

DETECCIÓN DE FAKE NEWS

USANDO INTELIGENCIA ARTIFICIAL

INTRODUCCIÓN

Este proyecto busca abordar el problema de las noticias falsas mediante el análisis de patrones en los datos. Utilizando diversos enfoques y modelos de clasificación, exploramos métodos que permitan identificar información confiable, promoviendo así un acceso más seguro y transparente al contenido en medios digitales.



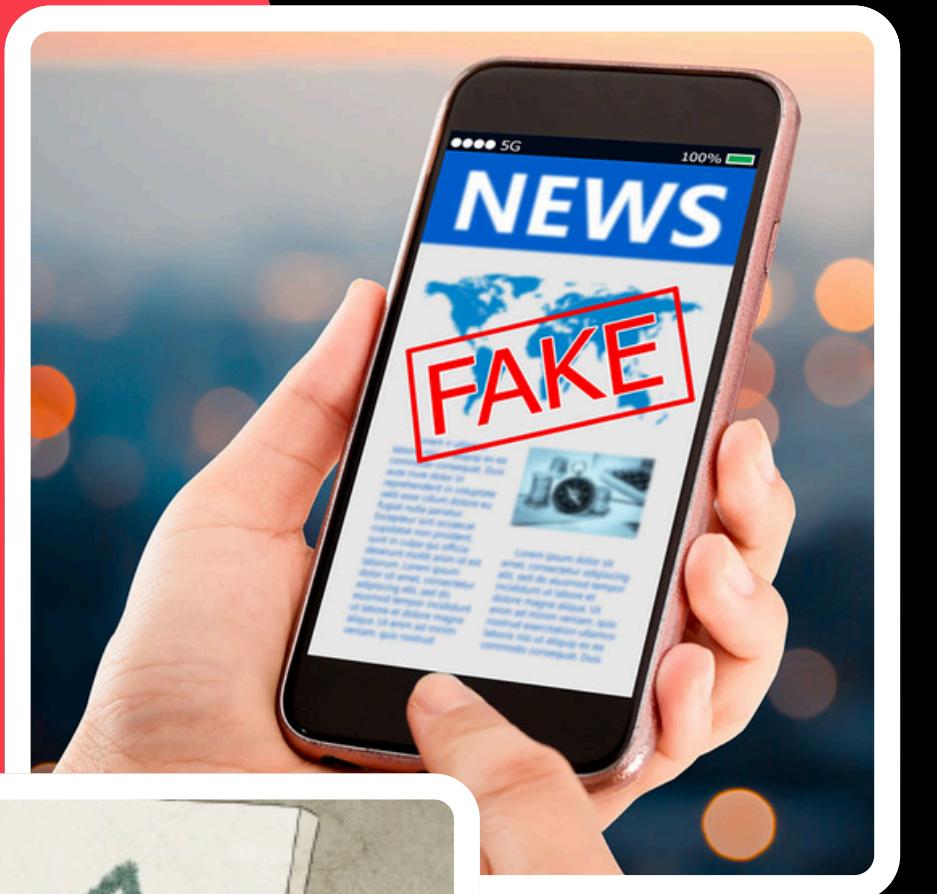
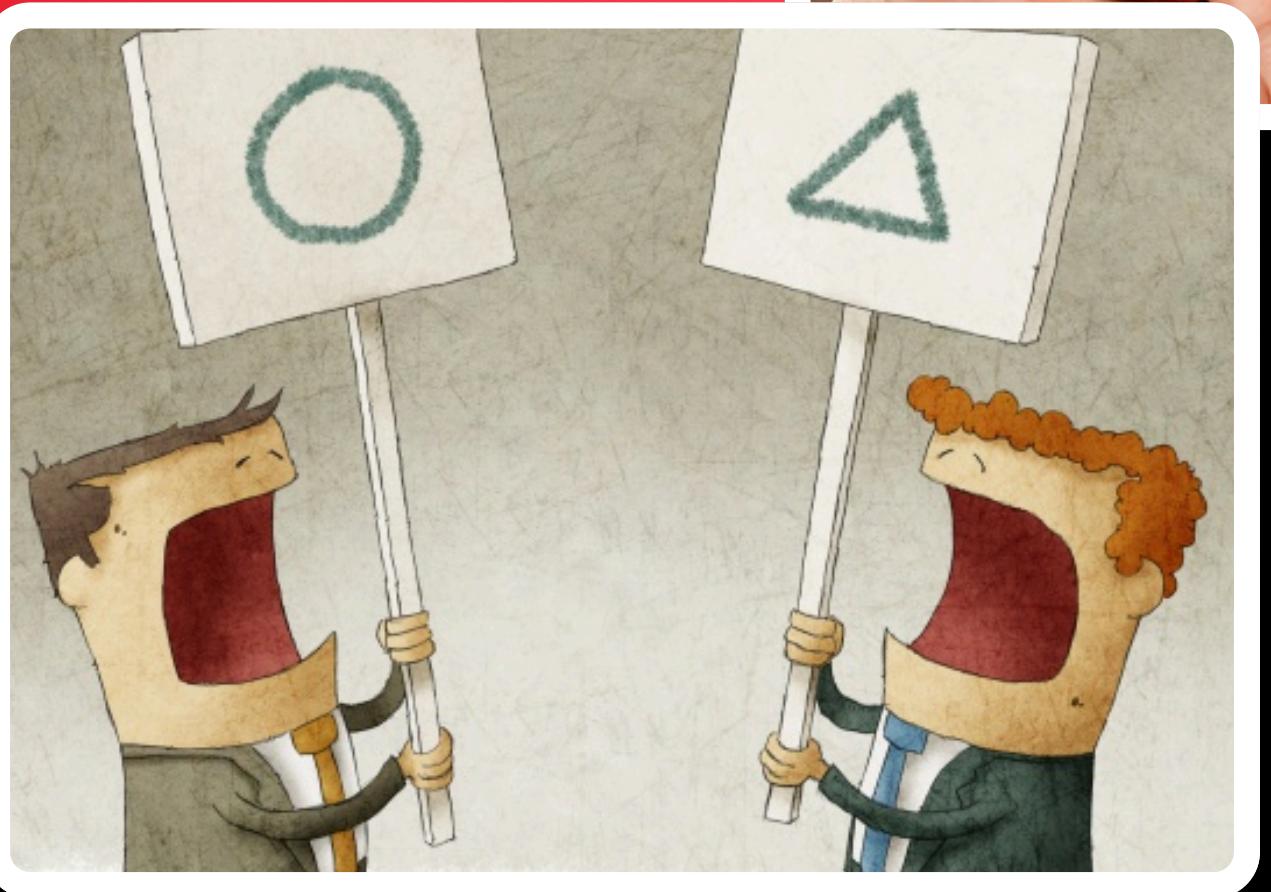
PLANTEAMIENTO

Consideramos que es un gran problema la distribución de noticias falsas, ya que actualmente se distribuyen enormes volúmenes de información, esto tiene como consecuencia que las Fake News se distribuyan considerablemente, dado que es inviable que las personas verifiquen personalmente la veracidad de las noticias.



OBJETIVO

Desarrollar un modelo de inteligencia artificial que analice el contenido de diferentes noticias recopiladas a apartir de medios de todo el mundo y aprenda a distiguir las noticias verdaderas y falsas, así como evaluar el sesgo político que tienen los autores de las noticias.



DATASET

FNDD

Fake News Detection Datasets: Kaggle

Conjunto de noticias reales y falsas de diversos medios reales alrededor del mundo. Las noticias falsas fueron recolectadas de distintos medios, cuyos dominios fueron categorizados como no factibles por PolitiFact, mientras que las verdaderas fueron recabadas mayormente de Reuters.

Title	Text	Subject	Date
Título de la noticia	Contenido de la noticia	Tópico de la noticia	Fecha de publicación de la noticia

GOSSIP COP

Fake News Net en GitHub

El dataset GossipCop contiene noticias centradas en el mundo de las celebridades y el entretenimiento, clasificadas como reales o falsas. Está diseñado para el análisis de desinformación y cómo esta se propaga en plataformas digitales.

ID	news_url	title	tweet_ids
Identificador único para cada noticia.	Enlace a la noticia original.	Título del artículo de la noticia.	IDs de los tweets relacionados con la noticia.

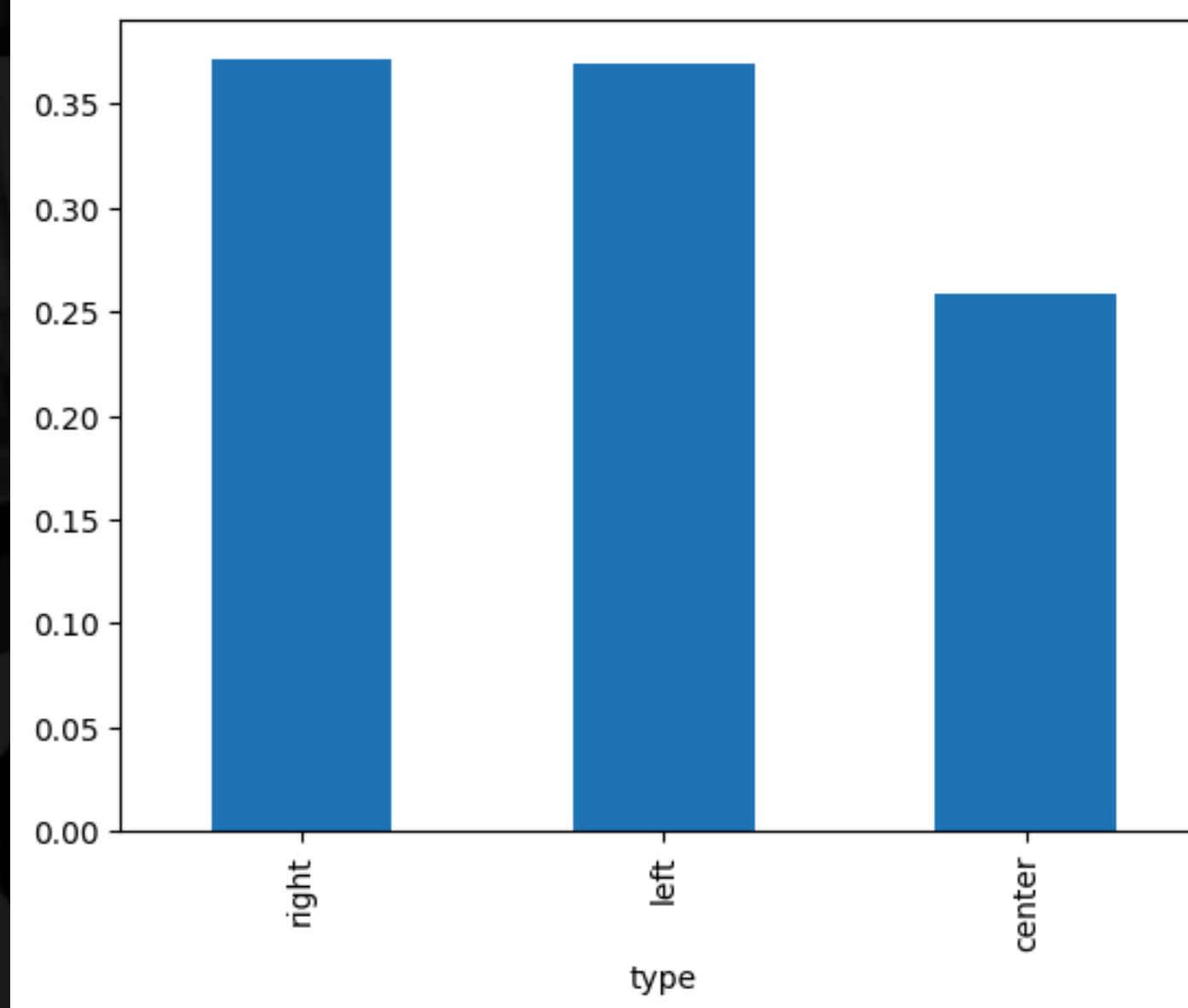
BABE

Media Bias Dataset: Annotations By Experts

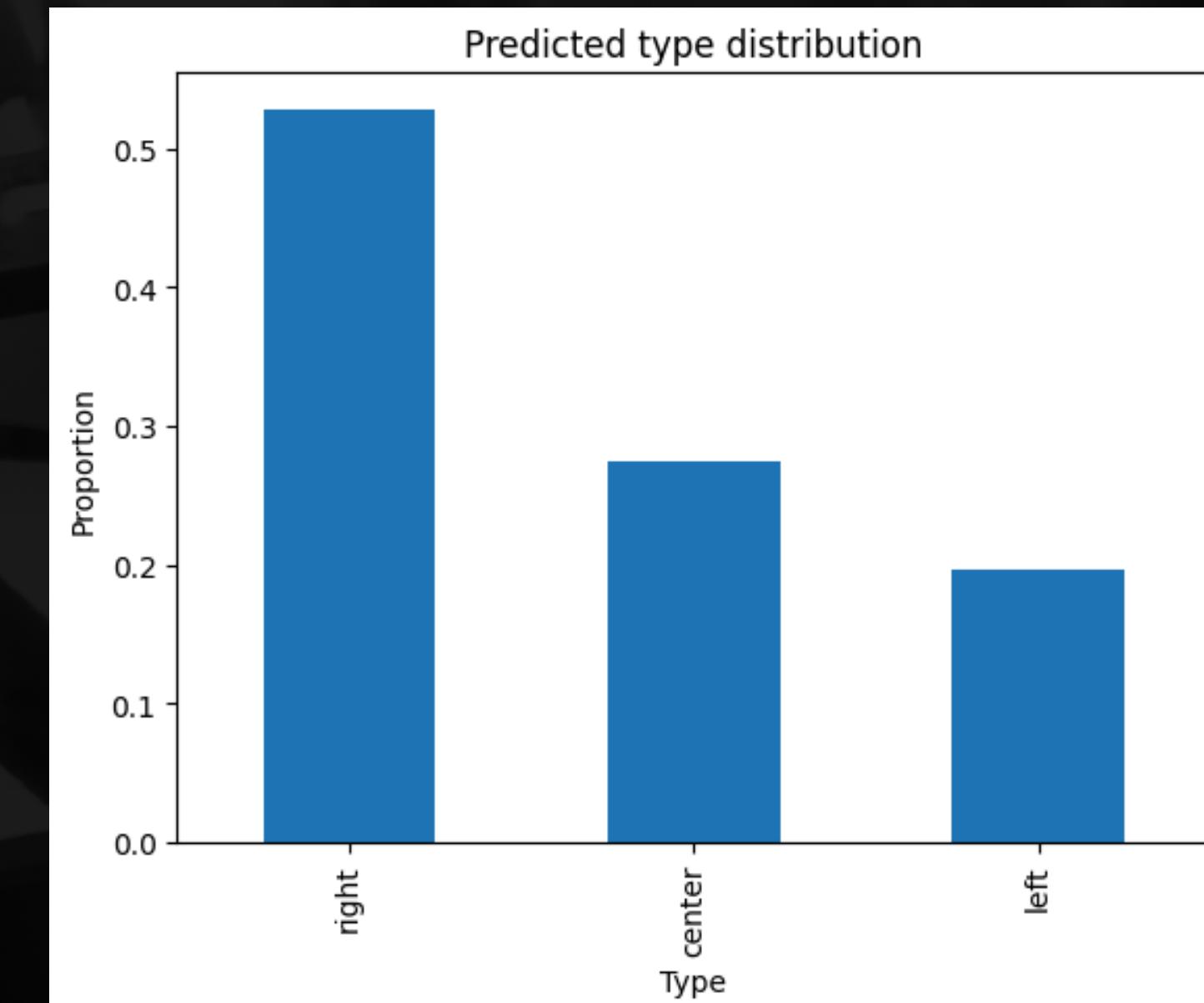
Dataset con noticias de diversos medios. Categorizados con su posición política y si el texto se encuentra sesgado, con anotaciones hechas por expertos.

text	outlet	topic	type	biased_words	label_bias
contenido de la noticia	medio publicador	tema de la noticia	izquierda, centro o derecha	lista de palabras sesgadas	si la noticia contiene sesgo

Sin llenar null values



Llenando null values



PROPÓSITO

El propósito del clustering es identificar patrones y agrupar elementos similares dentro de un conjunto de datos, sin necesidad de etiquetas predefinidas. Es una técnica esencial para análisis exploratorios y descubrimiento de estructuras ocultas en los datos.



Home

About

Contact

CLUSTERING

OBJETIVOS

- **identificar relaciones entre noticias**

Agrupar titulares que comparten características similares, como temáticas, vocabulario o estilo.

- **Enriquecer el modelo supervisado**

Las etiquetas de cluster generadas se integran como características adicionales, ayudando al modelo supervisado a mejorar su capacidad de clasificación.



