

Inteligencia Artificial como alternativa en la detección de noticias falsas

Artificial Intelligence as an alternative in the detection of false news

Quevedo Cortés, Angélica ¹ . Escobar Cerón, Dario ²

Citar este documento:

Quevedo-Cortés, Angélica. Escobar-Cerón, Darío. Inteligencia Artificial como alternativa en la detección de noticias falsas. Revista Technol.Investig.Academia TIA, ISSN: 2344-8288, Volumen 10 Numero 1, pp. 20-37. Bogotá-Colombia.

¹ Matemática Egresada Fundación Universitaria Konrad Lorenz, Estudiante de la Especialización en Ingeniería de Software – Universidad Distrital Francisco José de Caldas. adquevedoc@correo.udistrital.edu.co .

² Ingeniero de sistemas, Estudiante de la Especialización en Ingeniería de Software – Universidad Distrital Francisco José de Caldas. dfescobar@correo.udistrital.edu.co. Bogotá - Colombia

Resumen

El objetivo del presente artículo es analizar cómo la Inteligencia Artificial ha servido como herramienta para la detección de noticias falsas. Se trata de un artículo de revisión bibliográfica, descriptivo, exploratorio cuya unidad de análisis está representada por artículos publicados en bases de datos tales como: Scielo, Dialnet, Science Direct y Scopus. Se obtuvo como resultado que existen distintos mecanismos de Inteligencia Artificial en el subcampo del Aprendizaje Automático, el Aprendizaje Profundo y Procesamiento de Lenguaje Natural (NLP), como el uso de máquinas de soporte vectorial, el clasificador Naive Bayes y distintos tipos de modelos de Redes de Neuronas Artificiales; como las Redes Neuronales Recurrentes (RNN), las Redes Neuronales Convolucionales (CNN), que son bastante populares en el Procesamiento de Lenguaje Natural, especialmente en el uso de Redes de Memoria a Corto y Largo Plazo (LSTM), las cuales permiten establecer mecanismos confiables y precisos para detectar patrones dentro de un objetivo con contenido textual.

Palabras Clave: noticias falsas, Inteligencia Artificial, Aprendizaje Automático, Procesamiento de Lenguaje Natural.

Abstract

The objective of this article is to analyze how artificial intelligence has served as a tool to stop fake news. It is a descriptive, exploratory bibliographic review article whose unit of analysis is represented by articles published in databases such as: Scielo, Dialnet, Science Direct and Scopus. It was obtained as a result that there are different Artificial Intelligence mechanisms in the subfield of Machine Learning, Deep Learning and Natural Language Processing (NLP), such as the use of vector support machines, the Naive Bayes classifier and different types of Network models of Artificial Neurons; such as Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), which are quite popular in Natural Language Processing, especially in the use of Short-Term Memory Networks (LSTM), which allow establishing reliable mechanisms and accurate to detect patterns within a textual target.

Key Words: False News, Artificial Intelligence, Machine Learning, Natural Language Processing, Neuron Networks, Deep Learning.

I. Introducción

Las noticias falsas o *fake news* hacen alusión a la información que careciendo de veracidad es emitida a través de los medios de comunicación tradicionales y por medio de internet, principalmente desde las redes sociales, y es recibida de forma masiva por usuarios que tienen el riesgo de ser manipulados en beneficio de intereses de algunos líderes u organizaciones, debido a que la desinformación puede modificar las creencias, pensamientos, decisiones y comportamientos, sin una validez suficiente, lo que hace que las decisiones sean desinformadas y beneficien a determinados actores [1].

Además de la desinformación, las noticias falsas también son un obstáculo para la democracia en la sociedad, para la libertad y la buena reputación tanto de las personas como de las organizaciones. Esta problemática es difícil de solucionar teniendo en cuenta un contexto en donde cada vez es más fácil producir contenido falso [2], que puede ser distribuido masivamente hacia receptores a través de internet quienes no siempre cuentan con educación suficiente para diferenciar entre noticias falsas y verdaderas o simplemente inciden en el error de dar como certera, información no comprobada. Asimismo, el discernimiento se ve afectado por la subjetividad que acompaña la naturaleza del ser humano la cual incide en los juicios e interpretaciones que se van a hacer sobre ese tipo de contenido [3].

Es importante reconocer que a partir de los vertiginosos avances tecnológicos se han podido generar y evaluar mecanismos de software capaces de reconocer entre una noticia falsa y una verdadera; entre los ejemplos más notables está la aplicación de técnicas de Aprendizaje Automático, como el Procesamiento de Lenguaje Natural (NLP), Máquinas de Soporte Vectorial, Redes Neuronales, Naive Bayes [4], entre otros, todo esto haciendo parte de lo que se conoce como Inteligencia Artificial (IA), la cual constituye uno de los principales avances tecnológicos que se han desarrollado recientemente, siendo incorporada en distintas áreas debido a sus múltiples aplicaciones y utilidades para resolver problemáticas.

