

Generación de conjuntos de datos de Tweets de verificadoras acreditadas por la IFCN en Ecuador**Generating Tweet Datasets from IFCN-accredited Fact-Checkers in Ecuador**

Katty Nancy Lino Castillo, Mariuxi Del Carmen Toapanta Bernabé, David Fernando Ramos Tomalá, Ivan Leonel Acosta Guzmán & Cercado Ruíz Jenny Melissa

DIMENSIÓN CIENTÍFICA**Enero - junio, V°7 - N°1; 2026****Recibido:** 26-01-2026**Aceptado:** 27-01-2026**Publicado:** 28-01-2026**PAIS**

- Ecuador, Guayaquil
- Ecuador, Guayaquil
- Ecuador, Guayaquil
- Ecuador, Guayaquil
- Ecuador, Guayaquil

INSTITUCION

- Universidad de Guayaquil
- Universidad de Guayaquil
- Universidad de Guayaquil
- Universidad de Guayaquil
- Universidad de Guayaquil

CORREO:

- ✉ katty.linoc@ug.edu.ec
- ✉ mariuxi.toapantab@ug.edu.ec
- ✉ david.ramost@ug.edu.ec
- ✉ ivan.acostag@ug.edu.ec
- ✉ jenny.cercador@ug.edu.ec

ORCID:

- 🌐 <https://orcid.org/0000-0002-0345-3246>
- 🌐 <https://orcid.org/0000-0002-4839-7452>
- 🌐 <https://orcid.org/0009-0007-2702-8926>
- 🌐 <https://orcid.org/0000-0002-1589-1825>
- 🌐 <https://orcid.org/0009-0006-3815-6045>

FORMATO DE CITA APA.

Lino, K., Toapanta, M., Ramos, D., Acosta, I. & Jenny, C. (2026). Generación de conjuntos de datos de Tweets de verificadoras acreditadas por la IFCN en Ecuador. *Revista G-ner@ndo*, V°7 (N°1). Pág. 880 – 898.

Resumen

El presente artículo tiene como propósito crear conjuntos de datos mediante los procesos de recolección y extracción de tweets publicados por verificadores aprobados por la International Fact-Checking Network (IFCN) en Ecuador, con la finalidad de que estos sirvan como base en proyectos futuros relacionados al ámbito de la verificación de hechos. Para llevar a cabo esta investigación, se empleó Tweepy que es una biblioteca de Python, permite interactuar con la API de X de manera eficiente. Los resultados obtenidos muestran que Tweepy es una herramienta poderosa para la recopilación de grandes volúmenes de datos de X, y que los conjuntos de datos construidos a partir de estos tweets pueden ser utilizados para diversos propósitos, como el estudio de tendencias de la desinformación, identificación de las fuentes más comunes de información falsa, y valoración de la efectividad de las estrategias de verificación de hechos.

Palabras clave: Verificación de hechos, X, IFCN, Tweepy, Conjunto de datos.

Abstract

The purpose of this article is to create datasets through the processes of collecting and extracting tweets published by fact-checkers approved by the International Fact-Checking Network (IFCN) in Ecuador, in order to serve as a basis for future studies related to the field of fact-checking. To carry out this research, Tweepy was used, which is a Python library, which allows interaction with the X API efficiently. The results obtained show that Tweepy is a powerful tool for collecting large volumes of X data, and that the datasets constructed from these tweets can be used for various purposes, such as studying disinformation trends, identifying the most common sources of false information, and assessing the effectiveness of fact-checking strategies.

Keywords: Fact-checking, X, IFCN, Tweepy, Dataset.

Introducción

Hoy en día, las redes sociales son consideradas los principales medios de información a nivel mundial, y X, antes Twitter, sobresale entre las plataformas más conocidas para noticias. El acceder a las noticias en tiempo real, se ha convertido en una actividad cotidiana. Sin embargo, esto también ha generado que trasciendan noticias falsas.

Este aspecto ha provocado que se desarrollen diversos estudios relacionados a desinformación en las redes sociales en diferentes ámbitos, siendo el tema político más relevante, debido a la influencia que ejerce sobre las personas, más aún en tiempos de campaña electoral.

En este sentido, (Kauk et al., 2025) en su trabajo de investigación analizaron un conjunto de datos a gran escala de Twitter (ahora "X") con aproximadamente 2 millones de tuits en 123 historias verificadas, encontrando que tanto la ambigüedad (falta de claridad) como la falsedad originan que la información reaparezca o se repitan en las redes sociales, lo que resalta la importancia de analizar la naturaleza de la (des)información en línea y el beneficio de crear dataset para investigaciones prácticas.

Investigaciones como las de (Számely et al., 2025) demuestran que las métricas de redes sociales en línea de fácil acceso pueden utilizarse para identificar posibles difusores de contenido poco confiable en X, resaltando el beneficio de poseer datos estructurados con la finalidad de efectuar estudios comparativos entre diversas agrupaciones de usuarios y patrones de publicación.