Home

About

Contact

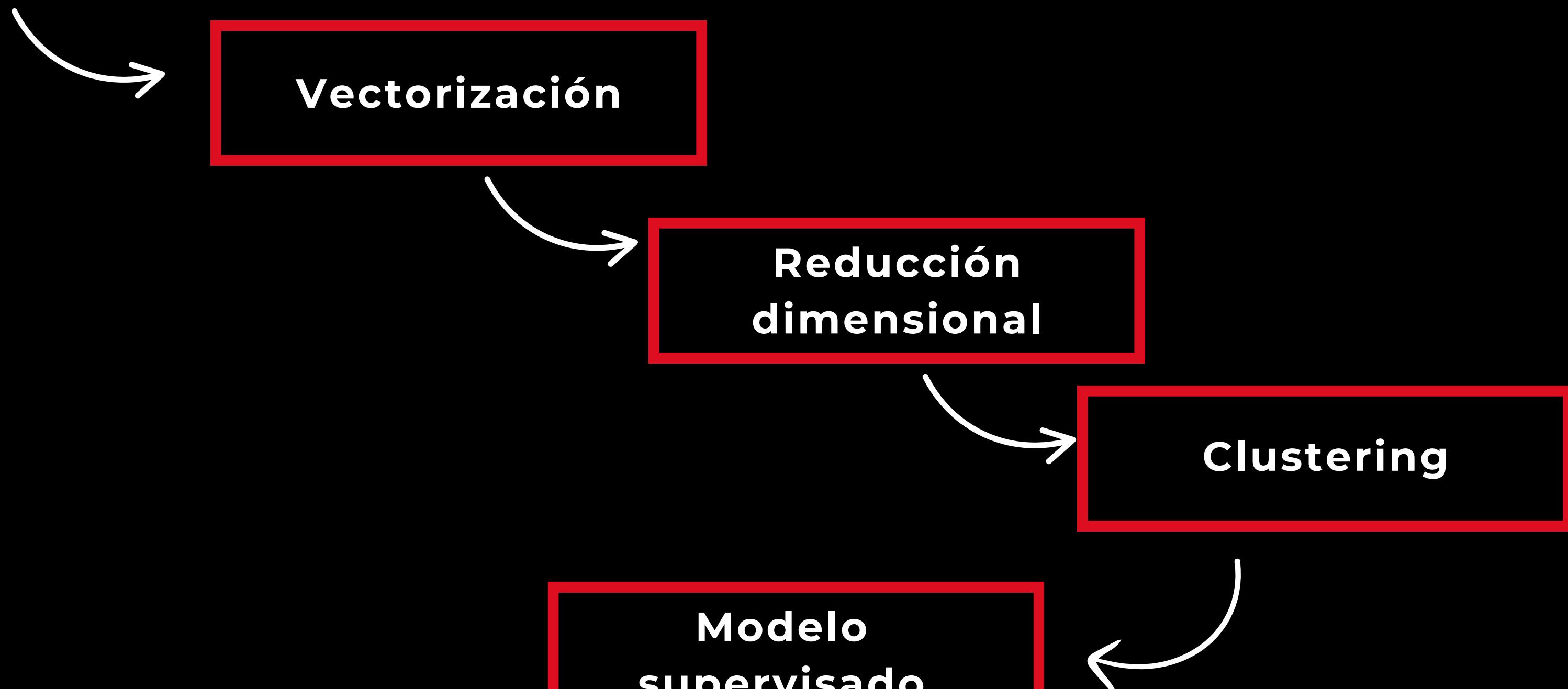
Preprocesamiento

Vectorización

Reducción
dimensional

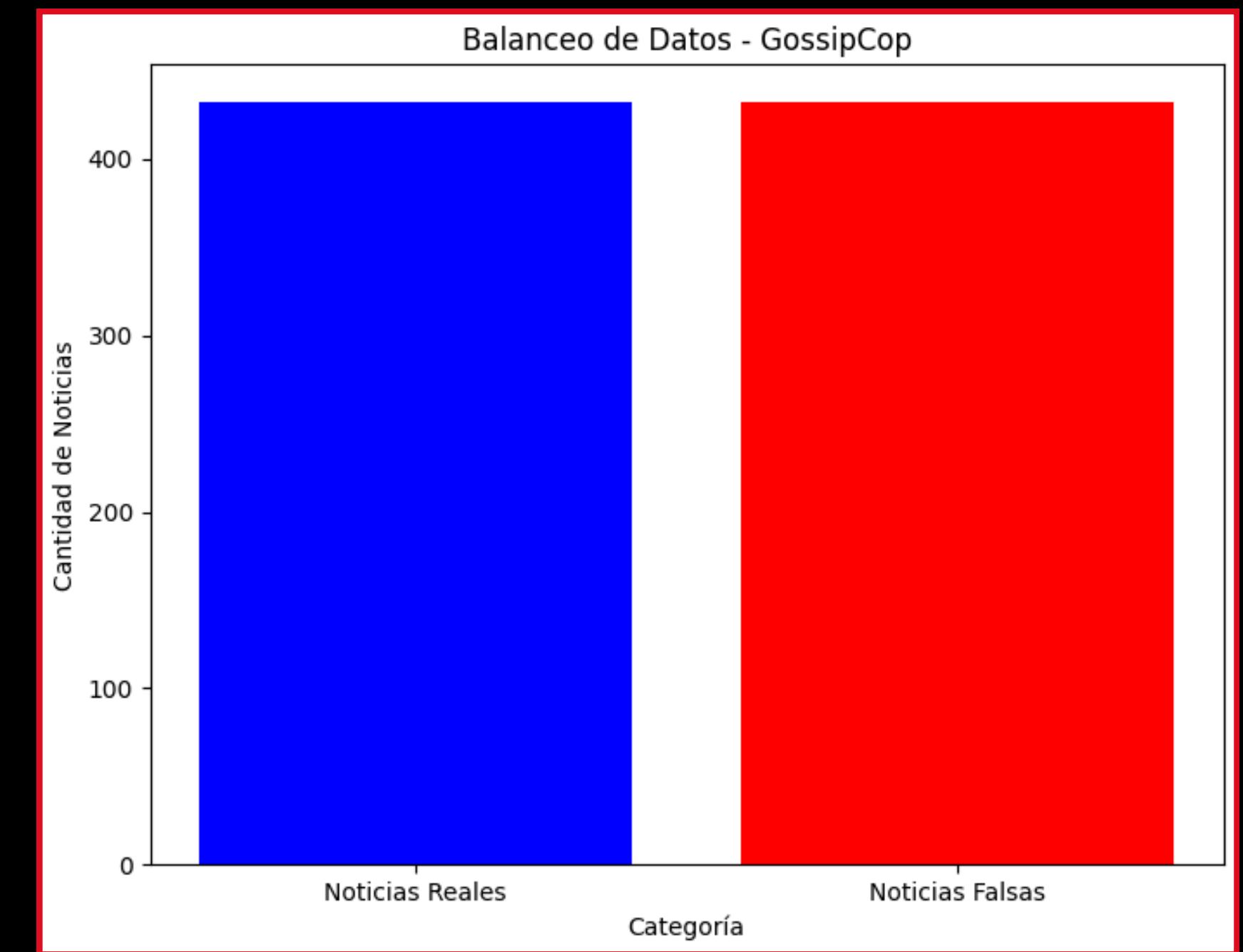
Clustering

Modelo
supervisado



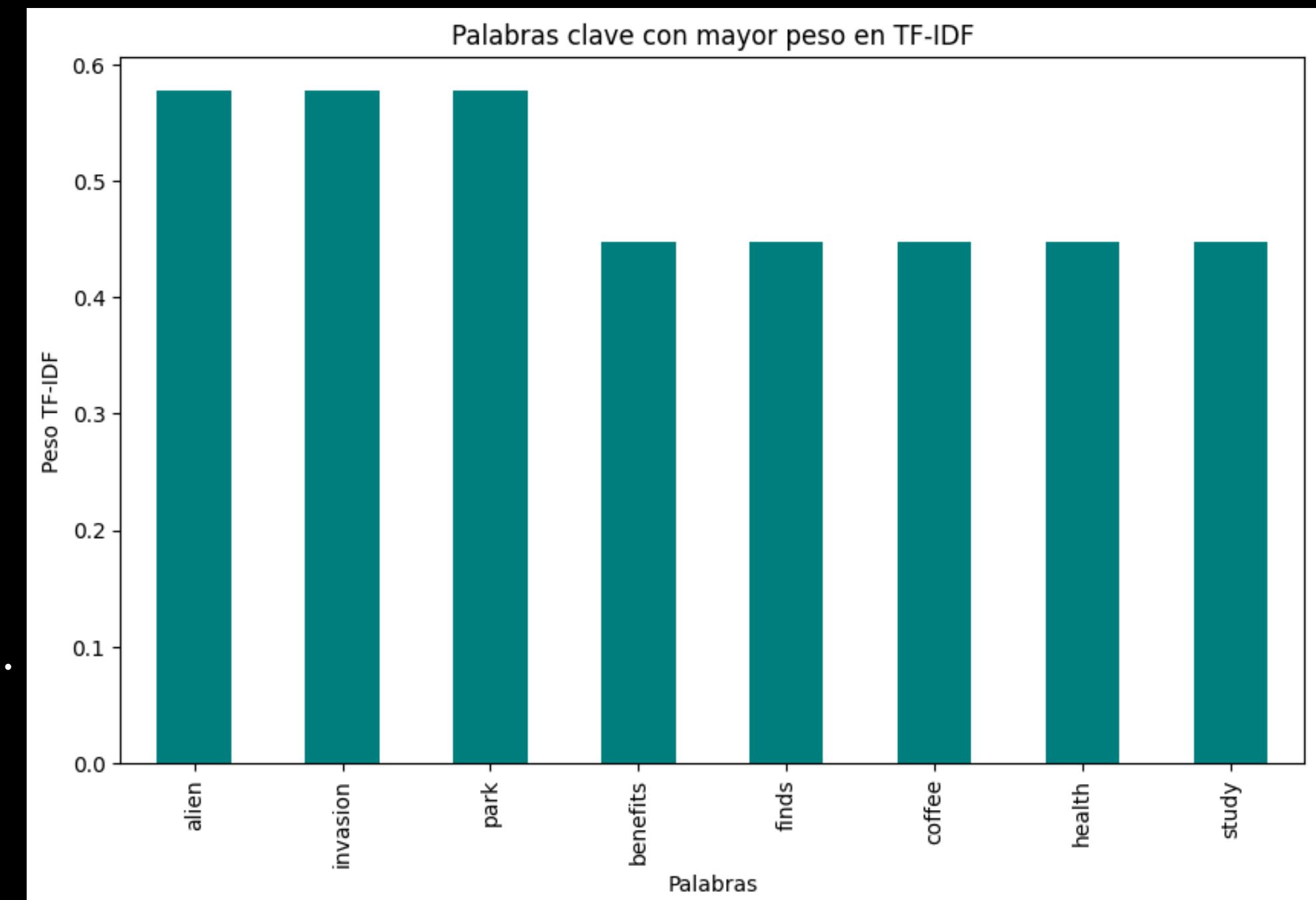
PREPROCESAMIENTO DE DATOS

- **Noticias reales: 432**
- **Noticias falsas: 432 (datos balanceados)**

[Read More](#)

VECTORIZACIÓN Y REPRESENTACIÓN

- **Técnica utilizada:** TF-IDF para transformar el texto en vectores numéricos basado en frecuencia.
- **Dimensión final:** (864,2000)

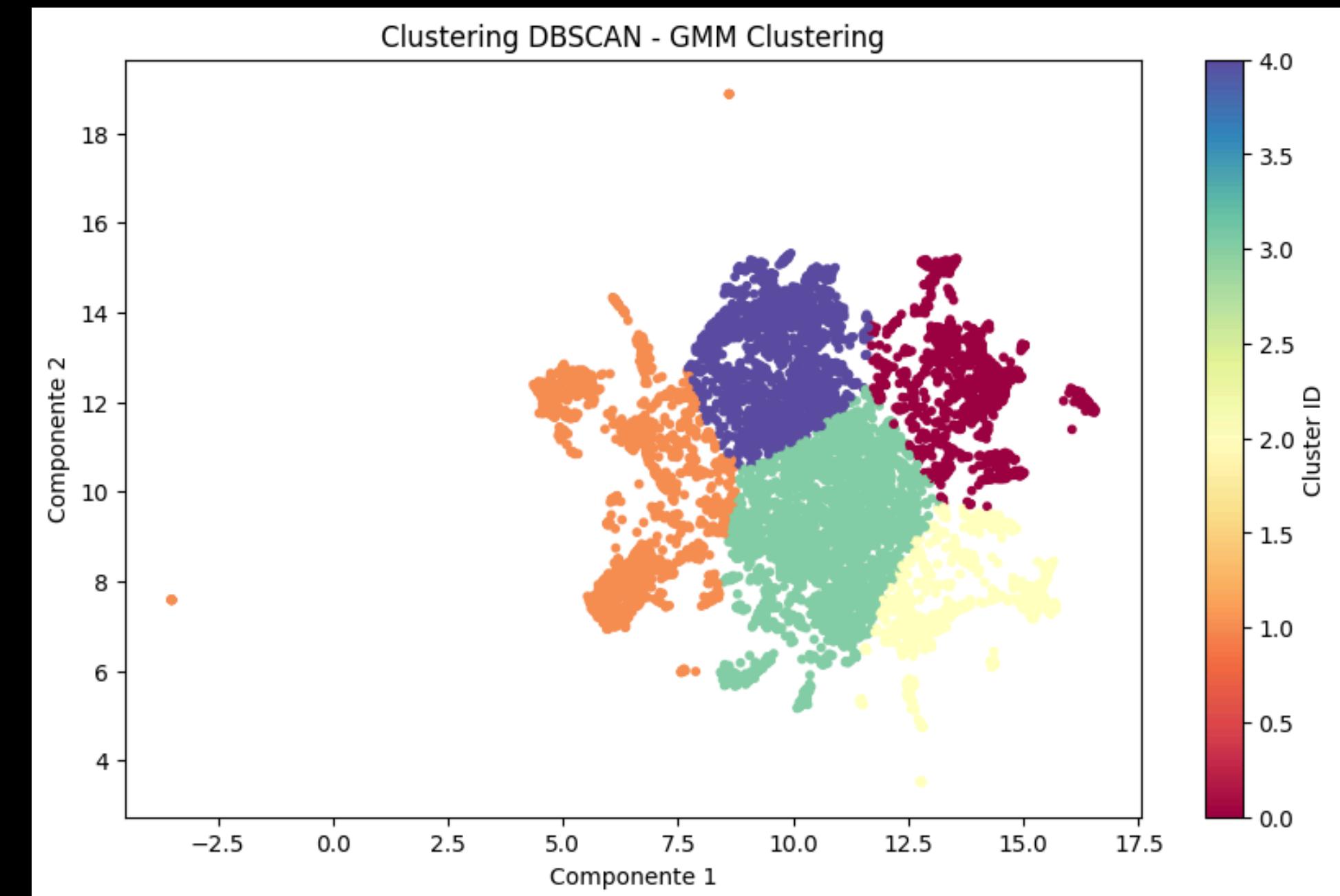


REDUCCIÓN DIMENSIONAL

- **UMAP (Uniform Manifold Approximation and Projection):** proyectar los datos en un espacio de menor dimensión (2D) para visualización y clustering.
- **Dimensiones finales: (864, 2)**

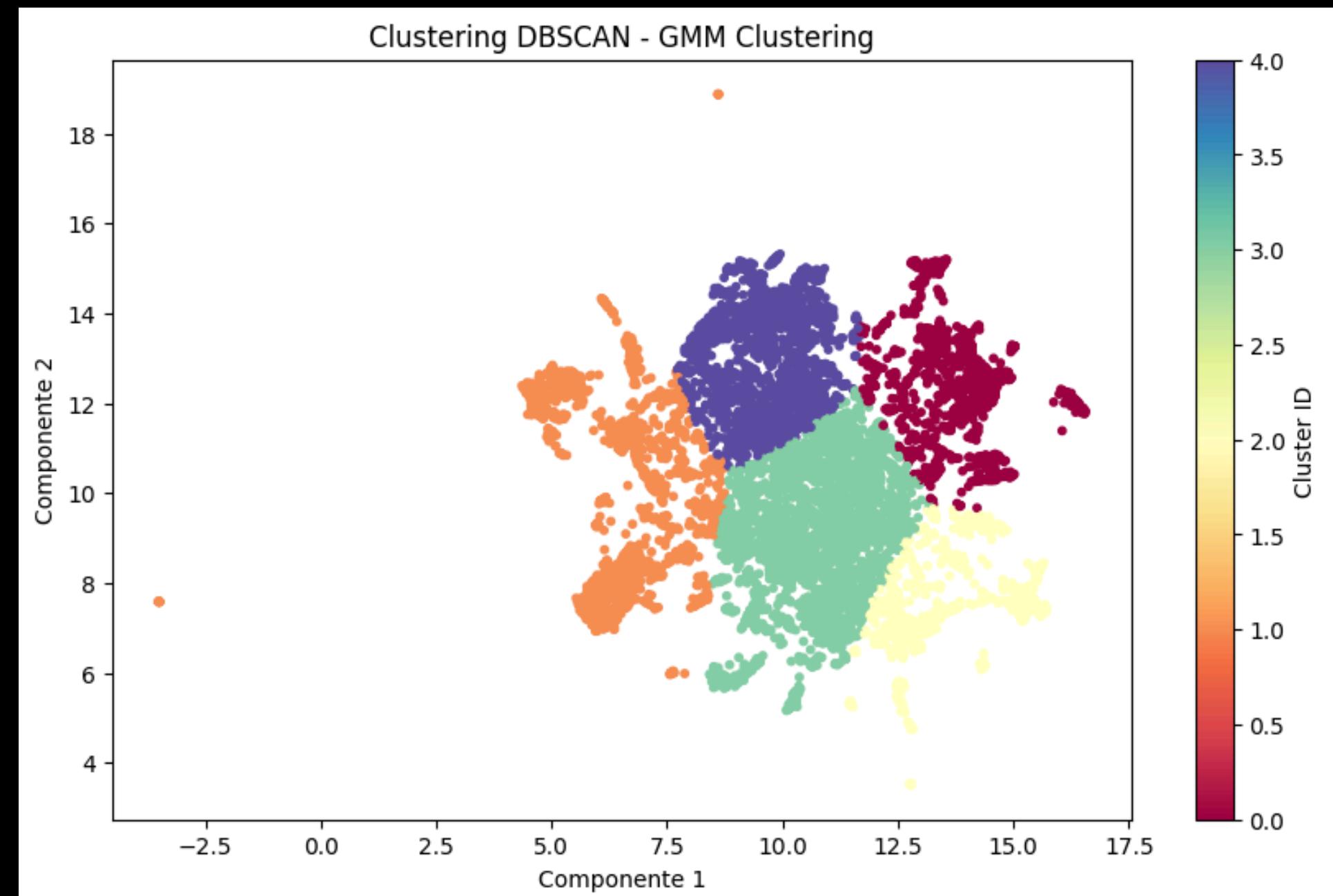
CLUSTERING CON DBSCAN

- **Parámetros utilizados:**
eps=0.1 min_samples=10
- **Dimensiones finales:**
(864, 2)



