



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA



Departamento de Ciencias de la Computación

Carrera de Ingeniería en Software

Trabajo de titulación previo a la obtención del título de Ingeniero en Software

“Interfaz de programación de aplicaciones para la detección de fake news aplicando técnicas de machine learning, en la red social Twitter caso práctico covid-19 Ecuador.”

AUTORES:

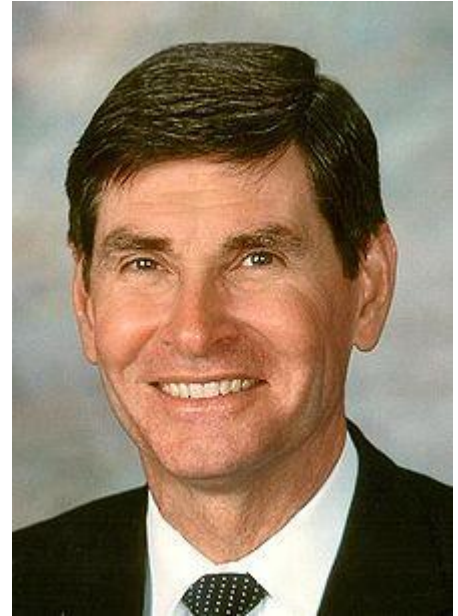
**ALTAMIRANO PEREZ, BRAYAN OMAR
VALENCIA VALENCIA, ODALIS DAYANNA**

Dr. Carrillo Medina Jose Luis, Director
Dr. Diego Miguel Marcillo Parra, Codirector



«La motivación es lo que te pone en marcha, el hábito es lo que hace que sigas»

Jim Ryun



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Índice

01	Introducción	<ul style="list-style-type: none">• La red social como una herramienta de comunicación• El uso de las redes sociales en plataformas de entretenimiento• Red social Twitter• Detección de la información errónea
02	Background	<ul style="list-style-type: none">• El impacto del internet en la vida diaria• Evaluación de fuentes de información en Internet.
03	Descripción del sistema	<ul style="list-style-type: none">• Diseño del software aplicado
04	Desarrollo del sistema	<ul style="list-style-type: none">• Proceso de desarrollo (6 Procesos)
05	Resultados y discusión	<ul style="list-style-type: none">• Análisis y comprobación del funcionamiento• Conclusiones y trabajos futuros



Abstract

- En el mundo siempre han existido las noticias engañosas, se ha vuelto mucho más sencilla la divulgación de falsa información en las nuevas tecnologías de comunicación, permitiendo que los usuarios sean además de consumidores, también productores de discursos que circulan y que muchas veces son de dudosa fuente de información, denominado como “Fake News”

- Las noticias falsas se difunden a través de medios engañosos como la red social Twitter que permite a los usuarios leer y compartir la información, aumentando su alcance e impacto, con ese fin, hay una variedad de formas de mitigar estas áreas a través de la API (Application Programming Interfaces)



Introducción



Introducción

La red social

Es una herramienta de comunicación que permite a la web participar en un entorno abierto para publicar e intercambiar datos que está en constante desarrollo

El uso de las redes sociales

Se ha adaptado a personas de todo el mundo y ahora es parte de la rutina diaria, al igual que las marcas y empresas las mismas que comparten información e interactúan directamente con los clientes a través de estos medios de difusión

Red social Twitter

Es una plataforma social, y su principal servicio es la comunicación bidireccional con el que puedes compartir información de diverso tipo de una forma rápida, sencilla y gratuita

Información errónea

Es aquella información que ha sido publicada con la intención de engañar y confundir al lector, los medios que fabrican estas noticias imitan la forma de publicación de los medios fiables, ignorando los procesos de verificación o directamente adulterando intencionadamente el contenido.



Introducción

Mitigar el problema

Para mitigar los problemas relacionados con la difusión de noticias falsas se usa una interfaz de programación de aplicaciones permitiendo la detección de Fake News en la red social twitter, a través de algoritmos de aprendizaje automático, que contribuirá a reducir la desinformación de usuarios con este tipo de redes.

