

Hindawi
Programación científica
Volumen 2021, número de artículo 3434458, 10
páginas <https://doi.org/10.1155/2021/3434458>

Artículo de investigación

Detección de noticias falsas en línea mediante la combinación Aprendizaje conjunto

Arvin Hansrajh, Timothy T. Adeliyi y Jeanette Wing

Grupo de Investigación sobre TIC y Sociedad, Universidad de Durban de Technology, Durban 4001, Sudáfrica

La correspondencia debe dirigirse a Timothy T. Adeliyi; timothy2@dut.ac.za

Recibido el 1 de mayo de 2021; Aceptado el 21 de julio de 2021; Publicado 29 de julio de 2021

Editor académico: Shah Nazir

Copyright © 2021 Arvin Hansrajh et al. (es un artículo de acceso abierto distribuido bajo la licencia de atribución de Creative Commons, que permite el uso, distribución y reproducción sin restricciones en cualquier medio, siempre que el trabajo original se cite correctamente.

no se están aplicando varios modelos avanzados de aprendizaje automático, aunque estudios recientes demuestran la eficacia del enfoque de aprendizaje automático por conjuntos; por lo tanto, el propósito de este estudio es ayudar en la detección automática de noticias falsas. Se adopta un enfoque de conjunto para ayudar a resolver la brecha identificada. (Este estudio propuso un modelo de conjunto de aprendizaje automático combinado desarrollado a partir de regresión logística, máquina de vectores de soporte, análisis discriminante lineal, descenso de gradiente estocástico y regresión de cresta, que luego se utiliza en un conjunto de datos disponible públicamente para predecir si un informe de noticias es verdadero o no. (El modelo propuesto se evaluará con los modelos clásicos de aprendizaje automático populares, mientras que métricas de rendimiento como AUC, ROC, recuperación, exactitud, precisión y aunque estudios recientes están demostrando la eficacia del enfoque de aprendizaje automático por conjuntos; por lo tanto, el propósito de este estudio es ayudar en la detección automática de noticias falsas. Se adopta un enfoque de conjunto para ayudar a resolver la brecha identificada. (Este estudio propuso un modelo de conjunto de aprendizaje automático combinado desarrollado a partir de regresión logística, máquina de vectores de soporte, análisis discriminante lineal, descenso de gradiente estocástico y regresión de cresta, que luego se utiliza en un conjunto de datos disponible públicamente para predecir si un informe de noticias es verdadero o no. (El modelo propuesto se evaluará con los modelos clásicos de aprendizaje automático populares, mientras que métricas de rendimiento como AUC, ROC, recuperación, exactitud, precisión y Se adopta un enfoque de conjunto para ayudar a resolver la brecha identificada. (Este estudio propuso un modelo de conjunto de aprendizaje automático combinado desarrollado a partir de regresión logística, máquina de vectores de soporte, análisis discriminante lineal, descenso de gradiente estocástico y regresión de cresta, que luego se utiliza en un conjunto de datos disponible públicamente para predecir si un informe de noticias es verdadero o no. (El modelo propuesto se evaluará con los modelos clásicos de aprendizaje automático populares, mientras que métricas de rendimiento como AUC, ROC, recuperación, exactitud, precisión y Se adopta un enfoque de conjunto para ayudar a resolver la brecha identificada. (Este estudio propuso un modelo de conjunto de aprendizaje automático combinado desarrollado a partir de regresión logística, máquina de vectores de soporte, análisis discriminante lineal, descenso de gradiente estocástico y regresión de cresta, que luego se utiliza en un conjunto de datos disponible públicamente para predecir si un informe de noticias es verdadero o no. (El modelo propuesto se evaluará con los modelos clásicos de aprendizaje automático populares, mientras que métricas de rendimiento como AUC, ROC, recuperación, exactitud, precisión y que luego se utiliza en un conjunto de datos disponible públicamente para predecir si un informe de noticias es verdadero o no. (El modelo propuesto se evaluará con los modelos clásicos de aprendizaje automático populares, mientras que métricas de rendimiento como AUC, ROC, recuperación, exactitud, precisión y que luego se utiliza en un conjunto de datos disponible públicamente para predecir si un informe de noticias es verdadero o no. (El modelo propuesto se evaluará con los modelos clásicos de aprendizaje automático populares, mientras que métricas de rendimiento como AUC, ROC, recuperación, exactitud, precisión ySe utilizará modelos de aprendizaje.

1. Introducción

(El uso cada vez mayor de Internet junto con las plataformas de redes sociales ha permitido que incluso más personas obtengan noticias de una amplia variedad de fuentes en lugar de los medios de comunicación tradicionales. Las personas que pasan mucho tiempo en línea tienen más probabilidades de obtener noticias y actualizaciones a través de redes sociales con un mayor riesgo de exposición a información errónea a gran escala [1] (proporciona un terreno fértil para noticias falsas a medida que proliferan artículos de noticias, engaños, reseñas, rumores, sátiras, anuncios, reseñas y afirmaciones exageradas. - la cantidad de noticias falsas es capaz de producir efectos extremadamente adversos en las personas y la humanidad [2]. Ahora se ha convertido en parte de la vida cotidiana escuchar sobre el empeoramiento de las crisis climáticas, la violencia política, la intolerancia entre las personas de

diferentes orígenes étnicos y culturales, e incluso influyen en cuestiones de salud pública. (Esto se hace a menudo para promover o poner en circulación ciertas ideas y a menudo se realiza con agendas políticas. (Por lo tanto, todos los gobiernos de todo el mundo están tratando de rastrear y abordar este problema [3]. Las noticias falsas no son una novedad concepto [4] o un producto de la era de la comunicación digital [5]. Ha salido a la luz recientemente durante las elecciones presidenciales de EE. UU. e influenciado por estas historias.

Facebook ha estado en el epicentro de la polémica de los medios de comunicación por apuntar a la población y mostrar los mensajes a su apoyo [6]. Se ha alegado que una noticia falsa podría haber sido decisiva en la presidencia de Estados Unidos de 2016.

elección [7]. Sin embargo, podemos sostener que las noticias falsas y la desinformación en general se han convertido en un gran problema, que puede tener un costo social significativo en el futuro. (La naturaleza ubicua de Internet permite a cualquier persona difundir información falsa y sesgada fácilmente. Es prácticamente imposible prevenir o controlar la creación o difusión de noticias falsas. En consecuencia, tanto las plataformas en línea como los investigadores son muy proactivos en la detección de posibles noticias falsas. es un problema complejo, ya que las noticias falsas pueden presentarse de múltiples formas, lo que dificulta su identificación de manera eficiente tanto física como automáticamente [8].

Los titulares en forma de clickbait se utilizan para atraer a los usuarios a ver artículos probablemente subjetivos para obtener ganancias. Según Wang [9], "(el problema de la detección de noticias falsas es más desafiante que la detección de críticas engañosas, ya que el lenguaje político en las entrevistas de televisión y publicaciones en Facebook y Twitter son en su mayoría declaraciones breves". de soluciones automáticas para la detección de noticias falsas es imperativo y exigente [3]. Trabajos anteriores han utilizado muchos modelos clásicos. Sin embargo, varios modelos de aprendizaje no convencionales no se aplican aunque han demostrado ser los mejores en numerosos problemas de clasificación de texto [10]. Se propone un enfoque para ayudar a resolver la brecha identificada.

