[[1]](#footnote-0)

RoBERTa For Fake News

Miguel S. Barrera, Joselin G. Arancibia

**Resumen - Este trabajo se enfoca en la detección de noticias falsas en español, un desafío crítico en la era digital. Para abordar este problema, se utiliza el modelo RoBERTa junto con técnicas de procesamiento de lenguaje natural. El proceso implica la recopilación y curación de un conjunto de datos en español que abarque noticias auténticas y falsas, considerando diversos temas y estilos de escritura. Una vez que se cuenta con el conjunto de datos adecuado, se configuran hiper parámetros específicos del idioma y se lleva a cabo la tokenización adecuada para adaptar el modelo RoBERTa al español. Luego, se procede a entrenar el modelo utilizando técnicas de aprendizaje profundo, como el ajuste fino, con el objetivo de desarrollar un sistema de detección de noticias falsas eficaz en español. El proceso de entrenamiento y evaluación involucra la separación de datos en conjuntos de entrenamiento, validación y prueba para ajustar los hiper parámetros y evitar el sobreajuste. Una vez que se ha logrado un rendimiento satisfactorio, el modelo entrenado se puede implementar en aplicaciones o sistemas, contribuyendo de esta manera a la mitigación de la difusión de información falsa y engañosa en el ámbito de habla hispana.**

**Índice de Términos - Artificial intelligence, Transfer learning, Machine learning, Deep learning, Big Data applications**

# introducción

En la era de la datificación y digitalización masiva de las noticias y en general todo tipo de información, la facilidad de acceder a noticias falsas, de manera intencionada o no, y además generar noticias con contenido falso a través de las plataformas online [1] ha resultado en la propagación de noticias falsas como un desafío crítico que afecta la confianza en los medios y puede influir en la percepción de la realidad afectando el comportamiento de los individuos receptores [2]. Los riesgos que traen las noticias falsas han generado la necesidad de desarrollar métodos efectivos para su detección teniendo en cuenta la tecnología actual, algoritmos y hardware, considerando el potencial de la inteligencia artificial para detectar patrones en grandes conjuntos de datos [3].

En este trabajo se presenta un enfoque basado en el procesamiento de lenguaje natural (NLP) y el uso del modelo RoBERTa para la identificación de noticias falsas porque la detección de estas no solo es esencial para garantizar la precisión de la información, sino que también tiene un impacto significativo en la toma de decisiones informadas.

El objetivo principal de este trabajo es evaluar la eficacia del modelo RoBERTa en la detección de noticias falsas a través de la adaptación y el entrenamiento específico en el idioma español, comenzando con el mapeo de conjuntos de datos y la selección de estos según las características que nos interesan para luego entrenar el modelo pre entrenado de RoBERTa con estos nuevos conjuntos de datos, esto es más conocido como aprendizaje por transferencia,, en el cual usaremos herramientas como las librerías pyspark para el proceso de ETL de los conjuntos de datos, librerías de hugging-face para la tokenización de los conjuntos de datos , entrenamiento del modelo, hasta el despliegue del modelo en los espacios de hugging face para el uso y prueba de este. Se aborda el problema de la detección de noticias falsas en el contexto del idioma español, un área que requiere enfoques y modelos especializados debido a las particularidades lingüísticas y culturales.

En resumen, este trabajo se centra en abordar el problema de las noticias falsas en español utilizando técnicas de NLP y el modelo RoBERTa, con el objetivo de mejorar la detección de noticias falsas en línea en este idioma.

# Objetivos

* 1. *Objetivo General*

El objetivo general de este trabajo es entrenar el modelo RoBERTa en español con el propósito de mejorar la detección de fake news en español, contribuyendo así a la mitigación de la difusión de información falsa y engañosa en el ámbito de habla hispana.

* 1. *Objetivos Específicos*

Recopilar y curar un conjunto de datos en español que incluya noticias auténticas y noticias falsas, asegurando la representación adecuada de diversos temas y estilos de escritura.

Configuración de hiper parámetros específicos del idioma y la tokenización adecuada.

Entrenar el modelo RoBERTa en el conjunto de datos en español utilizando técnicas de aprendizaje profundo, como el ajuste fino, con el objetivo de desarrollar un modelo de detección de fake news eficaz en español.

# Plan Detalle

Este plan detallado tiene como objetivo abordar la tarea específica de recopilar y curar un conjunto de datos que será fundamental para el entrenamiento y evaluación de nuestro modelo, el cual se basa en el modelo RoBERTa.