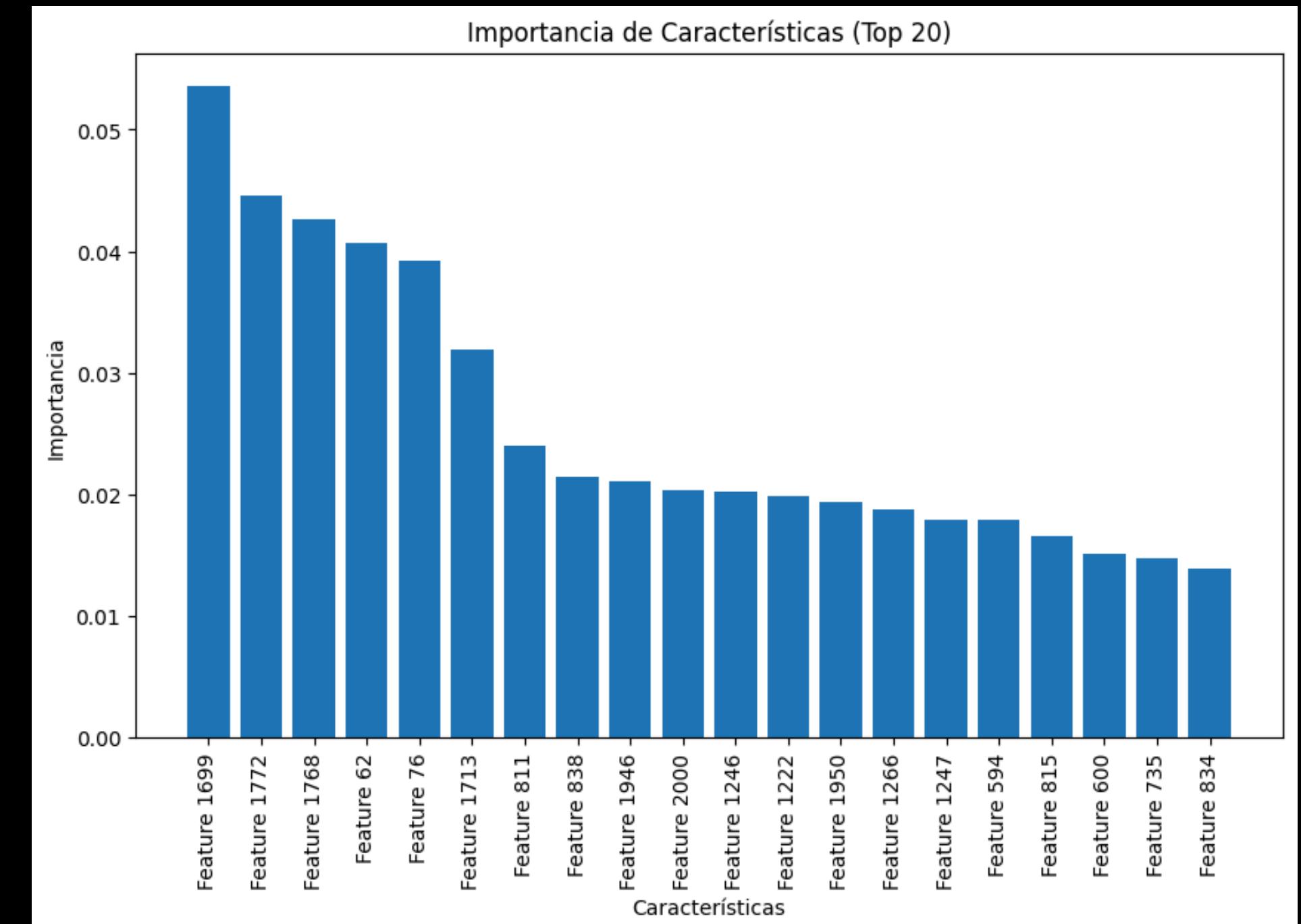
EVALUACIÓN DE CLUSTERING

- **Silhouette Score:** -0.31
(clusters débiles con solapamiento.)
- **Calinski-Harabasz Index:** 1.77 (baja separación entre clusters)



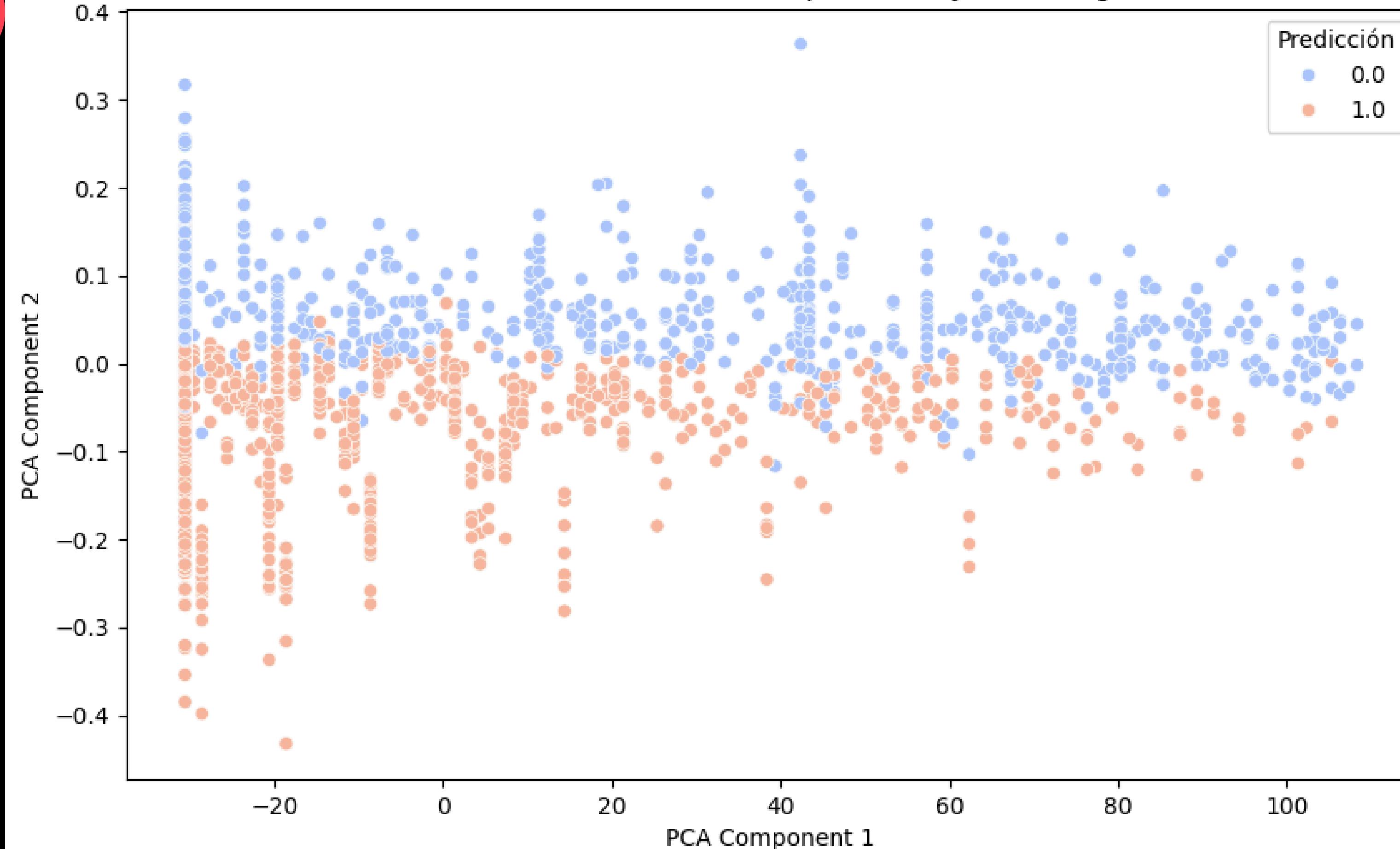
MODELO SUPERVISADO

- **Modelo aplicado:** Random Forest para clasificación
- **Configuración:** etiquetas de clustering agregadas como nueva característica.



Precisión: 98%

Distribución con modelo supervisado y clustering



¿QUÉ LOGRAMOS?

- Incorporar clustering **mejoró la capacidad del modelo** para clasificar noticias reales y falsas.
- **Limitaciones:** Clustering con DBSCAN presentó desafíos debido a la baja densidad del dataset.

TF-IDF



VECTORIZACIÓN Y NAIVE BAYES

- Se hizo uso de TF-IDF para analizar el contenido de las noticias
- Por medio del uso de Naive Bayes Multinomial se calculó la probabilidad de que la noticia fuera verdadera o falsa

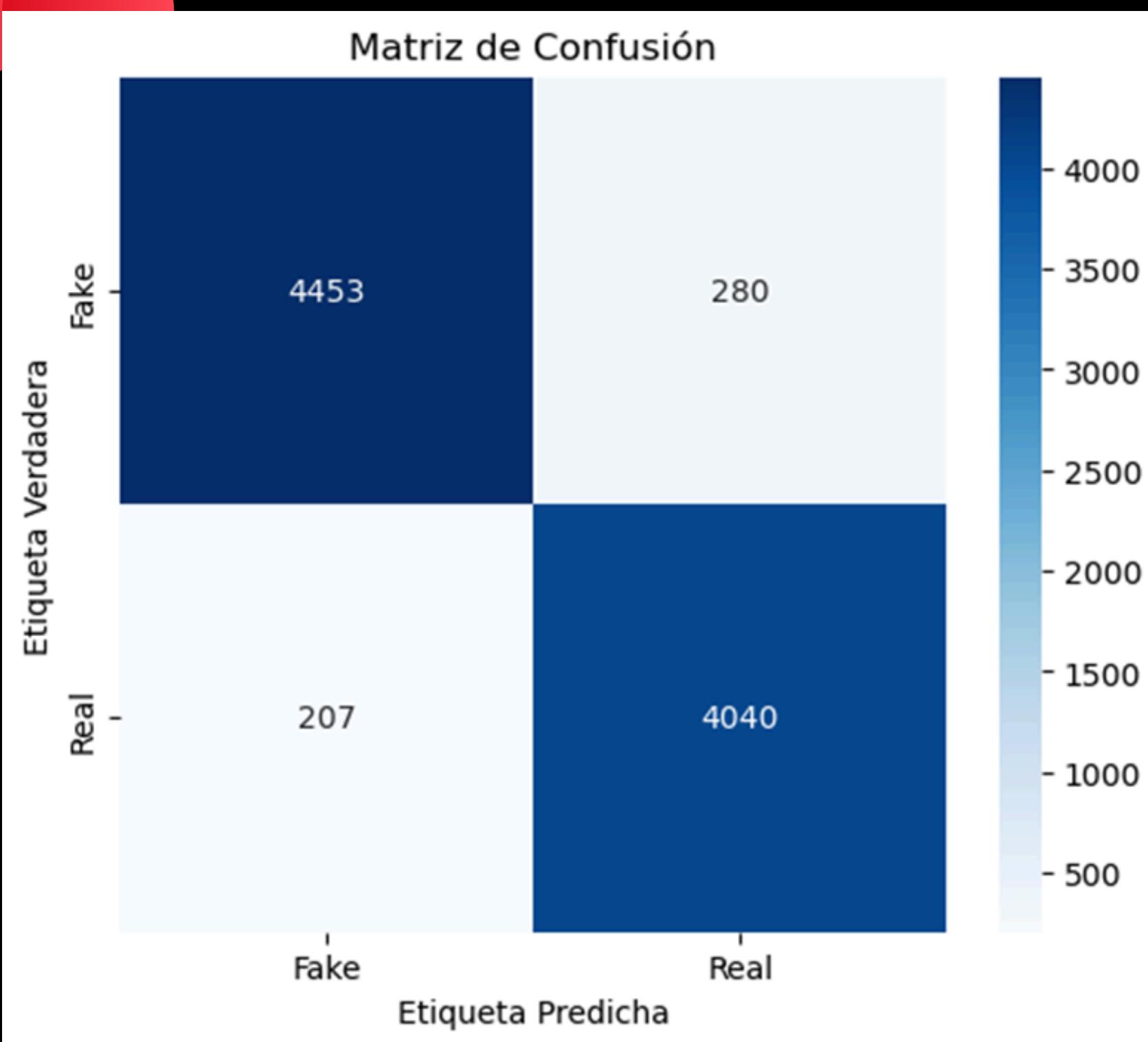
```
data = data[['text', 'label']]  
data['text'] = data['text'].astype(str).apply(clean_text)
```

```
X_train, X_test, y_train, y_test = train_test_split(data['text'],  
data['label'], test_size=0.2, random_state=42)
```

```
vectorizer = TfidfVectorizer(max_features=5000)  
X_train_vec = vectorizer.fit_transform(X_train)  
X_test_vec = vectorizer.transform(X_test)
```

```
model = MultinomialNB()  
model.fit(X_train_vec, y_train)
```