Estudios recientes están demostrando la efectividad de los enfoques de aprendizaje por conjuntos con resultados prometedores [11]. (Este estudio investigará cómo las técnicas de procesamiento del lenguaje natural y el aprendizaje automático se pueden combinar en un enfoque de conjunto de combinación para crear un modelo que utilizará los datos de informes de noticias anteriores y predecirá un informe de noticias como verdadero o no. compararse con los modelos clásicos de aprendizaje automático utilizando métricas de rendimiento, por ejemplo, AUC, ROC, recuperación, exactitud, precisión y F1 puntuación. (Estas medidas se utilizarán para medir el rendimiento del modelo.

(El resto de este documento se describe a continuación. (La literatura relacionada se discute en la Sección 2. La Sección 3 presenta los materiales y métodos de estudio, mientras que el análisis de resultados se presenta en la Sección 4 y el documento se concluye sucintamente en la Sección 5.

2. Obras relacionadas

Los humanos son bastante poco impresionantes al reconocer el engaño. La mayoría de las personas creen que la información que obtienen es objetiva y confiable. (Tienden a ser inexplicablemente perceptivos al conocimiento que no se comprende completamente [12]. El sesgo de confirmación influye en las personas para que capten solo lo que quieren percibir [13]. (Por lo tanto, la proliferación y propagación de noticias falsas es una preocupación importante debido a Su capacidad para generar consecuencias devastadoras. Se utilizan diversos enfoques de aprendizaje automático para combatirlo. Sin embargo, la mayoría se centró en una categoría específica de noticias sin utilizar varios métodos avanzados [13, 14].

Se han aplicado numerosas redes neuronales y modelos basados en el aprendizaje automático para detectar noticias falsas. Los modelos se desarrollaron con características diseñadas para conjuntos de datos específicos. Yun y Ahn [15] detectaron noticias falsas en Corea con aprendizaje automático y minería de texto utilizando un enfoque de dos pasos.

Inicialmente, los contenidos de las noticias se convierten a valores aplicando minería de texto, y luego los clasificadores son entrenados en estos valores. Aphiwongsophon y Chongstitvatana [16] basaron sus modelos en la identificación de noticias falsas utilizando datos seleccionados de Twitter. Es probable que estos enfoques sean víctimas del sesgo del conjunto de datos y posiblemente funcionen mal en una categoría diferente de noticias [10]. Gilda [17] exploró algunos enfoques tradicionales de aprendizaje automático. Ahmed y col. [18] investigó y comparó seis técnicas de clasificación diferentes usando análisis de n-gramas en un solo conjunto de datos usando extracción de características. (Los modelos se evaluaron de forma independiente y el clasificador de máquina de vector de soporte lineal logró la mejor puntuación. Sin embargo, varios modelos de aprendizaje avanzado son no aplica a pesar de que han sobresalido en la clasificación de textos [10].

La investigación que utiliza el aprendizaje profundo para identificar trabajos de noticias falsas ha logrado resultados alentadores [10]. Rashkin y col. [19] utilizó análisis de características lingüísticas y logró el notable resultado de la memoria a corto plazo. Wang [9] construyó un modelo híbrido utilizando una red neuronal convolucional que superó a otros modelos de aprendizaje tradicionales. Singhanian y col. [20] aplicó una red de atención de tres niveles que incorpora oraciones, palabras y titulares. (ota et al. [21] presentó una red neuronal para pronosticar la postura utilizando el titular y el cuerpo del artículo. Wang [9] presentó un conjunto de datos de referencia llamado Mentiroso e investigó utilizando modelos actuales. (La evaluación insinúa cómo se comportan los diferentes tipos de modelos en los datos estructurados. Además, algunos modelos eran propensos a sobreajustarse.

Ruchansky y col. [22] construyó un modelo CSI (captura, puntuación e integración) que usaba texto, respuesta al artículo y características del comportamiento de los usuarios. Ajao y col. [23] desarrolló un marco para clasificar e identificar noticias falsas en publicaciones de Twitter utilizando un híbrido de redes neuronales. (La táctica identificó intuitivamente las características pertinentes sin tener en cuenta los conocimientos previos. Lu y Li [24] desarrollaron (GCAN) Redes de atención conjunta conscientes de gráficos para determinar si un tweet es falso mediante el uso de la secuencia asociada de usuarios de retweet. Khan et al. [10] analizó el rendimiento de enfoques diferentes en tres conjuntos de datos y mostró que Naive Bayes puede lograr un resultado similar al de los modelos de redes neuronales cuando se trabaja con un conjunto de datos que contiene menos de 100 mil artículos Vijayaraghavan et al. [25] aplicó diferentes modelos para detectar noticias falsas y afirmar que las redes neuronales generalmente funcionan de manera consistente y sirven como un poderoso aproximador universal. Sin embargo, la pérdida y la precisión se producen después de usar demasiadas épocas y, por lo tanto, entra en juego el problema del sobreajuste. Además, un modelo más simple que usa regresión logística también arrojó buenos resultados de desempeño. En consecuencia, no se sigue necesariamente que cuanto más complicado sea el modelo, mejor será el rendimiento. Además, el aprendizaje profundo "consume tiempo y recursos" [26]. En consecuencia, no se sigue necesariamente que cuanto más complicado sea el modelo, mejor será el rendimiento. Además, el aprendizaje profundo "consume tiempo y recursos" [26]. En consecuencia, no se sigue necesariamente que cuanto más complicado sea el modelo, mejor será el rendimiento. Además, el aprendizaje profundo "consume tiempo y recursos" [26].

Los investigadores han estudiado numerosos algoritmos para la clasificación de textos que ofrecen un buen rendimiento. Sin embargo, algunos algoritmos funcionan mejor en algunos conjuntos de datos, pero pueden dar incluso un rendimiento medio en otros conjuntos de datos. (Por lo tanto, en lugar de usar un solo clasificador, es mejor usar un grupo de clasificadores y tomar una decisión colectiva o en equipo, en lugar de

basar una decisión en un clasificador individual [27, 28]. (Este enfoque llamado enfoque de conjunto supera la debilidad de un clasificador por la fuerza de otros clasificadores y ofrece un mejor rendimiento que un clasificador individual. (La naturaleza diversa del enfoque y el mantenimiento de la varianza bajo control contribuyen en gran medida a su éxito. el aprendizaje puede resultar en esquemas de clasificación más robustos.

Roy y col. [3] desarrolló modelos basados en una memoria bidireccional a corto plazo y una red neuronal convolucional. (La salida de ambos modelos se ingresó en un modelo de perceptrón multicapa para obtener el resultado final. Al-Ash et al. [29] utilizaron un clasificador de bosque aleatorio que consiste en un clasificador de árbol de decisión como un clasificador de conjunto para detectar noticias falsas de Indonesia . Reddy et al. [26] presentaron un enfoque híbrido para la detección de datos falsos usando un modelo de conjunto. Ahmad et al. [30] exploraron diferentes propiedades textuales en un enfoque de conjunto para detectar noticias falsas. Gutiérrez-Espinoza et al. [11] evaluó el rendimiento del aprendizaje por conjuntos utilizando diferentes técnicas de aprendizaje automático para la clasificación con el fin de identificar información falsa en línea.

