Detección de noticias falsas en tweets árabes durante la Pandemia de COVID-19

Ahmed Redha Mahlous[1]
Facultad de Ciencias de la Información y la Computación
Universidad Prince Sultan
Riyadh, Arabia Saudita

Ali Al-Laith[2]
Centro de Ingeniería del Lenguaje -
Universidad de Ingeniería y Tecnología KICS
Lahore, Pakistán

Abstracto—En marzo de 2020, la Organización Mundial de la Salud declaró que el brote de COVID-19 era una pandemia. Poco después, la gente comenzó a compartir millones de publicaciones en las redes sociales sin considerar su confiabilidad y veracidad. Si bien ha habido una extensa investigación sobre COVID-19 en el idioma inglés, hay una falta de investigación sobre el tema en árabe. En este artículo, abordamos el problema de la detección de noticias falsas relacionadas con COVID-19 en tweets árabes. Recopilamos más de siete millones de tweets árabes relacionados con la pandemia del virus corona desde enero de 2020 hasta agosto de 2020 utilizando los hashtags de tendencia durante el tiempo de la pandemia. Confiamos en dos verificadores de hechos: la Agencia de Prensa de Francia y la Autoridad Antirrumores de Arabia Saudita para extraer una lista de palabras clave relacionadas con la desinformación y los temas de noticias falsas. Se extrajo un pequeño corpus de los tweets recopilados y se anotó manualmente en clases falsas o genuinas. Usamos un conjunto de características extraídas de los contenidos de los tweets para entrenar a un conjunto de clasificadores de aprendizaje automático. El corpus anotado manualmente se utilizó como base para construir un sistema para detectar automáticamente noticias falsas a partir de texto árabe. La clasificación del conjunto de datos anotado manualmente logró una puntuación F1 de 87,8% utilizando Regresión logística (LR) como clasificador con la frecuencia de documento inverso de frecuencia de término (TF-IDF) de nivel n-gramo como característica, y un 93,3% F1 - puntúe en el conjunto de datos anotado automáticamente utilizando el mismo clasificador con la función de vector de recuento. El sistema y los conjuntos de datos introducidos podrían ayudar a los gobiernos, los responsables de la toma de decisiones y el público a juzgar la credibilidad de la información publicada en las redes sociales durante la pandemia de COVID-19. Usamos un conjunto de características extraídas de los contenidos de los tweets para entrenar a un conjunto de clasificadores de aprendizaje automático. El corpus anotado manualmente se utilizó como base para construir un sistema para detectar automáticamente noticias falsas a partir de texto árabe. La clasificación del conjunto de datos anotado manualmente logró una puntuación F1 de 87,8% utilizando Regresión logística (LR) como clasificador con la frecuencia de documento inverso de frecuencia de término (TF-IDF) de nivel n-gramo como característica, y un 93,3% F1 - puntúe en el conjunto de datos anotado automáticamente utilizando el mismo clasificador con la función de vector de recuento. El sistema y los conjuntos de datos introducidos podrían ayudar a los gobiernos, los responsables de la toma de decisiones y el público a juzgar la credibilidad de la información publicada en las redes sociales durante la pandemia de COVID-19. Usamos un conjunto de características extraídas de los contenidos de los tweets para entrenar a un conjunto de clasificadores de aprendizaje automático. El corpus anotado manualmente se utilizó como base para construir un sistema para detectar automáticamente noticias falsas a partir de texto árabe. La clasificación del conjunto de datos anotado manualmente logró una puntuación F1 de 87,8% utilizando Regresión logística (LR) como clasificador con la frecuencia de documento inverso de frecuencia de término (TF-IDF) de nivel n-gramo como característica, y un 93,3% F1 - puntúe en el conjunto de datos anotado automáticamente utilizando el mismo clasificador con la función de vector de recuento. El sistema y los conjuntos de datos introducidos podrían ayudar a los gobiernos, los responsables de la toma de decisiones y el público a juzgar la credibilidad de la información publicada en las redes sociales durante la pandemia de COVID-19. El corpus anotado manualmente se utilizó como base para construir un sistema para detectar automáticamente noticias falsas a partir de texto árabe. La clasificación del conjunto de datos anotado manualmente logró una puntuación F1 de 87,8% utilizando Regresión logística (LR) como clasificador con la frecuencia de documento inverso de frecuencia de término (TF-IDF) de nivel n-gramo como característica, y un 93,3% F1 - puntúe en el conjunto de datos anotado automáticamente utilizando el mismo clasificador con la función de vector de recuento. El sistema y los conjuntos de datos introducidos podrían ayudar a los gobiernos, los responsables de la toma de decisiones y el público a juzgar la credibilidad de la información publicada en las redes sociales durante la pandemia de COVID-19. El corpus anotado manualmente se utilizó como base para construir un sistema para detectar automáticamente noticias falsas a partir de texto ára

Palabras clave-Noticias falsas; Gorjeo; medios de comunicación social; Corpus árabe

I. INTRODUCCIÓN

El auge de lo social redes como Facebook, Twitter, y muchos otros han permitido la rápida difusión de información. Cualquier usuario en las redes sociales puede publicar lo que quiera sin considerar la veracidad y confiabilidad de la información publicada, lo que presenta desafíos en el aseguramiento de la confiabilidad de la información. Twitter es una de las plataformas de redes sociales más populares. Está diseñado para permitir a los usuarios enviar información en forma de textos cortos, conocidos como tweets, con no más de 280 caracteres, y cada usuario en Twitter puede seguir tantas cuentas como quiera. Hoy en día, y con el estallido de la pandemia COVID-19, diariamente se generan millones de tuits, lo que ha provocado algunos efectos adversos que impactan a las personas y la sociedad. Por ejemplo, la difusión de información errónea sobre los síntomas del COVID-19 puede dañar a las personas [1]. Por ejemplo, podría provocar ansiedad en una persona que experimenta síntomas similares al COVID-19, incluso si no ha sido infectada con el virus. Los términos noticias falsas e información errónea están estrechamente relacionados y, a menudo, utilizado indistintamente. Los autores en [2] definieron los rumores como: "una hipótesis ofrecida en ausencia de información verificable sobre circunstancias inciertas que son importantes para aquellos individuos que posteriormente están ansiosos por su falta de control como resultado de esta incertidumbre". Otra definición presentada en [3] es: "declaraciones de información no verificadas e instrumentalmente relevantes en circulación que surgen en contextos de ambigüedad, peligro o amenaza potencial, y que funcionan para ayudar a las personas a tener sentido y manejar el riesgo".

La detección de noticias falsas en tweets en inglés es un área de investigación activa y se han publicado muchos estudios y conjuntos de datos durante la pandemia de COVID-19 [4]. En árabe, la detección de noticias falsas es bastante nueva y queda un largo camino por recorrer para alcanzar el nivel alcanzado en otros idiomas, especialmente en inglés. Por lo tanto, la lucha contra las noticias falsas requiere un sistema que ayude automáticamente a verificar la veracidad de la información compartida sobre la pandemia de COVID-19 en las redes sociales. La detección de noticias falsas es una tarea muy desafiante, especialmente con la falta de conjuntos de datos disponibles relacionados con la pandemia. Es necesario un sistema automatizado de detección de noticias falsas mediante la utilización de técnicas de anotación humana, aprendizaje automático / profundo y procesamiento del lenguaje natural [5].
[6].

En este artículo, abordamos el problema de la detección de noticias falsas en Twitter durante el período de la pandemia de COVID-19. Nuestro objetivo es crear un conjunto de datos anotado manualmente para la detección de noticias falsas desde la plataforma de redes sociales de Twitter. Dependemos de fuentes de verificación de hechos para anotar manualmente un conjunto de datos de muestra. La consideración de estas fuentes de verificación de hechos podría ayudar a reducir la difusión de información errónea [7], [8], [9], [10]. Como la anotación manual es cara y requiere mucho tiempo [11], también desarrollamos un sistema para expandir el conjunto de datos anotado manualmente mediante la anotación automática de un conjunto de datos grande y sin etiquetas. Usamos una clasificación de aprendizaje supervisado para entrenar y probar los conjuntos de datos anotados manual y automáticamente para asegurar la calidad de nuestra anotación. Usamos seis algoritmos de aprendizaje automático diferentes, cuatro características diferentes con cada algoritmo, y tres técnicas de preprocesamiento. El resto del artículo está organizado de la siguiente manera. En la Sección 2, cubrimos el trabajo relacionado. La Sección 3 presenta nuestra metodología para anotar y detectar automáticamente noticias falsas relacionadas con COVID-19. En la Sección 4, presentamos los resultados y la discusión. Finalmente, la conclusión y el trabajo futuro se presentan en la Sección 5.