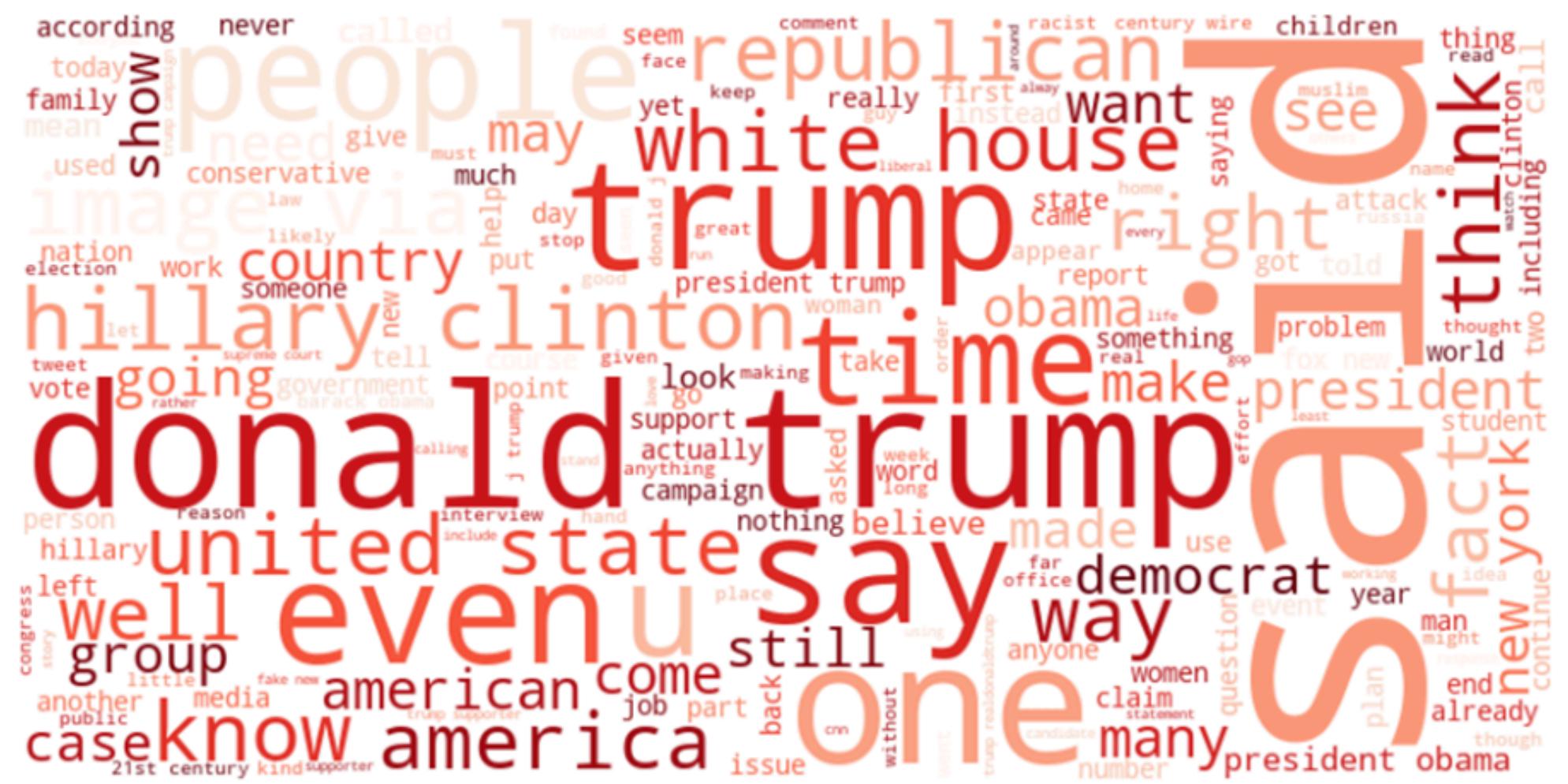
Matriz de Confusión



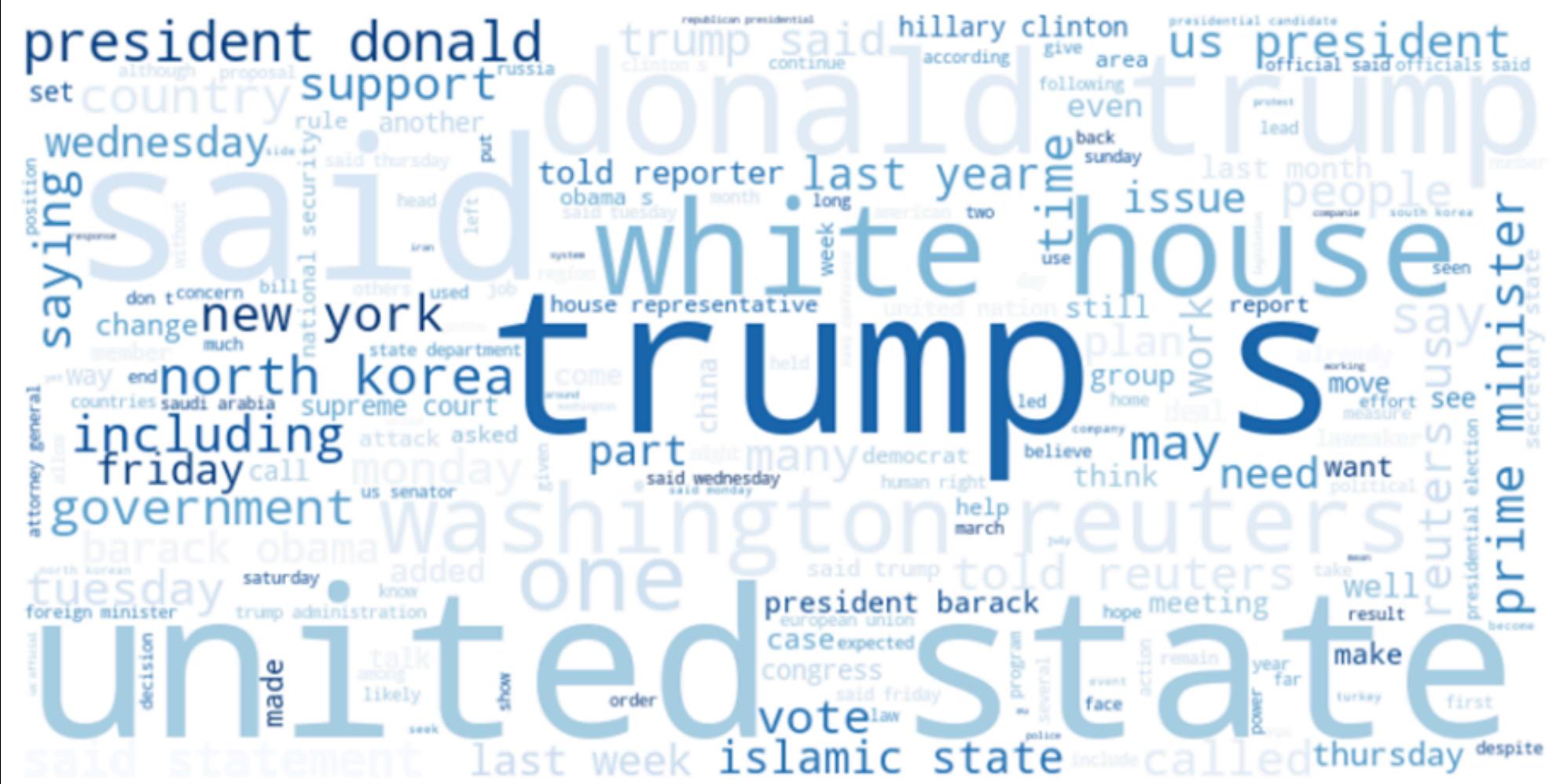
Reporte de Clasificación

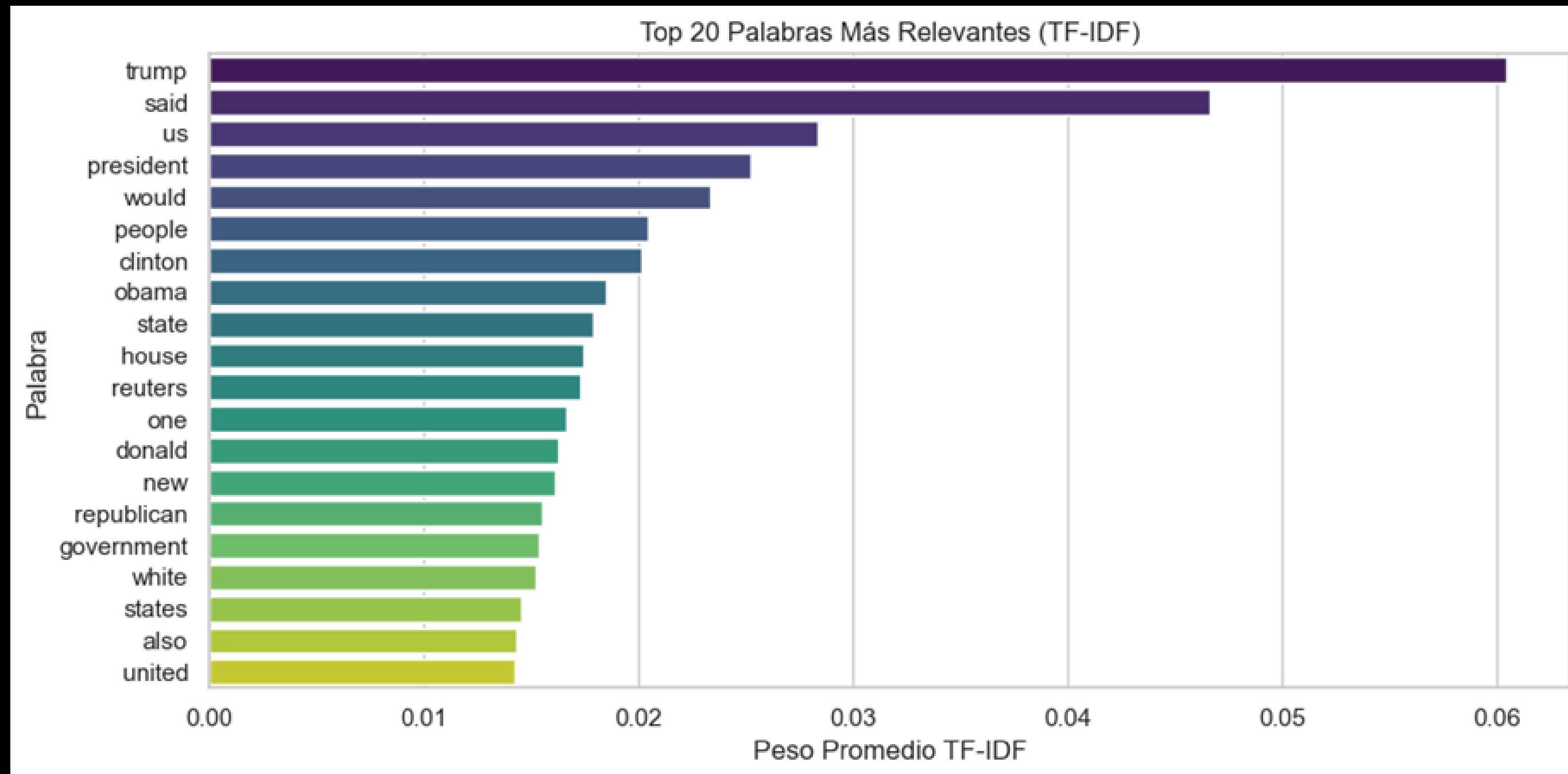
	precision	recall	f1-score	support
Fake	0.96	0.94	0.95	4733.0
Real	0.94	0.95	0.94	4247.0
accuracy			0.95	
macro avg	0.95	0.95	0.95	8980.0
weighted avg	0.95	0.95	0.95	8980.0

Nube de Palabras - Noticias Falsas



Nube de Palabras - Noticias Reales





REGRESIÓN



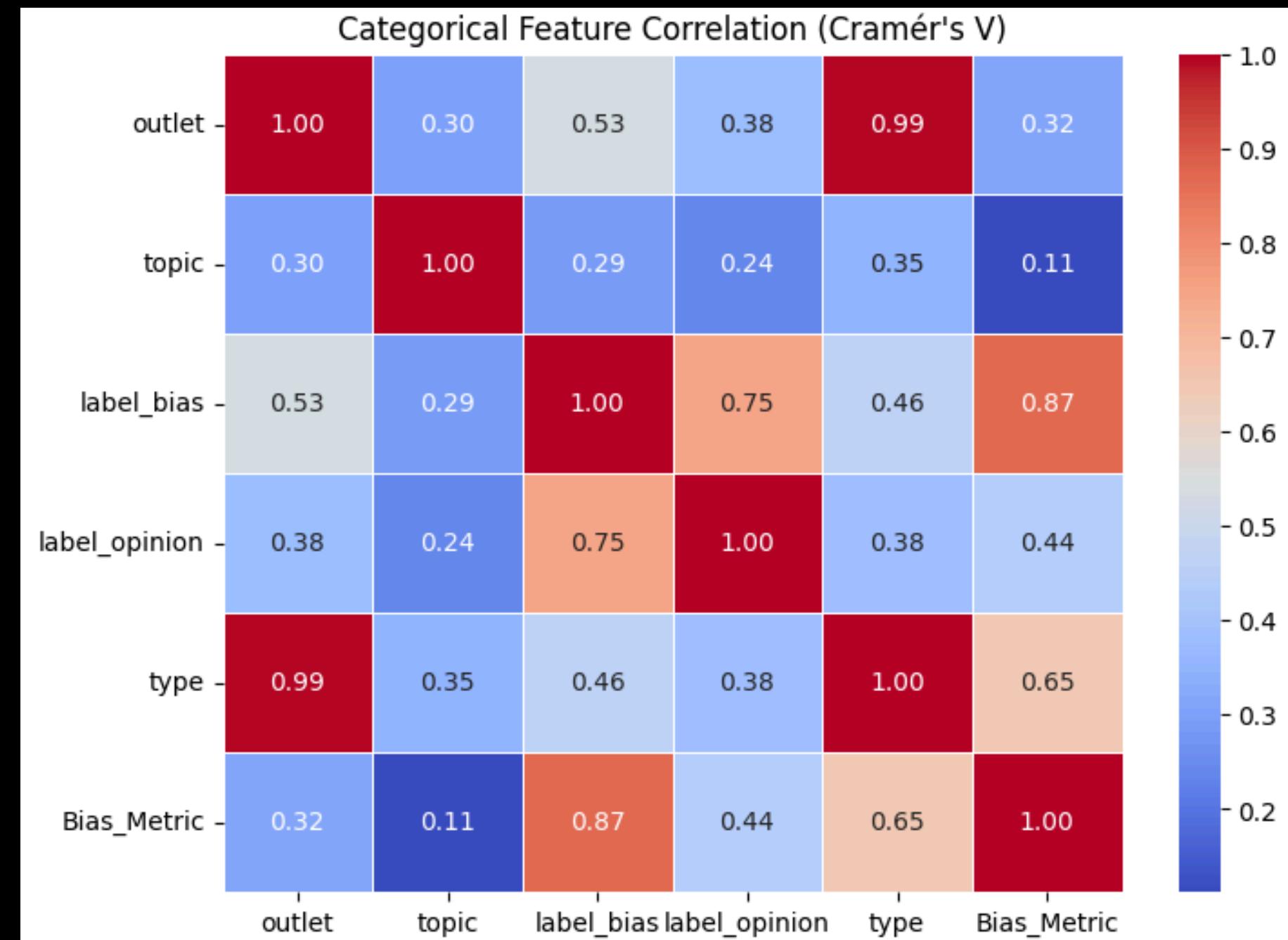
MEDICIÓN DE SESGO



Definimos el nivel de sesgo como

*bias_score = position * # of biased words * is biased*

CORRELACIONES

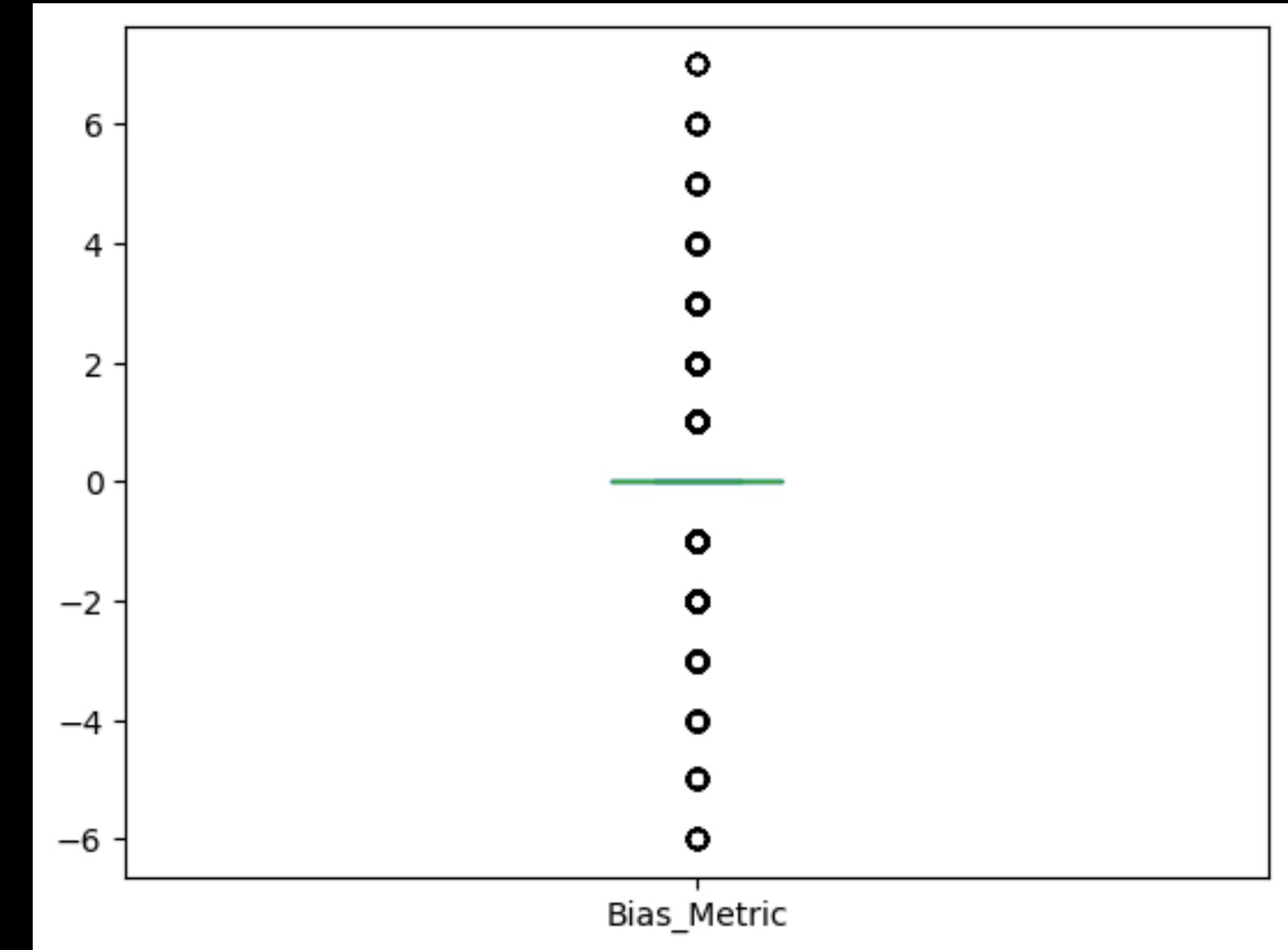
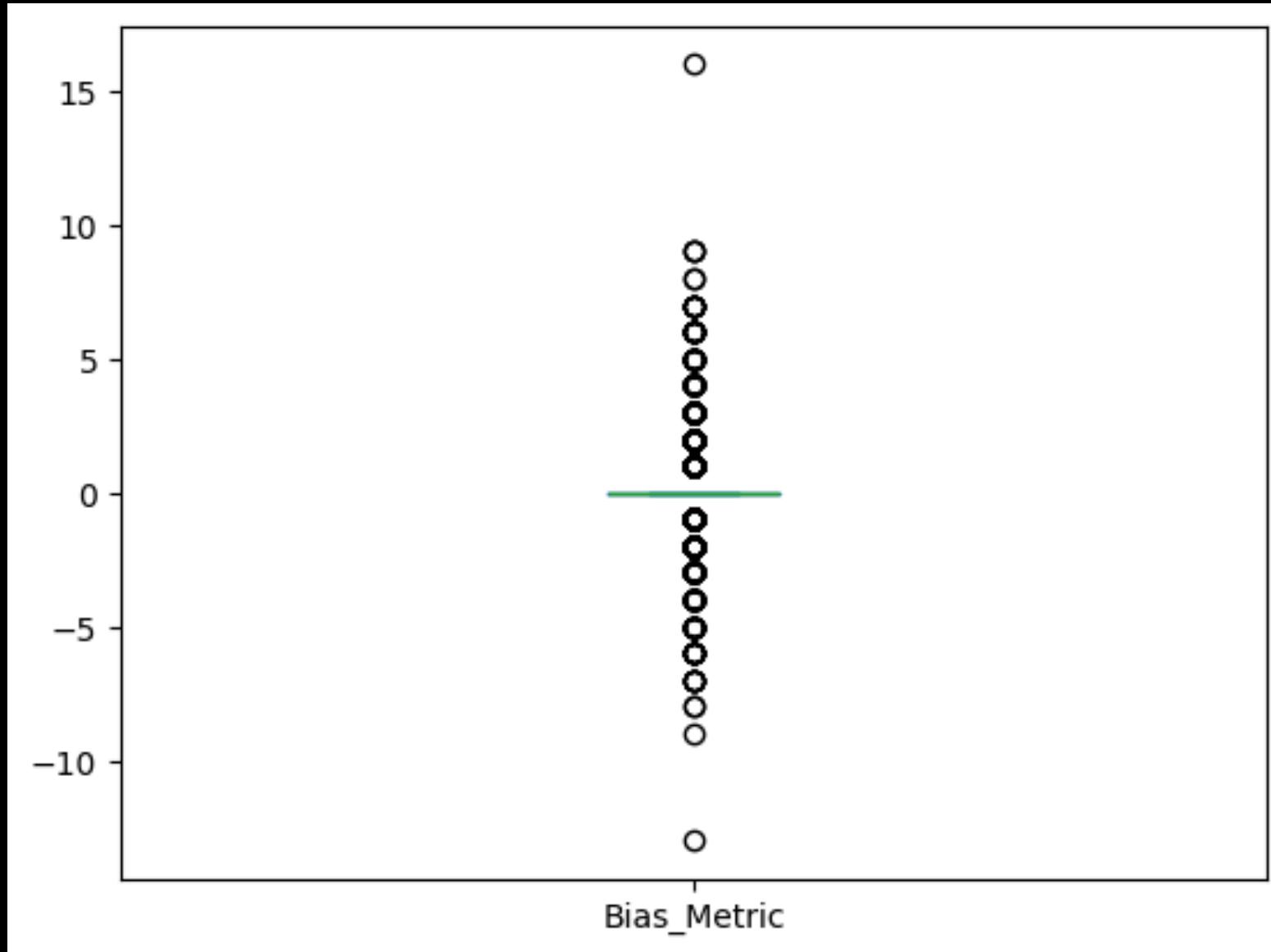


