito) y SVM (RBF) y XGBoost solo acertaron en un etiquetado (9% de éxito). Al ser estos tres últimos mayoría, los resultados unificados logran tan solo un 27% de éxito conjunto. Llama la atención la noticia T1-10, clasificada incorrectamente

Tabla 3. Resultados de clasificación de noticias. Un color rojo significa que la etiqueta es incorrecta, y un color verde, que la etiqueta es correcta.

		Clasificadores					Unificación		
Grupo	**Noticia**	**SVM (lineal)**	**SVM (RBF)**	**RF**	**XGBoost**	**LR**	**Real**	**Falsa**	**Correctitud**
T1	T1-1	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-2	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-3	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-4	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-5	Falsa	Real	Falsa	Real	Falsa	2	3	Sí
	T1-6	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-7	Falsa	Falsa	Real	Real	Falsa	2	3	Sí
	T1-8	Falsa	Real	Real	Real	Falsa	3	2	No
	T1-9	Falsa	Real	Falsa	Falsa	Falsa	1	4	Sí
	T1-10	Real	Real	Real	Real	Real	5	0	No
	T1-11	Falsa	Real	Real	Real	Falsa	3	2	No
"Falsas" T1		10	1	2	1	10			27%
"Reales" T1		1	10	9	10	1			
Correctas T1		91%	9%	18%	9%	91%			
T2	T2-1	Falsa	Falsa	Real	Real	Falsa	2	3	No
	T2-2	Falsa	Real	Real	Real	Falsa	3	2	Sí
	T2-3	Falsa	Falsa	Real	Real	Real	3	2	Sí
	T2-4	Real	Real	Real	Real	Real	5	0	Sí
	T2-5	Falsa	Real	Real	Real	Falsa	3	2	Sí
	T2-6	Real	Real	Falsa	Falsa	Real	3	2	Sí
	T2-7	Real	Real	Real	Real	Real	5	0	Sí
	T2-8	Real	Real	Falsa	Real	Real	4	1	Sí
	T2-9	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T2-10	Falsa	Real	Real	Real	Falsa	3	2	Sí
	T2-11	Real	Real	Real	Real	Real	5	0	Sí
"Falsas" T2		6	2	3	1	5			82%
"Reales" T2		5	9	8	10	6			
Correctas T2		45%	82%	73%	91%	55%			
T3	T3-1	Real	Real	Falsa	Real	Real	4	1	Sí
	T3-2	Falsa	Real	Real	Real	Falsa	3	2	Sí
	T3-3	Real	Real	Falsa	Falsa	Real	3	2	Sí
	T3-4	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-5	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-6	Real	Real	Falsa	Falsa	Real	3	2	Sí
	T3-7	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-8	Real	Real	Falsa	Real	Real	4	1	Sí
	T3-9	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-10	Falsa	Real	Falsa	Real	Falsa	2	3	No
	T3-11	Falsa	Real	Falsa	Real	Falsa	2	3	No
"Falsas" T3		7	0	10	2	7			45%
"Reales" T3		4	11	1	9	4			
Correctas T3		36%	100%	9%	82%	36%			
"Falsas" Total		23	3	15	4	22			
"Reales" Total		10	30	18	29	11			
Correctas Total		19 (57,6%)	21 (63,6%)	11 (33,3%)	20 (60,6%)	20 (60,6%)			

por todos los modelos como real. A nivel de rasgos lingüísticos, se trata de una noticia breve, exenta de léxico emocional, que en general está conformada por afirmaciones y que referencia a otras fuentes; por lo demás, no se observa en ella marcadores discursivos textuales contraargumentativos, todo lo cual justifica la dificultad de clasificarla como falsa.

Para el grupo T2 (noticias reales de rectificación), los resultados se invierten. En este caso son los modelos XGBoost, SVM (RBF) y RF, en este orden, los más exitosos (91%, 82% y 73% de éxito, respectivamente), quedando LR y SVM lineal relegados a bajos 55% y 45%, respectivamente. Por la misma razón anterior, en este caso los modelos exitosos son la mayoría, por lo que se obtiene un éxito conjunto del 82%.

Finalmente, para el grupo T3 (noticias reales sobre Hawking), se da un fenómeno distinto a los dos anteriores. Aquí SVM (RBF) y XGBoost funcionan muy bien (82% y 100% de éxito, respectivamente), igual que en el caso anterior, pero RF queda muy mal posicionado, con solo un 9% de éxito. Por esto, al solo haber dos clasificadores exitosos, el éxito compartido es bajo, es decir, de solo un 45%.