II. Rexaltado Work

Recientemente, se han realizado muchos trabajos para abordar el tema de la detección de fake news, rumores, desinformación o desinformación en las redes sociales. La mayoría de estos estudios se pueden clasificar en enfoques de aprendizaje supervisados y no supervisados. Además, son menos los trabajos que abordan el problema mediante técnicas semisupervisadas.

Para el enfoque supervisado, un sistema basado en técnicas de aprendizaje automático para detectar noticias falsas o rumores en el idioma árabe en las redes sociales durante la pandemia de COVID-19 se presenta en [12]. Los autores recopilaron un millón de tweets árabes utilizando la API de transmisión de Twitter. Los tuits recopilados se analizaron identificando los temas discutidos durante la pandemia, detectando rumores y prediciendo el origen de los tuits. Una muestra de 2000 tweets se etiquetó manualmente como información falsa, información correcta y no relacionada. Se aplicaron diferentes clasificadores de aprendizaje automático, incluidos Support Vector Machine, Logistic Regression y Na¨ıve Bayes. Obtuvieron un 84% de precisión en la identificación de rumores. Las limitaciones de esta investigación incluyen la falta de disponibilidad del conjunto de datos y el hecho de que se basa en una única fuente de rumores:

En [13] se ha propuesto identificar los rumores de noticias de última hora en Twitter. Los autores construyeron un modelo word2vec y un modelo LSTM-RNN para detectar rumores de noticias publicadas en las redes sociales. El modelo propuesto es capaz de detectar rumores basados en el texto de un tweet, y los experimentos demostraron que el modelo propuesto supera a los clasificadores de última generación. Como los rumores se pueden considerar más tarde como verdaderos o falsos, su modelo es incapaz de memorizar los hechos a lo largo del tiempo; solo mira el tweet en el momento actual. La detección de rumores de tweets árabes utilizando características extraídas del usuario y contenido se ha propuesto en [14]. Los autores obtuvieron temas sobre rumores y no rumores de los sitios web anti-rumores y de Ar-Riyadh. Se recopilaron más de 270.000 tweets, que contenían 89 y 88 eventos de rumores y no rumores, respectivamente.
78,6%. La limitación de esta investigación es que el conjunto de datos propuesto no se verifica utilizando ninguno de los conjuntos de datos de referencia.

En [15], se propone un enfoque de aprendizaje supervisado para la detección de credibilidad en Twitter. Se utilizó un conjunto de características que incluyen características basadas en contenido y fuentes, para capacitar a cinco clasificadores de aprendizaje automático. El clasificador Random Forest superó a los demás clasificadores cuando se usó con un conjunto combinado de características. Se anotaron manualmente un total de 3.830 tweets en inglés con clases creíbles o no creíbles. Las características textuales no se estudiaron para examinar su impacto en la detección de credibilidad. En [16] se propuso otro enfoque de aprendizaje automático supervisado para detectar rumores de revisiones comerciales. Se utilizó un conjunto de datos disponible públicamente para realizar experimentos de detección de rumores. Se utilizaron diferentes clasificadores de aprendizaje supervisado para clasificar las revisiones comerciales. Los resultados experimentales mostraron que el clasificador Na¨ıve Bayes logró la mayor precisión y superó a tres clasificadores, a saber, el Clasificador de vectores de soporte, K-Vecinos más cercanos y Regresión logística. La limitación de este trabajo es el pequeño tamaño del conjunto de datos utilizado para capacitar a los clasificadores de aprendizaje automático.

En [17] se propuso la detección de noticias falsas mediante el análisis de n-gramas y técnicas de aprendizaje automático. Dos características diferentes Se investigaron y compararon técnicas de extracción y seis algoritmos de aprendizaje automático en función de un conjunto de datos de artículos políticos que se recopilaron de Reuters.com y kaggle.com para noticias reales y falsas. Otro corpus árabe para la tarea de detectar noticias falsas en YouTube se presenta en [18]. Los autores presentaron un corpus que cubría los temas más preocupados por los rumores. Se recopilaron más de 4.000 comentarios para construir el corpus. Se utilizaron tres clasificadores de aprendizaje automático diferentes (Máquina de vectores de soporte, Árbol de decisión y Bayes ingenioso multinomial) para diferenciar entre comentarios de rumor y no rumor con la función TF-IDF n-gram. El clasi fi cador SVM logró los mejores resultados. Los autores de [19] propusieron identificar noticias falsas en las redes sociales. Utilizaron varios pasos de preprocesamiento en los datos textuales, y luego usó 23 clasificadores supervisados con la función de ponderación TF. El preprocesamiento de texto combinado y los clasificadores supervisados se probaron en tres conjuntos de datos en inglés del mundo real diferentes, incluidos BuzzFeed Political News, Random Political News e ISOT Fake News.

En [4] se ha propuesto un enfoque automático para detectar noticias falsas de tweets en árabe e inglés utilizando clasificadores de aprendizaje automático. Los autores desarrollaron un conjunto de datos extenso y continuo para noticias falsas en árabe e inglés durante la pandemia de COVID-19. La información compartida en sitios web oficiales y cuentas de Twitter se consideró una fuente de información real. Junto con los datos recopilados de sitios web oficiales y cuentas de Twitter, también se basaron en varios sitios web de verificación de datos para construir el conjunto de datos. Se utilizó un conjunto de 13 clasificadores de aprendizaje automático y otras siete técnicas de extracción de características para crear modelos de noticias falsas. Estos modelos se utilizaron para anotar automáticamente el conjunto de datos en información real y falsa. El conjunto de datos se recopiló durante 36 días, desde el 4 de febrero hasta el 10 de marzo de 2020.

En [11] se ha propuesto un gran corpus para luchar contra los medios infodemiconso- ciales de COVID-19. Los autores desarrollaron un esquema que cubre varias categorías que incluyen consejo, cura, llamado a la acción o hacer una pregunta. Consideraron que estas categorías eran útiles para los periodistas, los responsables de la formulación de políticas o incluso para la comunidad en su conjunto. El conjunto de datos recopilado contiene tweets en árabe e inglés. Se utilizaron tres clasificadores para realizar experimentos de clasificación utilizando tres representaciones de entrada: basadas en palabras, FastText y BERT. Los autores solo hicieron públicos 210 de los tuits clasificados.

También se han construido dos corpus árabes, sin anotación manual. En [20], se recopilaron más de 700.000 tweets árabes de Twitter durante el período COVID-19. El corpus cubre temas prevalentes discutidos en ese período y está disponible públicamente para permitir la investigación en diferentes dominios, como la PNL, la recuperación de información y las redes sociales computacionales. Utilizaron la API de Twitter para recopilar los tweets a diario, que abarca el período comprendido entre el 27 de enero de 2020 y el 31 de marzo de 2020.

El segundo corpus se presenta en [21]. Los tuits fueron recopilados durante el período de la pandemia COVID-19 para estudiar la pandemia desde una perspectiva social. El corpus se desarrolló para identificar a los que influyen en la información durante el mes de marzo de 2020 y contiene casi cuatro millones de tweets. Se utilizaron diferentes algoritmos para analizar la influencia de la difusión de información y comparar la clasificación de los usuarios.

Para la detección de noticias falsas en otros idiomas, hay muchos corpus que están disponibles públicamente para abordar la propagación de información falsa. En [22] se ha introducido un conjunto de datos de noticias de verificación de hechos multilingües entre dominios para COVID-19. El conjunto de datos recopilado cubrió 40 idiomas y se basa en artículos verificados de 92 sitios web de verificación de hechos diferentes para anotar manualmente el conjunto de datos. El conjunto de datos está disponible en GitHub. Otro conjunto de datos disponible públicamente llamado "TweetsCOV19" se introdujo en [23]. Este conjunto de datos contiene más de ocho millones de tweets en inglés sobre la pandemia de COVID-19. El conjunto de datos se puede utilizar para entrenar y probar una amplia gama de métodos de aprendizaje automático y de PNL y está disponible en línea. Un nuevo conjunto de datos de Twitter se presenta en [24], que fue desarrollado para caracterizar las comunidades de desinformación COVID-19. Los autores clasificaron los tweets en 17 clases, que incluyen cura falsa, tratamiento falso y hechos o prevención falsos. Realizaron diferentes tareas en el conjunto de datos desarrollado, incluida la identificación de comunidades, análisis de redes, detección de bots, análisis sociolingüístico y postura de vacunación. Las limitaciones de este estudio son que solo una persona realizó la anotación, los análisis son correlacionales y no causales, y los datos recolectados cubrieron un período corto de solo tres semanas. MM-COVID es un repositorio de datos de noticias falsas multilingüe y multidimensional Las limitaciones de este estudio son que solo una persona realizó la anotación, los análisis son correlacionales y no causales, y los datos recolectados cubrieron un período corto de solo tres semanas. MM-COVID es un repositorio de datos de noticias falsas multilingüe y multidimensional Las limitaciones de este estudio son que solo una persona realizó la anotación, los análisis son correlacionales y no causales, y los datos recolectados cubrieron un período corto de solo tres semanas. MM-COVID es un repositorio de datos de noticias falsas multilingüe y multidimensional [25]. El conjunto de datos contiene 3981 contenidos de noticias falsos y 7192 genuinos de inglés, español, portugués, hindi, francés e italiano. Los autores exploraron el conjunto de datos recopilados desde diferentes perspectivas, incluidas las interacciones sociales y los perfiles de usuario en las redes sociales.