API

Se trata de un conjunto de definiciones y protocolos que se utiliza para desarrollar e integrar el software de las aplicaciones, permitiendo la comunicación entre dos aplicaciones de software a través de un conjunto de reglas

API en Twitter

La API de Twitter permite a los desarrolladores interactuar con la plataforma y acceder a sus datos, mediante este recurso, los desarrolladores pueden crear aplicaciones que se integren con Twitter y ejecutar acciones como publicar tweets, recopilar datos de usuarios, revisar tendencias y en general, analizar la actividad en la plataforma



Background



Background



El impacto del internet en la vida diaria

- La comunicación a través de internet afecta las relaciones sociales.
- Internet libera al individuo de las restricciones geográficas
- El internet y su evolución
- Manipulación de códigos
- Generación de contenidos



Evaluación de fuentes de información en Internet


- Internet como proceso social
- Constante desarrollo y mutación
- Desarrollo de mecanismos para su uso
- Garantizar a sus usuarios un orden relativo
- Elaboración de instrumentos que permitan controlar la calidad de las informaciones



Descripción del sistema



Esquematización de enfoque en dos secciones

- 
- El diseño del software aplicado, esquematiza un enfoque en dos secciones: el entrenamiento de la neurona y la estructura que conllevan las extensiones web, generando un stack tecnológico centrado en el óptimo desempeño de una aplicación de inteligencia artificial

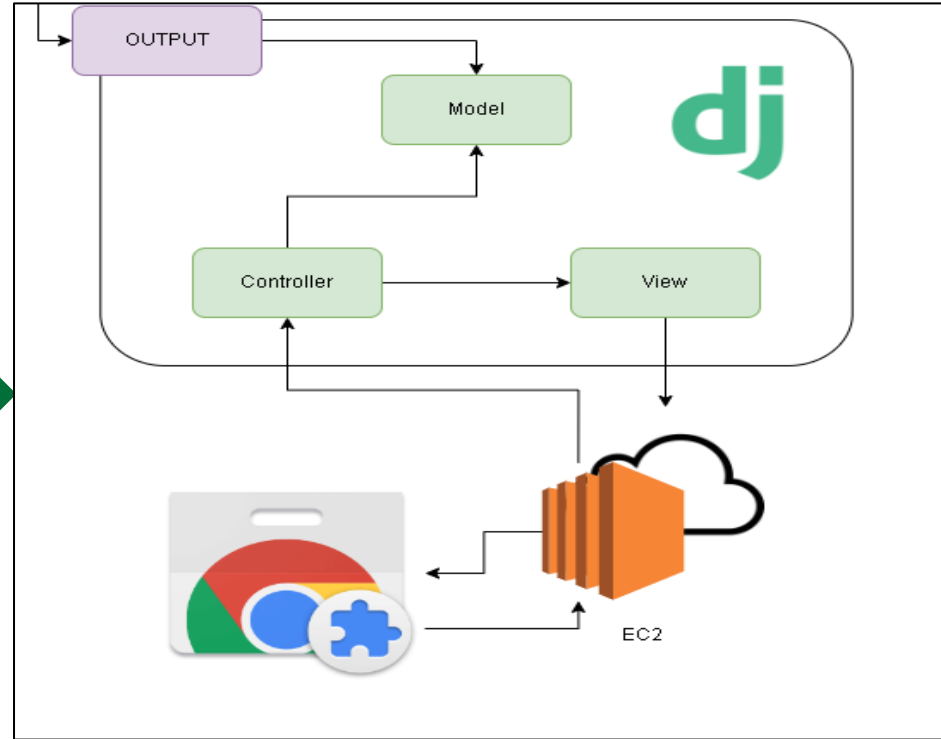
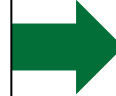
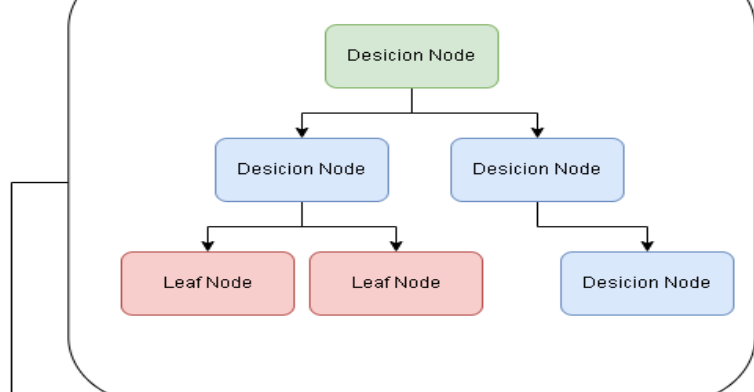


Esquematización de enfoque en dos secciones



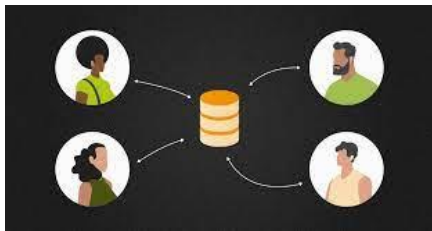
Each news item will be processed in a decision tree in order to determine if a news item is false.