WORDCLOUD

Palabras más frecuentes en noticias

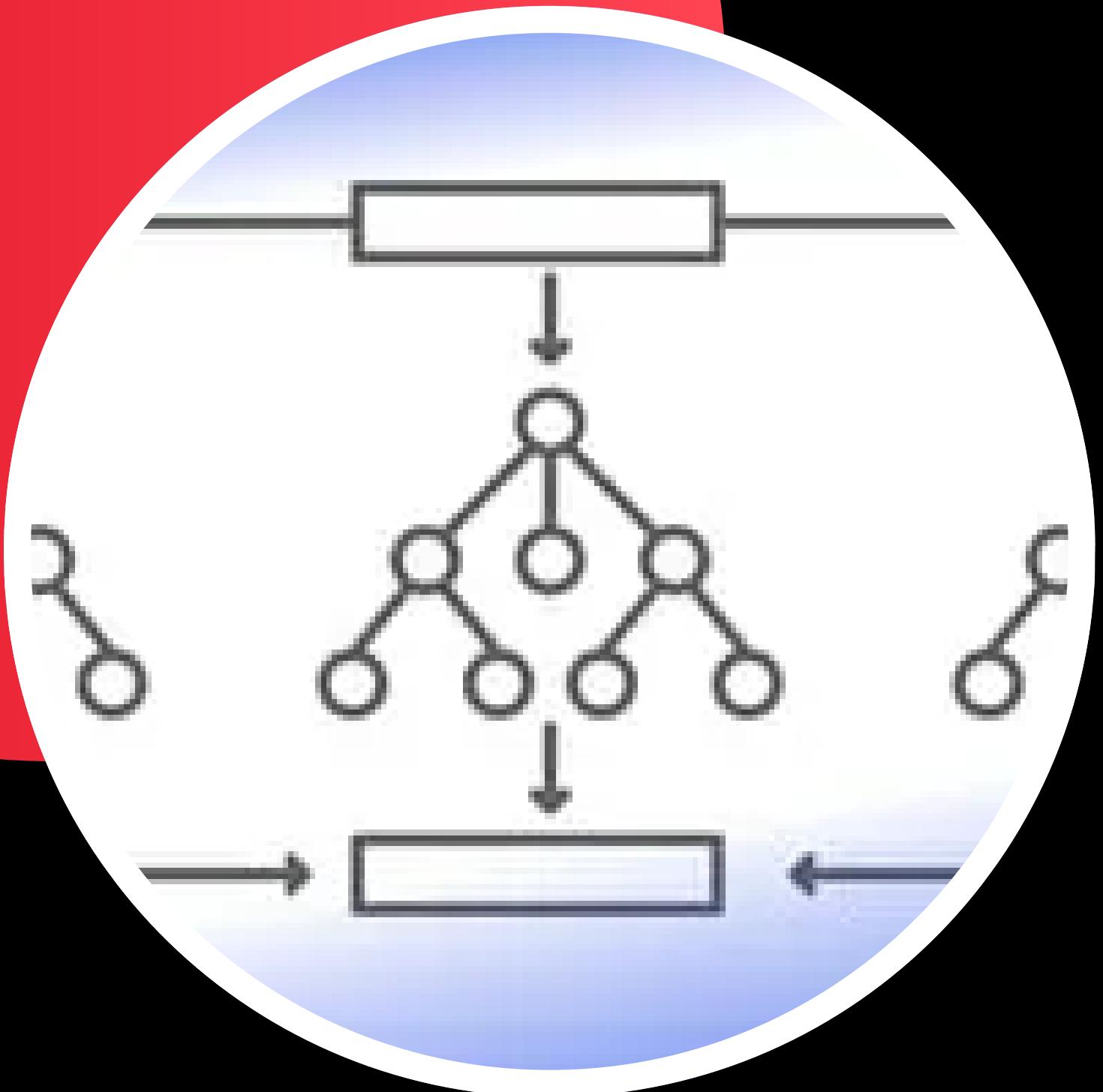


OUTLIERS



RANDOM FOREST REGRESSOR

Con el texto y otras características, determinamos que tan sesgada es la noticia.



RANDOM FOREST REGRESSOR

```
preprocessor = ColumnTransformer([
    ('text', TfidfVectorizer(max_features=5000), 'text'),
    ('cat', OneHotEncoder(handle_unknown='ignore'), ['outlet', 'topic']),
    ('num', StandardScaler(), ['word_count', 'char_count'])
])

model = Pipeline([
    ('preprocessor', preprocessor),
    ('regressor', RandomForestRegressor(n_estimators=500,
                                         random_state=42, max_depth=10, min_samples_leaf=4,
                                         min_samples_split=10))
])
```

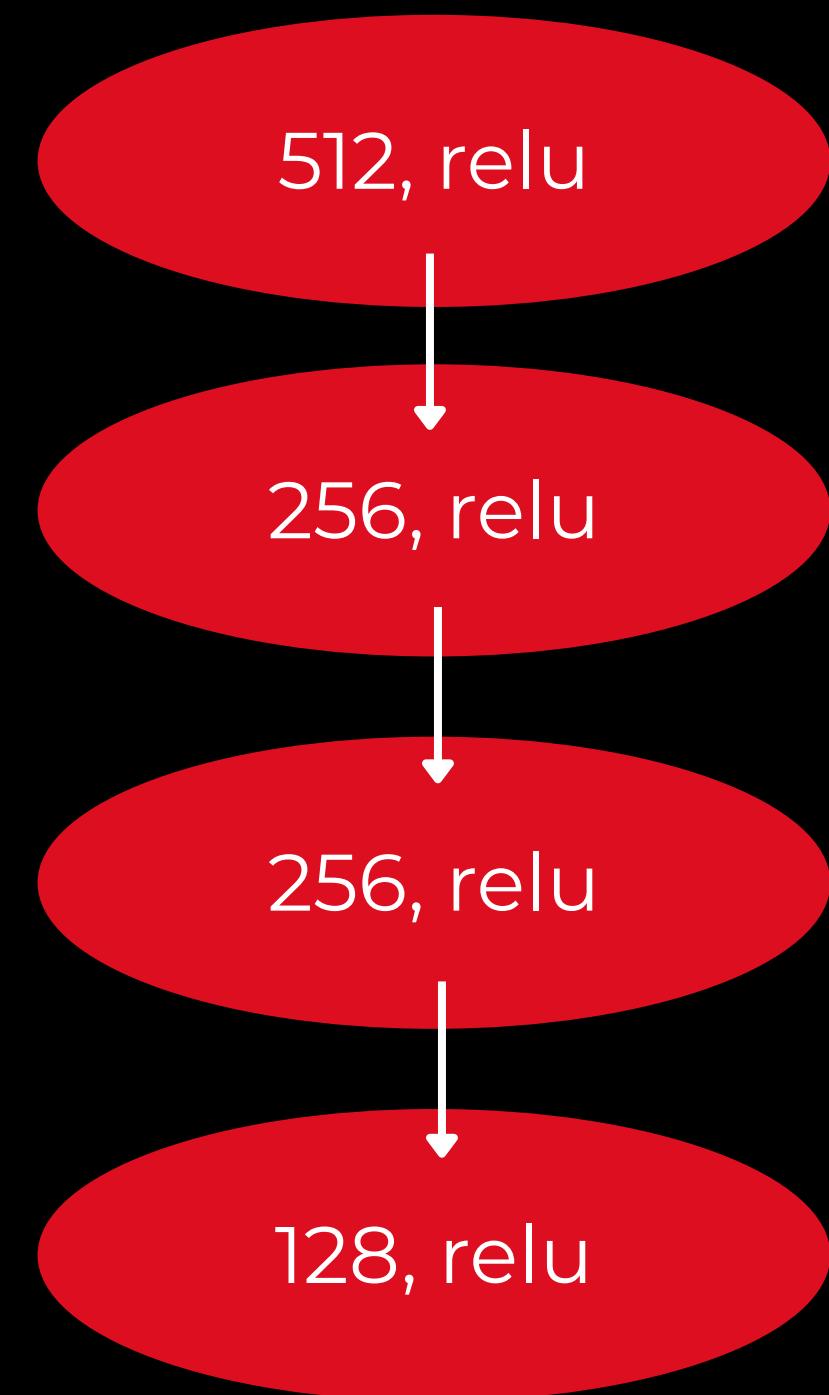
r2 score de 0.51

MAE de 0.77

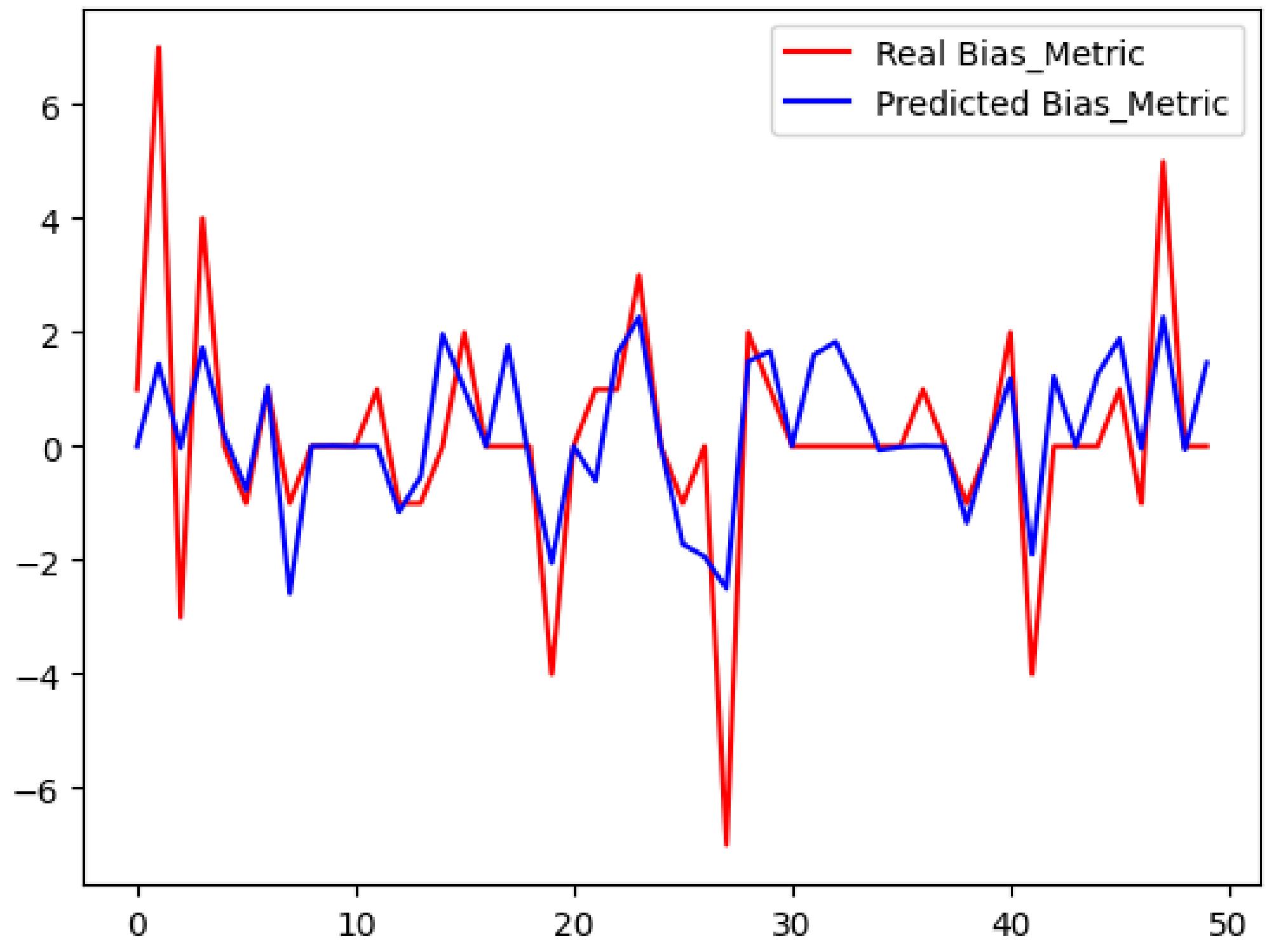
CONVOLUTIONAL NN

r2 score de 0.51

MAE de 0.73



Bias Metric Prediction



CONCLUSIONES Y MEJORAS

Los modelos logran explicar un poco de la varianza del comportamiento del sesgo segun el texto, teniendo un MAE decente. Con implementaciones como BERT, diferentes configuraciones de la red neuronal, o una definicion distinta de nivel de sesgo, las puntuaciones se pueden mejorar.