Mahabub [31] utilizó un método distinto para detectar noticias falsas en el desarrollo de un clasificador de votación en conjunto que incorpora muchos algoritmos de aprendizaje automático familiares. Kaur y col. [32] diseñó un modelo de votación con múltiples niveles para automatizar la detección de noticias falsas experimentando con varios modelos. Saeed y col. [1] incorporó un enfoque de en-semble para detectar spam de textos árabes. Li y col. [33] aplicó una canalización para identificar noticias falsas teniendo en cuenta el título y el texto del artículo en un conjunto apilado. En todos estos estudios, el enfoque de conjunto arrojó un mejor desempeño en comparación con el modelo individual en la detección de información engañosa. (Por lo tanto, se adopta un enfoque de conjunto avanzado para detectar noticias falsas. (La estrategia integrará la combinación y el aprendizaje automático con el procesamiento del lenguaje natural para extender y mejorar los enfoques actuales.

3. Materiales y métodos

En esta sección, presentamos los conjuntos de datos, el marco propuesto, la explicación de los algoritmos y las métricas que se utilizan para la evaluación del desempeño. Se han seleccionado dos conjuntos de datos para nuestros experimentos que incluyen noticias de una variedad de categorías diferentes y una combinación de artículos falsos y veraces. Ambos conjuntos de datos están disponibles públicamente y son fácilmente accesibles en la web. La categorización de noticias como "noticias falsas" puede ser "una tarea muy desafiante y que requiere mucho tiempo" [34]. Por lo tanto, en este estudio se utilizan conjuntos de datos existentes. (El mayor desafío para identificar noticias falsas es la accesibilidad y el calibre de los conjuntos de datos [35]. Además, encontrar un corpus de artículos relacionados con noticias es particularmente problemático debido a preocupaciones sobre derechos de autor [17]. (e Liar [9] e ISOT [36] se utilizan conjuntos de datos.

(El conjunto de datos del mentiroso es de acceso público y se ha utilizado con éxito [37]. Comprende 12836 declaraciones breves etiquetadas depolitifact.com. (Hay seis etiquetas para calificar la veracidad de una declaración: "pantalones-fuego", "falso", " apenas-

verdadero ", " cierto a medias ", " mayormente cierto "y" verdadero ". Nos enfocamos en clasificar las noticias como verdaderas o falsas. Para la clasificación binaria de las noticias, transformamos estas etiquetas en dos etiquetas. "Pantalones de fuego", "falso" y "apenas cierto" se consideran falsos y "medio verdadero", "mayormente verdadero" y "verdadero" se consideran verdaderos. (Este conjunto de datos se centra en gran medida en la política que contiene declaraciones de republicanos y demócratas, además de una cantidad sustancial de publicaciones de las redes sociales [10].

(El conjunto de datos de noticias falsas de ISOT comprende artículos de noticias falsos y verdaderos procedentes de varios dominios [30]. (Los artículos verdaderos se obtuvieron principalmente de reuters.com, un sitio de noticias muy conocido en la web. de numerosas fuentes, principalmente de sitios web que han sido marcados por politifact.com. (El conjunto de datos consta de 44.898 artículos, 23.481 son artículos falsos y 21.417 artículos verdaderos. Cada punto de datos consta de un título, texto, asunto y fecha. (e El texto es el artículo de noticias real, y el tema o categoría es cualquiera de Medio Oriente, noticias del gobierno, noticias de EE. UU., noticias del mundo, noticias de política, noticias de izquierda, política y noticias.

En el marco propuesto, como se muestra en la Figura 1, estamos ampliando la literatura actual mediante la introducción de técnicas de aprendizaje por conjuntos que incorporan la combinación. Los artículos de noticias de varios dominios se clasifican como verdaderos o falsos al trabajar con diferentes conjuntos de características. En nuestro enfoque se utilizan técnicas de combinación de conjuntos con frecuencia de término, frecuencia de documento invertida de frecuencia de término y n-gramas.

Los textos sin procesar de las noticias deben procesarse previamente antes de introducirlos en los modelos. Se aplicarán técnicas de procesamiento del lenguaje natural para ayudar a mejorar la precisión. (Las siguientes operaciones se llevarán a cabo durante el procesamiento previo del conjunto de datos:

- (i) Limpieza de datos: elimine los datos irrelevantes que no se requieren para el análisis.
- (ii) Compruebe si faltan valores que puedan tener un efecto adverso en el resultado final.
- (iii) Convierta el texto a minúsculas para que haya coherencia.
- (iv) Elimine todos los signos de puntuación.
- (v) Elimine las palabras vacías del conjunto de datos textuales. (Estas son palabras que no proporcionan un significado semántico añadido y no tienen importancia durante el procesamiento del lenguaje natural.
- (vi) La derivación (o lematización) implica convertir las palabras a su estructura original y, por lo tanto, reducir las clases o tipos de palabras presentes en el conjunto de datos. Por ejemplo, "Dancing", "Dance" y "Dancer" se reducirán a "dance". La derivación hace que la clasificación sea más eficiente y rápida [18]. (Se utilizará el algoritmo Porter Stemmer debido a su precisión.

El diseño de las funciones juega un papel clave en el rendimiento de los modelos de aprendizaje automático. (La extracción de las palabras más relevantes o importantes y su uso como características puede ser extremadamente útil. Término Frecuencia, Término Frecuencia Invertida

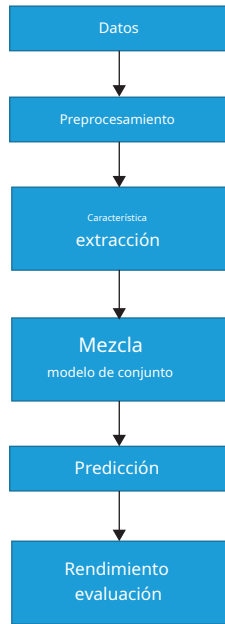


Figura 1: Modelo propuesto.

La frecuencia del documento y los n-gramas se utilizarán en la extracción de características del conjunto de datos. (Este enfoque ha sido elegido sobre la incrustación de palabras en base a los resultados experimentales obtenidos en estudios previos. (ota et al. [21] obtuvieron mejores resultados utilizando n-gramas sobre la incrustación de palabras. Vijayaraghavan et al. [25] utilizaron la incrustación de Word2Vect y mostró que tuvo el peor desempeño en comparación con los modelos TF-IDF. Resultados similares también fueron con fi rmados por Smitha y Bharath [38]. Término Frecuencia utiliza los recuentos de palabras presentes en los documentos para determinar la semejanza entre los documentos. Un vector con una dimensión igual que contiene el recuento de palabras se asocia con cada documento Término Frecuencia La frecuencia invertida del documento es una métrica que se utiliza con frecuencia en el procesamiento del lenguaje natural y la recuperación de información. Mide la importancia de un término en un documento incluido en el conjunto de datos. (Los gramas basados en palabras se utilizan para representar el contexto del documento y para la generación de características que pueden ser útiles en la clasificación de un documento como real o falso. (Este enfoque se ha utilizado con éxito con unigramas y bigramas en la detección de noticias falsas. [10].