Además, (Sharifpoor et al., 2025) realizaron un estudio basado en modelos de aprendizaje automático para identificar el enfoque más eficaz en detectar y clasificar la información fiable y la desinformación sobre el contenido sanitario compartido en X en

relación con el COVID-19, demostrando que las Redes neuronales conversacional basadas en texto, fueron el método más eficaz.

Ante esta problemática, actualmente existen muchas investigaciones como las de (Toapanta M. et al., 2024) que muestran análisis comparativos de modelos de aprendizaje profundo para catalogar las noticias falsas.

Es así que, surge la red International Fact Checking Network (IFCN), encargada de impulsar la fiabilidad de la información en la pugna mundial en oposición la desinformación, además de patrocinar a los fact-checkers certificados mediante la formación de nuevas redes. Actualmente, cuenta con una red sólida que tiene mayor presencia mundial, encargados de la creación de políticas sobre fact-checking a nivel internacional (Moreno-Gil & Salgado-De Dios, 2023).

En Ecuador existen dos organizaciones acreditadas: Ecuador Chequea y Ecuador Verifica, quienes confirman la autenticidad de contenidos con potencial de desinformación que se divulga en las redes sociales, por ser aquellos a los que el público está más expuesto y que pueden generar un impacto negativo en la toma de decisiones (Ecuador Verifica, 2020). Estas verificadoras realizan un seguimiento constante a las redes sociales, especialmente X debido a la propagación acelerada de información (Ecuador Chequea, 2025). Los tweets son estudiados para revelar patrones y tendencias en la circulación de información, además de determinar la legitimidad de las noticias publicadas.

Considerando lo anteriormente expuesto, se destaca la importancia de crear datasets que contengan Tweets de fact-checkers certificados por la IFCN en Ecuador. Si bien es cierto, este recurso no será de libre acceso, su creación satisface la necesidad de disponer datos estructurados y confiables que ayuden en proyectos futuros. Además, en Ecuador existe escasa producción científica enfocada en la generación de conjuntos de

datos estructurados de cuentas de verificadoras acreditadas, evidenciando una carencia de rigor en el ámbito nacional.

Es así que, esta investigación tiene como objetivo generar datasets con los tweets difundidos por los factcheckers certificados por la IFCN, para permitir su aprovechamiento en proyectos futuros. Estos Tweets luego de ser recopilados y extraídos de la plataforma X en tiempo real, son selectos conforme a diferentes parámetros de búsqueda determinados por los fact-checkers certificados en Ecuador, mediante herramientas que facilitan la extracción de datos. Estos datasets ayudarán a identificar comportamientos en la difusión de noticias engañosas y estudio de la validez de los fact-checkers certificados en la detección de información incorrecta. Para desarrollar esta investigación se utilizan metodologías de recopilación y limpieza de datos con el objetivo de confirmar la calidad y relevancia de los datos obtenidos (Subirats et al., 2019).

En conclusión, esta investigación constituye un avance significativo en el empleo de técnicas de analítica de datos, para la batalla contra la desinformación en redes sociales. Además, contribuir al fortalecimiento del entorno vinculado a la verificación de hechos en Ecuador y de promover una sociedad más informada y analítica en relación con las noticias que se divulgan en las redes. La implementación de estas tecnologías y metodologías proporcionará una perspectiva más profunda sobre los patrones de verificación en Ecuador, presentando herramientas valiosas para investigadores e interesados en general. La integración de la tecnología y comunicación es fundamental para enfrentar los retos que expone la propagación de información.

Métodos y Materiales

El desarrollo de esta investigación posee un enfoque cualitativo y cuantitativo. Según Edsson & Ticona (2025) argumentan que “La investigación cualitativa ayuda a

entender los fenómenos, indagándolos bajo la perspectiva de los componentes en un entorno natural relacionado con el hecho, mientras que la investigación cuantitativa se orienta en determinar hechos por medio de indicadores, utilizando técnicas estadísticas que ayuden a establecer patrones y vínculos numéricos“. En este sentido, se utilizará el enfoque cualitativo para analizar los contenidos de los tweets y cuantitativo en la recopilación de los datos numéricos relacionados con la cantidad de tweets, likes, frecuencias de publicación y otros.

Asimismo, la naturaleza de esta investigación es de tipo descriptiva, debido a que son datos provenientes de un objeto en estudio, que luego proporcionará información detallada para generar un conjunto de datos donde se pueda identificar patrones y tendencias de las publicaciones realizadas por las Verificadoras. Como lo señala (Alban et al., 2020), “La finalidad de la investigación descriptiva es comprender los aspectos preponderantes por medio de la explicación puntual del problema”.

Recopilación de datos

En este proceso, inicialmente se procedió con la identificación y seguimiento de las cuentas de los Fact-Checkers Ecuador Chequea y Ecuador Verifica, para luego obtener sus tweets publicados dentro del espacio geográfico ecuatoriano y correspondientes al período de marzo a agosto del 2023.