Home

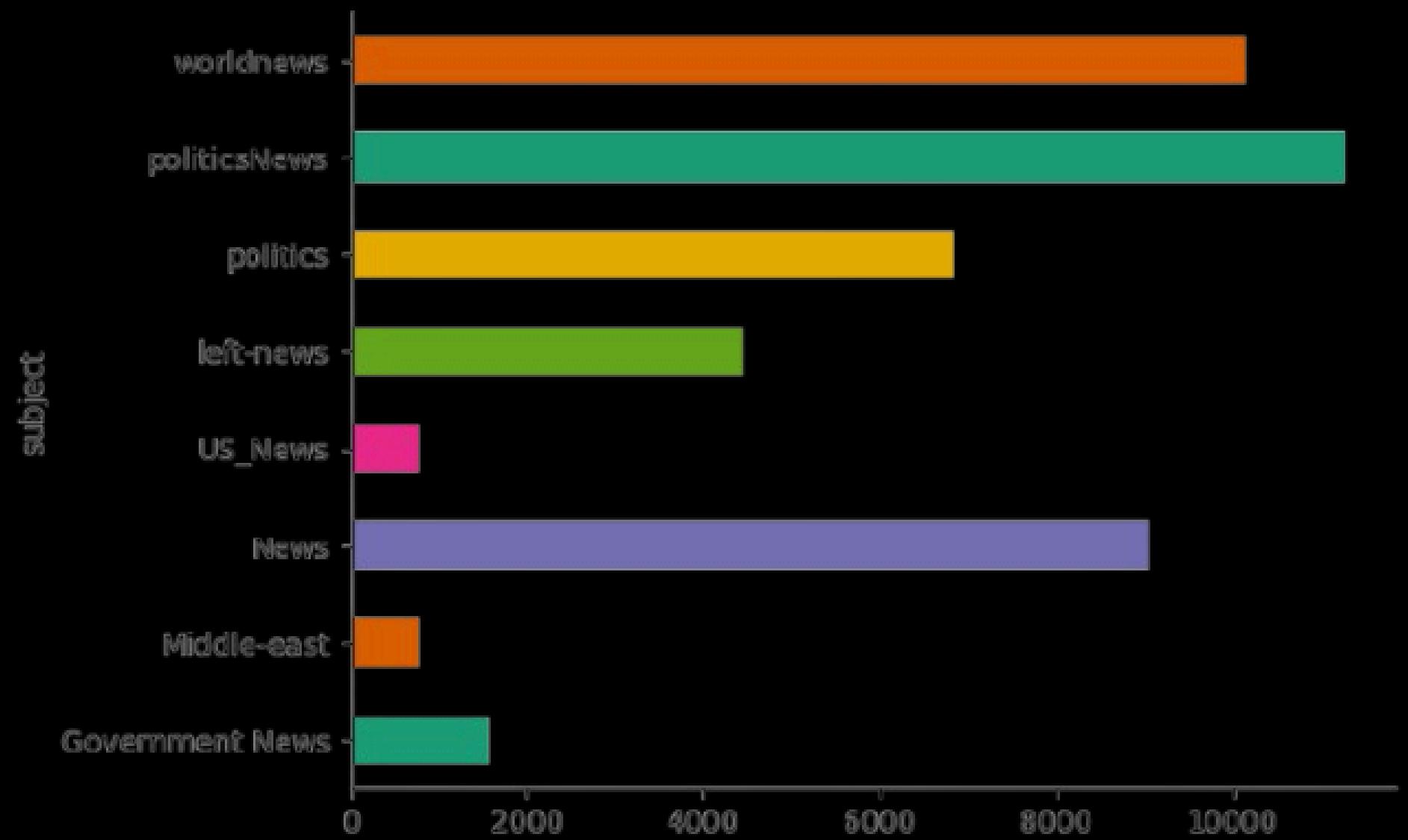
About

Contact

CLASIFICACIÓN

Análisis de sentimientos para evaluar la veracidad de las noticias

44,689 noticias



$$\text{IDF}(\text{palabra}) = \log \left(\frac{\text{Número total de documentos en el corpus lingüístico}}{\text{Cantidad de documentos con la palabra}} \right)$$

TF-IDF

F1-SCORE

$$2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

**Regresión
Logística**



0.9785 ± 0.0010

0.9851 ± 0.0021



**Máquina
de Soporte
Vectorial**

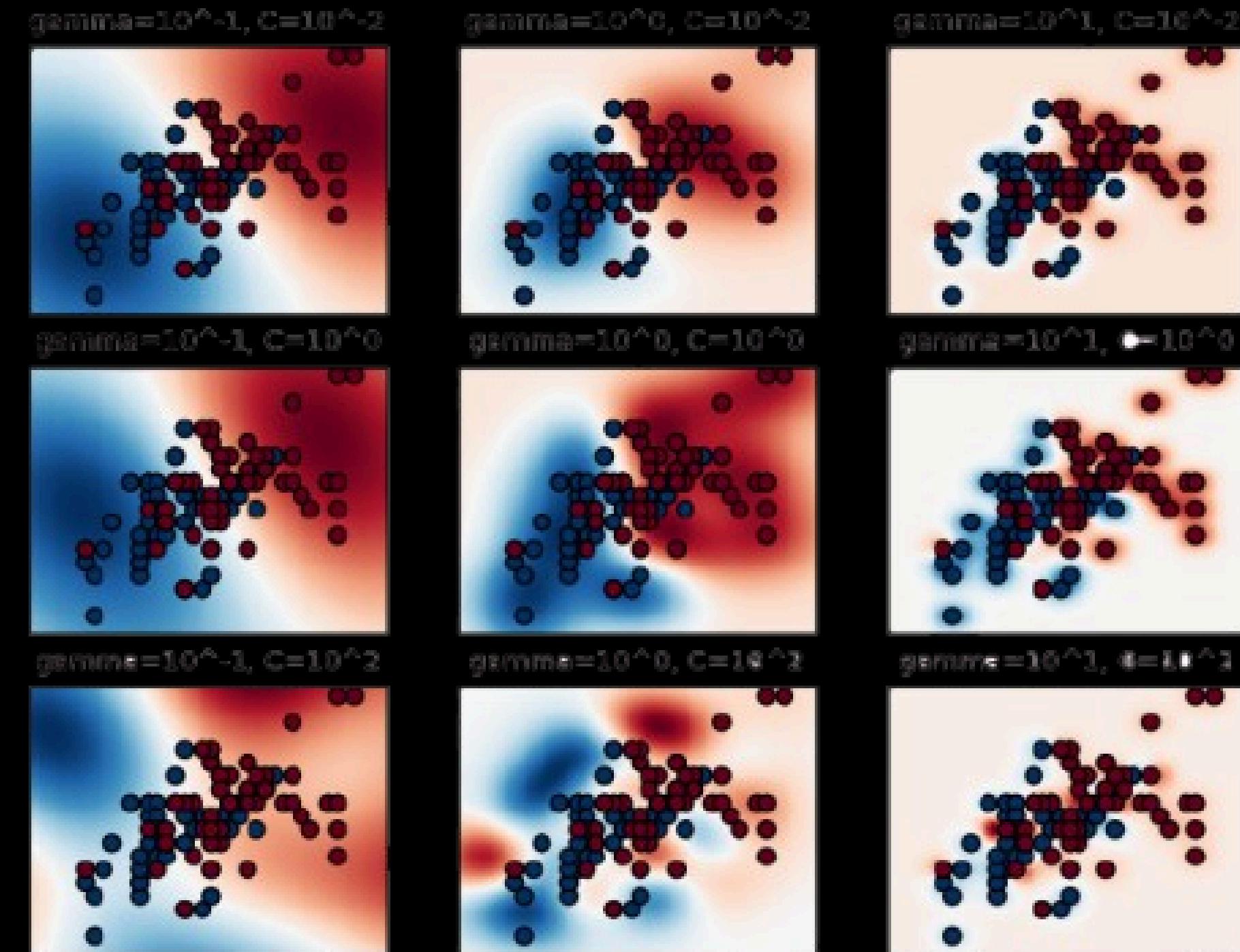
**Bosque
Aleatorio**



0.9782 ± 0.0026

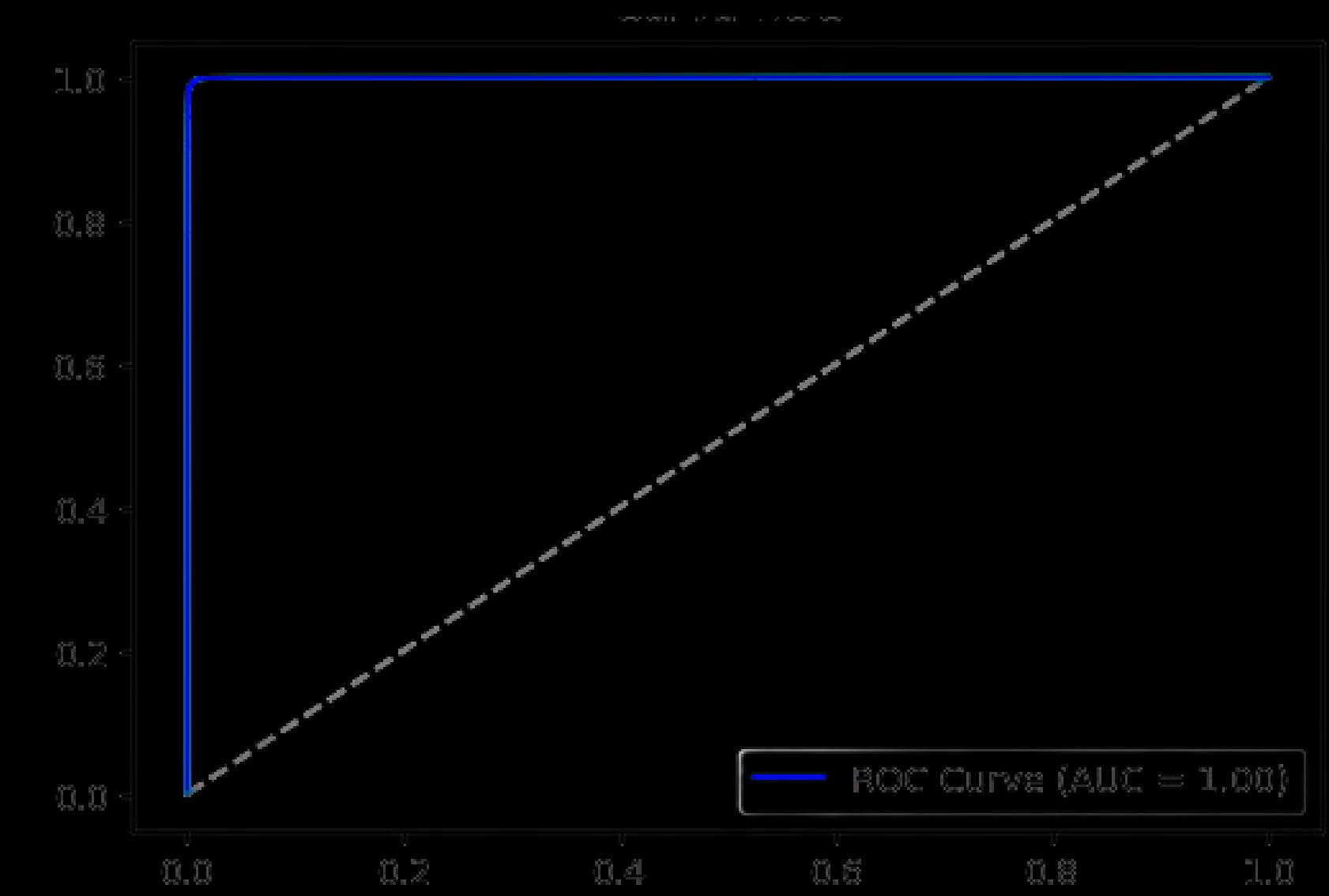
GRID SEARCH CV

- C: 1
- Kernel: Linear



scikit-learn. RBF SVM parameters

Exactitud: **0.9942**



NEWS

Trump tariffs trigger steepest US stocks drop since 2020 as China, EU vow to hit back

Nike and Apple were among brands worst hit, but Trump maintained the US economy would ultimately "boom".

31 mins ago | US & Canada



'So crazy' or a 'necessary evil'? - Americans react to Trump's tariffs

Five Americans share their view on the sweeping import taxes the president has

3 hrs ago | US & Canada



Denmark and Greenland show united front against US 'annexation' threats

The Danish Prime Minister Mette Frederiksen says the US "cannot annex other countries"

1 hr ago | World



Three National Security Council officials fired by Trump

The firings come after a meeting between far-right activist Laura Loomer and President Donald Trump at the White House.

37 mins ago | US & Canada



Israeli strike on Gaza City school kills 27, health ministry says

Palestinian authorities say children were among the dead, while Israel says it hit a Hamas command-and-control centre.

ago | Middle East



Influencers 'new' threat to uncontacted tribes, warns group after US tourist arrest

The man allegedly landed on North Sentinel Island and filmed his visit.

4 hrs ago | Asia

► Indonesia volcano eruption creates huge column of ash

Mount Marapi erupted on Thursday, sending a column of ash towering into the sky.

7 hrs ago | Asia

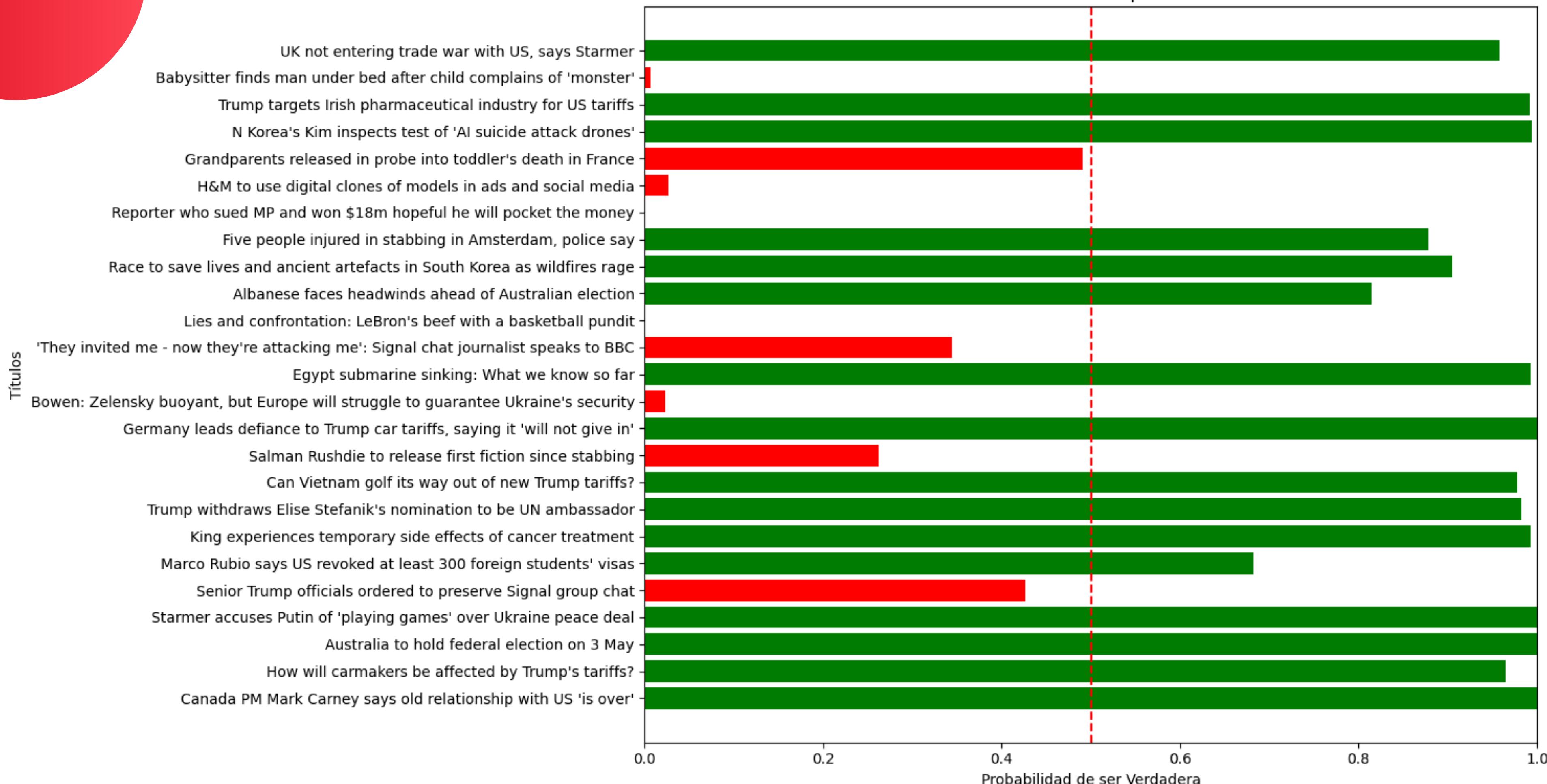
Deadly storms kill at least four across US South

A series of storms that brought high winds and flash floods is expected to last for days.