En suma, podemos concluir que ninguno de los modelos es exitoso para todos los tipos de noticias. De los cinco, los modelos SVM y XGBoost parecen ser los mejores, aunque sirven para clasificar tipos de noticias distintas. De entre los tres, se debe considerar además que XGBoost no tiene tan buenas métricas de rendimiento como los modelos SVM (Tabla 1).

Por otra parte, se observa una marcada diferencia en la correctitud de los modelos para las noticias tipo T1 y tipo T3, lo que da cuenta de que, a criterio de los clasificadores automáticos, las noticias falsas no intencionadas son percibidas de distinta manera que las reales, pese que todas ellas hablan acerca de la muerte de una celebridad y están redactadas de manera similar. Esta dificultad para distinguir entre realidad y falsedad es mermada (solo en parte) gracias al análisis de rasgos lingüísticos como el realizado en la sección anterior.

INTERPRETACIÓN DE LOS RESULTADOS

En los resultados se destaca la considerable variabilidad en la precisión de los clasificadores automáticos según el tipo de noticia, reflejando la complejidad de distinguir entre información verdadera y falsa en un contexto de distribución y consumo rápido. Esto pone de manifiesto la vulnerabilidad de la esfera pública digital ante la desinformación, subrayando una crisis de confianza en los medios de comunicación y en las instituciones tradicionales que tiene alcance nacional [41] y global [42]. La dificultad para clasificar correctamente las noticias falsas no intencionadas muestra cómo la rápida difusión de información no verificada, impulsada por la competencia en el campo periodístico y la economía de la atención, exacerba el problema de la desinformación. Además, en el contexto del uso y masificación a través de plataformas de redes sociales, dificulta rastrear la fuente inicial de la información, así como advertir si esta se desmiente o se corrobora. Sociológicamente, conmina a preguntarnos por el efecto en la discusión pública y las formas en que nos relacionamos comunicativamente. En este sentido, también nos insta a problematizar el poder de los medios sobre la constitución de la verdad.

Asimismo, la dependencia de los modelos de aprendizaje automático y procesamiento de lenguaje natural del contexto en el que fueron entrenados limita su capacidad para generalizar a nuevos contextos o temáticas, indicando que la tecnología por sí sola no puede resolver el problema de la desinformación. Nótese cómo la ausencia de una pregunta desde el "cómo" limita las posibilidades de comprensión de la problemática. Esto sugiere que se requiere una comprensión profunda del contexto social y cultural en el que se produce y consume la información. Es decir, ir desde el fenómeno al campo en el cual se constituye.

Respecto a la práctica periodística, durante la búsqueda de las noticias, se evidenció que muchos medios recolectan noticias de múltiples fuentes y no eliminan las noticias falsas una vez estas ya han sido desmentidas. Más aún, hay medios que mantienen las noticias falsas que desmienten ellos mismos, incluso cuando han sido escritas por el/la mismo/a periodista. Esto potencia las dificultades para distinguir entre realidad y falsedad, y evidencian aquel fenómeno de "posverdad" discutido al inicio de este artículo. De hecho, queda la interrogante de qué ocurre cuando una noticia falsa ha sido desmentida masivamente, pero aun así un medio de noticias decide por una u otra razón mantener la información errónea que

ayudó a amplificar dicha falsedad. Y es que ante esta falta de corrección, dicha noticia falsa, inicialmente no intencionada, se transforma en intencionada, pasando de ser *misinformation* a *disinformation*. He ahí otro ejemplo que evidencia las enormes dificultades para poder distinguir automáticamente, incluso mediante técnicas de inteligencia artificial avanzadas, entre realidad y falsedad.

Respecto a lo anterior, cabe destacar la controversia que existe respecto a que estas informaciones falsas, sean intencionadas o no intencionadas, no son eliminadas de las plataformas digitales de medios de comunicación, cuando las editoriales argumentan su derecho a la libertad de expresión, por encima del denominado "derecho al olvido" [43] asociado con noticias que pudieran perjudicar la imagen de las personas, como es el caso de la falsa muerte de una personalidad pública, que fue revisada en esta investigación. Sin embargo, el tema del derecho al olvido digital sigue siendo controversial, ya que si bien está siendo abordado rigurosamente en Europa, en países como Estados Unidos, "los motores de búsqueda no ostentan ningún tipo de responsabilidad sobre la información que aparece en sus resultados de búsqueda" [43]. Por su parte, en el caso de Chile, en que el sistema legal todavía no es enfático ni categórico respecto al tema, se encuentra en proceso un proyecto de ley que podría permitir eliminar información "que se considera obsoleta por el transcurso del tiempo, o que de alguna manera afecte el desarrollo de alguno de los derechos fundamentales" [44].