Decision Tree



Obtención de un data set

Para poder obtener un data set entrenado enfocado en detectar noticias falsas, se procedió a:



Recopilar y preparar datos de Twitter

Las cuales incluyen las noticias verdaderas como falsas, donde se pueda utilizar la API de Twitter o bibliotecas de terceros para recopilar datos, con los datos recopilados se procedió a realizar su respectiva preparación donde se eliminó los irrelevantes, se normalizo y se filtró según las especificaciones base



División de datos

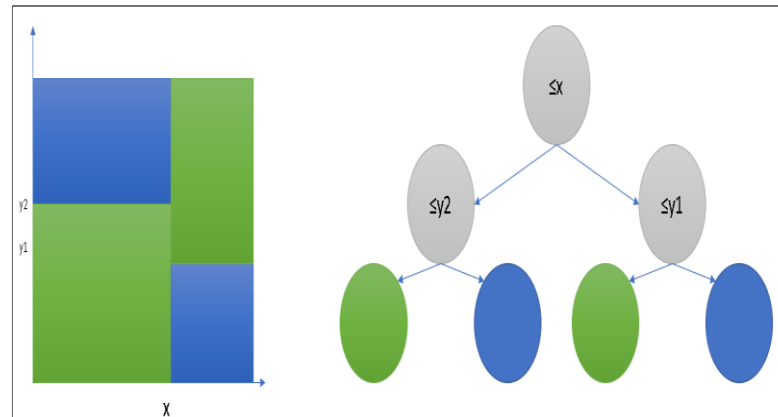
Con los datos preparados se procede a dividirlos en dos conjuntos, uno para entrenar el modelo y otro para probar su precisión. Por lo general, se utiliza una proporción de 70/30, lo que significa que el 70% de los datos se utilizan para entrenar el modelo y el 30% se utilizan para probar la precisión del modelo.



Modelo de árbol de decisión

Aplicación del modelo

- Con el modelo entrenado se procede a evaluar su precisión utilizando el conjunto de prueba. Esto te permitirá saber cómo se desempeña el modelo en datos que no ha visto antes mediante el uso de árboles de decisión.
- Una vez se obtuvo el modelo entrenado, se puede detectar cuando es o no una noticia falsa, la misma que se configuró utilizando un API creada en DJANGO que será la encargada de la gestión de tweets.



- Se realiza la conexión con el order bot diseñado en Watson, el mismo se encargará de gestionar la información remitida por la API y determinará si la noticia obtiene un visto bueno para ser transmitida

Desarrollo del sistema



Desarrollo del sistema

Proceso de desarrollo

- Diseño y desarrollo del dataset
- Criterios de validación y algoritmos
- Selección de algoritmos de decisión
- Integración del Modelo Entrenado y API:
- Creación de la extensión para Chrome
- Comunicación con el Order bot:



Diseño y desarrollo del Dataset

- Al ser un conjunto de datos, el objetivo del dataset es presentar un conjunto completo enfocado en las noticias suscitadas en idioma español que afectaron al país tanto de manera local como internacional.

- Para ello se procedió a investigar y tomar noticias de fuentes confiables internas del país y de canales dudosos, comparando las características que hacen a una noticia real de una falsa.