42 mins ago | US & Canada

31 Contribuyentes & 41 Etiquetas

Probabilidades Predichas para Cada Título

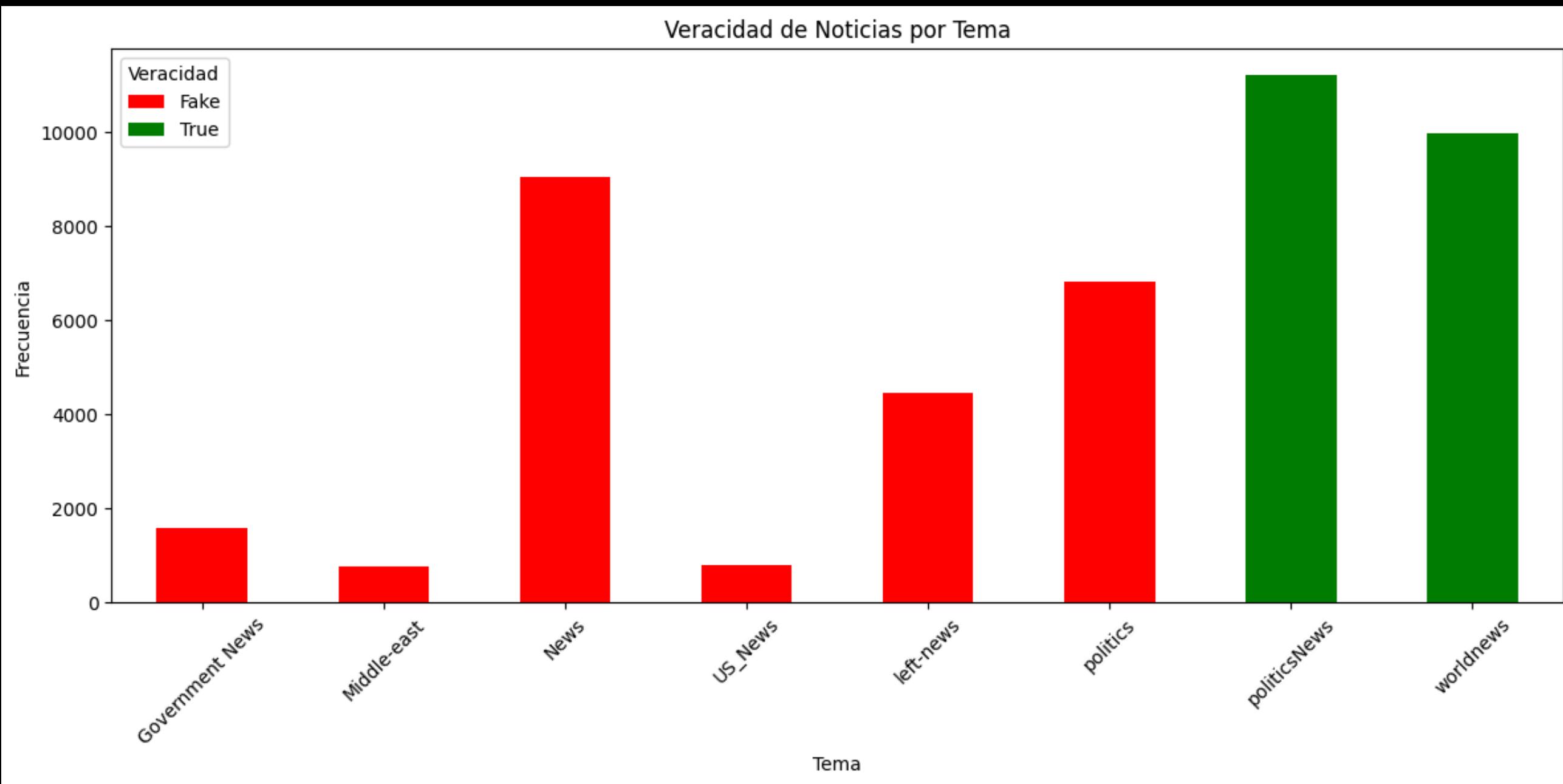




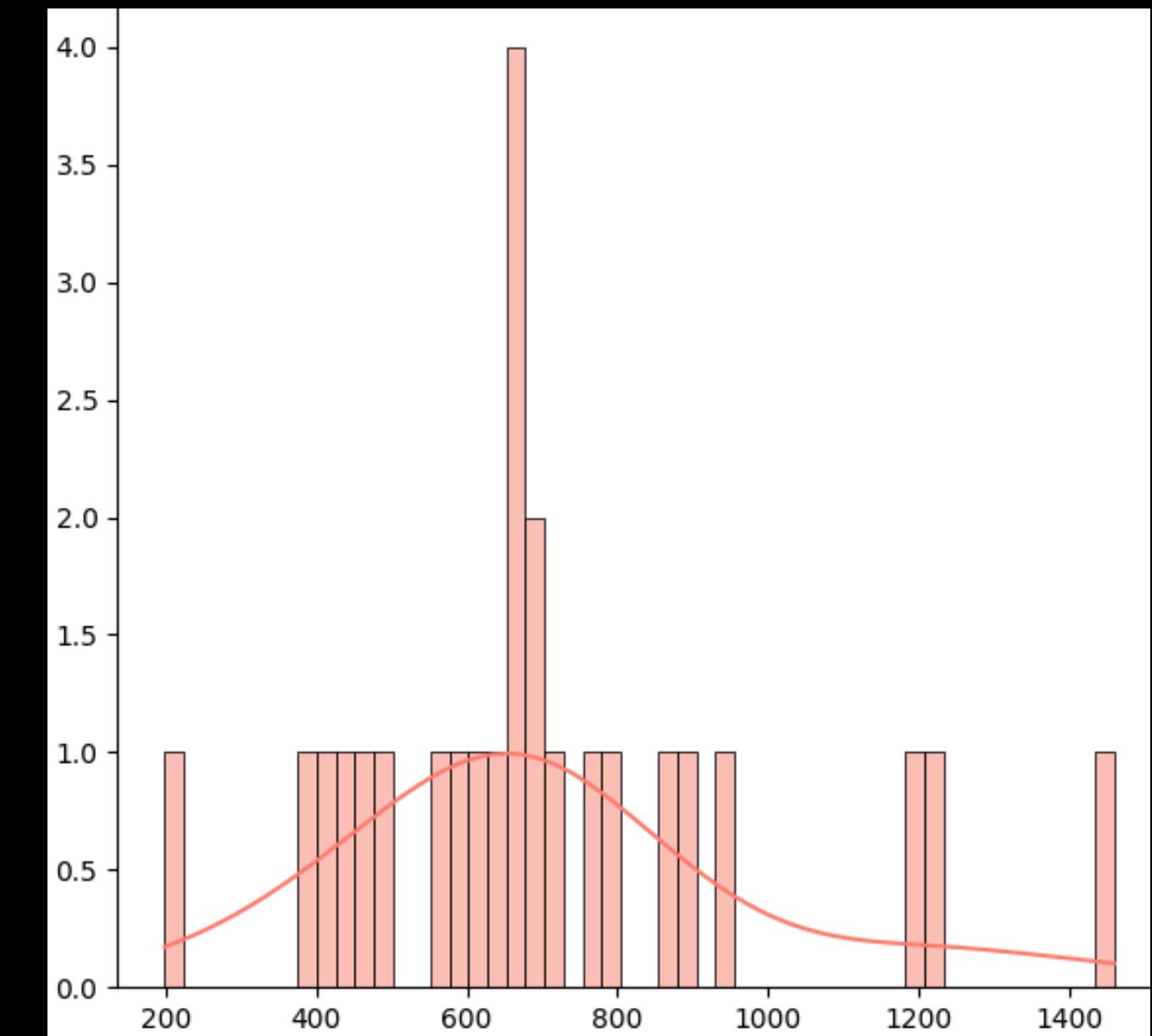
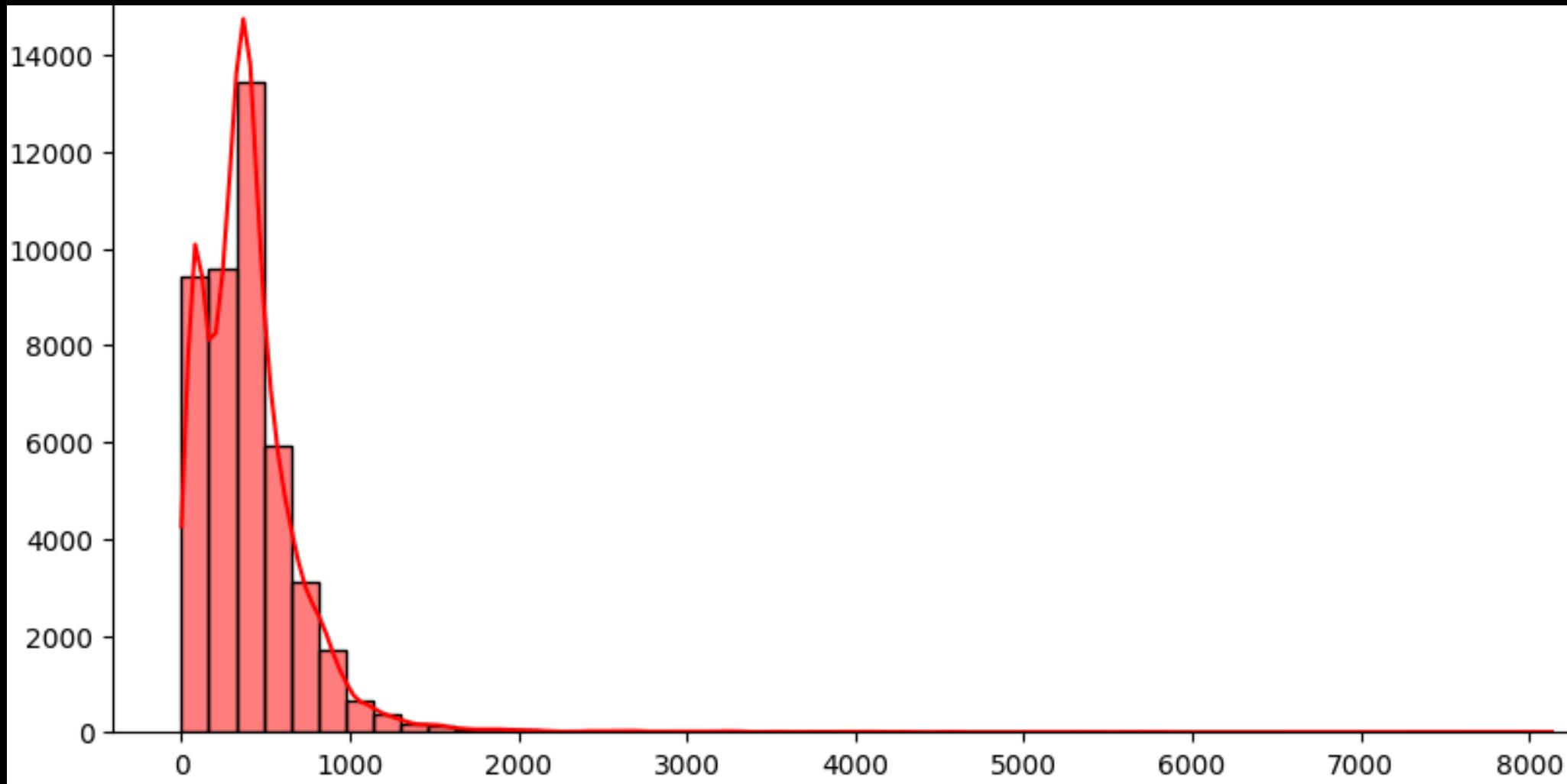
Lies and confrontation: LeBron's beef with a basketball pundit



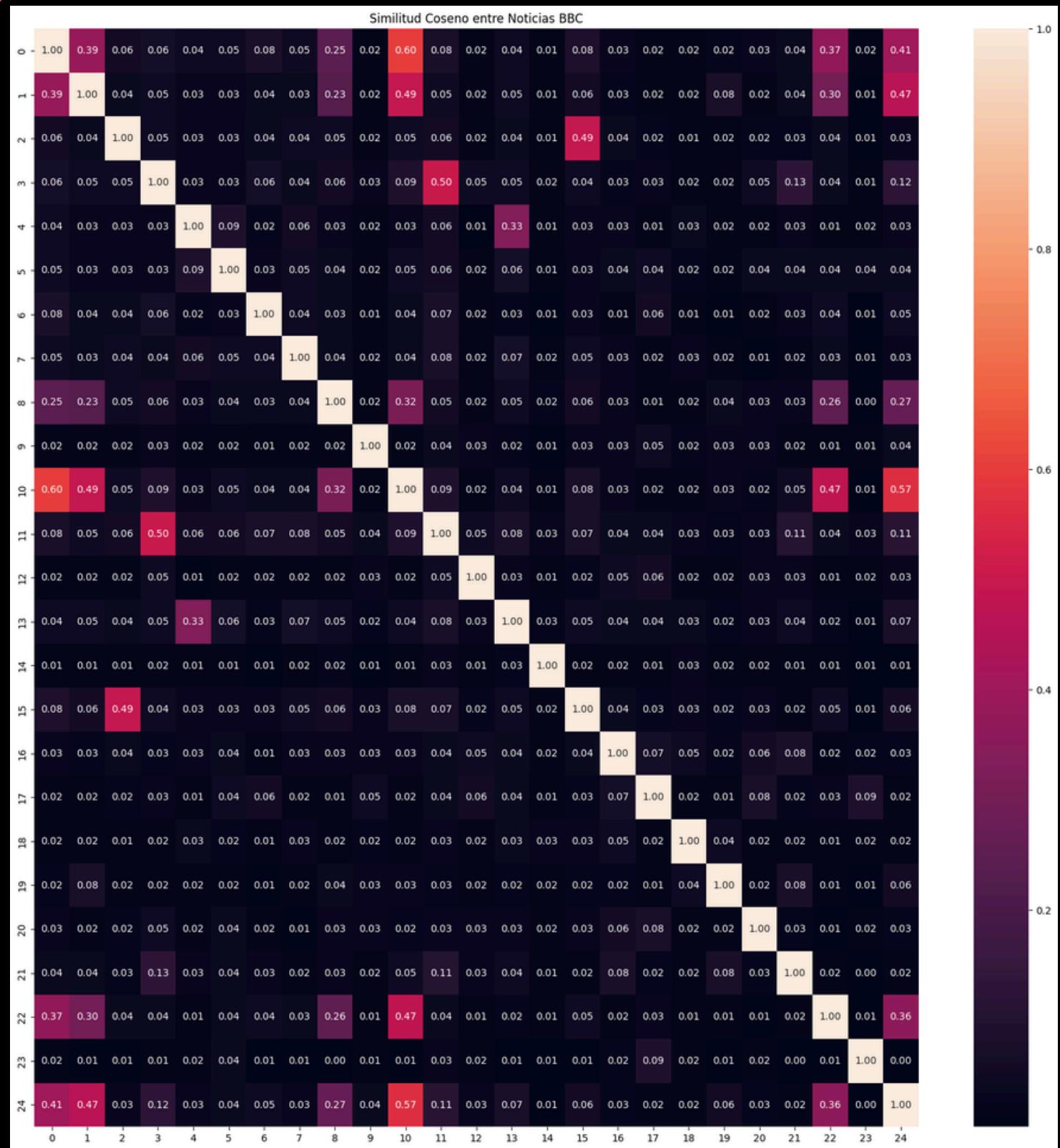
H&M to use digital clones of models in ads and social media



Longitudes de Texto



Babysitter finds man under bed after child complains of 'monster'



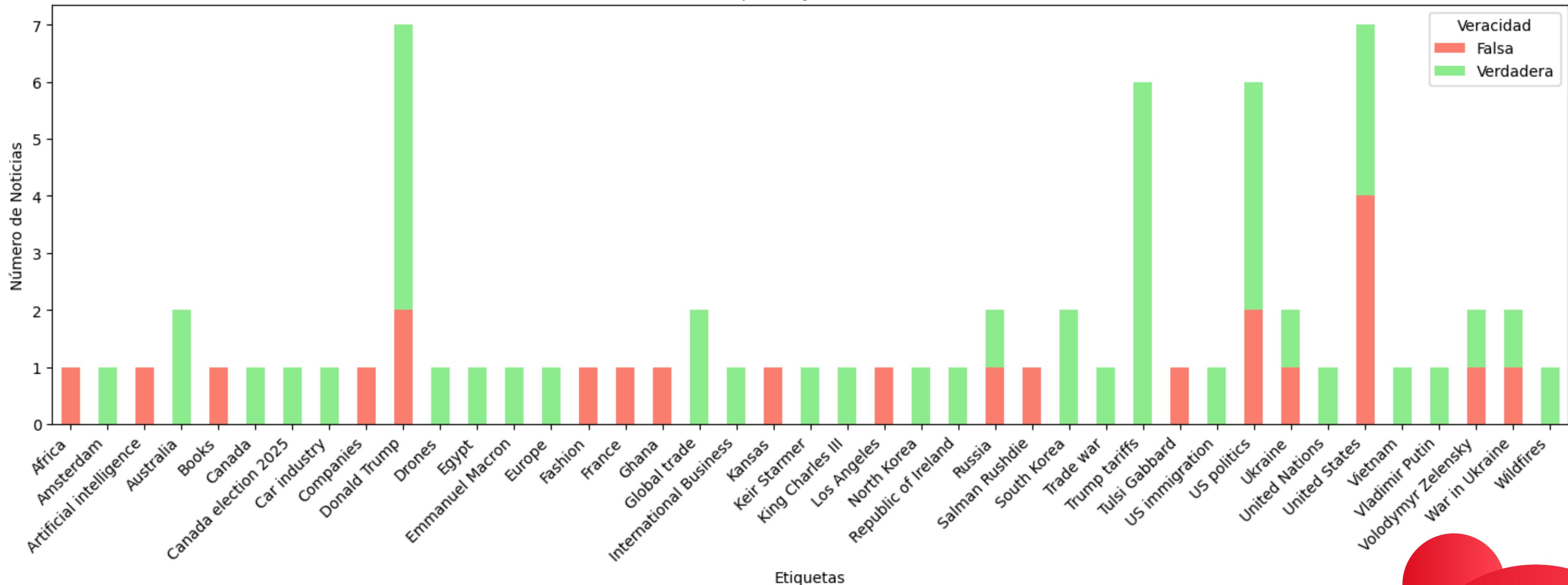
Similitud Coseno

Bowen: Zelensky buoyant, but Europe will struggle to guarantee Ukraine's security