En relación a esto destaca la tendencia existente en los últimos años sobre la utilización de la misma para la detección de noticias falsas, las cuales surgen por distintos motivos de tipo socio económicos, culturales, políticos, entre otros. En relación a estos últimos, los acontecimientos ocurridos a lo largo del mundo han provocado un aumento de la popularidad, así como la difusión de noticias falsas o *fake news* (FN).

Se han realizado esfuerzos para automatizar el proceso de su detección, los intentos más notorios han sido la creación de las denominadas listas negras de fuentes y autores que se ha comprobado que no son completamente confiables. Aunque estos instrumentos resultan de gran utilidad, para establecer una solución definitiva es necesario tomar en cuenta los casos más arduos donde las consideradas fuentes confiables y autores reconocidos publican noticias falsas [5].

Un estudio reciente del Instituto Tecnológico de Massachusetts evidenció que las FN poseen 70% más de posibilidades de ser tuiteadas. Igualmente, en una encuesta del Barómetro de Edelman Trust, 59% de los individuos expresó que no estaba seguro si alguna historia presentada era en realidad "verdadera" o "falsa". En relación con esto, destaca la afirmación: "el acceso a información confiable y de calidad debe ser un derecho de cualquiera, dondequiera que viva" [6]. En este sentido, el presente artículo tiene como objetivo realizar una revisión de los mecanismos formulados para detectar noticias falsas propuestos por la Inteligencia Artificial, para ello a continuación se desglosan las secciones: marco teórico, metodología, principales hallazgos y conclusiones.

II. Marco teórico

Para tener un contexto de la situación problema desde la perspectiva de autores reconocidos en la materia, a continuación, se expone un breve marco teórico que aporta carácter científico al artículo desarrollado, el cual se centra en las tres variables principales bajo estudio: noticias falsas (*fake news*), Inteligencia Artificial y Aprendizaje Automático.

a. Noticias falsas (Fake news)

En líneas generales, las noticias falsas o *fake news* (FN) se definen como todas aquellas "informaciones publicadas deliberadamente en medios digitales que no han sido comprobadas ni verificadas, que carecen de fuentes identificadas y que no cuentan con la supervisión de un editor" [7]. En otras palabras, son los artículos informativos redactados con el propósito de confundir o engañar a los lectores, cuyo contenido puede comprobarse que es intencionalmente falso [8].

En este orden de ideas, se plantea que las *fake news* no representan una problemática nueva, al contrario, siempre han existido en el ámbito del periodismo. Lo que ha llevado el tema al campo de investigación científica, es que en la actualidad estas noticias resultan más visibles debido principalmente a la popularidad creciente y exponencial de las redes sociales (RR. SS.), que es donde usualmente se hacen virales logrando alcanzar un público objetivo numeroso [9].

Este hecho perjudica a la sociedad en general, pues la desinformación o información equivocada puede conllevar a la toma de decisiones que perjudiquen individual y colectivamente a los ciudadanos.

b. Inteligencia artificial (IA)

La Inteligencia Artificial (IA) es un área de estudio que pretende desarrollar una serie de paradigmas innovadores que requieren las máquinas de toda clase para realizar un conjunto de tareas específicas, que eventualmente las personas ejecutan con mayor eficacia y control. En este sentido, un sistema que posea IA debe ser capaz de cumplir una serie de funciones, tales como: guardar conocimiento y posteriormente aplicarlo para solventar alguna problemática existente, de manera que pueda adquirir conocimientos nuevos a través de las experiencias vividas en el día a día [10] [11].