- Finalmente se obtuvo un listado ordenado a través de un .csv, recompilando alrededor de 310 fuentes repartidas entre noticias falsas 186 y reales 124



Diseño y desarrollo del Dataset

Presentación del Dataset

```
detencion_fake_news_radio > deteccion_fake_news > files_model_dtrees > dataset.csv
1 Id,Texto,Me gusta,Retweets,Respuestas,Citas,Usuario,Seguidores,Verificado,Número Tweets,Filtro
2 1490526119573000196,noticia karelim lópez presentó dos pruebas falsas covid19 asistir comisión fiscalización httpstc
3 1278123219262869507,gotas nanoparticulas citricos habló hace días secretaria gobernación molgascordero creación cien
4 1454604259555745796,gente confía vacuna srs vacuna hace inmune vacuna solo baja riesgo mortalidad puede dar covid ma
5 1454605385843810308,creen covid influido bastante problemas cardíacos deportistas httpstcozijwe6bq5,0,0,0,Josuesi
6 1454606547565285384,capital china ordenó sábado cierre parte cines mientras país enfrenta rebrote epidémico covid19d
7 1454606548291014657,capital china ordenó sábado cierre parte cines mientras país enfrenta rebrote epidémico covid19d
8 1454608923688833025,tras casi dos años pandemia golpeará puerta unidades terapia intensiva quito hemos librado guerra
9 1454611187539320836,hoy brindó atención puntos vacunación municipal iniciativa impulsada alcaldesa cynthiaviteri6 pe
10 1454611806962454530, covid19 china donó 500 mil dosis sinovac ecuadorla5pata nosotroslabuscamos ecuador china sinov
11 1454614191466237953, estudio tercera dosis vacuna pfizer indica efectiva reducir riesgo hospitalización covid19 > htt
```

	Id	Texto	Me gusta	Retweets	Respuestas	Citas	Usuario	Seguidores	Verificado	Número Tweets	Filtro
0	1490526119573000196	noticia karelim lópez presentó dos pruebas fal...	619	291	81	19	willaxtv	677528	False	197145	0
1	1278123219262869507	gotas nanoparticulas citricos habló hace días ...	5761	1846	4798	1854	martibatres	541435	True	40843	1
2	1454604259555745796	gente confía vacuna srs vacuna hace inmune vac...	0	0	0	0	te_informoYo	57	False	9789	0
3	1454605385843810308	creen covid influido bastante problemas cardia...	0	0	0	0	Josuesiv18	639	False	54367	0
4	1454606547565285384	capital china ordenó sábado cierre parte cines...	8	2	0	0	tctelevision	1696032	True	202551	1



Crterios de validación y algoritmos

Para una correcta validación e implementación se tomaron en cuenta varios algoritmos de entrenamiento, obteniendo mejores resultados y procesamiento de dos, arboles de decisión, Random forest y k-Nearest Neighbors(k-NN).

Crterio 1 (Validez)

- Fidedignidad y confiabilidad de las informaciones

Crterio 2 (Precisión)

- Estrechamente vinculada a la validez, se refiere a la corrección de las informaciones

Crterio 3 (Autoridad y reputación de la fuente)

- Especialidad y status del productor.

Crterio 4 (Singularidad)

- Cantidad de información primaria no disponible en otras fuentes

Crterio 5 (Acabado)

- Grado de acabado o terminación de la información disponible

Crterio 6 (Cobertura)

- Profundidad y amplitud de la fuente



Selección de algoritmos de decisión

k-Nearest Neighbors (k-NN):



- Con los algoritmos seleccionados se procedió a realizar las respectivas pruebas utilizando un conjunto de datos preparado, detallando a continuación los resultados obtenidos tanto para el algoritmo k-NN como para Random forest y arboles de decisión

Confusion Matrix KNN



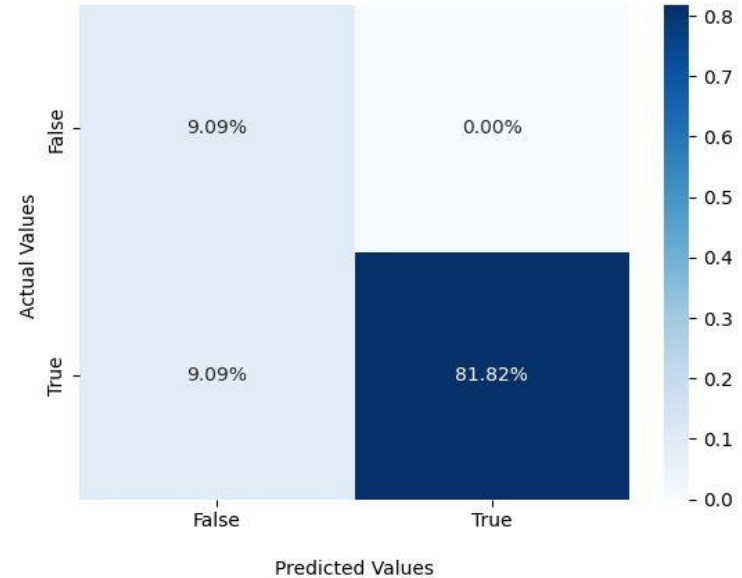
Selección de algoritmos de decisión

Random Forest



- Método de aprendizaje conjunto para clasificación, regresión y otras tareas que opera mediante la construcción de una multitud de árboles de decisión en tiempo de entrenamiento.