3.1. Mezcla EnsembleModel.La mezcla está muy relacionada con el apilamiento. El apilamiento (generalización apilada) implica un algoritmo de aprendizaje entrenado para agrupar las predicciones de varios otros algoritmos de aprendizaje. Todos los algoritmos se entrenan con los datos disponibles. Finalmente, se utiliza un algoritmo combinador para la predicción final teniendo en cuenta las predicciones de los otros algoritmos [39]. (El conjunto de mezcla es una variación del apilamiento. (La variación del conjunto de mezcla de predicción se basa en una validación de conjunto de datos de reserva que se utilizó en este estudio para ajustarse al metamodelo en lugar de las predicciones fuera del pliegue. (El modelo aprende a combinar las predicciones de varios conjuntos contribuyentes

modelos base. En la formación del conjunto se utilizan modelos que implementan regresión logística, máquina de vectores de soporte, análisis discriminante lineal, descenso de gradiente estocástico y regresión de crestas. (El algoritmo de conjunto de mezcla se da como sigue (Algoritmo 1).

3.1.1. Regresión logística.Se utiliza un modelo de regresión logística ya que el texto se está clasi fi cando dando como resultado una salida binaria (0/1 o verdadero / falso o verdadero / falso). (La función de hipótesis puede ser definido matemáticamente de la siguiente manera:

$$h_{\theta}(X) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 X))}. \quad (1)$$

Una función sigmoidea transforma la salida en una probabilidad. (El objetivo es lograr una probabilidad óptima mediante minimizando la función de costo como se muestra como sigue [30]:

$$\text{Porque } t(h_{\theta}(X), y) = \begin{cases} \text{Iniciar sesión}(h_{\theta}(X), & y = 1, \\ \text{-registro}(1 - h_{\theta}(X), & y = 0. \end{cases} \quad (2)$$

Por lo tanto, la regresión logística produce una curva logística que se restringe a valores entre 0 y 1 mediante la utilización de una función sigmoidea.

3.1.2. Máquinas de vectores soporte.La máquina de vectores de soporte crea un hiperplano para aislar y agrupar entidades. Los vectores de soporte se crean a cada lado del hiperplano para calcular el hiperplano óptimo con cada vector maximizando la distancia entre ellos. (Una mayor distancia del vector alrededor del hiperplano da como resultado un límite de decisión más preciso entre las características de la categoría [25]. (Los puntos de datos se clasifican en clases distintas dependiendo de su posición en el hiperplano. (El motivo clave es maximizar las brechas que existen entre el hiperplano y los puntos de datos (el margen se maximiza mediante la función de pérdida. (El hiperplano se define por

$$w \cdot x + B = 0, \quad (3)$$

dónde w es el vector de peso y B es el sesgo.

$$L(w) = \max_{x_i} \{0, 1 - y_i (w \cdot x_i + b) / \|w\| \} + \lambda \|w\|. \quad (4)$$

(Los errores se calculan por el primer término en la función de pérdida. (La función de regularización está representada por el segundo término y se usa para evitar el sobreajuste [27].

3.1.3. Análisis discriminante lineal.El análisis discriminante lineal (LDA) es un algoritmo de aprendizaje automático para la clasificación. Calcula las estadísticas de resumen, por ejemplo, la desviación estándar y la media, asociadas con las características de entrada por la etiqueta de clase. (Las estadísticas indican lo que ha aprendido el modelo al consumir los datos de entrenamiento. Las predicciones se basan en estimaciones de probabilidad de un nuevo ajuste de caso en una etiqueta de clase establecida para cada valor de entidad de entrada. (La clase que tiene la mayor probabilidad se asigna a el nuevo caso. LDA puede verse como un sencillo

(1) Divida el conjunto de datos.
(El conjunto de datos se divide en conjuntos de prueba y de tren.)

(2) Construya los modelos base.

(3) Entrene al conjunto de mezcla.

Repetir
 Encaja en el conjunto de entrenamiento.
 Hacer predicciones en el set de reserva.
 Almacene las predicciones como entrada para la combinación hasta el final de los modelos base.

Construya una matriz 2D utilizando las predicciones almacenadas.

Crea el modelo de fusión.
 Ajuste el modelo de fusión en las predicciones de los modelos base.

(4) Haga predicciones con el conjunto de mezcla.

Repetir
 Haga predicciones con el modelo base en el conjunto de prueba.
 Almacene la predicción hasta el final de los modelos base. Construya una matriz 2D utilizando las predicciones almacenadas.

(5) Evalúe las predicciones.

ALGORITMO 1: algoritmo de conjunto de fusión.

aplicación del teorema de Bayer dirigido a la clasificación.
(El proceso se puede resumir mediante tres pasos importantes [37]:

- (i) Calcule la varianza entre clases (separabilidad) entre las diferentes clases. (se expresa por la siguiente fórmula:

$$S_B = \frac{1}{I} \sum_{i=1}^I \text{gramo}(\bar{X}_i - \bar{X})^2 \quad (5)$$

- (ii) Calcule la varianza dentro de la clase utilizando la siguiente fórmula baja:

$$S_W = \frac{1}{I} \sum_{i=1}^I \text{gramo} \left(\frac{1}{N_i} \sum_{j=1}^{N_i} (X_{yo,j} - \bar{X}_i)^2 \right) \quad (6)$$

- (iii) Cree el espacio de dimensiones inferiores para maximizar la varianza entre clases y minimizar la varianza dentro de la clase. (La proyección espacial de dimensiones inferiores (criterio de Fisher) viene dada por la siguiente ecuación:

$$PAG_{ida} = \argmax_{PAG} \frac{\sum_{i=1}^I \text{PD}_{B,PAG}}{\sum_{i=1}^I S_{WP}} \quad (7)$$

LDA asume que tenemos variables de entrada numéricas distribuidas normalmente y tenemos el mismo margen (varianza). De lo contrario, puede ser necesario transformar o normalizar los datos antes de modelar. (El modelo es multiclase. Admite problemas de clasificación de clase doble y clasificación multiclase sin modificación.

3.1.4. Descenso de gradiente estocástico. El descenso de gradiente estocástico utiliza un método iterativo para optimizar una función objetivo a través de propiedades de suavidad apropiadas, como diferenciable o subdiferenciable. (El método consume muestras seleccionadas aleatoriamente para medir la

gradientes. (Por lo tanto, el descenso de gradiente estocástico "se puede considerar como una aproximación estocástica de la optimización del descenso de gradiente" [6]. (e gradiente es principalmente la pendiente o inclinación de una función. Es la cantidad "de cambio de un parámetro con el cantidad de cambio en otro parámetro " [40]. (Cuanto mayor es el gradiente, más pronunciada es la pendiente. El descenso de gradiente se aplica iterativamente para encontrar los valores de los parámetros de una función que minimizarán el valor de la función con la cantidad máxima. (Por lo tanto, el objetivo es determinar los valores óptimos de los parámetros en orden para obtener el valor mínimo de la función de costo.