A la par, se solicitó una cuenta de desarrollador en la plataforma X Developer, una vez aprobada se obtuvieron las claves y tokens para acceder a la aplicación y obtener los datos de estudio de esta investigación.

Para realizar la obtención de los datos se utilizó Tweepy, librería de Python que permite acceder a la API de X (Documentación de Tweepy, 2023). Esto fue realizado el

entorno de Colab donde se instaló Tweepy, luego se importaron algunas librerías encargadas del tratamiento de los datos.

Figura 1. *Instalación de la librería Tweepy.*

```
pip install git+https://github.com/tweepy/tweepy.git
```

Nota. Instalación de librería Tweepy en el entorno de Google Colab.

Figura 2. *Librerías Tweepy.*

```
import json
import tweepy
import pandas as pd
import requests
import re
import seaborn as sns
import matplotlib.pyplot as plt
```

Nota. Importación de librería Tweepy y otras librerías necesarias para la extracción de datos.

A continuación, se muestra la configuración de las credenciales, conexiones con Tweepy, definición de criterios de búsqueda: @EcuadorChequea y @EcuadorVerifica y cantidad máxima de tweets a descargar (max_tweets). Para efectos de esta investigación se consideró una población de 100 registros.

Figura 3. *Credenciales de autenticación.*

```
consumer_key = "zfclj85wXaOvmZdhqhJeXtK0D"
consumer_secret =
"HBlaZiNSDZlPMkHotp5VCyqWNNYo6ZIgaUC68vxUnLdTLyweyK"
bearer_token =
"AAAAAAAAAAAAAAAAAAAAAL1siQEAAAAADac39K3BlpGiIsykyrlW1D%2FNdXQ%3Dqy2u
5n4xNAohkFeboQS8nMtqfpbXkArgLz9JFkQSgnGcATwy5F"
access_token = "1123776505618866176-duCQIHYhwSp9jPj fA67cyzhqBqwXJs"
access_token_secret = "Nzp43d83voQVJCWNvP5qL0HVuoItzaKGp4zSv89oHSeKK"
```

Nota. Autenticación con credenciales de desarrollador de X.

Figura 4. *Parámetros de búsqueda.*

```
response = client.search_recent_tweets(query=query,
    tweet_fields=
    "id,created_at,author_id,text,public_metrics,entities",
    expansions="author_id,attachments.media_keys",
    user_fields="username,name,profile_image_url",
    media_fields="public_metrics,url,height,width,alt_text",
    max_results=100)
```

Nota. Definición de Parámetros de búsqueda y cantidad de registros a extraer.

Los registros obtenidos fueron guardados en diferentes objetos con la finalidad de extraer los atributos necesarios. Finalmente, se crearon dos DataFrames: “Tweet_dict” y “Tweets_list”. A continuación, se muestra los campos que contienen cada uno de ellos.

Figura 5. *DataFrame data_tweets.*

```
tweets_list = []
media_list = []
tweet_links = []
for tweet in tweets_data['data']:
    author_id = tweet['author_id']
    # Buscar el objeto del autor en el campo 'includes'
    author = next((user for user in tweets_data['includes']['users']
    if user['id'] == author_id), None)
    tweet_dict = {
        "ID": tweet['id'],
        "Fecha de Creacion": tweet['created_at'],
        "ID del Autor": author_id,
        "Texto": tweet['text'],
        "Retweets": tweet['public_metrics']['retweet_count'],
        "Respuestas": tweet['public_metrics']['reply_count'],
        "Likes": tweet['public_metrics']['like_count'],
        "Citas": tweet['public_metrics']['quote_count'],
        "Nombre de Usuario del Autor": author['username'] if author
    else None,
        "Nombre del Autor": author['name'] if author else None,
        "URL de la Imagen de Perfil del Autor":
    author['profile_image_url'] if author else None,
        "Fuente": tweet['entities']['source'] if 'entities' in tweet
    and 'source' in tweet['entities'] else None
    }
```

Nota. Creación del DataFrame tweet_dict.

Figura 6. *DataFrame Tweets_List.*

```
tweets_list.append(tweet_dict)
if 'attachments' in tweet and 'media_keys' in
tweet['attachments']:
    for media_key in tweet['attachments']['media_keys']:
        media_info = next((media for media in
    tweets_data['includes']['media'] if media['media_key'] == media_key),
    None)
        if media_info:
            media_dict = {
                "ID del Tweet": tweet['id'],
                "Tipo de Media": media_info['type'],
                "URL de la Media": media_info.get('url'),
                "Alt Text": media_info.get('alt_text'),
                "Metricas de la Media": media_info.get
            ('public_metrics')
            }
            media_list.append(media_dict)
```

Nota. Creación del DataFrame tweets_list.

Procesamiento y Limpieza de Datos

En esta fase se realizó la depuración de datos, que consistió en la identificación y eliminación de valores duplicados, valores nulos y consistencia en los tipos de datos, ya que estos pueden sesgar el análisis (Subirats et al., 2019).