El análisis de sentimiento también se ha utilizado en la detección de noticias falsas. También se ha facilitado. En [26], los autores utilizaron el análisis de sentimientos para eliminar los tweets neutrales. Afirmaron que los tweets relacionados con noticias falsas son más negativos y tienen una fuerte polaridad de sentimiento en comparación con las noticias genuinas. El principal problema al utilizar este enfoque para detectar noticias falsas a partir de texto árabe es la falta de recursos sobre sentimientos árabes, incluidos léxicos y corpus de sentimientos [27]. La prueba de si las emociones juegan un papel en la formación de creencias en la desinformación política en línea se presenta en [28]. Los autores exploran las respuestas emocionales como un mecanismo poco explorado de creencia en la desinformación política.

La clasificación de texto mediante machine / deep learning proporciona buenos resultados en muchas aplicaciones de PNL, entre las que se incluyen el análisis de sentimientos [30], [31], la detección de emociones [32], la detección de discursos de odio [33], la detección de sarcasmo [34] y otras aplicaciones.

En resumen, la mayoría de los conjuntos de datos existentes se orientan al idioma inglés, y solo unos pocos se orientan al árabe. Además, la mayoría de los conjuntos de datos árabes relacionados con COVID-19 se publican sin anotaciones. Los conjuntos de datos que se anotan se anotaron automáticamente y se recopilaron durante un corto período de tiempo. Además, no todos estos conjuntos de datos están disponibles públicamente. En esta investigación, abordamos estos problemas empleando tres anotadores para realizar manualmente la tarea de anotación.

III. METROEtodologia

La figura 1 presenta la arquitectura del sistema de detección de noticias falsas propuesto. En el primer paso del marco, recopilamos datos de Twitter utilizando la API de transmisión de Twitter. En el segundo paso, realizamos la extracción de tweets que discuten rumores o temas de noticias falsas durante la pandemia, anotamos una pequeña muestra de tweets manualmente y desarrollamos un sistema para anotar un gran conjunto de datos de tweets sin etiquetar automáticamente.

En el último paso, almacenamos el conjunto de datos en una base de datos y lo usamos para realizar nuestros experimentos y análisis. Esta investigación pretende construir un corpus de noticias falsas en árabe que se pueda utilizar para analizar la propagación de noticias falsas en los medios sociales durante la pandemia de COVID-19. Para abordar esta necesidad, realizamos los siguientes cuatro pasos: 1) recopilación de datos, 2) extracción de palabras clave de rumores / desinformación, 3) procesamiento previo de datos y 4) anotación de noticias falsas.

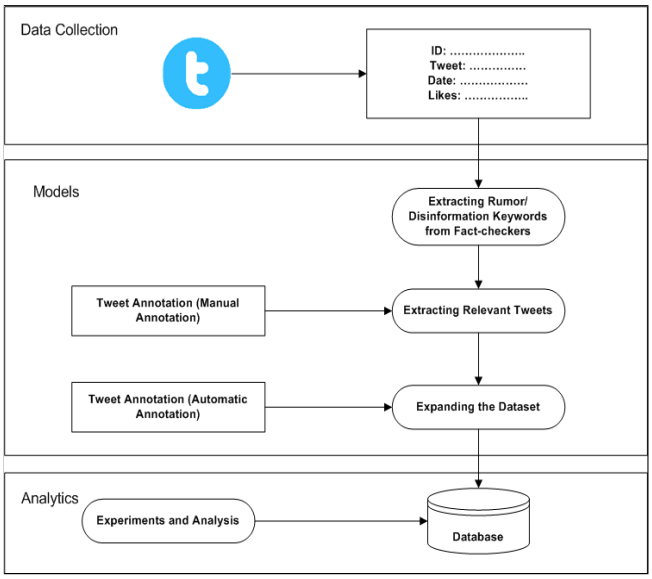

Fig. 1. Nueva arquitectura de detección falsa.

A. Recolección de datos

En esta sección, describimos el proceso de recopilación de datos de Twitter. En la primera instancia, preparamos una lista de hashtags que aparecieron durante el brote de COVID-19, como se muestra en la Tabla I. Armados con la biblioteca Tweepy Python y usando la API de Twitter, procedimos a recopilar tweets árabes relacionados con COVID-19 de enero Desde el 1 de enero de 2020 hasta el 31 de mayo de 2020. Luego, buscamos tweets que contengan uno o más de los hashtags definidos en el texto del tweet. Este paso nos permitió recopilar más de siete millones de tweets únicos. Después de aplicar algunos filtros como eliminar los tweets cortos y repetidos, los tweets restantes son 5,5 millones de tweets. Sin embargo, como algunos de los tweets recopilados eran irrelevantes, decidimos mantener solo aquellos tweets relevantes para la pandemia de COVID-19 y que contienen palabras clave de noticias falsas.

B. Extracción de palabras clave de noticias falsas

Para recopilar una lista de palabras clave relevantes para los rumores que circulan durante la pandemia, utilizamos dos fuentes:

- Agence France-Presse (AFP)[1] con su equipo de investigación sanitaria recién formado que tiene la responsabilidad

---

[1] https://factuel.afp.com/ar/CORNA%20COMPILATION%202-20

TABLA I. LIST DE HASHTAGS UTILIZADOS PARA COLLETE EL DATASET

#	Hashtag	Traducción en inglés
1	AKðPñ »#	Corona
2	AKðPñ »_ ?? ðQ ¯¯#	coronavirus
3	Yj. J ?? ÖÏ @_ AKðPñ »#	Nueva Corona
4	19_ YJ ¯¯ñ »#	COVID-19
5	Reino Unido. AJÉ @_ ?? ðQ	coronavirus
6	®É @ # úÍQ ÖÏ @_ Qj. mi'@#	Cuarentena en casa
7	új ?? É @_ Qj. mi'@#	Cuarentena
8	ú «AÒJk. B @ _ Y «ÁJ. JÉ @ #	Distanciamiento social
9	È ðñ ?? Ó_ AJÊ¿ #	Todos somos responsables
10	ÕºËPAJǿ ¯@ ñ®K. @ #	Quédate en tus casas

de tratar grandes cantidades de noticias falsas en varios idiomas e indicar su error o inexactitud.

- La Autoridad Anti-Rumores (Sin Rumores)[2], un proyecto independiente establecido en 2012 para abordar y contener los rumores y la sedición para evitar que causen algún daño a la sociedad.

Después de leer y analizar los rumores y la información errónea circulada en las redes sociales utilizando las fuentes mencionadas anteriormente, se extrajo una lista de 40 palabras clave y se usó para preparar nuestro conjunto de datos, como se muestra en la Tabla II. Estas palabras clave cubren una variedad de temas asociados con noticias falsas, rumores, racismo, métodos de cura no probados, información falsa. Por ejemplo, hubo un rumor de que el té de hierbas se usa para tratar COVID-19. Otro tema que circuló fue que Cristiano Ronaldo ofreció transformar sus hoteles en hospitales y dar tratamiento gratuito a los pacientes con COVID-19. Uno alegó que el virus corona se dirige solo a aquellos que tienen la piel amarilla y a los asiáticos para reducir la densidad de población. Otros temas incluyen la conversión de no musulmanes al Islam.

Extrajimos un corpus de más de 37,000 tweets únicos relacionados con rumores y temas de desinformación durante la pandemia de COVID-19. Los tweets fueron escritos por 24,117 usuarios con un promedio de 1.5 tweets por usuario. Los detalles de la información estadística sobre el corpus se presentan en la Tabla III.