Matemáticamente, los detalles se pueden expresar (para la clasificación) de la siguiente manera: dado un conjunto de ejemplos de entrenamiento $(X_1, y_1), \dots, (X_{norte}, y_{norte})$, donde $X_i \in \mathbb{R}_{metro}$ y $y_i \in \{-1, 1\}$, el objetivo es aprender una función de puntuación lineal. $f(x) = w^T x + B$ con parámetros de modelo $w \in \mathbb{R}_{metro}$ e interceptar $B \in \mathbb{R}$. Las predicciones para la clasificación binaria se hacen mirando el signo de $f(x)$. Para determinar los parámetros del modelo, el error de entrenamiento regularizado se minimiza y se muestra como sigue:

$$E(w, b) = \frac{1}{norte} \sum_{i=1}^{norte} L(y_i, f(x_i)) + \alpha \|w\|^2 \quad (8)$$

dónde L es una función de pérdida y R es un regularización (penalty) término utilizado para penalizar al modelo complejidad; $\alpha > 0$ es un hiperparámetro que controla la fuerza de la regularización [37].

3.1.5. Regresión de crestas. (El método de regresión sirve como base para el clasificador de crestas. Para la clasificación binaria, la variable de destino se convierte en +1 o -1 dependiendo de la clase a la que pertenece, y para los datos multiclase que utilizan regresión de múltiples salidas, el valor más grande para la predicción se reconoce como la clase de destino. La regresión de crestas es prácticamente igual que la regresión lineal, excepto que se introduce un pequeño sesgo. En consecuencia, la varianza se reduce significativamente. Por tanto, si se comienza con un ajuste algo peor, es posible realizar mejores predicciones a largo plazo. (e añadido

El sesgo se denomina penalización por regresión de crestas. Se calcula encontrando el producto de lambda y el peso al cuadrado asociado con cada característica. (La imposición de una penalización basada en el tamaño de los coeficientes se ocupa de algunos de los problemas de los mínimos cuadrados ordinarios. La regresión de crestas aborda algunos de los problemas de los mínimos cuadrados ordinarios al imponer una penalización sobre el tamaño de los coeficientes. suma residual de cuadrados se minimiza por el coeficientes de cresta:

$$\min_w \|Xw - y\|_2^2 + \alpha \|w\|_2^2 \quad (9)$$

$\alpha \geq 0$ es el parámetro de complejidad que regula a contracción. (mi mayor valor de α , cuanto mayor sea el contracción, y por lo tanto, los coeficientes resultan ser cada vez más robustos a la colinealidad.

4. Resultados y discusión

En este estudio, presentamos el análisis de rendimiento de los modelos tradicionales de aprendizaje automático y el conjunto de mezclas. (Se realiza tanto para los conjuntos de datos Liar como para ISOT. Se han utilizado seis medidas de desempeño en la comparación de los seis modelos. $F1$ puntuación, AUC, precisión, recuperación y exactitud. (Las métricas se calculan para las clases reales y falsas. La Tabla 1 resume los resultados experimentales para el conjunto de datos de Mentiroso medido en la detección de noticias falsas.

(El modelo base con mejor desempeño en el conjunto de datos de Mentiroso es el clasificador de regresión logística que logró las mejores puntuaciones para cuatro de las seis métricas de comparación. (Estos incluyen ROC AUC, AUC, precisión y exactitud. Sin embargo, en general, el conjunto de mezcla ha entregado el mejor rendimiento (las cuatro mejores puntuaciones obtenidas de las seis incluyen ROC AUC, AUC, recuerdo y precisión. La Tabla 2 resume los resultados experimentales para el conjunto de datos ISOT medidos en la detección de noticias falsas.

(El clasificador de máquina de vectores de soporte lineal es el modelo base con mejor rendimiento en el conjunto de datos ISOT con las mejores puntuaciones en cinco de seis métricas de comparación (que incluyen ROC AUC, $F1$ puntuación, AUC, recuerdo y precisión. Sin embargo, en general, el conjunto de mezcla es el modelo de mejor rendimiento. (Las cuatro mejores puntuaciones obtenidas incluyen ROC AUC, $F1$ puntuación, recuperación y precisión de las seis posibles métricas de rendimiento.

4.1. ROC Curve. (El desempeño de un problema de clasificación puede visualizarse o verificarse usando la curva ROC. (e tasa de verdaderos positivos (en el y-eje) se traza contra la tasa de falsos positivos (en el X-eje). Se considera una curva de probabilidad. (El área bajo la curva se considera una métrica clave para evaluar el desempeño de clasificación del modelo. Mide el desempeño del problema de clasificación en diferentes configuraciones de umbral e indica el "grado o medida de separabilidad" [6]. la capacidad

del modelo para distinguir entre diferentes clases. (Cuanto mayor sea la medición del área bajo la curva, mejor podrá distinguir el modelo entre artículos de noticias verdaderos y artículos de noticias falsos. (Las curvas ROC para los conjuntos de datos Mentiroso e ISOT se muestran en las Figuras 2 y 3, respectivamente. En ambos casos, observamos que el conjunto de mezcla es el modelo superior ya que el área encerrada debajo de la curva es la más grande. Por otro lado, el área bajo las curvas ROC para la LDA es la más pequeña (por lo tanto, podemos concluir que la LDA El modelo tiene el peor desempeño en ambos conjuntos de datos (las deducciones se pueden validar fácilmente verificando las respectivas puntuaciones AUC ROC en las tablas de métricas de desempeño.

4.2. Curva de recuperación de precisión. (La curva de recuperación de precisión para los conjuntos de datos Liar e ISOT se muestra en las Figuras 4 y 5, respectivamente. Se construye calculando y luego trazando la precisión (en el y-eje) contra el retiro (en el X-eje) para cada clasificador en varios umbrales. (La curva resume la compensación relativa a la tasa positiva verdadera y la etiqueta predictiva positiva (valor) para un modelo de clasificación (predictivo) mediante el consumo de umbrales de probabilidad variados.

Un buen clasificador mantiene tanto un alto recuerdo como una alta precisión en todo el gráfico y "abrazará" la esquina superior derecha en los gráficos de abajo [41]. (es evidente para el conjunto de datos ISOT que indica un desempeño sustancialmente mejor de todos los clasificadores en comparación con el conjunto de datos Mentiroso. (Esta observación se ve reforzada por los puntajes AUC en las tablas de métricas de desempeño. Una vez más, el conjunto de mezcla presenta características muy fuertemente en las gráficas de comparación. Tiene el puntaje AUC más alto en el conjunto de datos Mentiroso y el segundo mejor puntaje, solo 0.001 detrás, en el conjunto de datos ISOT.

4.3. Matriz de confusión. Se utiliza una matriz de confusión para el análisis de un modelo de aprendizaje automático. Refleja los datos en relación con los verdaderos positivos, falsos negativos, falsos positivos y verdaderos negativos [42]. Las Figuras 6 y 7 representan la matriz de confusión del conjunto de mezcla sobre las predicciones realizadas en los conjuntos de prueba para los conjuntos de datos Liar e ISOT, respectivamente.