Figura 7. *Valores Duplicados*

```
media_data = [dict(t) for t in {tuple(d.items()) for d in
media_data}]
tweets_data = [dict(t) for t in {tuple(d.items()) for d in
tweets_data}]
```

Nota. Eliminación de valores duplicados en el DataFrame.

Figura 8. *Valores Nulos*

```
media_data = [d for d in media_data if d is not None]
tweets_data = [d for d in tweets_data if d is not None]
```

Nota. Eliminación de valores nulos en el DataFrame.

Figura 9. *Tipos de Datos*

```
for tweet in tweets_data:
    tweet['retweet_count'] = int(tweet['retweet_count'])
```

Nota. Verificación de Tipos de datos correctos.

Figura 10. *Limpieza de textos*

```
for tweet in tweets_data:
    tweet['text'] = tweet['text'].replace('[^a-zA-Z0-9\s]', '',
regex=True)
```

Nota. Limpieza de textos.

Una vez realizada la limpieza de los datos, estos fueron almacenados en un nuevo archivo JSON, para su uso posterior.

Figura 11. Almacenamiento de datos limpios

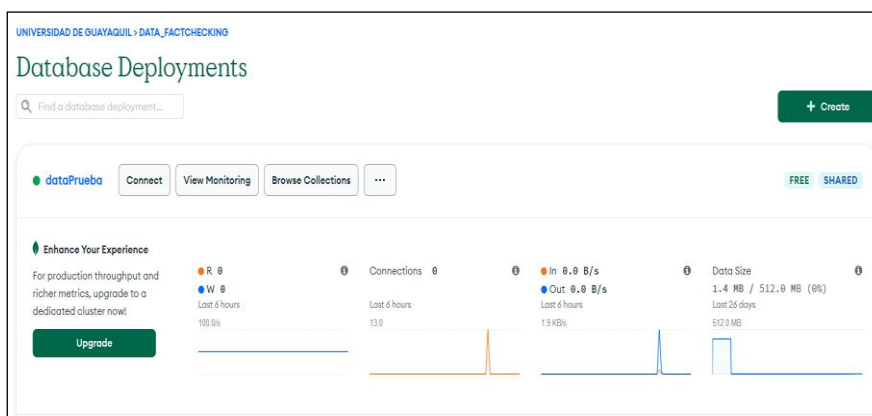
```
with open('cleaned_media_list.json', 'w') as media_cleaned_file:
    json.dump(media_data, media_cleaned_file, indent=4)

with open('cleaned_tweets_list.json', 'w') as tweets_cleaned_file:
    json.dump(tweets_data, tweets_cleaned_file, indent=4)
```

Nota. Almacenamiento de textos en archivos JSON.

Carga de datos

Posterior a lo realizado, se creó la Base de Datos en el Gestor MongoDB, para cargar los datos obtenidos en la fase anterior. MongoDB fue considerada para esta investigación por los beneficios que presenta, entre estos: escalabilidad, flexibilidad y alta disponibilidad relacionados al gran volumen de acceso (Sarasa Cabezuelo, 2016).

Figura 12. Gestor de Base de Datos MongoDB.

Nota. Creación de Base de Datos en MongoDB.

En el entorno Colab, se realizó la conexión a la Base de Datos, además de la invocación de la función `insertar_documentos_json`, encargada de subir el contenido del archivo JSON en la base de datos.

Figura 13. *Conexión con MongoDB*

```
def conectar mongodb cloud(usuario, contraseña, cluster, bd):
```

Nota. Función para conectar la base de datos en la nube.

Figura 14. *Insertar Documentos MongoDB*

```
def insertar documentos json(dataPrueba, tweets list):
```

Nota. Función para insertar documentos en MongoDB.

Figura 15. *Gestor de Base de Datos MongoDB*

```
with open(tweets_list, 'r') as file:  
    documentos = json.load(file)
```

Nota. Lectura de Archivo JSON y carga de contenido en documentos de MongoDB.

Figura 16. *Resultado de Ingreso de Datos*

```
resultado = dataPrueba.insert_many(documentos)  
return resultado.inserted ids
```

Nota. Devolución de Id de los documentos que se ingresaron exitosamente.

Finalmente, se utilizan otras funciones para consultar los registros de los documentos cargados en la base de datos.

Evaluación de datos

Para confirmar la calidad de los datos que van a ser utilizados en futuros análisis, informe o decisiones, se procedió a la verificación por medio de una matriz de Confusión, la cual muestra la relación entre aciertos y errores del sistema, donde los aciertos son las coincidencia y error las no coincidencias (Menoyo Ros et al., 2021).

El aprendizaje automático se ha transformado en un instrumento para la analítica de datos complejos. Los métodos de Machine Learning son muy preciados para estandarizar fuentes de datos disímiles o que carecen de valores, además de revelar

incoherencias, elementos que afectan de manera concisa el performance de los modelos predictivos (Wei et al., 2025).