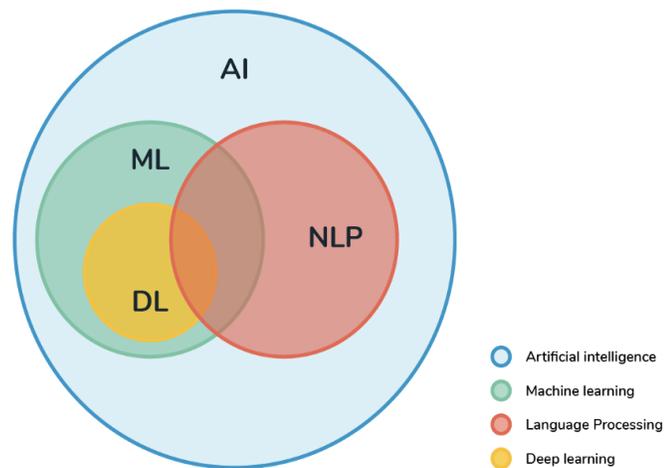
c. Clasificación de la IA

Tradicionalmente, para clasificar la IA se utilizan “representaciones simbólicas basadas en un número finito de primitivas y de reglas para la manipulación de símbolos (por ejemplo, redes semánticas, lógica de predicados, etc.), los cuales fueron y siguen siendo parte central de dichos sistemas” [12]. Otro tipo de representación es el denominado sub-simbólico, el mismo emplea representaciones numéricas o como su nombre lo indica “sub-simbólicas” del conocimiento. Aunque la mayoría de las publicaciones científicas referente a IA (Hebb, 1949, Minsky and Papert 1969) enfatizan únicamente la labor realizada por Rosenblatt y Widrow en la década de los 50 con Redes Neuronales en este tipo de representación [12].

Dicho enfoque está caracterizado por la creación de sistemas que poseen capacidad de aprendizaje. El mismo puede obtenerse a nivel de individuo imitando el funcionamiento del cerebro humano, lo que se conoce como “Redes Neuronales”, a nivel de especie, imitando la evolución. Recientemente, resultaba normal hablar de Algoritmos Genéticos (AG) en general, en lugar de identificar diferentes tipos de AE, puesto que el resto de los algoritmos se pueden interpretar como variaciones o mejoras de los AG [12].

Es importante reconocer que, lo que busca la IA es permitir que los computadores imiten la inteligencia humana, como la toma de decisiones, el procesamiento de texto, la percepción visual. Por esta razón, la IA posee una amplia cantidad de campos de aplicación que se contienen en varios subcampos como lo son: el Aprendizaje Automático, la Robótica, la visión por computador y el NLP. Para fines del artículo es importante tener en cuenta la siguiente relación plasmada gráficamente en la Figura 1.

Figura 1. Identificación subcampos Inteligencia Artificial



Fuente: Can (2020) [13]

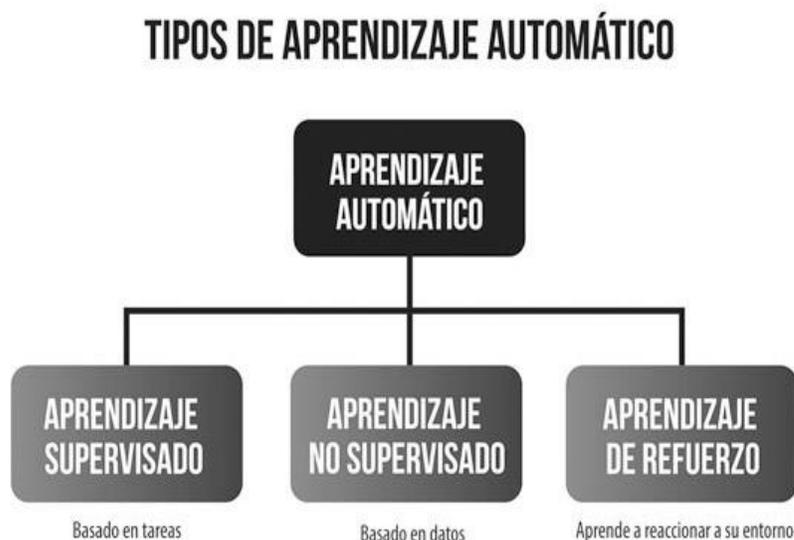
d. Aprendizaje Automático

El Aprendizaje Automático o *Machine Learning* (ML), es un término directamente relacionado con el concepto de la IA y que se constituye como subcampo de esta, puesto que hace referencia a la capacidad de una máquina o *software* para aprender automáticamente mediante la generación de algoritmos que responden a ciertos datos de entrada en el sistema. Originalmente, en los inicios de la Informática la única manera de lograr que un sistema informático ejecutara un proceso era escribiendo un algoritmo que definiera el contexto y detalles de cada acción que se debía seguir.

En contraste, los algoritmos usados en el desarrollo del ML realizan una gran parte de estas acciones de forma autónoma. Lo anterior, lo hacen obteniendo sus propios cálculos de acuerdo con los datos recopilados en el sistema, destacando que mientras más datos obtengan, mejores y más precisas serán las acciones resultantes de los mismos [14]. Es por ello que, el ML ha alcanzado gran relevancia a nivel de investigación, siendo utilizado en un gran número de disciplinas y especialidades.