Podemos hacer las siguientes deducciones basadas en la matriz de confusión del conjunto de datos de Mentiroso:

- (i) 1201 artículos de noticias falsos se han predicho correctamente como falsos
- (ii) 2702 artículos de noticias que son verdaderos (reales) se han predicho correctamente como verdaderos
- (iii) 1687 artículos de noticias falsas se han predicho incorrectamente como verdaderos
- (iv) 828 artículos de noticias verdaderas se han predicho incorrectamente como falsos

De manera similar, podemos hacer las siguientes inferencias basadas en la matriz de confusión del conjunto de datos ISOT:

- (i) 11646 artículos de noticias falsos se han predicho correctamente como falsos

Mesa 1: Métricas de rendimiento para el conjunto de datos Liar.

	Conjunto de datos mentiroso usando n-gram					
	AUC de la República de China	F1 puntuación	AUC	Precisión	Recordar	Precisión %
Clasificador de regresión logística (LR)	0,634	0,673	0,666	0,616	0,742	60.346
Clasificador de análisis discriminante lineal (LDA)	0.553	0.573	0.596	0.589	0.558	54.269
Clasificador de descenso de gradiente estocástico (SGDC)	0,616	0,682	0,648	0.592	0,805	58.725
Clasificador de crestas (RC)	0.598	0,685	0,626	0.580	0,836	57.682
Clasificador de máquina de vector de soporte lineal (SVM)	0,609	0,682	0,642	0.589	0,810	58.523
Conjunto de mezcla (BLD)	0,634	0,682	0,668	0,616	0,765	60.813

Mesa 2: Métricas de rendimiento para el conjunto de datos ISOT.

	Conjunto de datos ISOT usando n-gram					
	AUC de la República de China	F1 puntuación	AUC	Precisión	Recordar	Precisión %
Clasificador de regresión logística (LR)	0,998	0,979	0,997	0,981	0,978	98.036
Clasificador de análisis discriminante lineal (LDA)	0,997	0,977	0,997	0,986	0,969	97.875
Clasificador de descenso de gradiente estocástico (SGDC)	0,998	0,982	0,997	0,980	0,983	98.254
Clasificador de crestas (RC)	0,998	0,980	0,997	0,979	0,981	98.071
Clasificador de máquina de vector de soporte lineal (SVM)	0,998	0,983	0,998	0,983	0,983	98.387
Conjunto de mezcla (BLD)	0,998	0,984	0,997	0,984	0,984	98.481

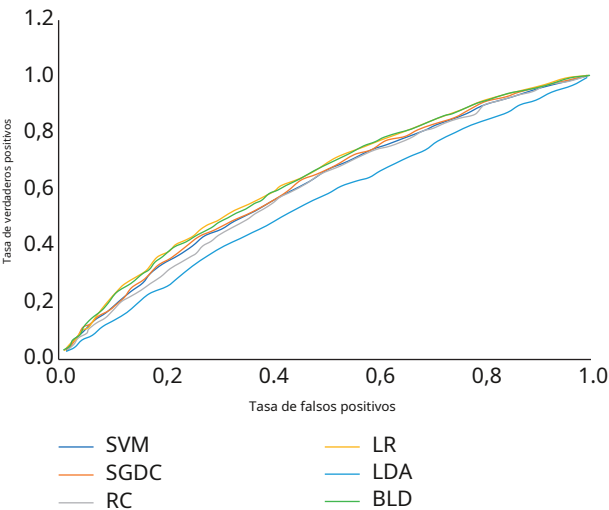


Figura 2: Curva de curva de funcionamiento del receptor (ROC) para el conjunto de datos Mentiroso.

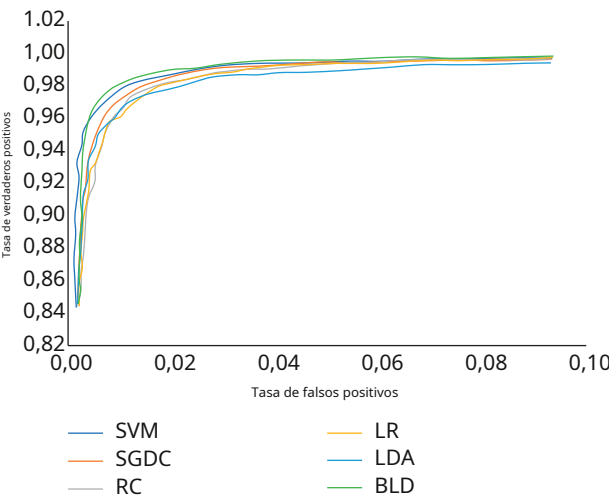


Figura 3: Curva de la curva de funcionamiento del receptor (ROC) para el conjunto de datos ISOT.

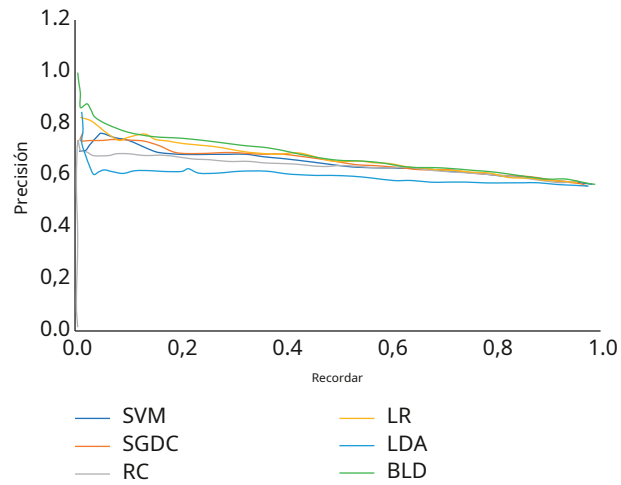


Figura 4: Curva de recuperación de precisión para el conjunto de datos Mentiroso.

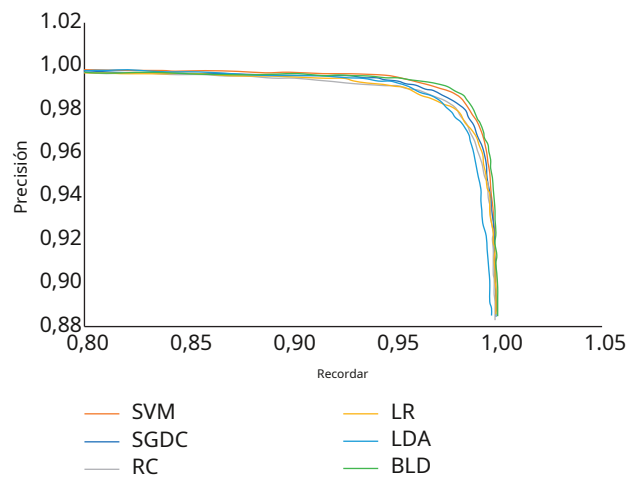


Figura 5: Curva de recuperación de precisión para el conjunto de datos ISOT.