CONCLUSIONES

En este trabajo se abordó el problema de la detección de noticias falsas no intencionadas, un tipo de noticias poco analizadas en la literatura. Para ello, se propusieron al menos tres novedades metodológicas. Primero, el uso de un método de clasificación híbrido, que combina aprendizaje automático supervisado con procesamiento de lenguaje natural, mediante el uso de rasgos lingüísticos específicos para el idioma español. Segundo, el uso de un conjunto de datos de validación exigente, que no considera noticias reales y falsas (intencionadas), como es lo usual, sino que noticias reales, falsas no intencionadas, y reales que desmienten a estas últimas; algo más cercano a la realidad de los medios de prensa tradicionales en línea. Finalmente, se emplea una aproximación interdisciplinaria desde la informática, la lingüística, la sociología y el periodismo para el análisis y discusión de los resultados.

Los resultados del caso de estudio analizado, sobre noticias de muertes de celebridades, dan cuenta de la enorme dificultad de clasificación de las noticias falsas no intencionadas. En términos lingüísticos, se observa que algunos rasgos lingüísticos, como los relacionados con el uso de adposiciones (preposiciones), el tamaño de las oraciones y de los textos, el uso de sustantivos propios, marcadores contraargumentativos, emociones positivas y negativas, entre otros, resultan relevantes para este propósito. En términos de los clasificadores, si bien algunos pudieron clasificar correctamente ciertos tipos de noticias, ninguno de ellos fue capaz de clasificar correctamente los tres tipos de noticias consideradas.

Una limitación importante en el problema de clasificación de noticias falsas, que también aplica a este trabajo, es el problema de la multicontextualidad, esto es, la dificultad de que los modelos sean robustos y mantengan su rendimiento evidenciado para los datos de entrenamiento y prueba, en los nuevos datos de validación. En efecto, la variación del contexto puede incidir negativamente en la correctitud de los modelos de clasificación automática.

Pese a las dificultades detectadas, este trabajo abre una nueva perspectiva en el análisis de los desórdenes de información, distinguiendo entre lo que son y representan las noticias falsas intencionadas y no intencionadas, las primeras más presentes en las plataformas de redes sociales, y las segundas en los medios de comunicación de masas y medios de prensa en línea tradicionales. Para poder continuar con experimentos en esta línea, se deja como trabajo futuro la generación de más datos para validación.

AGRADECIMIENTOS

Este proyecto fue apoyado por el "Fondo de estudios sobre el pluralismo en el sistema informativo nacional" PLU230017, ANID, Chile.

REFERENCIAS

[1] Unesco, *Journalism, fake news & disinformation: handbook for journalism education*

and training, C. Ireton, and J. Posetti, Eds. París, Francia: Unesco, 2018.

[2] K. Shu, S. Wang, D. Lee, and H. Liu, "Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements Disinformation," in *Misinformation, and Fake News in Social Media. Emerging Research Challenges and Opportunities*, K. Shu, S. Wang, D. Lee, and H. Liu, Eds. Cham, Switzerland: Springer International Publishing, 2020, pp. 1-19, doi: 10.1007/978-3-030-42699-6.

[3] E. Puraivan, E. Godoy, F. Riquelme, and R. Salas, "*Fake news detection on Twitter using a data mining framework based on explainable machine learning techniques*," in *11th International Conference of Pattern Recognition Systems (ICPRS 2021),* Talca, Chile, Mar. 17-19, 2021, pp. 157-162, doi: 10.1049/icp.2021.1450.

[4] I. Stavre and M. Puntí, "Fake news, something new?," *Sociology and Anthropology*, vol. 7, no. 5, pp. 212-219, 2019, doi: 10.13189/sa.2019.070504.

[5] S. van der Linden, C. Panagopoulos, and J. Roozenbeek, "You are fake news: political bias in perceptions of fake news," *Media, Culture & Society*, vol. 42, no. 3, pp. 460-470, Apr. 2020, doi: 10.1177/0163443720906992.

[6] N. Poulose, "Fake news in health and medicine," *Data Science for Fake News*, vol. 42, 2021, pp. 193-204, doi: 10.1007/978-3-030-62696-9_9.

[7] F. Olan, U. Jayawickrama, E.O. Arakpogun, J. Suklan, and S. Liu, "Fake news on social media: the impact on society," *Information Systems Frontiers*, vol. 26, no. 2, pp. 443-458, Jan. 2022, doi: 10.1007/s10796-022-10242-z.