No obstante, alcanzar un alto nivel de precisión demanda una gran cantidad de datos que en ocasiones resultan complicados, costosos e incluso poco prácticos de obtener. Es por lo que, la integración del conocimiento humano en el ML puede constituir una alternativa de solución para reducir los requisitos de datos, aumentando la confiabilidad y solidez del ML y construir sistemas de aprendizaje automático explicables [14]. Es importante comprender que el Aprendizaje Automático usa algoritmos para aprender los patrones de los datos y basados en cómo se realiza este proceso de reconocimiento de características el Aprendizaje Automático, se divide en 3 subconjuntos ilustrados a continuación en la Figura 2 [15]:

Figura 2. Tipos de aprendizaje automático



Fuente: Rouhianen (2018) [15]

e. Algoritmos y *netbots* en la IA

En las redes sociales, principal fuente de *fake news*, los algoritmos suelen ser de tipo “adaptativos”, esto significa que efectúan ligeros cambios sobre sí mismos constantemente, con el propósito de obtener mejores resultados o al menos intentarlo. Estos algoritmos mejorados, resultan más llamativos para el público y, por

ende, mucho más rentables, en ellos siempre hay un porcentaje de aleatoriedad [16]. Cuando un algoritmo les proporciona experiencias a unos individuos, resulta que la aleatoriedad que influye en la adaptación algorítmica también puede inducir a una especie de adicción.

En este escenario, el algoritmo se encuentra en la constante búsqueda de parámetros perfectos que permitan manipular el cerebro, mientras que este último, intentando encontrar un significado más profundo, cambia en respuesta a los experimentos del algoritmo; es una especie de juego del gato y al ratón, basado en matemática pura [16].

En este contexto, organizaciones, estados, universidades y empresas invierten grandes recursos económicos, tecnológicos y operativos en el desarrollo de algoritmos detectores de FN. Esta tecnología, todavía incipiente requiere una serie de detectives humanos o como son conocidos en inglés “*fact checkers*”, cuya labor es encontrar la información falsa que se encuentra circulando en el Internet [6]. En este orden de ideas, uno de los *bots* verificadores de información más conocido es *el Fake News Detector*, el mismo permite detectar y señalar *Fake News* y *Click Bait*s en noticias. Su funcionamiento, se basa en que, al clasificar una noticia, otras personas que usan el *bot* van a ver la clasificación asignada, de manera que estarán más atentas y al mismo tiempo podrán clasificar.

Esta información es almacenada en una base de datos y es leída por el robot diseñado para este propósito llamado: El Robinho. Este último, se fundamenta en la información dada por los desarrolladores y con la experiencia va aprendiendo a clasificar automáticamente una noticia como *Fake News*, *Click Bait*, etc., por el texto que contenga. Con ello, incluso noticias nuevas que nadie haya visto pueden ser clasificadas de forma rápida [6]. Otro *bot* ampliamente utilizado es el *Facterbot*, este es un *chatbot* de *Facebook Messenger*, el mismo está diseñado para enviar FN de gran alcance a las bandejas de entrada de los usuarios. Además de informar a los mismos sobre la información más reciente, su propósito es ayudar a los *fact checkers* a realizar su trabajo de forma más eficiente [6].

f. *Deep learning* o Aprendizaje Profundo

Todo ser vivo tiene la capacidad de aprender, de ver, percibir, actuar, comunicar e interpretar de forma rápida y eficiente que hasta el momento ningún método de aprendizaje de máquina ha podido simular. Esta tarea que yace en el cerebro es un proceso profundo, ya que implica una extensa cadena de conexiones sinápticas que involucran capas y capas de operaciones y procesos [17].

La gran tarea de la comunidad académica es determinar algoritmos que se acerquen a un procesamiento eficiente, tal que simulen tareas del cerebro animal. En este punto, se destacan los algoritmos de aprendizaje sin supervisión y supervisados que determinan un enfoque diferente para la alimentación de datos que se aplican según la necesidad y el entorno. Por esta razón, se deben generar algoritmos que sean capaces de seguir una arquitectura de jerarquías, es decir un modelo profundo de características para el reconocimiento [18].

Mediante las técnicas profundas se ha logrado potenciar el reconocimiento automático, el procesamiento de escritos, palabras e imágenes. Los avances en *deep learning* han generado altos costos en cuanto a procesamiento computacional, ya que implica el uso de múltiples parámetros, por lo cual una de las fijaciones ha estado en la reducción de dichos costos manteniendo los resultados obtenidos. Dentro del proceso de *deep learning* se encuentran las redes de neuronas profundas, que se basan en un preentrenamiento supervisado de datos que constan de ciertas etiquetas [19].