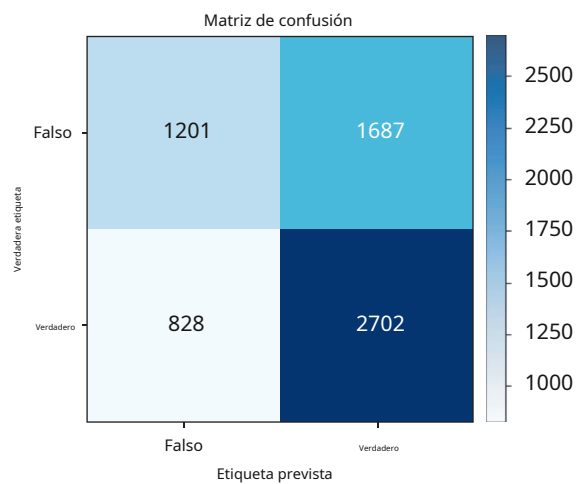


Figura 6: Matriz de confusión del conjunto de datos Liar.

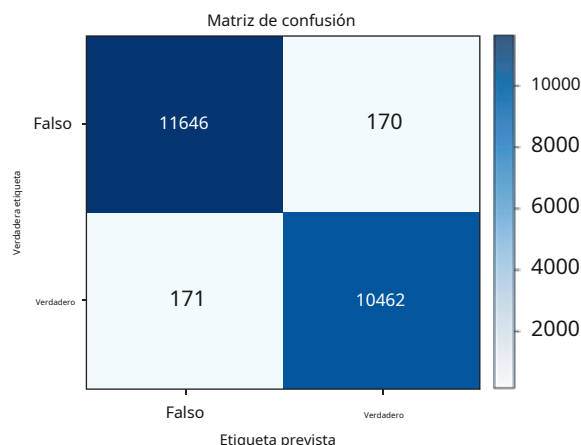


Figura 7: Matriz de confusión del conjunto de datos ISOT.

- (ii) 10462 artículos de noticias que son verdaderos (reales) se han predicho correctamente como verdaderos
- (iii) 170 artículos de noticias falsas se han predicho incorrectamente como verdaderos
- (iv) 17 artículos de noticias verdaderos se han predicho incorrectamente como falsos

5. Conclusiones

Nuestros planes incluyen experimentar con otros conjuntos de datos más grandes y variar el tipo, la combinación y la cantidad de modelos base para el conjunto. También consideraremos examinar las tendencias actuales en las redes sociales relacionadas con las noticias falsas para incorporarlas en nuestro modelo de detección. Sin embargo, las limitaciones asociadas son que los datos a menudo son inconsistentes, lo que se suma a los errores o anomalías del modelo de predicción.

Disponibilidad de datos

Todos los datos están disponibles de los autores.

Conflictos de interés

(Los autores declaran no tener ningún conflicto de intereses.

Referencias

- [1] U. Saeed, H. Fahim y F. Shirazi, *Perfiles de difusores de noticias falsas en Twitter*, Notebook para PAN en CLEF, Clerkenwell, Reino Unido, 2020.
- [2] X. Zhou, R. Zafarani, K. Shu y H. Liu, "Noticias falsas: teorías fundamentales, estrategias de detección y desafíos", en *Actas de la 12a Conferencia Internacional ACM sobre búsqueda web y minería de datos*, Melbourne, Australia, enero de 2019.
- [3] A. Roy, K. Basak, A. Ekbal y P. Bhattacharyya, "Un marco de conjunto profundo para la detección y clasificación de noticias falsas", 2018, <https://arxiv.org/abs/1811.04670>.
- [4] D. de Beer y M. Matthee, "Enfoques para identificar noticias falsas: una revisión sistemática de la literatura", en *Actas de la Conferencia Internacional sobre Ciencia Integrada*, Kep, Camboya, mayo de 2020.
- [5] V. L. Rubin, Y. Chen y N. K. Conroy, "Detección de engaños para noticias: tres tipos de falsificaciones", *Actas de la Asociación para la Ciencia y Tecnología de la Información*, vol. 52, no. 1, págs. 1–4, 2015.
- [6] Y. Shukla, N. Yadav y A. Hari, "Un enfoque único para la detección de noticias falsas mediante el aprendizaje automático" *Proceedings of the International Journal for Research in Applied Science and Engineering Technology*, vol. 7, no. 6, págs. 491–496, 2019.
- [7] H. Allcott y M. Gentzkow, "Redes sociales y noticias falsas en las elecciones de 2016", *e Journal of Economic Perspectives*, vol. 31, no. 2, págs. 211–236, 2017.
- [8] A. Bondielli y F. Marcelloni, "Una encuesta sobre técnicas de detección de rumores y noticias falsas", *Ciencias de la Información*, vol. 497, págs. 38–55, 2019.
- [9] W. Y. Wang, "Liar, liar pants on fire": un nuevo conjunto de datos de referencia para la detección de noticias falsas", 2017, <https://arxiv.org/abs/1705.00648>.
- [10] J. Y. Khan, M. Khondaker, T. Islam, A. Iqbal y S. Afroz, "Un estudio de referencia sobre métodos de aprendizaje automático para la detección de noticias falsas", 2019, <https://arxiv.org/abs/1905.04749>.
- [11] L. Gutierrez-Espinoza, F. Abri, A. S. Namin, K. S. Jones y D. R. Sears, "Detección de revisiones falsas a través del aprendizaje en conjunto", 2020, <https://arxiv.org/abs/2006.07912>.
- [12] G. Pennycook, J. A. Cheyne, N. Barr, D. J. Koehler y J. A. Fugelsang, "Sobre la recepción y detección de tonterías pseudoprofundas", *Juicio y toma de decisiones*, vol. 10, págs. 549–563, 2015.