* 1. *Actividades*
     1. *Mapeo y ETL:* Obtener datos desde diversas fuentes, como bases de datos, archivos, o servicios web. Limpiar, estructurar y enriquecer los datos para que sean coherentes y útiles. Insertar los datos transformados en una base de datos o almacén de datos de destino. Asegurarse de que los datos cargados sean precisos y coherentes.
     2. *Entrenar:* Separar los datos en conjuntos de entrenamiento, validación y prueba para evaluar el rendimiento del modelo. Alimentar el modelo con los datos de entrenamiento y ajustar sus parámetros para minimizar la función de pérdida, lo que implica la iteración a través de los datos varias veces (épocas).
     3. *Evaluar:* Evaluar el rendimiento del modelo en el conjunto de validación para ajustar hiper parámetros y evitar el sobreajuste. Evaluar el rendimiento del modelo en el conjunto de validación para ajustar hiper parámetros y evitar el sobreajuste.
     4. *Desplegar:* Desplegar el modelo entrenado en una aplicación o sistema para su uso práctico.
  2. *Asignación de responsabilidades*

Las responsabilidades están seccionadas según las actividades, la asignación resultó en lo siguiente:

Las actividades de “Mapeo y ETL” y “Entrenar” están asignadas a Miguel Barrera Huamanlazo.

Las actividades de “Evaluar” y “Desplegar” están asignadas a Joselin Arancibia Sedano.

* 1. *Recursos necesarios*

Los recursos necesarios para el desarrollo de las actividades del trabajo son los siguientes:

Conjuntos de datos etiquetados de noticias en español. Hardware para el procesamiento, entrenamiento y evaluación con especificaciones de 9.0 Gigabytes de RAM y 30.0 Gigabytes de almacenamiento.

Software para el procesamiento, entrenamiento y evaluación, entorno de desarrollo con librerías como pytorch, transformers, hugging-face, pyspark.

# Cronograma

Según el plan detallado anteriormente y la fecha de entrega presentada se juzgaron los tiempos estimados de la siguiente manera:  
ETL:

* Conseguir datasets de noticias en español(3 nov)
* Limpiar los datasets (6 nov)
* Cargar los datasets con pyspark(6 nov)

Training

* Cargar el modelo RoBERTa y evaluar las métricas claves (10 nov)
* Tokenizar las entradas y entrenar (14 nov)

Assessment

* Evaluar las métricas claves del modelo nuevo y compararlas (17 nov)
* Evaluar las respuestas con nuevos datos (17 nov)

Deploy

* Desplegar el modelo en Hugging Face (18 nov)
* Probar el modelo en Hugging Face (18 nov)

# Metodología

La metodología empleada en este proyecto se fundamenta en la aplicación de técnicas de procesamiento de lenguaje natural (NLP) y la utilización del modelo RoBERTa para la detección de noticias falsas en español. A continuación, detallamos las etapas clave del proceso:

1. Recopilación y Curación de Datos:

* Obtención de conjuntos de datos que contengan noticias auténticas y falsas en español desde diversas fuentes, como bases de datos, archivos y servicios web.
* Limpieza, estructuración y enriquecimiento de los datos para garantizar coherencia y utilidad.
* Aseguramiento de la precisión y coherencia de los datos transformados antes de su inserción en una base de datos o almacén de datos de destino.

1. Configuración de Hiperparámetros y Tokenización:

* Definición de hiperparámetros específicos del idioma español para la configuración del modelo RoBERTa.
* Aplicación de técnicas de tokenización adecuadas para adaptar el modelo al español.

1. Entrenamiento del Modelo:

* Separación de los datos en conjuntos de entrenamiento, validación y prueba para evaluar el rendimiento del modelo.
* Alimentación del modelo con datos de entrenamiento y ajuste de sus parámetros mediante técnicas de aprendizaje profundo, como el ajuste fino.
* Iteración a través de los datos durante varias épocas para mejorar la capacidad del modelo para detectar patrones en noticias falsas.

1. Evaluación del Modelo:

* Evaluación del rendimiento del modelo en el conjunto de validación para ajustar hiperparámetros y prevenir el sobreajuste.
* Evaluación del modelo en el conjunto de prueba para obtener métricas clave y comparar el rendimiento con otros modelos o enfoques.