Dentro del concepto de Redes Neuronales, una neurona se constituye como un conjunto de operaciones entre pesos sinápticos que seguida de una transformación ya sea lineal o de mayor complejidad, en conjunto generan una arquitectura multicapa de neuronas de entrada, capas ocultas y de salida, es decir un modelo profundo [19]. Gracias a las arquitecturas profundas, los algoritmos de Inteligencia Artificial pueden aprender múltiples niveles de representación y abstracción que les ayuden a crear sentidos visuales como las imágenes, o a reconocer textos, o secuencias como lo es el sonido.

El *deep learning* se basa en algoritmos computacionales complejos de aprendizaje automático, que como se mencionó anteriormente, intenta recrear o modelar abstracciones visuales y sensoriales usando arquitecturas que están compuestas de transformaciones no lineales múltiples. Es prudente acotar que las Redes Neuronales Convolucionales son usadas en este tipo de algoritmo. Algo interesante que busca el *deep learning* es el reconocimiento de ciertos patrones, sean aleatorios o no [19].

g. **Procesamiento de Lenguaje Natural (NLP)**

El Procesamiento del Lenguaje Natural (NLP) es una subdisciplina de la IA y rama de la Ingeniería Lingüística Computacional; su objetivo principal es construir sistemas y mecanismos que permitan la comunicación entre personas y máquinas por medio de lenguajes naturales [20]. Asimismo, se plantea que “es el procesamiento, o tratamiento por computadora, del lenguaje natural, es decir, los lenguajes humanos, en oposición a los lenguajes de programación” [21].

Igualmente, se define como el conjunto de herramientas informáticas que permiten analizar el lenguaje humano con el propósito de otorgarle sentido a lo que se encuentra escrito [22].

Ampliando esta definición, se “toma el NLP en un sentido amplio para cubrir cualquier tipo de manipulación informática del lenguaje natural. En un extremo, podría ser tan simple como contar frecuencias de palabras para comparar diferentes estilos de escritura. En el otro extremo, el NLP implica "comprender" las expresiones humanas completas, al menos hasta el punto de poder darles respuestas útiles” [23]. En este contexto, se destaca que el logro de que un computador aprenda a interpretar el lenguaje natural se debe a dos senderos, por una parte, el epistemológico y por otra, el heurístico:

- Epistemológico: Es aquel que define el espacio de conceptos que el programa puede aprender.
- El heurístico: Es el encargado de definir los algoritmos para el aprendizaje [20].

El primer avance obtenido en el NLP, ocurrió en la NASA, específicamente en el área del acceso a las bases de datos con el sistema lunar en el año 1973. El NLP procura poder crear programas capaces de analizar, entender y generar lenguajes que las personas utilizan normalmente, de forma que el usuario pueda llegar a ser capaz de comunicarse con la máquina o computador de la misma forma que lo haría con otro ser humano [20].

III. Metodología

El artículo desarrollado es de tipo revisión donde la unidad de análisis está representada por artículos publicados en bases de datos tales como: Scielo, Dialnet, Science Direct, latindex, Scopus, los cuales abordan la temática de la detección de noticias falsas utilizando la Inteligencia Artificial. Se utilizó una metodología descriptiva-exploratoria, para identificar cómo se detectan las noticias falsas utilizando Inteligencia Artificial, específicamente en el campo del Aprendizaje Automático.

IV. Principales hallazgos

Para contrarrestar el fenómeno de las FN en la actualidad, instituciones, organizaciones, universidades y gobiernos han venido promoviendo diversas iniciativas. Sin embargo, muchas de estas iniciativas recalcan en la IA que con la ciencia de los algoritmos desarrollan *netbots* y plataformas cuyo propósito es combatir las noticias falsas que resultan tóxicas para la sociedad [6]. En este sentido, para poder detectar las FN, y en definitiva diseñar algoritmos efectivos para cumplir con este propósito, “se debe hacer una distinción entre las diferentes clases de FN que existen: a) fabricaciones serias (descubiertas en los medios de comunicación

principales o participantes, prensa amarilla o tabloides); b) engaños a gran escala; c) falsificaciones humorísticas (sátira de noticias, parodia, programas de juegos)” [24]. A partir de este punto, es posible la creación de filtros más efectivos para cada uno de estos tipos de FN.

Existen algunos tipos de IA como los *bots*, que han sido diseñados y creados con el propósito de ayudar a las empresas de noticias en el proceso de verificación de la información, contribuyendo a que los lectores que la reciben, lo hagan de una manera totalmente confiada, creíble, que sirva para la formación de sus opiniones y toma de decisiones de toda índole: económicas, políticas, sociales u otras [6].