Figura 17. Creación de la matriz de confusión.

```
confusion = confusion_matrix(y_true, y_pred)
```

Nota. Creación de la Matriz de Confusión.

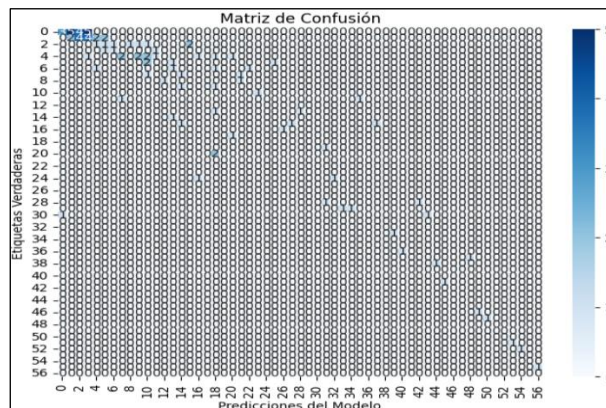
Esta validación de datos fue realizada con el propósito de identificar errores, incoherencias o valores anómalos que generen información incorrecta. Para la visualización de los resultados, se creó el siguiente código:

Figura 18. Visualización de la Matriz de Confusión.

```
plt.figure(figsize=(8, 6))
sns.heatmap(confusion, annot=True, cmap='Blues', fmt='g')
plt.xlabel('Predicciones del Modelo')
plt.ylabel('Etiquetas Verdaderas')
plt.title('Matriz de Confusión')
plt.show()
```

Nota. Configuración de etiquetas de la Matriz de Confusión.

Figura 19. Matriz de Confusión.



Nota. Matriz de Confusión con los resultados del análisis de los objetos.

Análisis de resultados

En esta fase, se presenta el producto final de esta investigación, los dos conjuntos de datos: `media_list.json` y `tweets_list.json`, que serán utilizados en futuras aplicaciones. Estos conjuntos de datos han sido desarrollados en base a la metodología propuesta en la tesis de (Cercado J., 2023).

Tabla 1. *Conjunto de datos tweets_list*

Campo	Descripción
Id	Id único del Tweet.
created_at	Fecha y hora de registro del Tweet.
author_id	Id único del autor del Tweet.
text	Contenido del Tweet.
retweet_count	Frecuencia de reenvíos.
reply_count	Cantidad de réplicas del Tweet.
like_count	Cantidad de "Me gusta".
quote_count	Número de veces que ha sido referenciado.
username	Usuario del creador del Tweet.
name	Nombre del creador del Tweet.
profile_image_url	URL de la imagen de perfil del creador del Tweet.
source	Origen de publicación del Tweet.

Nota. Información adaptada de Cercado (2023).

El conjunto de datos `media_list` reúne las propiedades de los medios o contenido multimedia compartidos en las publicaciones. Este dataset está compuesto por los siguientes atributos:

Tabla 2. *Conjunto de datos media_list*

Campo	Descripción
id	Identificador único de cada medio o contenido multimedia
type	Formato de contenido (imagen, video, etc.)
url	Dirección web para acceder al contenido multimedia.
alt_text	Texto alternativo para el contenido multimedia. (relacionado a personas invidentes)
public_metrics	Criterios concernientes con el contenido.

Nota. Información tomada de la investigación de campo. Elaborado por los autores.

Discusión

Como resultado final de este trabajo se determina que se han recopilado 100 tweets del período de marzo a agosto del 2023, correspondiente a las cuentas de X de las Verificadoras EcuadorChequea y EcuadorVerifica, lo cual atribuye a información verificada por las mismas.

Estos registros han sido recopilados con el fin de identificar comportamientos y evolución en la autenticidad de la información difundida en X, lo cual a su vez puede generar nuevos hallazgos relacionados a la forma de difundir y verificar la información.

A continuación, se muestran los registros de los dos conjuntos de datos generados, donde los datos se encuentran limpios, libres de duplicidad, valores nulos y anómalos, esto con la finalidad de asegurar la integridad y calidad de los mismos.

Figura 20. Datos de tweets_list.