1. Despliegue del Modelo:

* Despliegue del modelo entrenado en una aplicación o sistema para su uso práctico en la detección de noticias falsas en tiempo real.
* Pruebas del modelo desplegado en entornos controlados y evaluación de su eficacia en la detección de noticias falsas.

# Desarrollo y Análisis

1. Recopilación de Datos:

Se exploraron plataformas como Hugging Face y Kaggle en busca de conjuntos de datos relevantes creados por usuarios expertos. Los conjuntos de datos seleccionados para este proyecto incluyen:

"FakeNewsCorpusSpanish" por sayalaruano.

"Spanish Fake and Real News" por FABRICIO A. ZULES.

"Noticias falsas en español" por ARSENII TRETIAKOV.

2. Configuración de Hiperparámetros y Tokenización:

Hiperparámetros utilizados en el modelo:

Número de batches: 16, una cantidad estándar para mantener un equilibrio entre eficiencia y rendimiento.

Épocas: 2, optando por un entrenamiento simple debido a la naturaleza del proyecto.

Número de muestras: 20,000, considerando el balance entre positivos y negativos para mitigar el riesgo de sobreajuste.

Learning rate: 12^-5, siguiendo estándares recomendados.

Se aplicó la tokenización utilizando RoBERTa, una técnica de vectorización esencial para el procesamiento eficiente de los conjuntos de datos importados.

3. Entrenamiento del Modelo:

Utilizando el modelo pre entrenado de RoBERTa y la librería Transformers de Hugging Face, se llevó a cabo un entrenamiento con una duración aproximada de trece minutos. Este proceso permitió adaptar el modelo a las características específicas de los conjuntos de datos seleccionados.

4. Evaluación del Modelo:

El modelo demostró un impresionante nivel de precisión con un accuracy del 98.16%, indicando que es altamente acertado en la detección de noticias falsas. Esta métrica respalda la eficacia del modelo entrenado.

5. Despliegue del Modelo:

Para facilitar su utilización, el modelo se desplegó en el hub de Hugging Face, proporcionando opciones para su implementación mediante pipeline o su acceso a través de una API. Este despliegue asegura una accesibilidad y usabilidad práctica del modelo en diferentes contextos y aplicaciones.

# Conclusiones

* La utilización del modelo preentrenado de RoBERTa simplificó significativamente el proceso de entrenamiento, eliminando la necesidad de crear modelos específicos para fillmask o clasificación desde cero. Esta decisión estratégica no solo ahorró tiempo sino que también aprovechó la capacidad del modelo preentrenado para comprender el contexto y las complejidades del lenguaje español.
* Los resultados obtenidos, especialmente en términos de accuracy, subrayan la eficacia del modelo en la detección de noticias falsas en español. La elevada precisión respalda la capacidad del modelo para discernir con precisión entre información auténtica y falsa en el conjunto de datos utilizado.
* La evaluación del training loss y del validation loss revela que el modelo ha aprendido de manera efectiva y generalizada, con valores bajos en ambas métricas. Este bajo nivel de pérdida indica que el modelo ha logrado capturar patrones significativos sin caer en el sobreajuste, lo cual es crucial para garantizar su capacidad de generalización a nuevos datos.

References

1. E. C. Tandoc, Z. W. Lim y R. Ling, “Defining “Fake News””, Digit. Journalism, vol. 6, n.º 2, pp. 137–153, agosto de 2017. Accedido el 3 de noviembre de 2023. [En línea]. Disponible: <https://doi.org/10.1080/21670811.2017.1360143>
2. A. Gelfert, “Fake News: A Definition”, Informal Log., vol. 38, n.º 1, pp. 84–117, marzo de 2018. Accedido el 3 de noviembre de 2023. [En línea]. Disponible: <https://doi.org/10.22329/il.v38i1.5068>
3. Á. Figueira y L. Oliveira, “The current state of fake news: challenges and opportunities”, Procedia Comput. Sci., vol. 121, pp. 817–825, 2017. Accedido el 3 de noviembre de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.procs.2017.11.106>
4. Spain AI. NLP de 0 a 100 con 🤗 - Sesión 4: Transformers y Aprendizaje por transferencia. (24 de agosto de 2021). Accedido el 2 de noviembre de 2023. [Video en línea]. Disponible: <https://www.youtube.com/watch?v=XeFDEokiK3c>

1. Documento recibido el 2 de noviembre de 2023.

   S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar. colostate.edu).

   T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp). [↑](#footnote-ref-0)