C.Preprocesamiento de datos

Realizamos varios pasos de preprocesamiento de texto basados en el procedimiento descrito en [35] para desinfectar los tweets recopilados antes de la anotación y clasificación. Nuestro conjunto de datos, que es una mezcla de árabe estándar moderno y árabe dialéctico, requiere un filtrado adicional, como eliminar letras duplicadas, palabras extrañas y palabras que no sean árabes. La siguiente es una lista completa de los pasos realizados:

- Eliminar menciones, hipervínculos y hashtags.
- Eliminando palabras extrañas y no árabes.
- Normalización de texto.

[2]https://twitter.com/Sin ¯rumores

TABLA II. LIST DE VERIFICADO RUMORS Y METROISINFORMACIÓN TÓPTICA

#	Palabra clave	Traducción en inglés
1	?? ÓAm+ t'@ m.l ' @ HA¾j. ??	Redes 5G
2	úæ ?? Jm.l ' @ Qj. ºÉ @	Impotencia
3	h C «¬A ?? »@	Descubrir una cura
4	26 è P @ Qk ék. PX	Una temperatura de 26
5	?? ®JÉ @ ??. K	Aguantando tu respiración
6	èQ «Q^É @	Gárgaras
7	PñÖÏ @	Banana
8	ZAÖÏ @ PAm '.	Vapor de agua
9	áKCKð @ ½JÉ Ag ?? Jj ??	Diagnóstico en línea de su condición
10	úæ. ?? ºÉ @ øA ?? É @	Té de hierbas
11	ÑèÉ @ ñÖAK. àñ®ÉK	Tira su dinero
12	ÑèÉ @ ñÖ @ àñÓQK	Tira su dinero
13	ø YKB @ HÁ®®m. ×	Secadores de manos
14	AKðA ?? É @	Sauna
15	ú¾j. K ?? KQÉ @	El presidente esta llorando
17	é®ðj ¯ ¯X 15 èAJÖÏ @	Agua 15 minutos
18	àñ® ?? J. K	Escupir
19	AKðPñ »úG @ YKQÉ @	Corona Zindani
20	PñºKð B @	Uigures
21	H. A ?? « B @	Hierbas
22	¿¿CORRIENTE CONTINUA?? B @ Entrar al Islam	
23	úæj ?? É @ ?? KQÉ @	Presidente chino
24	Z @ Xñ ?? É @ èQà ?? J. É @	Piel oscura
25	àñ®JJºK	Abarcar
26	kA ?? Ó © KPñK	Distribuyendo el Corán
27	á ?? k Ð @ Y ??	Saddam Hussein
28	ðYÉ AKðP HAJ® ?? ?? Ó	Hospitales Ronaldo
29	é £ AKQ «à @ XB @	Llamado a la oración en Granada
30	àYJÉ à @ XB @	Llamado a la oración en Londres
31	áÒJÉ @ úÍ @ úæJ ?? H. ðQè	Los chinos escapan a Yemen
32	ñÒÉ ?? @	Ingresó al Islam
33	à AKXB @ © JÒg.	Todas las religiones
34	ñ »ñ ?? AK	Tasuco
35	AÓAK. D@	Obama
36	?? J «ÉJK.	Bill Gates
37	AJKAÖÏ @ à @ X @ © ¯ºP	Llamado a la oración en Alemania
38	àñÊ ??	Ellos rezan
39	ZAÓ ?? ÉÈ ¼ðQ Ó	Arriba hasta el cielo
40	á j. ?? » B @ ?? ®K	Falta de oxígeno
	HAÓAÒºÉ @ PY ?? H. QºÖÏ @	Marruecos exporta máscaras

- Eliminación de signos de puntuación y signos diacríticos árabes.
- Eliminando caracteres repetidos que añaden ruido e influyen en el proceso de minería.

Se utilizaron dos bibliotecas para realizar un preprocesamiento adicional en el texto del corpus, incluida la derivación y el enraizamiento. La primera biblioteca es NLTK, que se usó para realizar la derivación en el texto del corpus usando ISRIStemmer4[3]. La segunda biblioteca es

[3]https://www.kite.com/python/docs/nltk.ISRIStemmer

Tashaphyne[4], que se utilizó para obtener la raíz de cada palabra en el corpus.

IV. Corpus A NOTACIÓN

A. Anotación manual

Una muestra de 2500 tweets se anotó manualmente en clases falsas o genuinas. Desarrollamos una pequeña aplicación para facilitar el proceso de anotación, como se muestra en la Fig. 2. Participamos tres anotadores en la anotación del conjunto de datos de muestra. Dos de los anotadores realizaron anotaciones mientras que al tercero se le asignó la tarea de evaluar su salida y resolver los conflictos. Solicitamos a los anotadores que lean y comprendan la lista de pautas y les informamos que omitan los tweets en los que hay una mezcla de noticias falsas y temas genuinos y solo anoten tweets que tengan un tema de noticias falsas claro y distinto.

El fo A continuación se muestran las pautas:

- Los tweets generalmente se consideran falsos si se discute un tema de noticias falsas en el tweet.
- Los tweets se consideran no falsos si se discute un tema de noticias falsas en el tweet y el tema se niega.
- Se omiten los tweets que contienen una mezcla de noticias falsas y genuinas.

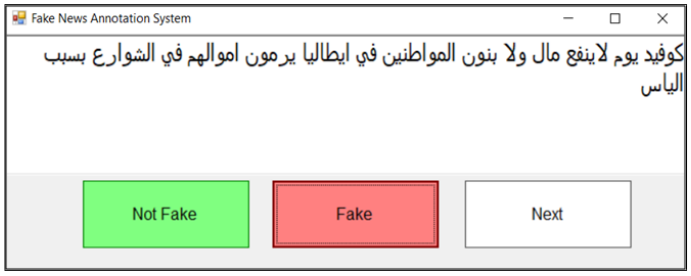

Fig. 2. Interfaz de anotación de noticias falsas.

El proceso de anotación resultó en un corpus que contenía 1.537 tweets (835 falsos y 702 genuinos), después de excluir los tweets duplicados, los tweets que contienen una mezcla de noticias falsas y genuinas y los tweets en los que las noticias falsas pretendían ser sarcasmo. La información estadística sobre el corpus anotado manualmente se muestra en la Tabla III. Utilizamos el coeficiente kappa de Cohen para medir la concordancia entre anotadores, obteniendo un valor de 0,91. La Tabla IV muestra un ejemplo de algunos tweets anotados.

B. Anotación automática

Inicialmente, capacitamos a diferentes clasificadores de aprendizaje automático en el corpus anotado manualmente y usamos el clasificador de mejor desempeño para predecir automáticamente las clases de noticias falsas de los tweets sin etiquetar restantes. El resultado del proceso de predicción es 34,529 tweets (19,582 falsos y 19,582 genuinos) como se muestra en la Tabla III.

Durante el proceso de anotación, los anotadores encontraron algunos tweets que contenían palabras clave de noticias falsas pero con sarcasmo. En este caso, se solicitó a los anotadores que los anotaran como genuinos. La Tabla V muestra una muestra de dichos tweets.

[4]https://pypi.org/project/Tashaphyne/

TABLA III. Corpus Sestadisticas

Corpus anotado manualmente

	Tweets falsos	Tweets no falsos
Tweets totales	835	702
Total de palabras	20,395	19,852
Palabras únicas	6.246	7.115
Caracteres totales	117.630	113,121

Corpus anotado automáticamente

	Tweets falsos	Tweets no falsos
Tweets totales	19.582	14,947
Total de palabras	479,349	463,768
Palabras únicas	79,383	88,037
Caracteres totales	2.855.454	2.680.067

TABLA IV. Fake y GRAMOenuine Tweets miXAMPLES

#	Pío	Clase
1	?? AJË @ úÎm 'úÎË @ AKðPñ »?? ðQ ¯¯úÎ« é «A ?? @ I. J» @ úæ ?? ®K úæ ?? Jm.Ì '@ ª ?? É @ I.. ?? AKðPñ »AïD¯¯Èñ¯?¯@ AïEñJK. ú¯¯é «ðPQÓ . AïDÓ IJ® ?? ñË úæk Quiero escribir un rumor sobre el COVID-19 que haría que la gente se quedara en sus casas y dijera que causa impotencia incluso si te recuperas.	Auténtico
2	áÓ Q »@ Éj. ¯ AKðPñ »á« á ?? k Ð @ Y ?? ÐC¿ ÉJÓ ¼Q. ®Ó AÖß.P . ÐA «áKQå ??« Quizás fabricado como el video de Saddam Hussein sobre COVID-19 hace más de 20 años.	Auténtico
3	. Ö®ºÉ @ ð úæ ?? Jm.Ì '@ ª ?? É @ I.. ?? AKðPñ »?? ðQ ¯¯úæJ ?? QKYm ' cаAdvertencia china: COVID-19 causa impotencia y esterilidad.	Falso
4	KñJ ?? B @ ð Z @ Q® ?? É @ èQå ??J. É @ H. ¿Soy ?? B © J ?? Ó ?? ðQ ®É @ AKðPñ »ø @ H. A ?? ÖÉ éK @ ÉjÉYÉ @ ð ?? PB @ úÎ «éjKA¾ ?? É @ é¯¯AJºÉ @ ?? JÉ®jÉ . Z @ Xñ ?? É @ èQå ??J. É @ øðX áÓ Yg @ COVID-19 está hecho para personas de piel amarilla y asiáticos para reducir la densidad de población, y la evidencia de eso es que ninguna persona de piel negra ha sido infectada.	Falso

V. Experimentos

En esta sección, presentamos los resultados de la noticias falsas clasificación después de describir las técnicas de extracción de características empleadas, la configuración experimental, el entrenamiento del modelo de clasi fi cador y las medidas de evaluación.