Confusion Matrix Random Forest



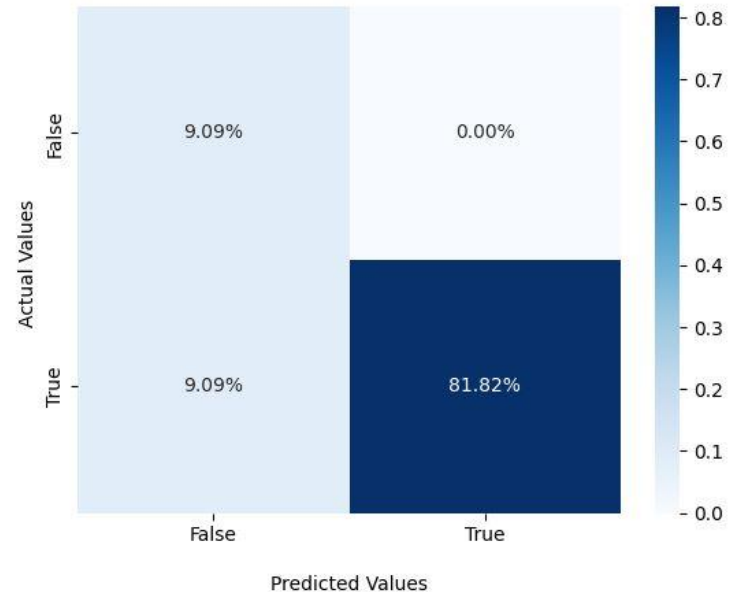
Selección de algoritmos de decisión

Decision Tree



- Método que se usa para realizar un análisis que consiste en delinear de forma gráfica los posibles resultados, costos y consecuencias de una decisión compleja.

Confusion Matrix Decision Tree



Selección de algoritmos de decisión



- Con la aplicación de cada uno de los algoritmos, se puede determinar el entrenamiento de la neurona, descartando KNN por un accuracy igual al 0,54 sobre 1



- Para determinar si una noticia es falsa o no y manteniendo como opciones los árboles de decisión como Random forest por su rendimiento superior al 90% de accuracy



- El paso final es la medición de los tiempos de respuestas, obteniendo como mejor resultado 68ms, tiempo de respuesta

Aplicación de los algoritmos

```
sorted(accuracies.items(), key=lambda x: x[1],reverse=True)
[ ]
... Classification models in descending order of accuracy:

[('Arbol de desicion', 0.9090909090909091),
 ('Random Forest', 0.9090909090909091),
 ('KNN', 0.5454545454545454)]
```

```
'Arbol de desicion': 0.9090909090909091 --- 68ms
'Random Forest', 0.9090909090909091 --- 377 ms
'KNN', 0.5454545454545454----- 2.6 segundos
```



Integración del Modelo Entrenado y API:

- Para poder enlazar el modelo se procede a realizar una API en Django

- La misma que procesará las solicitudes y hará el llamado al modelo para realizar la predicción

- Una vez que la API recibe los datos, limpia y analiza la información, y envía el resultado solicitado.



Creación de la extensión para Chrome:

- Con el API ejecutándose de manera correcta se procede a crear la extensión para Chrome.
- Creamos una instancia EC2 en AWS
- La extensión permitirá obtener los datos del tweet y se comunicará con el API en AWS.
- Se procede a publicar la extensión en Chrome Web Store
<https://chrome.google.com/webstore/detail/detecci%C3%B3n-fake-news/ibimbkdefdfhlbciigkmkekikmgifaid?hl=es&authuser=0>



Comunicación con el Order bot

- Una vez la comunicación con la api y la extensión es estable se procede a realizar la comunicación entre los resultados obtenidos y las solicitudes definidas en el order bot, para ello se diseñó e implemento un midelware encardado de realizar la comunicación entre ambos módulos, además se procedió a establecer las reglas y los secretos necesarios en IMB Watson



OrderBot



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Resultados y discusión



Detección Tweet

Detección Tweet al ser una extensión para navegador su funcionamiento es sencillo y el único requisito adicional aparte de tener la extensión en el navegador es estar loqueado en Twitter.