[8] E. Papadogiannakis, P. Papadopoulos, E.P. Markatos, and N. Kourtellis, "Who funds Misinformation? A systematic analysis of the ad-related profit routines of fake news sites," *Proceedings of the* ACM Web Conference, 2023, pp. 2765-2776, doi: 10.1145/3543507.3583443.

[9] J.A. Braun and J.L. Eklund, "Fake news, real money: Ad tech platforms, profit-driven hoaxes, and the business of journalism," *Digital Journal*, vol. 7, no. 1, pp. 1-21, Jan. 2019, doi: 10.1080/21670811.2018.1556314.

[10] J. Alghamdi, S. Luo, and Y. Lin, "A comprehensive survey on machine learning approaches for fake news detection," *Multimedia Tools and Applications*, vol. 83, no. 17, pp. 51009-51067, Jan. 2024, doi: 10.1007/s11042-023-17470-8.

[11] J. Rose, "To believe or not to believe: an epistemic exploration of fake news, truth, and the limits of knowing," *Postdigital Science and Education*, vol. 2, no. 1, pp. 202-216, Jan. 2020, doi: 10.1007/s42438-019-00068-5.

[12] T. Lelo and R. Fígaro, *A materialist approach to fake news*, in *Politics of Disinformation: The Influence of Fake News on the Public Sphere*, G. López-García, D. Palau, M.B. Torres, E. Campos, and P. Masip, Eds. New York, USA: Wiley-Blackwell, 2021, pp. 23-34, doi: 10.1002/9781119743347.ch2.

[13] J. Gray, L. Bounegru, and T. Venturini, "'Fake news' as infrastructural uncanny," *New Media & Society*, vol. 22, no. 2, pp. 317-341, Jan. 2020, doi: 10.1177/1461444819856912.

[14] P. Bourdieu, *Sobre la televisión*, Barcelona, España: Anagrama, 2006.

[15] P. Bourdieu, *Curso de sociología general 1. Conceptos fundamentales*, 1ra. ed. Buenos Aires, Argentina: Siglo XXI Editores, 2020.

[16] D.L. Swartz, "Spotlight," in *Journalism and Truth in an Age of Social Media*, 1st. ed., J. Katz and K. Mays, Eds. New York, USA: Oxford University Press, 2019, pp. 36-38, doi: 10.1093/oso/9780190900250.003.0003.

[17] E.C. Tandoc, "The facts of fake news: A research review," *Sociology Compass*, vol. 13, no. 9, Art. no. e12724, pp. 1-19, Jul. 2019, doi: 10.1111/soc4.12724.

[18] J. Compton, S. van der Linden, J. Cook, and M. Basol, "Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories," *Social and Personality Psychology Compass*, vol. 15, no. 6, Art. no. e12602, pp. 1-16, May 2021, doi: 10.1111/spc3.12602.

[19] E. Puraivan, R. Venegas, and F. Riquelme, "An empiric validation of linguistic features in machine learning models for fake news detection," *Data & Knowledge Engineering*, vol. 147, Art. no. 102207, Sep. 2023, doi: 10.1016/j.datak.2023.102207.

[20] J. Valenzuela, "Lingüística contrastiva inglés-español: una visión general," *Carabela*, no. 51. pp. 27-45, 2002. [En

línea]. Disponible: https://cvc.cervantes.es/ensenanza/biblioteca_ele/carabela/pdf/51/51_027.pdf

[21] B. Ch. Han, *Infocracia. La digitalización y la crisis de la democracia*, Bogotá, Colombia: Taurus, 2021.

[22] P. Bourdieu, "Social Space and Symbolic power," *Sociology Theory*, vol. 7, no. 1, pp. 14-25, 1989.

[23] S. Sismondo, "Post-truth?," *Social Studies Science*, vol. 47, no. 1, pp. 3-6, Feb. 2017, doi: 10.1177/0306312717692076.

[24] B. Çanakpınar, M. Kalelioğlu, and V. D. Günay, "Semiotics and political discourse in the post-truth era," *Language and Semiotic Studies*, vol. 10, no. 1, pp. 65-82, Jan. 2024, doi: 10.1515/lass-2023-0040.

[25] J. Angermuller, "Truth after post-truth: for a Strong Programme in Discourse Studies," *Palgrave Communications*, vol. 4, no. 1, Art. no. 30, Mar. 2018, doi: 10.1057/s41599-018-0080-1.

[26] H. Rao and H.R. Greve, "The Plot Thickens: A Sociology of Conspiracy Theories," *Annual Review of Sociology*, vol. 50, no. 1, Feb. 2024, doi: 10.1146/annurev-soc-030222-031142.

[27] D. Rivera López, "Verdad y catástrofe en el presente algorítmico," *Latin American Journal of Humanities and Educational Divergences*, vol. 3, no. 1, pp. 100-114, Jun, 2024, doi: 10.5281/zenodo.12629874.