- [13] VL Rubin, N. Conroy, Y. Chen y S. Cornwell, "¿Noticias falsas o verdad? Usar señales satíricas para detectar noticias potencialmente engañosas", en *Actas del segundo taller sobre enfoques computacionales para la detección de engaños*, San Diego, CA, EE. UU., Junio de 2016.
- [14] E. Tacchini, G. Ballarin, ML Della Vedova, S. Moret y L. de Alfaro, "A algunos les gusta el engaño: detección automatizada de noticias falsas en redes sociales", 2017, <https://arxiv.org/abs/1704.07506>.
- [15] TU Yun y H. Ahn, "Detección de noticias falsas para noticias coreanas mediante técnicas de minería de texto y aprendizaje automático", *Revista de aplicaciones y gestión de tecnologías de la información*, vol. 25, págs. 19–32, 2018.
- [16] S. Aphiwongsophon y P. Chongstitvatana, "Detecting fake news with machine learning method", en *Actas de la 15a Conferencia Internacional sobre Ingeniería Eléctrica / Electrónica, Computación, Telecomunicaciones y Tecnología de la Información*, págs. 18–21, Chiang Rai, (Tailandia, julio de 2018).
- [17] S. Gilda, "Evaluación de algoritmos de aprendizaje automático para la detección de noticias falsas", en *Actas de la 15a Conferencia de Estudiantes sobre Investigación y Desarrollo del IEEE 2017 (SCORED)*, págs. 13–14, Putrajaya, Malasia, diciembre de 2017.
- [18] H. Ahmed, I. Traore y S. Saad, "Detección de noticias falsas en línea mediante el análisis de n-gramas y técnicas de aprendizaje automático", en *Actas de la Conferencia internacional sobre sistemas inteligentes, seguros y fiables en entornos distribuidos y en la nube*, págs. 25–27, Vancouver, Canadá, octubre 2017.
- [19] H. Rashkin, E. Choi, JY Jang, S. Volkova y Y. Choi, "Truth of varying matices: analizando el lenguaje en noticias falsas y verificación de hechos políticos", en *Actas de la Conferencia de 2017 sobre métodos empíricos en el procesamiento del lenguaje natural*, págs. 7 a 11, Copenhague, Dinamarca, septiembre de 2017.
- [20] S. Singhanian, N. Fernandez y S. Rao, "3HAN: una red neuronal profunda para la detección de noticias falsas", en *Actas de la Conferencia Internacional sobre Procesamiento de Información Neural*, Guangzhou, China, noviembre de 2017.
- [21] A. Ota, P. Tilak, S. Ahluwalia y N. Lohia, "Detección de noticias falsas: un enfoque de aprendizaje profundo", *Revisión de ciencia de datos de SMU*, vol. 1, pág. 10, 2018.
- [22] N. Ruchansky, S. Seo y YC Liu, "Un modelo profundo híbrido para la detección de noticias falsas", en *Actas de la ACM de 2017 sobre la Conferencia sobre Gestión de la Información y el Conocimiento*, págs. 6 a 10, Singapur, noviembre de 2017.
- [23] O. Ajao, D. Bhowmik y S. Zargari, "Identificación de noticias falsas en Twitter con modelos híbridos de CNN y RNN", en *Actas de la Novena Conferencia Internacional sobre Redes Sociales y Sociedad*, págs. 18–20, Copenhague, Dinamarca, julio 2018.
- [24] YJ Lu y CT Li, "GCAN: redes de atención conjunta con reconocimiento de gráficos para la detección explicable de noticias falsas en los medios sociales", 2020, <https://arxiv.org/abs/2004.11648>.
- [25] S. Vijayaraghavan, Y. Wang, Z. Guo et al., "Detección de noticias falsas con diferentes modelos", 2020, <https://arxiv.org/abs/2003.04978>.
- [26] PBP Reddy, PK Reddy, GVM Reddy y K. Mehata, "Análisis y detección de datos falsos mediante algoritmo híbrido ensamblado", en *Actas de la 3a Conferencia Internacional de Metodologías de Computación y Comunicación de 2019*, págs. 27–29, Erode, India, marzo de 2019.
- [27] A. Agarwal y A. Dixit, "Detección de noticias falsas: un enfoque de aprendizaje conjunto", en *Actas de la Cuarta Conferencia Internacional de 2020 sobre Sistemas de Control e Informática Inteligente*, págs. 13–15, Madurai, India, mayo de 2020.
- [28] M. Granik, V. Mesyura y A. Yarovy, "Determinar declaraciones falsas hechas por figuras públicas por medio de inteligencia artificial", en *Actas de la XIII Conferencia Científica y Técnica Internacional sobre Ciencias de la Computación y Tecnologías de la Información*, Lviv, Ucrania, septiembre de 2018.
- [29] HS Al-Ash, MF Putri, P. Mursanto y A. Bustamam, "Enfoque de aprendizaje conjunto sobre la clasificación de noticias falsas de Indonesia", en *Actas del 3er Congreso Internacional de Informática y Ciencias Computacionales*, págs. 29–30, Semarang, Indonesia, octubre de 2019.
- [30] I. Ahmad, M. Yousaf, S. Yousaf y MO Ahmad, "Detección de noticias falsas mediante métodos de conjunto de aprendizaje automático", *Complejidad*, vol. 2020, ID de artículo 8885861, 11 páginas, 2020.
- [31] A. Mahabub, "Una técnica sólida de detección de noticias falsas usando clasificador de votación por conjuntos y comparación con otros clasificadores", *SN Ciencias Aplicadas*, vol. 2, págs. 1–9, 2020.
- [32] S. Kaur, P. Kumar y P. Kumaraguru, "Automatización del sistema de detección de noticias falsas mediante el modelo de votación de varios niveles", *Computación blanda*, vol. 24, no. 12, págs. 9049–9069, 2020.
- [33] S. Li, K. Ma, X. Niu et al., "Aprendizaje conjunto basado en apilamiento en características de baja dimensión para la detección de noticias falsas", en *Actas de la 21a Conferencia Internacional IEEE 2019 sobre Computación y Comunicaciones de Alto Rendimiento*, págs. 10–12, Zhangjiajie, China, agosto de 2019.
- [34] V. Agarwal, HP Sultana, S. Malhotra y A. Sarkar, "Análisis de clasificadores para la detección de noticias falsas", *Procedia Ciencias de la Computación*, vol. 165, págs. 377–383, 2019.
- [35] R. Oshikawa, J. Qian y WY Wang, "Una encuesta sobre el procesamiento del lenguaje natural para la detección de noticias falsas", 2018, <https://arxiv.org/abs/1811.00770>.
- [36] H. Ahmed, I. Traore y S. Saad, "Detectar mensajes no deseados de opinión y noticias falsas mediante la clasificación de texto", *Seguridad y privacidad*, vol. 1, no. 1, pág. e9, 2018.
- [37] Digital Vidya. <https://www.digitalvidya.com> 2021.
- [38] N. Smitha y R. Bharath, "Comparación de rendimiento de clasificadores de aprendizaje automático para la detección de noticias falsas", en *Actas de la 2a Conferencia Internacional de 2020 sobre investigación de inventos en aplicaciones informáticas*, págs. 15–17, Coimbatore, India, julio de 2020.
- [39] K. Kowsari, M. Heidarysafa, DE Brown, KJ Meimandi y LER Barnes, "Aprendizaje profundo multimodelo aleatorio para la clasificación", en *Actas de la 2da Conferencia Internacional sobre Sistemas de Información y Minería de Datos*, Lakeland, FL, Estados Unidos, abril de 2018.
- [40] H. Garg, A. Goyal y MA Joshi, "Técnicas de detección de noticias falsas", *Revista de Investigación Avanzada en Ingeniería de Instrumentación y Control*, vol. 7, págs. 8–11, 2020.
- [41] M. Mqadi, N. Naicker y T. Adeliyi, "Un enfoque de punto de datos de sobremuestreo basado en SMOTE para resolver el problema de desequilibrio de datos de tarjetas de crédito en la detección de fraude financiero", *Revista Internacional de Computación y Sistemas Digitales*, vol. 10, no. 1, págs. 277–286, 2021.
- [42] D. Chicco y G. Jurman, "Las ventajas del coeficiente de correlación de Matthews (MCC) sobre F1 puntaje y precisión en la evaluación de clasificación binaria", *BMC Genomics*, vol. 21, no. 1, págs. 6 a 13, 2020.
- [43] P. Mohale y WS Leung, "Detección de noticias falsas mediante aprendizaje automático conjunto", en *Actas de la 18a Conferencia Europea sobre Ciberseguridad y Seguridad*, Coimbra, Portugal, julio de 2019.