Asimismo, se ha evidenciado que los algoritmos de Aprendizaje Automático permiten desarrollar modelos capaces de detectar contenido falso y/o *bots* en redes sociales [25], siendo estas las principales fuentes de *fake news* que se convierten en virales en la web. En la actualidad, la gran mayoría de las actividades de consulta de información se realizan en línea. Los periódicos que antes eran preferidos en su formato impreso cada vez más son sustituidos por aplicaciones como Facebook, Twitter, Instagram, Telegram, WhatsApp y artículos de noticias en portales web. El creciente problema de las FN complica las cosas y al mismo tiempo intenta cambiar u obstaculizar la opinión y actitud de las personas hacia el uso de la tecnología digital.

Cuando un ser humano es engañado por las noticias reales, sucede que la gente empieza a creer que sus percepciones sobre un tema en particular son verdaderas como se supone. Así, con el fin de frenar el fenómeno, las personas han desarrollado su sistema de detección de FN que toma la entrada del usuario y lo clasifica como verdadero o falso [26].

Es menester hacer un recorrido por las técnicas y modelos de Aprendizaje Automático, Procesamiento de Lenguaje Natural y Aprendizaje Profundo desde los modelos más tradicionales que vislumbraban algunos problemas de detección; donde se hacía uso de modelos de IA poco profundos que produjeron representaciones de alta dimensionalidad lingüística, con problemas de la detrás de los datos subyacentes, los cuales principalmente se presentaron debido a que los datos de entrada eran frecuentemente escasos para que un modelo pudiera aprender. Además, debido a la introducción de palabras implícitas de bajo valor e integradas de forma adecuada en el lenguaje natural, los modelos se hacían ineficientes y poco precisos; por lo cual los modelos basados en Redes de Neuronas lograron un éxito rotundo en cuanto a tareas de Procesamiento de Lenguaje Natural.

El Procesamiento de Lenguaje Natural ha visto una tremenda mejoría con el uso del Aprendizaje Profundo con modelos como: Recurrent Neuronal Network (RNN) y otros mejorados como el Long Short Term Memory (LSTM) [27] y el Gated Recurrent Units (GRU) [28] que han demostrado ser promesas en la resolución del problema del NLP, especialmente en la principal cualidad de la NLP que es la capacidad de retener los resultados anteriores y utilizar esos datos en los cálculos actuales. Esta es la razón por la que los modelos RNN son los más adecuados para modelar datos secuenciales como frases y párrafos. A diferencia de otros modelos, que sólo toman el actual conjunto de características de entrada.

Teniendo en cuenta esto, los RNN se constituyen como los métodos más adecuados para las aplicaciones de NLP porque son capaces de capturar todo el significado y el contexto de un largo artículo de la corriente de palabras entrante [29]. Por esta razón, varias tareas de la NLP como el diálogo, generación, subtítulo de imágenes y traducción asistida por máquinas, etc. hacen un uso extensivo de los RNN.

El uso de Redes Neuronales Convolucionales y recurrentes, junto con LSTM, permite encontrar patrones complejos en datos textuales con los cuales se ha hecho análisis de texto sobre *twitter*, demostrando una superioridad en los modelos, frente a modelos convencionales [30], pero también se tiene la aplicación de LSTM bidireccionales para el análisis de patrones textuales de forma que se tenga una revisión en ambas direcciones de los datos, llegando a resultados bastante acertados [31].

También, se han desarrollado modelos de Redes Neuronales incorporando Redes de Memoria de Largo-Corto Plazo (LSTM) y Redes Neuronales Convolucionales (CNN), esto puesto a prueba para clasificar los mensajes de *twitter*. La parte lingüística se introdujo usando GloVe, el cual es una biblioteca de vectores pre entrenados [32], obteniendo muy buenos resultados.

En otro enfoque, se encuentra el análisis de clasificación, que se constituye como un desafío debido a que muchos investigadores han tratado de resolver el problema de detección de noticias falsas bajo métodos de clasificación y así mismo tratando de probar y corroborar que estos mismos son eficientes y funcionales para este tipo de problema, obteniendo resultados deseables [29]. En algunos estudios se han examinado los enfoques de detección de noticias falsas desde la perspectiva de la minería de datos, incluyendo la extracción de características y la construcción de modelos.

Como se mencionó anteriormente, los modelos de Redes Neuronales Profundas apuntan a ser uno de los mejores mecanismos para abordar este problema como por ejemplo con el uso de RNN para la detección de posiciones de noticias falsas, donde con el uso de este tipo de red permiten capturar patrones temporales de la actividad de los usuarios en torno a un determinado texto, entonces el comportamiento del usuario se utiliza para extraer las características de la fuente y usar esta información para construir un modelo de clasificación de noticias falsas [33].