A. Extracción de características

El siguiente paso después de realizar el preprocesamiento de texto es preparar las características para construir modelos de clasificación. Para lograr eso, usamos las siguientes características:

- Vector de recuento: el texto de nuestro corpus se convirtió en un vector de recuento de términos.
- TF-IDF a nivel de palabra: Cada término de nuestro corpus está representado en una matriz TF-IDF.

TABLA V. Anotación confusión

#	Pío	Traducción en inglés
1	Õæ ?? m 'á ?? k Ð @ Y ?? éjK. @ ñKYJ®Ë AK. . AKðPñ »á« AëYË @ ð ñKYJ¯¯ÈYg.	En el video, la hija de Saddam Hussein resuelve la controversia del video de su padre sobre COVID-19.
2	hA®Ë éË A ?? Óð ?? J «ÉJK. á «@ Q¯?¯@ àñJÊÓ 75 È @ © Êj. Óð AKðPñ »	Lea sobre Bill Gates, el problema de la vacuna COVID-19 y la cantidad de 75 millones [dólares]

TABLA VI. S ESTADÍSTICAS EN XPERIMENTAL Un corpus DE SABOR

anotado manualmente

	Tweets falsos	Tweets no falsos	Tweets totales
Capacitación	668	562	1.230
Pruebas	167	140	307
Tweets totales	835	720	1,537

Corpus anotado automáticamente

	Tweets falsos	Tweets no falsos	Tweets totales
Capacitación	15,666	11,958	27,624
Pruebas	3.926	2,989	6,905
Total T Weets	19.582	14,947	34,529

- TF-IDF de nivel N-gramo: Utilizamos modelos unigrama, bigrama y trigrama en nuestros experimentos. Luego representamos estos términos en una matriz que contiene las puntuaciones de TF-IDF.

- TF-IDF a nivel de personaje: Representamos puntuaciones de carácter TF-IDF para cada tweet en nuestro corpus.

Estas Las funciones se utilizan para entrenar a múltiples clasificadores con el fin de construir modelos de aprendizaje automático con la capacidad de decidir la categoría más probable para los tweets nuevos y no vistos.

B. Configuración experimental

Esta sección describe las configuraciones experimentales utilizadas para realizar la tarea de clasificación de texto. Diseñamos un conjunto de experimentos con el objetivo de validar y asegurar la calidad de las anotaciones generadas manual y automáticamente. También exploramos la detección de noticias falsas como un problema de clasificación binaria (falso y genuino). El total de tweets en nuestro conjunto de datos de noticias falsas es 1.537 y 34.529 tweets en corpus anotados tanto manual como automáticamente, respectivamente. Dividimos ambos conjuntos de datos en 80% para entrenamiento y 20% para pruebas. La Tabla VI muestra detalles sobre los conjuntos de datos anotados manuales y automáticos.

Se utilizaron seis clasificadores de aprendizaje automático para realizar la clasificación de noticias falsas para ambos conjuntos de datos: Na¨ıve Bayes, Regresión logística (LR), Máquina de vectores de soporte (SVM), Perceptrón multicapa (MLP), Modelo de ensacado de bosque aleatorio (RF) y Gradiente extremo Modelo de impulso (XGB). Los siguientes son los hiperparámetros utilizados con cada clasificador:

- NB: alfa = 0,5
- LR: con valores predeterminados
- SVM: c = 1.0, kernel = lineal, gamma = 3
- MLP: función de activación = ReLU, iteraciones máximas = 30, tasa de aprendizaje = 0,1
- RF: con valores predeterminados
- XGB: con valores predeterminados

C.Entrenamiento de modelos

Una vez que se completó la forma numérica de los tweets textuales, el marco de datos que contiene el vector de conteo, TF-IDF de nivel de palabra, TF-IDF de nivel de n-grama y TF-IDF de nivel de carácter para cada tweet de nuestro corpus se utilizaron para entrenar a seis clasificadores diferentes. Usamos scikit-learn, una biblioteca de Python para la implementación del clasificador y la predicción de las clases del conjunto de datos sin etiquetar. Se utilizó la validación cruzada de K-fold para seleccionar el clasificador que proporciona los resultados más altos y muestra la mejor capacidad para generalizar. La colección se dividió en cinco partes, cuatro de las cuales se utilizaron para entrenamiento en cada iteración y la quinta para evaluación.

D. Medidas de evaluación

La evaluación se llevó a cabo utilizando tres medidas: precisión, recuperación y puntuación F1 de la siguiente manera:

$$Precisión = \frac{TruePositive}{TruePositive + FalsePositive)} \quad (1)$$

$$Recordar = \frac{TruePositive}{TruePositive + FalseNegative)} \quad (2)$$

$$F1 - puntuación = 2 * \frac{Precisión * Recordar}{Precisión + Recuperación)} \quad (3)$$

Dónde:

- Verdadero positivo: la cantidad de tweets falsos que se predice correctamente como tweets falsos.
- Verdadero negativo: la cantidad de tweets genuinos que se predicen correctamente como tweets genuinos.
- Falso positivo: la clase real es genuina, pero la clase predicha es falsa.
- Falso negativo: la clase real es falsa, pero la clase predicha es genuina.

E. Resultados experimentales

Presentamos los resultados experimentales sobre el conjunto de datos de noticias falsas árabes. Se utilizaron seis clasificadores de aprendizaje automático (NB, LR, SVM, MLP, RF y XGB) para realizar nuestros experimentos en los conjuntos de datos anotados manual y automáticamente. Usamos vector de conteo y vectorización TF-IDF (nivel de palabra, nivel de n-gramo y nivel de carácter) para entrenar a los clasificadores. La precisión, el recuerdo y la puntuación F1 son las medidas que se han utilizado para evaluar a cada clasificador mediante una validación cruzada de cinco veces. Los valores en negrita indican qué configuración produjo el mejor rendimiento de clasificación de tweets falsos.

Los resultados mostraron que el uso del clasi fi cador LR con la característica n-grama TF-IDF y sin aplicar más preprocesamiento en el texto (como derivación o enraizamiento) produjo un rendimiento de clasificación significativamente mejor. El clasificador dio un resultado de clasificación de 87,8% de puntuación F1 con el corpus anotado manualmente, como se muestra en la Tabla VII. El mismo clasi fi cador, con la característica de conteo de palabras y sin aplicar derivación o enraizamiento, obtuvo el mejor desempeño de clasificación cuando se aplicó al corpus anotado automáticamente, como se muestra en la Tabla VIII. Logró una puntuación F1 del 93,3%.