Para iniciar con su uso, lo primero que se debe buscar una noticia dentro de la plataforma, una vez encontrada, la extensión notificará si se desea realizar el análisis de la información, en caso de aceptar dará inicio al proceso de análisis de la noticia

← Tweet

 La República EC
@larepublica_ec

Ecuador tiene 29 días sin muertes por COVID-19



larepublica.ec
Ecuador tiene 29 días sin muertes por COVID-19
El 16 de marzo de 2020, en Ecuador, el tiempo se detuvo. Ese día marcó el inicio de una cuarentena que apareció tres meses . Las calles estaban vacías...

Buscar e

Detección Tweet



TWEET VERDADERO

Personas

 La República EC
@larepublica_ec
Información
Artículos de política, cultura y deportes. Periodismo Independiente.

Tendencias para ti

Tendencia en Ecuador ...
#URGENTE
32,8 mil Tweets

Tendencia en Ecuador ...
#MasterChefEcuador
1.183 Tweets

Tendencia en Ecuador ...
Esmeraldas
3.295 Tweets

Tendencia en Ecuador ...

← Tweet

 Wilson Romero J.
@WilsonRomeroJ

Neoliberales nefastos, cobardes mentirosos al estilo mafia, destruyen, desprestigian al oponente y luego se aprovechan de los bienes del pueblo, plaga enquistada en el poder político-económico de las naciones en desarrollo desde el siglo anterior, es más mortal que el Covid 19.

12:01 a. m. · 31 oct. 2021

Buscar e

Detección Tweet



TWEET FALSO

Personas

 Wilson Romero J.
@WilsonRomeroJ
Economía

Tendencias para ti

Solo en Twitter · Tendencia ...

O rd	Id tweet	Result ado en newtr ales	Predic ción
1	https://twitter.com/Africamar/status/1634288274998022191	Fake	True
2	https://twitter.com/dejanirasilveir/status/1632479592438022145	Fake	True
3	https://twitter.com/MakerlyYenny/status/1582335544755093504	Fake	Fake
4	https://twitter.com/DIWWOOO/status/1603774311034015746	Fake	Fake
..
13	https://twitter.com/Newtral/status/1595073374304481281	True	True
...
19	https://twitter.com/Newtral/status/1621184368675880961	True	True
20	https://twitter.com/Newtral/status/1624390092625149954	True	True

Realizado el proceso de análisis la extensión presentara dos estados, el primero notificando si la noticia ingresada es verdadera y el segundo que se presentara en caso de ser la noticia falsa, enviando dichos resultados al order bot que determinara el valor de la noticia e informara al locutor de radio sobre la misma

La prueba uno dio como resultado 5 aciertos sobre 10 para noticias falsas y 10 de 10 aciertos para las verdaderas



Ord	Id tweet	Resultado en newtrales	Predicción
1	https://twitter.com/Africamar/status/1634288274998022191	Fake	Fake
2	https://twitter.com/dejanirasilveir/status/1632479592438022145	Fake	Fake
3	https://twitter.com/MakerlyYenny/status/1582335544755093504	Fake	Fake
4	https://twitter.com/DIWWOOO/status/1603774311034015746	Fake	Fake
..
13	https://twitter.com/Newtral/status/1595073374304481281	True	True
...
19	https://twitter.com/Newtral/status/1621184368675880961	True	True
20	https://twitter.com/Newtral/status/1624390092625149954	True	True

Con una mejor calibración se procedió a realizar más pruebas, obteniendo como resultados finales los valores presentados en la tabla



Conclusiones y trabajos futuros





- La aplicación de herramientas de inteligencias artificial son un paso hacia la libertad de expresión y la veracidad de la información, las mismas en la actualidad se han expandido, por lo que combatir la falta de información y las publicaciones mal intencionadas son un medio de defensa para el inocente, además poder implementar IA en la actualidad cada vez es más accesible para la población en general, permitiendo un crecimiento constante en su uso.



- Se planea mejorar el order bot a fin de permitir una comunicación directa con la audiencia de la radio, permitiendo la automatización de la programación y generando un medio de emergencia en caso de no existir personal activo o se suscite alguna emergencia



Gracias por su atención