Es importante señalar que incluso un modelo de red simple y directo puede superar a los modelos complejos. Así que, claramente la complejidad del modelo no es la solución óptima en este caso, si no también la elección correcta de los parámetros y datos esenciales como entrada de los modelos [33]. Desafortunadamente, la detección de noticias falsas tiene muchas complicaciones que no pueden ser resueltas usando solo NLP. Es imposible determinar la autenticidad de un artículo con sólo leerlo, incluso para un ser humano. Para confirmar la validez de una noticia más allá de toda duda, la comprobación de los hechos es una necesidad. Por lo tanto, un modelo de Aprendizaje Profundo que considera sólo el aspecto del NLP del artículo, conlleva a enormes limitaciones en el mundo real para resolver este problema [29].

Es evidente que se han hecho muchos intentos, todos ellos con la aplicación de técnicas diferentes y unos más complejos que otros, pero el tema aún no está resuelto. Hay mucho espacio para el desarrollo y la investigación en esta área, especialmente porque las declaraciones de las noticias han involucrado distintas variables adjuntas a ellas: sarcasmo, abreviatura, metáforas, etc. Sin embargo, se han hecho esfuerzos para encontrar la mejor aproximación técnica guiada por la IA para solventar el problema de la propagación de noticias falsas.

V. Conclusiones

Las *fake news* pueden generarse en cualquier ámbito informativo: económico, educativo, social, político; en cada uno de los cuales se han desarrollado iniciativas específicas a través de la creación de algoritmos de Aprendizaje Automático para evitar o minimizar la propagación de las FN.

La gran cantidad de *fake news*, o en general, elementos de desinformación, que se puede detectar diariamente en las redes sociales, hace muy cuestionable su uso como fuente de información, y condiciona de forma determinante la opinión pública [34].

Como se ha señalado previamente, es a través de redes sociales en donde se difunden las FN y en este caso haciendo énfasis en *Twitter*, dada su influencia en temas sociopolíticos y de otros temas de interés, que se ha convertido en uno de los principales puntos de encuentro de los usuarios para discutir y debatir sobre temas de interés colectivo.

La detección de noticias falsas está inmensamente apoyada en el uso de diferentes algoritmos de los cuales existen múltiples avances, investigaciones, modelos, que se han plasmado en artículos y *papers*. Claramente, estos avances no son definitivos y con el pasar del tiempo y apoyados en el crecimiento tecnológico es posible perfeccionar técnicas actuales, vislumbrar nuevos métodos y acrecentar este conocimiento en pro de este objetivo, que en conjunto se integran en retos de fundamental importancia que se deben afrontar.

A partir de lo anterior, es posible afirmar que las noticias falsas tienen una tendencia al crecimiento y son insuficientes las acciones para poder ejercer un control, por lo que esta problemática seguirá generando desinformación en los ciudadanos, obstaculizando la democracia y afectando injustamente la reputación de líderes, empresarios y entidades a nivel nacional e internacional, representando riesgos en las organizaciones para conservar su capacidad operativa en el corto, mediano y largo plazo [31]. Para evitar que este pronóstico se materialice a futuro, es necesaria la implementación de mecanismos confiables y precisos que permitan la detección y propagación de las FN, es aquí donde la IA juega un papel importante presentando un sin fin de modelos y algoritmos capaces de satisfacer esta tarea.

VI. Referencias

- [1] El Tiempo, «Europa busca regular las redes sociales,» [En línea]. Available: <https://www.eltiempo.com/tecnosfera/novedades-tecnologia/como-sera-la-regularizacion-de-las-redes-sociales-en-europa-347622>. [Último acceso: 16 noviembre 2020].
- [2] E. Hunt, «What is fake news? How to spot it and what you can do to stop it,» *The Guardian*, p. 17, 2016.
- [3] N. Fernández García, «Fake news: una oportunidad para la alfabetización mediática,» *NUSO*, n° 269
- [4] S. Aphiwongsophon y P. Chongstitvatana, «Detecting Fake News with Machine Learning Method,» *15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 528-531, 2018.