[28] J. Buschman, "Fake news as systematically distorted communication: an LIS intervention," *Journal of Documentation*, vol. 80, no. 1, pp. 203-217, Jul. 2024, doi: 10.1108/JD-03-2023-0043.

[29] O.L.T. Mai *et al.*, "Factors Affecting Students' Fake News Identification during Covid-19 in Vietnam: Access from sociological study and application of PLS-SEM model," *Wseas Transactions on Business and Economics*, vol. 20, Art. no. 126, pp. 1422-1438, Jun. 2023, doi: 10.37394/23207.2023.20.126

[30] E. Aïmeur, S. Amri, and G. Brassard, "Fake news, disinformation and misinformation in social media: a review," *Social Network Analisys and Mining*, vol. 13, no. 1, pp. 13-30, Feb. 2023, doi: 10.1007/s13278-023-01028-5.

[31] K. Hunt, P. Agarwal, and J. Zhuang, "Monitoring Misinformation on Twitter During Crisis Events: A Machine Learning Approach," *Risk Analysis*, vol. 42, no. 8, pp. 1728-1748, Nov. 2020, doi: 10.1111/risa.13634.

[32] Y. Zhao, J. Da, and J. Yan, "Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches," *Information Processing & Management*, vol. 58, no. 1, pp. 1-24, Jan. 2021, doi: 10.1016/j.ipm.2020.102390.

[33] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, New York: Springer, 2021, doi: 10.1007/978-1-0716-1418-1.

[34] F. Riquelme *et al.*, "FalleDesinfo_ES: Dataset de noticias en español reales y falsas sobre fallecimientos de celebridades," Zenodo, 2024, doi: 10.5281/ZENODO.13117808.

[35] S. Bradshaw *et al.*, "Country Case Studies Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation," demtech.oii.ox.ac.uk. Accessed: Sep. 2024. [Online]. Available: https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2021/03/Case-Studies_FINAL.pdf

[36] S. Bradshaw *et al.*, "Industrialized Disinformation 2020 Global Inventory of Organized Social Media Manipulation. Computational Propaganda Research Project," demtech.oii.ox.ac.uk. [Online]. [Online]. Available: https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2021/01/CyberTroop-Report-2020-v.2.pdf

[37] E. Cryst and I. García Camargo, "#VivaJOH o #FueraJOH. An analysis of Twitter's takedown of Honduran accounts," *Standford Internet Observatory*, Apr. 2020, doi: 10.25740/vz964sv5963.

[38] E. Culliford, "Facebook says it removed troll farm run by Nicaraguan government," reuters.com. [Online]. Available: https://www.reuters.com/world/americas/facebook-says-it-removed-troll-farm-run-by-nicaraguan-government-2021-11-01/

[39] V. Bergengruen, "Honduras Shows How Fake News Is Changing Latin American Elections," time.com. [Online]. Available: https://time.com/6116979/honduras-political-disinformation-facebook-twitter/

[40] E. Providel, D. Toro, F. Riquelme, M. Mendoza, and E. Puraivan, "CLNews:

The First Dataset of the Chilean Social Outbreak for Disinformation Analysis," *Proceedings of the 31st ACM International Conference on Information & Knowledge Management,* Atlanta, GA, USA, Oct. 2022, doi: 10.1145/3511808.3557560.

[41] C. Mellado y A. Cruz, "Consumo de noticias y evaluación del periodismo en Chile", periodismopucv.cl. Accedido: septiembre de 2024. [En línea]. Disponible en:https://periodismopucv.cl/app/uploads/2024/03/Informe-2024-noticias-y-periodismo.pdf

[42] N. Newman, R. Fletcher, C.T. Robertson, A. Ross Arguedas, and R. Kleis Nielsen, "Digital News Report 2024," reutersinstitute.politics.ox.ac.uk. [Online]. Available: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2024-06/RISJ_DNR_2024_Digital_v10%20lr.pdf

[43] A. Moreno, "El derecho al olvido digital: una brecha entre Europa y Estados Unidos," *Revista de Comunicación*, vol. 18, no 1, pp. 259-276, Feb. 2019, doi:10.26441/RC18.1-2019-A13.

[44] S. Pérez, "Los datos y la imagen personal como límite a la libertad de información", vti.uchile.cl. [En línea]. Disponible en: https://vti.uchile.cl/derecho-al-olvido-en-internet/

[45] PACTE, "Revisa estadísticas de tus textos así de rápido", redilegra.com. [En línea]. Disponible en: http://www.redilegra.com/pacte