	ID	Fecha de Creación	ID del Autor	Texto	Retweets	Respuestas	Likes	Citas	Nombre de Usuario del Autor	Nombre del Autor	URL de la Imagen de Perfil del Autor	Fuente
0	1691284169001713664	2023-08-15T03:02:23.000Z	777891400297897985	🚫 #FALSO/r/n/r/nLa cadena France 24 no ha publ...	2	0	8	0	ECUADORCHEQUEA	Ecuador Chequea	https://pbs.twimg.com/profile_images/169111615...	NaN
1	1691269797927919616	2023-08-15T02:05:16.000Z	777891400297897985	#AlGrano Ecuador Chequea realizó 11 verific...	0	0	3	0	ECUADORCHEQUEA	Ecuador Chequea	https://pbs.twimg.com/profile_images/169111615...	NaN
2	1691256706162917376	2023-08-15T01:13:15.000Z	777891400297897985	La pacificación fue una de las causas de la re...	1	0	5	0	ECUADORCHEQUEA	Ecuador Chequea	https://pbs.twimg.com/profile_images/169111615...	NaN
3	1691253766450798592	2023-08-15T01:01:34.000Z	777891400297897985	🔥 #DOMINGO Edición especial de #LaReglaDelPom...	1	0	5	0	ECUADORCHEQUEA	Ecuador Chequea	https://pbs.twimg.com/profile_images/169111615...	NaN
4	1691238694198431744	2023-08-15T00:01:41.000Z	777891400297897985	ec#ESPECIAL: Ecuador es presa de la violencia ...	2	0	6	0	ECUADORCHEQUEA	Ecuador Chequea	https://pbs.twimg.com/profile_images/169111615...	NaN
...
95	1690053056564469763	2023-08-11T17:30:23.000Z	777891400297897985	#LoMásVisto El político fue liberado por fal...	2	0	5	1	ECUADORCHEQUEA	Ecuador Chequea	https://pbs.twimg.com/profile_images/169111615...	NaN
96	1690007753182384128	2023-08-11T14:30:21.000Z	777891400297897985	⚠️#ATENCIÓN: Este domingo 13 de agosto a las 19...	8	0	11	0	ECUADORCHEQUEA	Ecuador Chequea	https://pbs.twimg.com/profile_images/169111615...	NaN
97	168996584840302080	2023-08-11T13:03:19.000Z	777891400297897985	#LoMásVisto Este fenómeno, está ligado a las ...	7	2	11	0	ECUADORCHEQUEA	Ecuador Chequea	https://pbs.twimg.com/profile_images/169111615...	NaN
98	1689845040565133313	2023-08-11T03:43:48.000Z	777891400297897985	#AlGrano El efecto en la percepción de segur...	10	0	14	0	ECUADORCHEQUEA	Ecuador Chequea	https://pbs.twimg.com/profile_images/169111615...	NaN

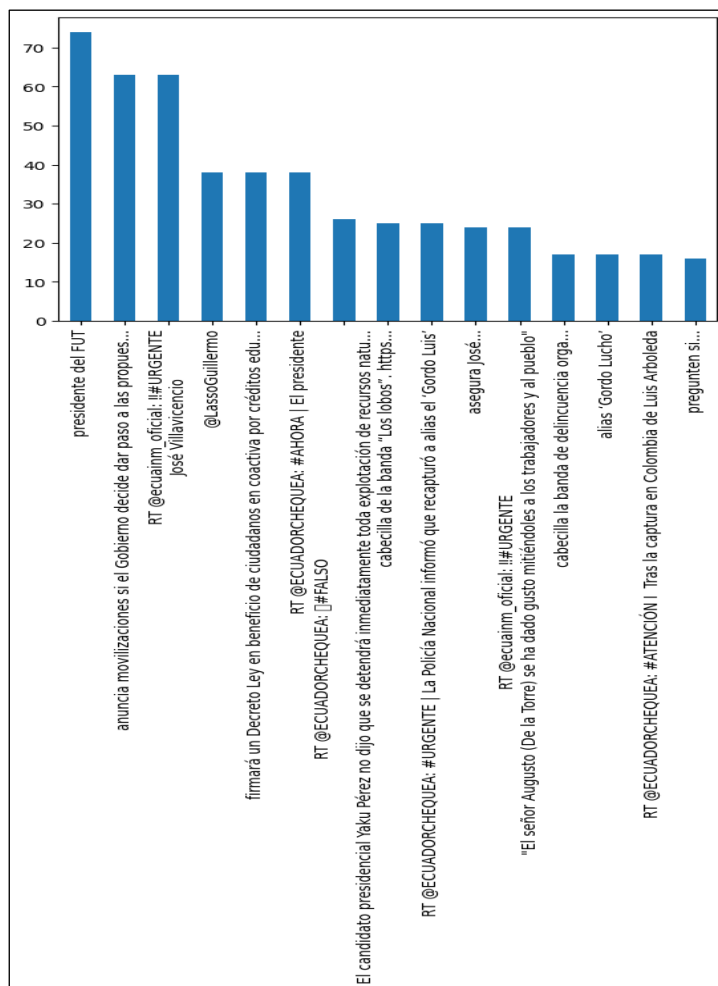
Nota. Registros del conjunto tweets_list.

Figura 21. Datos de media_list.