Como se muestra en la Fig. 3, el valor de precisión más alto se obtuvo utilizando la función TF-IDF n-gramo con el clasi fi cador LR (87,8%) y la función vectorial de conteo con el clasi fi cador LR (93,4%) anotado de forma manual y automática. corpora, respectivamente. Los resultados obtenidos con el texto sin procesar son mejores que con el texto del corpus después de aplicar la raíz y el enraizamiento. Podemos concluir que realizar un preprocesamiento adicional no

TABLA VII. PAGRECISIÓN (P), RECALL (R), Y F1-PUNTAJE (F1) CLASIFICACIÓN RESULTADOS (METROANUAL ANOTADO DATASET)

Característica		El recuento de palabras			TF-IDF (nivel de palabra)			TF-IDF (nivel de n-gramo)			TF-IDF (nivel de personaje)		
Clasificador	La medida	Texto sin formato	Derivado	Enraizamiento	Texto sin formato	Derivado	Enraizamiento	Texto sin formato	Derivado	Enraizamiento	Texto sin formato	Derivado	Enraizamiento
nótese bien	PAG	77,4	78,1	75,65	82,61	80,9	78,4	80,4	85,8	82.07	83,8	83,3	77,72
	R	77,1	77,5	74,68	73,38	75,6	70,13	74	77,5	73,38	72,7	74,9	69,16
	F1	77,2	77,6	74,98	75,13	76,7	71,9	75,5	78,9	75.06	75,2	76,5	71.04
LR	PAG	84	79,9	79,8	82,1	74	81,4	87,8	76	80,9	81,3	68,7	78
	R	84	79,8	79,9	81,8	74	81,2	87,7	76	79,9	80,5	68,2	77,3
	F1	84	79,9	79,9	81,5	74	81,2	87,8	76	80,1	80,2	68,3	77,5
SVM	PAG	80,8	76,3	79,2	78,5	79,9	82,6	85,7	81,6	86,1	80,6	78,1	76,7
	R	79,9	76	79,2	78,6	79,9	82,5	85,7	81,2	85,7	80,5	77,9	76
	F1	80,1	76,1	79,2	78,4	79,9	82,4	85,7	81	85,7	80,3	77,7	76
MLP	PAG	83,6	78,64	78,37	78,7	76,91	76,21	80,4	78.22	79.01	86,4	77,14	77
	R	83,1	78,57	77,92	78,6	75,65	75,97	78,6	78.25	75,32	86,4	77,27	76,62
	F1	83,2	78,6	78,1	78,6	76.06	76.09	78,9	78.23	76.06	86,4	77.15	76,79
RF	PAG	81,16	80,44	79.21	74,96	74,44	78,31	77,45	75,35	77,43	79,39	73,79	75,92
	R	77,6	79,55	78,57	75	74.03	78.25	77,27	74,68	75,65	79.22	73,38	75
	F1	78,27	79,69	78,73	74,96	74.12	78,27	77,33	74,81	76.04	79,28	73,47	75.23
XGB	PAG	74,99	78,2	72.08	73,26	79,53	73,34	79,77	77,39	75,82	76,59	76,66	74.05
	R	74.03	76,95	70,78	71,75	76,95	73.05	77,6	75	75	75,97	75	73.05
	F1	74,26	77.23	71	72.11	77,44	73.11	78	75,46	75,13	76,13	75,35	73.21

TABLA VIII. PAG RECISIÓN (PAG), RECALL (R), Y F1-PUNTAJE (F1) CLASIFICACIÓN RESULTADOS (AUTOMÁTICO ANOTADO DATASET)

Característica		Recuento de palabras			TF-IDF (nivel de palabra)			TF-IDF (nivel de n-gramo)			TF-IDF (nivel de carácter)		
Clasificador	La medida	Texto sin formato	Derivado	Enraizamiento	Texto sin formato	Derivado	Enraizamiento	Texto sin formato	Derivado	Enraizamiento	Texto sin formato	Derivado	Enraizamiento
nótese bien	PAG	76,7	75,6	72,6	76,2	76	74	74,4	75,9	75,1	72,8	73,6	70,1
	R	76,4	75,5	72,5	76,2	74,8	69,4	74,4	76	75	72,8	73,6	70
	F1	76,6	75,4	72,4	76,2	75,1	70,5	74,4	75,9	74,9	72,8	73,6	70
LR	PAG	93,4	84,6	76,3	91,8	85,8	77,1	90,8	85,8	79,7	84,3	83,1	76,9
	R	93,3	84,6	76	91,7	85,8	77	90,7	85,7	79,6	84,2	83,1	76,9
	F1	93,3	84,6	76,1	91,7	85,8	77,1	90,7	85,7	79,6	84,3	83,1	76,9
SVM	PAG	92,1	82,9	78,3	91,2	84,2	79,8	90,6	85,4	82	89,7	83,8	76
	R	92	82,8	78	91,2	84,2	79,6	90,5	85,3	81,8	89,1	83,1	75,1
	F1	92	82,8	78	91,2	84,2	79,7	90,5	85,3	81,9	89,4	83,3	75,4
MLP	PAG	88,8	79,6	70,1	87,1	77	70,4	71,7	73,2	72,2	80,5	78	72,3
	R	88,5	79,5	70	87,1	77	70,4	71,7	73,2	72,2	80,5	78	72,3
	F1	88,6	79,5	70,1	87,1	77	70,4	71,7	73,2	72,2	80,5	78	72,3
RF	PAG	84,9	82,5	77,7	84,6	82,1	77,7	84,9	81,5	78,3	78,3	79	76,1
	R	84,7	82,3	77,5	84,7	82,1	77,6	84,9	81,6	78,3	78,3	78,9	76,1
	F1	84,7	82,4	77,6	84,6	82,1	77,6	84,9	81,5	78,3	78,3	78,9	76,1
XGB	PAG	82,8	82,1	76,2	82,8	81,7	75,8	82,3	82	76,3	80,1	80,1	76,5
	R	80,2	80,4	74,7	80,9	80,5	75,2	79,9	80,4	75,2	79,3	79,3	75,8
	F1	80,7	80,7	75,1	81,2	80,7	75,3	80,4	80,7	75,5	79,5	79,5	76

mejorar los medios de clasificación.

resultados con el texto de las redes sociales

Como se muestra en la Fig.4, se obtuvo el recuerdo más alto utilizando la función de vector de conteo TF-IDF con el clasi fi cador LR (87,7%) y la función de vector de conteo con el clasi fi cador LR (93,3%) en corpus anotados manual y automáticamente, respectivamente. La puntuación F1 más alta, como se muestra en la figura 5, se obtuvo utilizando la función TF-IDF de nivel n-gramo con el clasi fi cador MLP (87,8%) y la función de vector de conteo con el clasificador LR (93,3%) de forma manual y automática. corpus anotados, respectivamente.

VI. DISCUSIÓN

El objetivo principal de esta investigación fue construir un conjunto de datos de referencia para noticias falsas en árabe relacionadas con la pandemia de COVID-19. Presentamos un nuevo corpus de noticias falsas en árabe, recopilado de Twitter. De los resultados experimentales se desprende claramente que el corpus anotado manualmente se puede utilizar como base para futuras investigaciones en el ámbito de las noticias falsas y la desinformación. Como no queda conjunto de datos de referencia para la detección de noticias falsas en árabe relacionadas con la pandemia COVID-19, este corpus ayudará a la comunidad de investigación una vez que el conjunto de datos esté disponible públicamente. El corpus propuesto fue anotado manualmente por tres anotadores para asegurar la calidad y utilidad del corpus desarrollado. Usamos un conjunto de clasificadores de aprendizaje automático para entrenar diferentes modelos de aprendizaje automático en el corpus anotado manualmente. Se seleccionó el mejor modelo para predecir las clases de noticias falsas de tweets sin etiquetar (más de 35,000 tweets). El análisis estadístico mostró valores más bajos de precisión, recuperación y puntuación F1 en la clasificación del corpus anotado manualmente, mientras que el corpus anotado automáticamente mostró mejores resultados. De los resultados presentados en la sección anterior,

El uso de métodos de aprendizaje automático para clasificar el corpus de noticias falsas utilizando funciones basadas en contenido da mejores resultados que las funciones basadas en el usuario. El corpus se puede ampliar aún más utilizando dos métodos: 1) aumentando el número de