Starmer accuses Putin of 'playing games' over Ukraine peace deal

Exactitud: 0.64

Relación entre Etiquetas y Veracidad de las Noticias



BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS



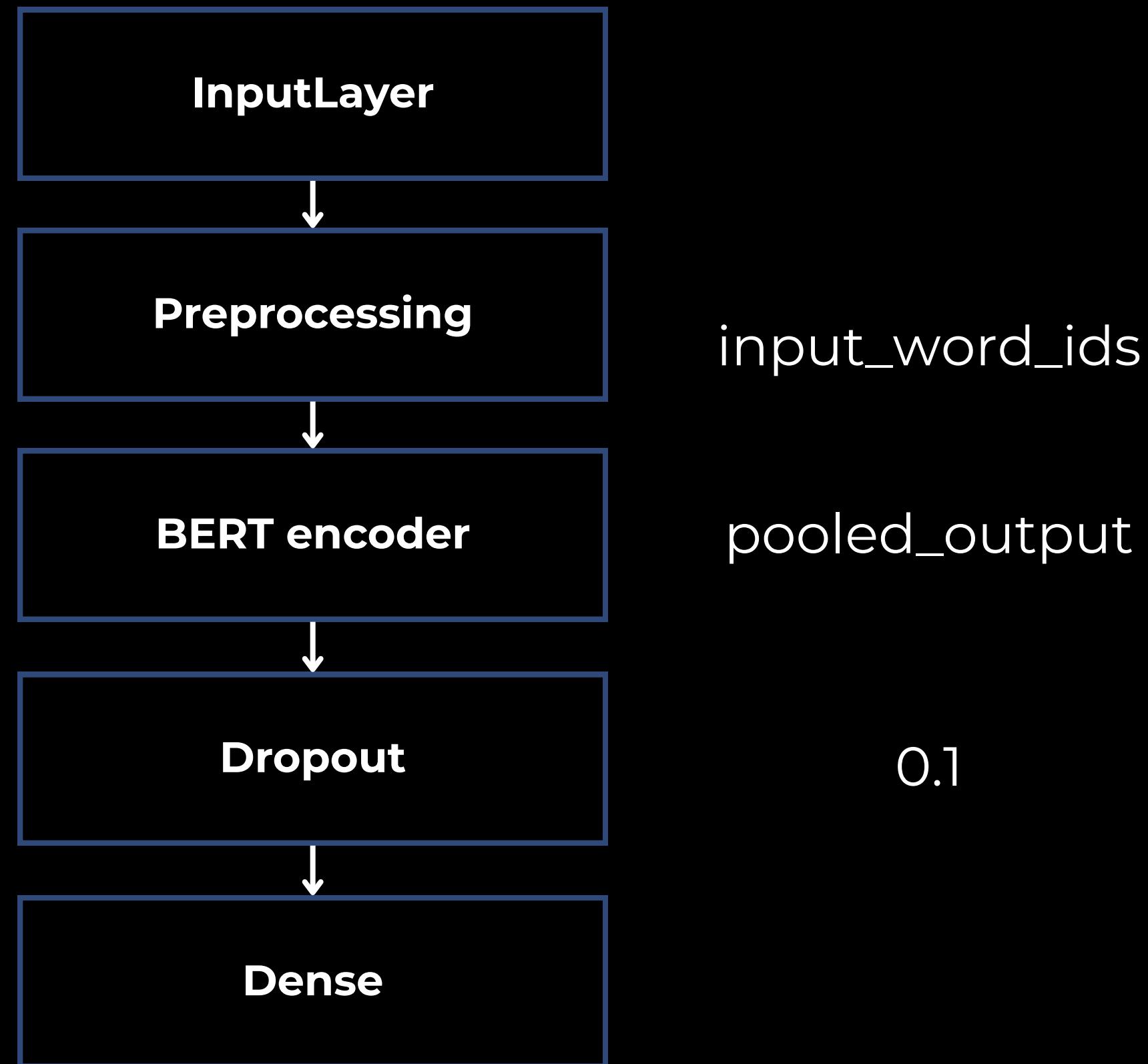
El pájaro tiene un **pico** afilado

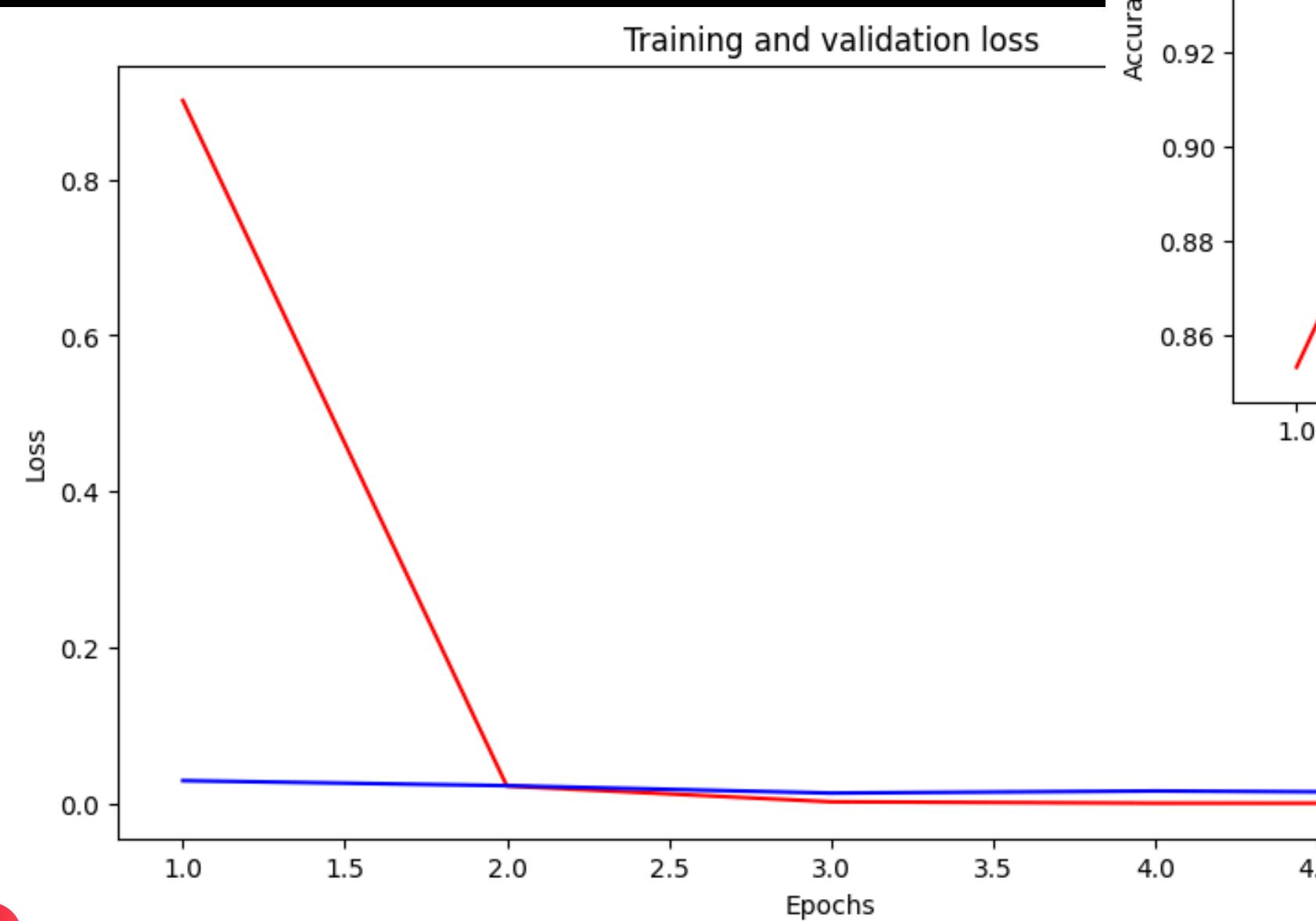
El **pico** de la montaña estaba cubierto de nieve



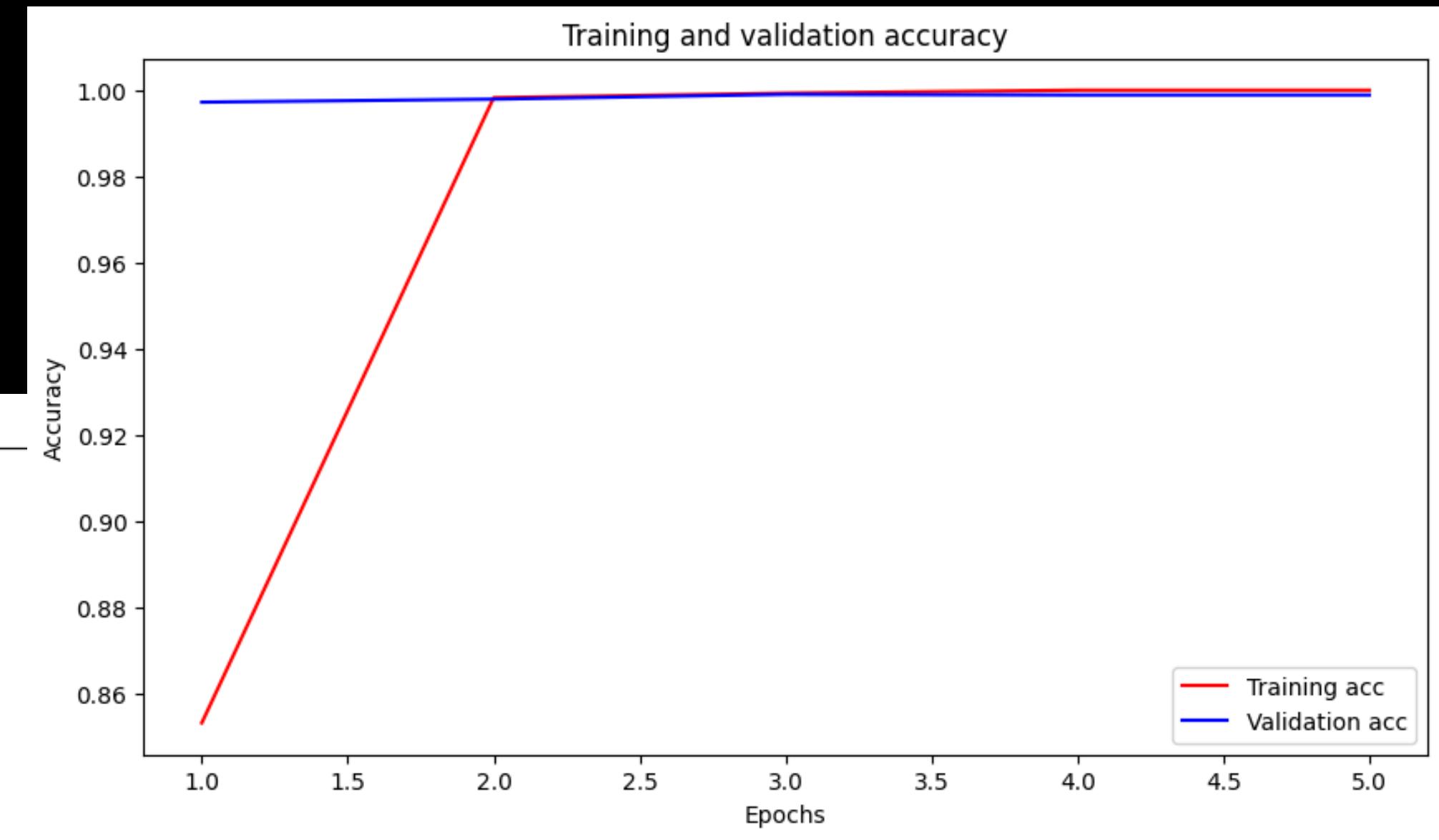
SMALL BERT

- L-4
- H-512
- A-8

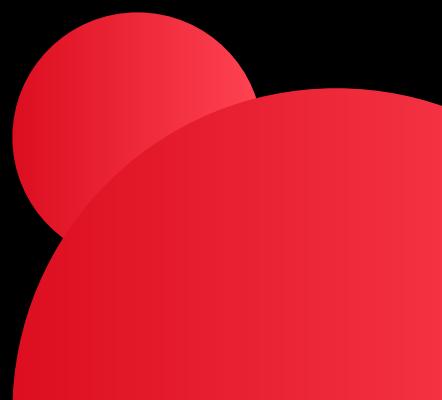
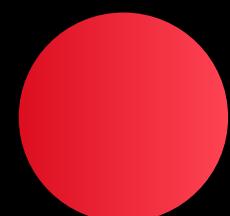
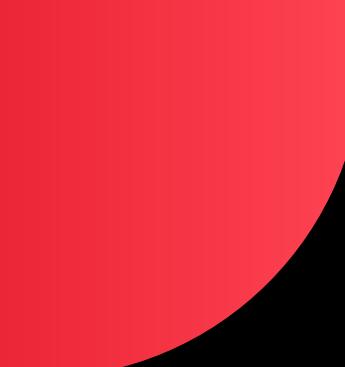




Exactitud: **0.9982**



Pérdida: **0.0197**



4745

9

7

4217

CONCLUSIONES

- Logramos obtener una estimación de sesgo en noticias con MAE de 0.73, y se puede mejorar en cuanto al R2.
- SVC destacó en noticias políticas, pero falló al generalizar. BERT, en cambio, ofrece mayor robustez
- Explorar nuevos datasets para clusterización y modelos más avanzados.

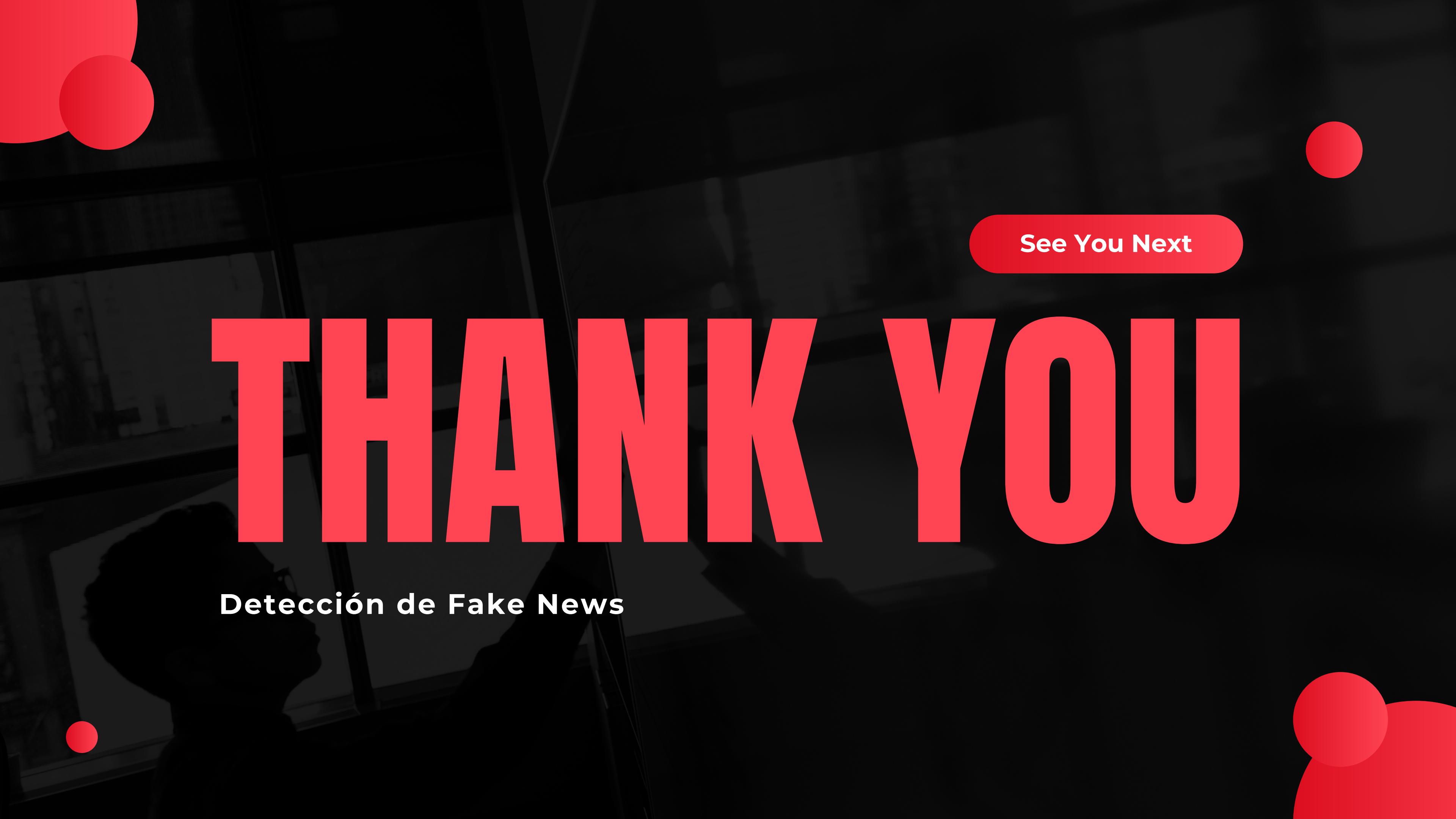


Si se desea consultar más a fondo el desarrollo del proyecto, consultar el siguiente repositorio

<https://github.com/jzarcoo/FakeNewsDetection>

BIBLIOGRAFÍA

- Google, (2025). Classification: Accuracy, recall, precision, and related metrics. Google. Recuperado el 13 de marzo del 2025 de <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=es-419>
- Shaikh, J., & Patil, R. (s.f.). Fake news detection using machine learning. Departamento de Electrónica y Telecomunicaciones, Facultad de Ingeniería K.J. Somaiya, Mumbai, India. Recuperado el 1 de febrero del 2025 de <https://svu-naac.somaiya.edu/C3/DVV/3.4.5/Confernce+and+Book+Chapter/234.pdf>
- Scikit-learn, (s.f.) RBF SVM parameters. Recuperado el 3 de abril del 2025 de https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html
- TensorFlow Core, (2020). Making BERT Easier with Preprocessing Models From TensorFlow Hub. TensorFlow. Recuperado el 15 de marzo del 2025 de https://blog.tensorflow.org.translate.goog/2020/12/making-bert-easier-with-preprocessing-models-from-tensorflow-hub.html?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc
- TensorFlow, (2024). Solve GLUE tasks using BERT on TPU. TensorFlow. Recuperado el 15 de marzo del 2025 de https://www.tensorflow.org/text/tutorials/bert_glue
- TensorFlow, (2024). Classify text with BERT. TensorFlow. Recuperado el 15 de febrero del 2025 de https://www.tensorflow.org/text/tutorials/classify_text_with_bert#about_bert



See You Next

THANK YOU

Detección de Fake News