Index	ID del Tweet	Tipo de Media	URL de la Media	Alt Text	Métricas de la Media	Métricas de la Media/view_count
0	1691284169001713664	photo	https://pbs.twimg.com/media/F3il-3rWKAIEVcu.jpg	NaN	NaN	NaN
1	1691284169001713664	photo	https://pbs.twimg.com/media/F3il-3rWKAIEHDj.jpg	NaN	NaN	NaN
2	1691269797927919616	photo	https://pbs.twimg.com/media/F3iY6eCXwAEHSSJ.jpg	NaN	NaN	NaN
3	1691256706162917376	photo	https://pbs.twimg.com/media/F3iNqTWsAEIoZD.jpg	NaN	NaN	NaN
4	1691253766450798592	photo	https://pbs.twimg.com/media/F3iKvcsW0AERNw.jpg	NaN	NaN	NaN
5	1691238694198431744	photo	https://pbs.twimg.com/media/F3i8nLXQAIlluz.jpg	NaN	NaN	NaN
6	1691234454168023041	photo	https://pbs.twimg.com/media/F3i4xJ-XkAIVGQQ.jpg	NaN	NaN	NaN
7	1691200610618523648	photo	https://pbs.twimg.com/media/F3iZ_V-WkAAKINB.jpg	NaN	NaN	NaN
8	169119899825412096	photo	https://pbs.twimg.com/media/F3iYh6XQAEdBaZ.jpg	NaN	NaN	NaN
9	169118550967344896	photo	https://pbs.twimg.com/media/F3iHMQbCXwAEje0l.jpg	NaN	NaN	NaN
10	1691182718212538368	photo	https://pbs.twimg.com/media/F3iJt_UksAAdSAr.jpg	NaN	NaN	NaN
11	1691147024329048064	photo	https://pbs.twimg.com/media/F3i3gQOoWMAgKmwN.jpg	NaN	NaN	NaN
12	169114702560524295	video	NaN	NaN	NaN	3182.0
13	169113717442225674	photo	https://pbs.twimg.com/media/F3i3gTBGWMACr43k.jpg	NaN	NaN	NaN
14	169107982627277696	photo	https://pbs.twimg.com/media/F3i6s1yWAAADnK.jpg	NaN	NaN	NaN
15	1690932340485201920	photo	https://pbs.twimg.com/media/F3iJ_yvXIAAIYe5.jpg	NaN	NaN	NaN
16	1690932340485201920	photo	https://pbs.twimg.com/media/F3iJ_yvWUAAz2v9.jpg	NaN	NaN	NaN
17	1690929915703500801	video	NaN	NaN	NaN	1332.0
18	1690929718994874369	photo	https://pbs.twimg.com/media/F3iJnFxCAAUEF.jpg	NaN	NaN	NaN
19	1690929718994874369	photo	https://pbs.twimg.com/media/F3iJnFW0AAi-h9.jpg	NaN	NaN	NaN
20	1690927761089581056	video	NaN	NaN	NaN	1756.0
21	1690927035936366592	video	NaN	NaN	NaN	1352.0
22	1690925594609872896	video	NaN	NaN	NaN	1148.0

Nota. Visualización de Datos de media_list.

Figura 22. *Tweets más reaccionados*



Nota. Visualización de tweets más reaccionados.

Si bien es cierto, esta investigación consistió en la creación de los conjuntos de datos, también se realizó un breve análisis a los mismos, donde se evidenció la preferencia de los usuarios en la comprobación de noticias concernientes con la política nacional, debido a que ésta afecta los diferentes aspectos de la vida (economía, salud, educación, etc.). La comprobación de noticias políticas generalmente está relacionada con campañas electorales, debates, entre otros. Dado que la política, establece uno de los mayores desafíos para la comunicación en el siglo XXI, se hace necesario combatir la proliferación de información tergiversada, manipulada o falsa.

Conclusiones

Para el desarrollo de esta investigación se realizó la selección de tweets publicados por los fact checkers certificados en Ecuador por la IFCN, mediante el uso de la biblioteca Tweepy en Python, con la finalidad de crear conjuntos de datos organizados e íntegros, que permitan estudiar la credibilidad de la información.

En la fase de recopilación, se consideraron parámetros definidos por Ecuador Chequea y Ecuador Verifica, con la finalidad de avalar la importancia de los datos obtenidos. Algunos de estos temas están relacionados a áreas de la salud, ciencia y política, lo cual resultó significativo, debido a la amplia variedad en el contexto informativo.

La fase de depuración de datos fue primordial en la generación de los conjuntos de datos, se descartaron datos duplicados, nulos o anómalos, garantizando la integridad y calidad de los mismos. La reestructuración de los datos en DataFrames, favoreció su manejo y posterior análisis.

El almacenamiento de los datos se realizó en el Gestor de Base de Datos MonogDB, por sus múltiples beneficios al ser una solución confiable para muchas empresas, es una herramienta potente y altamente escalable.

En la fase de comprobación de datos se realizó mediante la generación de la Matriz de Confusión, la cual permitió analizar la efectividad de los procesos anteriores.

El análisis de los resultados reveló información relevante relacionada a tendencias en la confirmación de información en Ecuador, además de temas que se comprueban habitualmente.