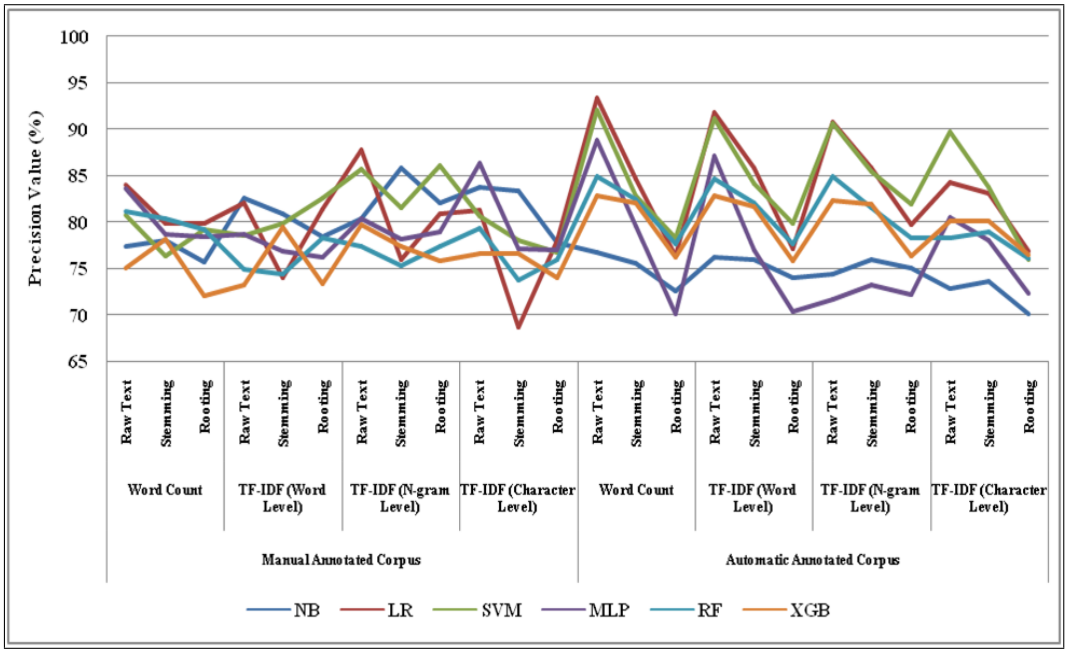

Fig. 3. Resultados de precisión.

rumores verificados o temas de desinformación, o 2) realizar una clasificación en más tweets sin etiquetar relacionados con la pandemia de COVID-19. Después de eso, el enfoque de aprendizaje profundo puede ser solía hacerlo mejorar la clasificación de noticias falsas.

VII. Conclusión y Future Work

En esta periódico, presentamos un nuevo árabe de cuerpo noticias falsas que se hará pública disponible para fines de investigación en este enlace: (https://github.com/yemen2016/FakeNewsDetection), después preparar los ID de los tweets y sus clases asociadas. Detallamos cómo explicó el proceso de recopilación de noticias falsas y dio durante la seleccionamos los rumores y los temas de desinformación se pandemia COVID-19. La tarea de clasificación Regresión, Máquina de realizaron utilizando seis clasificadores (Na¨ıve Bayes, Logistic Random vectores de soporte, Perceptrón multicapa, prueba la posibilidad de Forest Bagging y eXtreme Gradient Boosting) para utilizar cuatro tipos reconocer tweets falsos y genuinos. TF-IDF a nivel de n-grama y TF-IDF de características: vector de conteo, TF-IDF a nivel de palabra, que el El a nivel de carácter. Notamos características y clasificadores utilizados. rendimiento alcanzado varía en función del texto como entrada para Además de considerar el crudo se utilizan dos métodos de los clasificadores de aprendizaje automático; además, las técnicas no preprocesamiento: despalillado y enraizamiento. El texto de ambos lograron mejorar los resultados de clasificación como los dialectos y los corpus se recopiló de Twitter, que incluye varios procedimientos de errores de lenguaje. Por lo tanto, la derivación y la conclusión de que enraizamiento que no arrojaron resultados correctos. El estudio podemos lograr un mayor rendimiento con más datos anotados.

En el futuro, planeamos expandir nuestro corpus con temas adicionales de rumores verificados y desinformación. También esperamos investigar el rendimiento de nuevos métodos de clasificación como el aprendizaje profundo. En esta investigación, solo utilizó funciones basadas en contenido para clasificar y analizar noticias falsas, aunque también se pueden utilizar funciones basadas en el usuario.

Aagradecimiento

“Este trabajo fue apoyado por el proyecto de investigación PSU-COVID19 Emergency Research Program; Universidad Prince Sultan; Arabia Saudita [PSU-COVID19 Emergency Research Program-CCIS-2020-57] ”.

Referencias

[1] J. AlHumaid, S. Ali e I. Farooq, "Los efectos psicológicos de la pandemia del covid-19 y cómo afrontarlos en arabia saudita". *Trauma psicológico: teoría, investigación, práctica y política,* vol. 12, no. 5, pág. 505, 2020.

[2] HB Dunn y CA Allen, "Rumores, leyendas urbanas y engaños de Internet", en *Actas de la Reunión Anual de la Asociación de Educadores Colegiados de Marketing,* 2005, pág. 85.

[3] N. DiFonzo y P. Bordia, "Rumor, gossip and urban legends", *Dio- genes,* vol. 54, no. 1, págs. 19–35, 2007.

[4] MK Elhadad, KF Li y F. Gebali, “Covid-19-fakes: un conjunto de datos de Twitter (árabe / inglés) para detectar información engañosa sobre covid-19 "en *Congreso Internacional sobre Redes Inteligentes y Sistemas Colaborativos.* Springer, 2020, págs. 256–268.

[5] R. Oshikawa, J. Qian y WY Wang, "Una encuesta sobre el procesamiento del lenguaje natural para la detección de noticias falsas", *preimpresión arXiv arXiv: 1811.00770,* 2018.

[6] MK Elhadad, KF Li y F. Gebali, "Detección de noticias falsas en las redes sociales: una encuesta sistemática", en *Conferencia IEEE Paci fi c Rim de 2019 sobre comunicaciones, computadoras y procesamiento de señales (PACRIM).* IEEE, 2019, págs. 1–8.

[7] PL Liu y LV Huang, "Desinformación digital sobre covid-19 y el efecto en tercera persona: examinar las diferencias de canal y los resultados emocionales negativos", *Ciberpsicología, comportamiento y redes sociales,* vol. 23, no. 11, págs. 789–793, 2020.

[8] NM Krause, I. Freiling, B. Beets y D. Brossard, "Verificación de hechos como comunicación de riesgos: el riesgo de información errónea de múltiples capas en tiempos de covid-19", *Revista de investigación de riesgos,* vol. 23, no. 7-8, págs. 1052-1059, 2020.

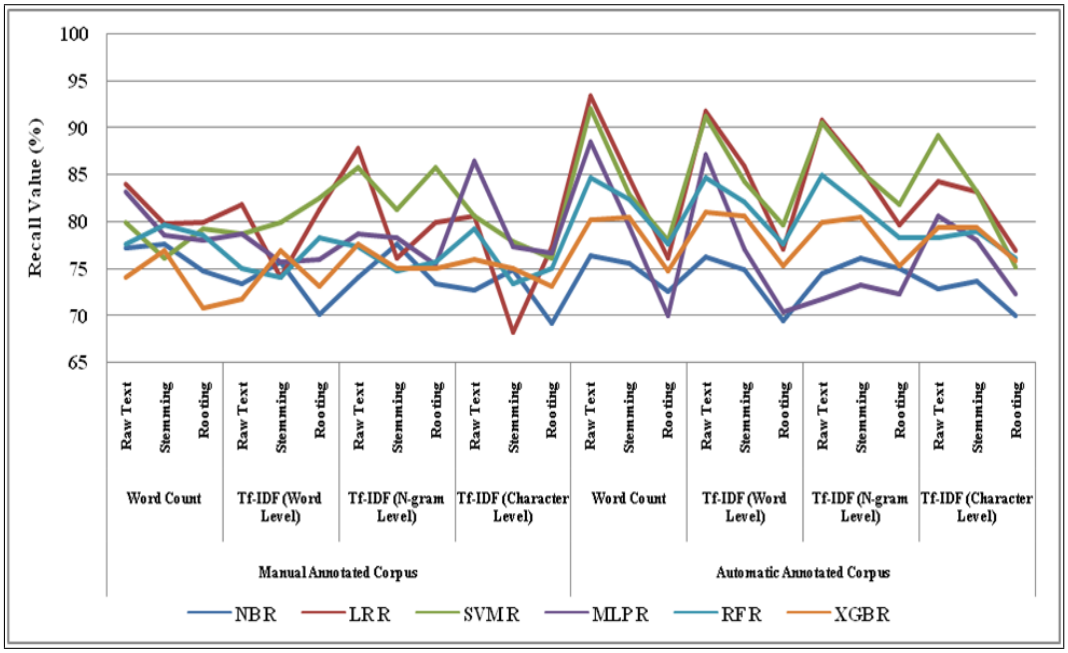

Fig. 4. Recuperar resultados.

[9] J. Donovan, "Recomendaciones concretas para eliminar la información errónea durante la pandemia del covid-19", 2020.

[10] M. Luengo y D. Garc´ıa-Mar´ın, "El desempeño de la verdad: políticos, periodismo de verificación de hechos y la lucha para abordar la desinformación del covid-19", *Revista Estadounidense de Sociología Cultural,* vol. 8, no. 3, págs. 405–427, 2020.