- [5] O. Nicole, *Machine Learning for Detection of Fake News*, Massachusetts: Massachusetts Institute of Technology, 2018.
- [6] J. Flores y F. Botelho, *Inteligencia Artificial: Cómo los bots pueden contrarrestar las 'fake news' que afectan a las organizaciones de noticias*, Madrid: UNED, 2019.
- [7] C. Pauner, «Noticias falsas y libertad de expresión e información. El control de los contenidos informativos en la red.,» *Teoría y Realidad Constitucional*, vol. 41, pp. 297-318, 2018.
- [8] H. Allcott y M. Gentzkow, «Social media and fake news in the 2016 election.,» *Journal of economic perspectives*, vol. 31, nº 2, pp. 211-236, 2017.
- [9] M. González, «Las fake news siempre han existido,» *La Gaceta*, vol. 14, 2018.
- [10] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Ontario, 1999.
- [11] M. Laster, *Teach the Way the Brain Learns: Curriculum Build Neuron Networks.*, North America: R&L Education., 2009.
- [12] J. Ponce, A. Torres, F. Quezada, A. Silva, E. Martinez y A. Casali, *Inteligencia Artificial*, 1a ed. - Iniciativa Latinoamericana de Libros de Texto Abiertos (LATIn), 2014. 225, 2014.
- [13] B. C. Tayiz, «Alternative NLP Method,» 6 enero 2020. [En línea]. Available: <https://becominghuman.ai/alternative-nlp-method-9f94165802ed>. [Último acceso: 25 noviembre 2020].
- [14] D. Changyu, J. Xunbi, R. Colton y Z. Jianyu, *Machine Learning with Human Knowledge*, ISCIENCE, 2020.
- [15] L. Rouhianen, *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*, Barcelona: Alienta Editorial, 2018.
- [16] J. Lanier, *Diez razones para borrar tus redes sociales de inmediato.*, Madrid, 2018.
- [17] E. De la Rosa Montero, *El aprendizaje profundo para la identificación de sistemas no lineales*, Centro de Investigación y de estudios avanzados del Instituto Politécnico nacional., 2014.
- [18] Yannn lecun, «Yannn lecun,» [En línea]. Available: <http://Yannn.lecun.com/>. [Último acceso: 25 noviembre 2020].

- [19] B. Chandra, «Fast learning in Deep Neural Networks.,» *Neurocomputin*, vol. 171, pp. 1205-1215, 2016.
- [20] P. Benavides y S. Rodríguez, *Procesamiento del lenguaje natural en la recuperación de información*, Bogotá: Universidad de La Salle, 2007.
- [21] K. Cohen y L. Hunter, «Natural language processing and systems biology.,» *Artificial intelligence methods and tools for systems biology*, pp. 147-173, 2004.
- [22] J. Prada, *Natural language processing como herramienta para detectar fake news: aplicado a las elecciones de alcaldía de Cali para el periodo 2020-2023*, Santiago de Cali: Universidad Icesi, 2019.
- [23] S. Bird , E. Klein y E. Loper, «Natural language processing with Python: Analyzing text with the natural language toolkit.,» *O'Reilly Media, Inc*, 2009.
- [24] V. Rubin, Y. Chen y N. Conroy, «Deception detection for news: Three types of fakes,» *Information Science with Impact: Research in and for the Community*, vol. 83, 2015.
- [25] W. Cuenca, *Machine Learning aplicado a seguridad*, Catalunya: Universidad Oberta de Catalunya, 2019.
- [26] U. Sharma, S. Sidarth y P. Shankar, «Fake News Detection Using Machine Learning Algorithms,» *International Journal of Creative Research Thoughts (IJCRT)*, vol. 8, n° 6, pp. 1394-1402, 2020.
- [27] S. Hochreiter y J. Schmidhuber, «Long short-term memory,» *Neural computation*, vol. 9, 1997.
- [28] J. Chung, C. Gulcehre y K. C. Yo, *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*, Nueva York: Cornell University, 2014.
- [29] S. Deepak y C. Bhadrachalam, «Deep neural approach to Fake-News identification, » *Procedia Computer Science*, vol. 167, pp. 2236-2243, 2020.
- [30] H. Mohammad, R. Goldani y S. Safabakhsh, «Convolutional neural network with margin loss for fake news detection,» *Information Processing & Management*, vol. 58, 2021.
- [31] B. Pritika Bahad, S. Preeti y K. Raj , «Fake News Detection using Bi-directional LSTM-Recurrent Neural Network,» *Procedia Computer Science*, vol. 165, pp. 74-82, 2019.

- [32] S. Volkova, K. Shaffer, J. Jang y N. Hodas, «Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter,» *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 2, pp. 647-653, 2017.
- [33] N. Ruchansky, S. Seo y Y. Liu, «Csi: A hybrid deep model for fake news detection,» *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 797-806, 2017.
- [34] A. Ferrer, Desarrollo y aplicación de un sistema de Inteligencia Artificial para la detección, KPI RISK ETHICS & COMPLIANCE, 2019.

Publicación Facultad de Ingeniería y Red de Investigaciones de Tecnología Avanzada – RITA

REVISTA

TIA