Por último, esta investigación reafirma la relevancia de comprobar información dado que la difusión de noticias engañosas puede incidir en la toma de decisiones de las personas. La creación de un conjunto de datos fidedignos aportará a investigadores y analistas relacionados al campo de la información y comunicación. Además, este artículo resalta el uso de herramientas tecnológicas y las técnicas de analítica de datos, a fin de estudiar situaciones complicadas y aportar al desarrollo de una sociedad bien informada y analítica. También, este estudio ayudará en la creación de otros proyectos del mismo ámbito.

Referencias bibliográficas

- Alban, G. P. G., Arguello, A. E. V., & Molina, N. E. C. (2020). Metodologías de investigación educativa (descriptivas, experimentales, participativas, y de investigación-acción). *RECIMUNDO*, 4(3), 163–173. [https://doi.org/10.26820/recimundo/4.\(3\).julio.2020.163-173](https://doi.org/10.26820/recimundo/4.(3).julio.2020.163-173)
- Cercado J. (2023). Recopilación y Extracción de Tweets Realizados por las Verificadoras Acreditadas en el Ecuador por la IFCN para la Generación de Conjunto de Datos usando Tweepy. <https://ev.turnitin.com/app/carta/es/?s=1&u=1133714290&o=2146644106&lang=es>
- Edsson, J., & Ticona, B. (2025). DISEÑO METODOLÓGICO DE LA INVESTIGACIÓN CIENTÍFICA PARA POSGRADO IMPRESORES. <https://ucbcba.academia.edu/JoseEdssonBarreroTicona>
- Instalación — documentación de tweepy 4.14.0. (n.d.). Retrieved January 24, 2026, from <https://docs.tweepy.org/en/stable/install.html>
- Kauk, J., Kreysa, H., & Schweinberger, S. R. (2025). Large-scale analysis of fact-checked stories on Twitter reveals graded effects of ambiguity and falsehood on information reappearance. *PNAS Nexus*, 4(2). <https://doi.org/10.1093/PNASNEXUS/PGAF028>
- Menoyo Ros, D., García López, E., & García Cabot, A. (2021). Fundamentos de la ciencia de datos.
- METODOLOGÍA - Ecuador Verifica. (n.d.). Retrieved January 24, 2026, from <https://ecuadorverifica.org/metodologia/>
- Moreno-Gil, V., & Salgado-De Dios, F. (2023). El cumplimiento del código de principios de la International Fact-Checking Network en las plataformas de verificación españolas. Un análisis cualitativo. *Revista de Comunicación*, 22(1), 293–307. <https://doi.org/10.26441/RC22.1-2023-2971>
- Nuestra metodología - Ecuador Chequea. (n.d.). Retrieved January 24, 2026, from <https://ecuadorchequea.com/metodologia-ecuador-chequea/>
- Sarasa Cabezuelo, A. (2016). Introducción a las bases de datos NoSQL usando MongoDB [Libro electrónico].
- Sharifpoor, E., Okhovati, M., Ghazizadeh-Ahsaee, M., & Avaz Beigi, M. (2025). Classifying and fact-checking health-related information about COVID-19 on Twitter/X using machine learning and deep learning models. *BMC Medical Informatics and Decision Making*, 25(1). <https://doi.org/10.1186/S12911-025-02895-Y>
- Subirats, L., Diego, M., Pérez, O., Mireia, T., & González, C. (n.d.). Introducción a la limpieza y análisis de los datos PID_00265704.
-

- Számely, J., Galeazzi, A., Koltai, J., & Omodei, E. (2025). Easy-access online social media metrics are associated with misinformation sharing activity. *Scientific Reports* 2025 15:1, 15(1), 41288-. <https://doi.org/10.1038/s41598-025-25049-6>
- Toapanta M., García M., & Ureña L. (2024). Vista de Uso de modelos en idioma español para la detección de noticias falsas y verificación de hechos en tuits de Ecuador Chequea y Ecuador Verifica. <https://rte.espol.edu.ec/index.php/tecnologica/es/article/view/1219/817>
- Wei, S., Dong, H., Yao, W., Chen, Y., Wang, X., Ji, W., Zhang, Y., & Guo, S. (2025). Machine learning models for predicting in-hospital mortality from acute pancreatitis in intensive care unit. *BMC Medical Informatics and Decision Making*, 25(1). <https://doi.org/10.1186/S12911-025-03033-4>
- Tweepy. (2020). API. <https://docs.tweepy.org/en/stable/api.html>
- X. (2023). Información sobre los diferentes tipos de Post. <https://help.x.com/es/using-x/types-of-posts>
- X. (2023). Política de desarrolladores. <https://developer.x.com/es/developer-terms/policy>
- X. (2023). Las reglas de X. <https://help.x.com/es/rules-and-policies/x-rules>
- Verificat. (2021). ¿Qué es el fact-checking? <https://www.verificat.cat/escola/articulo/que-es-el-fact-checking-cast>
- Toapanta, M. (2023). Investigaciones de tesis doctoral como estudiante de la Universidad de Jaén y cursante del doctorado en Tecnologías de Información y Comunicación.
-