[11] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, GDS Martino, A. Abdelali, H. Sajjad, K. Darwish *et al.*, "Combatir la infodemia covid-19 en las redes sociales: una perspectiva holística y un llamado a las armas" *preimpresión de arXiv arXiv: 2007.07996,* 2020.

[12] L. Alsudias y P. Rayson, "Covid-19 y twitter árabe: ¿Cómo pueden los gobiernos del mundo árabe y las organizaciones de salud pública aprender de las redes sociales?" en*Actas del 1er taller sobre PNL para COVID-19 en ACL 2020,* 2020.

[13] SA Alkhodair, SH Ding, BC Fung y J. Liu, "Detectando rumores de noticias de última hora sobre temas emergentes en las redes sociales", *Procesamiento y gestión de la información,* vol. 57, no. 2, pág. 102018, 2020.

[14] SM Alzanin y AM Azmi, "Detección de rumores en tweets árabes usando expectativa-maximización semi-supervisada y no supervisada", *Sistemas basados en el conocimiento,* vol. 185, pág. 104945, 2019.

[15] NY Hassan, WH Gomaa, GA Khoriba y MH Haggag, "Enfoque de aprendizaje supervisado para la detección de credibilidad en Twitter", en *2018 XIII Congreso Internacional de Ingeniería y Sistemas Informáticos (ICCES).* IEEE, 2018, págs. 196–201.

[dieciséis] A. Habib, S. Akbar, MZ Asghar, AM Khattak, R. Ali y U. Batool, "Detección de rumores en reseñas de empresas mediante aprendizaje automático supervisado", en *2018 5ta Conferencia Internacional sobre Comportamiento, Computación Económica y Sociocultural (BESC).* IEEE, 2018, págs. 233–237.

[17] 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 *Internacional conferencia sobre entornos sistemas seguros y confiables en sistemas distribuidos inteligentes y en la nube.* Springer, 2017, págs. 127-138.

[18] M. Alkhair, K. Meftouh, K. Sma¨ili y N. Othman, "Un corpus árabe de noticias falsas: recopilación, análisis y clasificación", en *Conferencia Internacional sobre Procesamiento de la Lengua Árabe.* Springer, 2019, págs. 292–302.

[19] FA Ozbay y B. Alatas, "Detección de noticias falsas en las redes sociales en línea mediante algoritmos de inteligencia artificial supervisados" *Physica A: Mecánica estadística y sus aplicaciones,* vol. 540, pág. 123174, 2020.

[20] F. Haouari, M. Hasanain, R. Suwaileh y T. Elsayed, "Arcov-19: The first arabic covid-19 twitter dataset with propagation networks", en *Actas del Sexto Taller de Procesamiento del Lenguaje Natural Árabe,* 2021, págs. 82–91.

[21] S. Alqurashi, A. Alashaikh y E. Alanazi, "Identificación de los super difusores de información de covid-19 a partir de tweets árabes", *Preimpresiones,* 2020.

[22] GK Shahi y D. Nandini, "Fakecovid: un conjunto de datos de noticias de verificación de hechos entre dominios multilingües para covid-19", *preimpresión de arXiv arXiv: 2006.11343,* 2020.

[23] D. Dimitrov, E. Baran, P. Fafalios, R. Yu, X. Zhu, M. Zloch y S. Dietze, "Tweetscov19: una base de conocimientos de tweets con anotaciones semánticas sobre la pandemia del covid-19", en *Actas de la 29a Conferencia Internacional ACM sobre Gestión de la Información y el Conocimiento,* 2020, págs. 2991–2998.

[24] SA Memon y KM Carley, "Caracterización de comunidades de información errónea de covid-19 mediante un nuevo conjunto de datos de Twitter", *preimpresión de arXiv arXiv: 2008.00791,* 2020.

[25] Y. Li, B. Jiang, K. Shu y H. Liu, "Mm-covid: un repositorio de datos multilingüe y multidimensional para combatir el falso nuevo covid-19", *preimpresión de arXiv arXiv: 2011.04088,* 2020.

[26] L. Cui y D. Lee, "Coaid: Conjunto de datos de desinformación de la atención médica Covid-19", *preimpresión de arXiv arXiv: 2006.00885,* 2020.

[27] O. Oueslati, E. Cambria, MB HajHmida y H. Ounelli, "A review of sentiment analysis research in arabic language", *Sistemas informáticos de futura generación,* vol. 112, págs. 408–430, 2020.

[28] L. Rosenzweig, B. Bago, AJ Berinsky y D. Rand, "Desinformación y emociones en nigeria: el caso de las noticias falsas covid-19", 2020.

[29] E. Cambria, D. Das, S. Bandyopadhyay y A. Feraco, "Computación afectiva y análisis de sentimientos", en *Una guía práctica para el análisis de sentimientos.* Springer, 2017, págs. 1-10.

[30] A. Al-Laith y M. Shahbaz, "Seguimiento del sentimiento hacia las entidades noticiosas de las noticias árabes en las redes sociales", *Sistemas informáticos de futura generación,* vol. 118, págs. 467–484, 2021.

[31] A. Al-Laith, M. Shahbaz, HF Alaskar y A. Rehmat, "Arasencorpus: Un enfoque semi-supervisado para la anotación de sentimientos de un gran corpus de texto árabe", *Ciencias Aplicadas,* vol. 11, no. 5, pág. 2434, 2021.

[32] A. Al-Laith y M. Alenezi, "Monitoreo de las emociones y síntomas de las personas a partir de tweets árabes durante la pandemia del covid-19", *Información,* vol. 12, no. 2, pág. 86, 2021.

[33] MZ Ali, Ehsan-Ul-Haq, S. Rauf, K. Javed y S. Hussain, "Improving

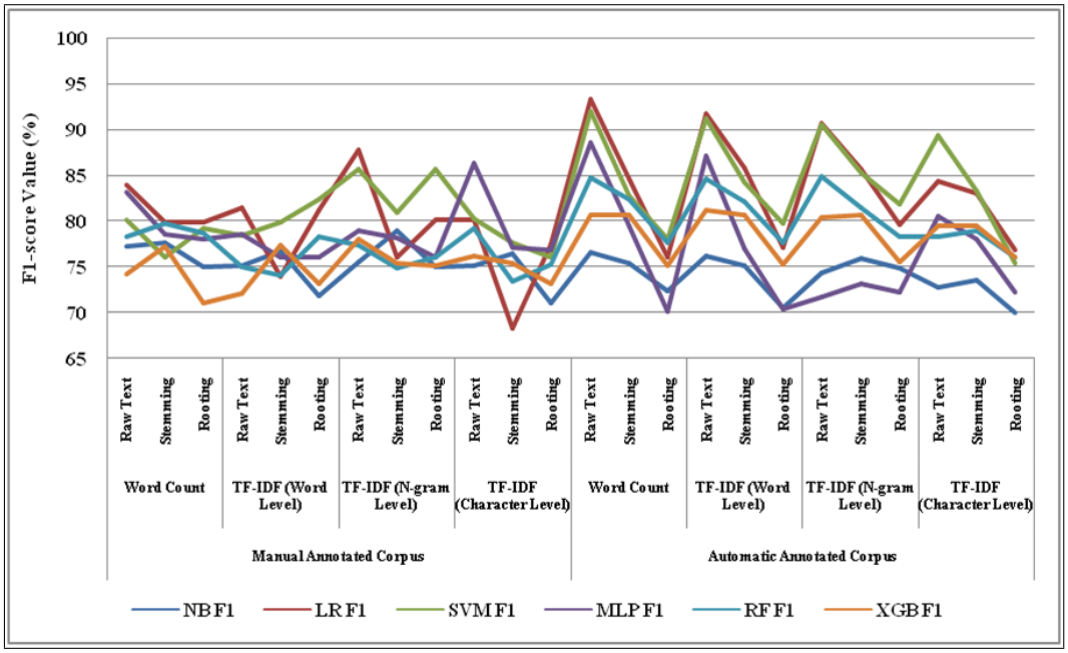

Fig. 5. Resultados de la puntuación F1.

detección de incitación al odio en tuits en urdu mediante el análisis de opiniones " *Acceso IEEE,* vol. 9, págs. 84 296–84 305, 2021.

[34] A. Allaith, M. Shahbaz y M. Alkoli, "Enfoque de red neuronal para la detección de ironía del texto árabe en las redes sociales". en*FUEGO (Trabajando Notas),* 2019, págs. 445–450.

[35] SR El-Beltagy, ME Kalamawy y AB Soliman, "Niletmrg at semeval-2017 task 4: Arabic sentiment analysis", *preimpresión arXiv arXiv: 1710.08458,* 2017.