

# Detección de FakeNews en tiempo real

Fabián Barquero

Lead University

2024

---

**Objetivo del Proyecto:** El objetivo de este proyecto de machine learning es desarrollar un sistema de detección de noticias falsas (fake news) relacionadas con pandemias, con un enfoque específico en la pandemia de COVID-19. El propósito es identificar de manera precisa y eficiente las noticias que contienen información engañosa o potencialmente perjudicial sobre la pandemia, ayudando así a mitigar la difusión de desinformación.

**Definición del Enfoque:** Este proyecto utilizará técnicas de aprendizaje supervisado para clasificar las noticias como verdaderas o falsas. Se emplearán modelos de procesamiento de lenguaje natural (NLP) para analizar el contenido de las noticias y extraer características relevantes que ayuden en la clasificación. Se enfocará en identificar patrones lingüísticos y características específicas que puedan indicar la veracidad o falsedad de una noticia.

La orientación del análisis NLP será totalmente científico, dejando de lado cualquier sesgo que por una u otra razón muchas veces nubla la veracidad de la información que circula por distintos medios. Parte del enfoque es generar un sistema de alerta y prevención para los organismos correspondientes, generando así, una acción proactiva a los problemas de la desinformación.

## Tecnologías a Utilizar:

- Lenguaje de Programación: Python
- Bibliotecas de Machine Learning: scikit-learn, TensorFlow/Keras (para modelos de NLP)
- Infraestructura: Google Cloud Platform (GCP) o Amazon Web Services (AWS) para almacenamiento de datos y despliegue del modelo.
- Herramientas de Procesamiento de Lenguaje Natural (NLP): NLTK, spaCy
- Herramientas de Scraping de Datos: BeautifulSoup, Scrapy

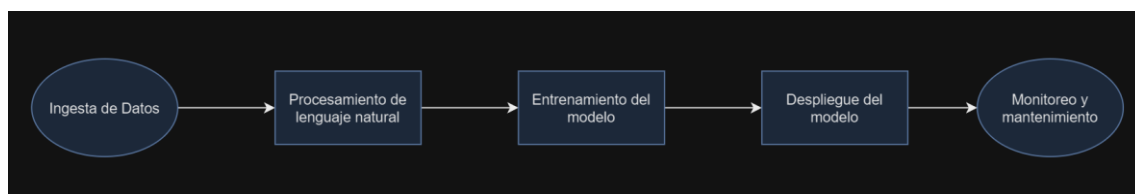
## Roles del proyecto:

- **Líder del Proyecto:** Encargado de coordinar el proyecto, establecer objetivos y gestionar la comunicación con el cliente.

- **Científico de Datos:** responsable del desarrollo y entrenamiento de los modelos de machine learning utilizando técnicas de NLP.
- **Ingeniero de Software:** Encargado del desarrollo del servicio web para el despliegue del modelo y la integración con otras herramientas.
- **Especialista en NLP:** responsable del preprocesamiento de datos y la extracción de características relevantes utilizando técnicas de NLP.

**Diseño y Diagrama de la Arquitectura:** La arquitectura del sistema incluirá los siguientes componentes:

1. **Ingesta de Datos:** Se recolectarán noticias relacionadas con la pandemia de fuentes confiables y de redes sociales. Estos datos se almacenarán en un repositorio de datos en la nube.
2. **Procesamiento de Lenguaje Natural (NLP):** Se utilizarán técnicas de NLP para preprocesar y analizar el contenido de las noticias, incluida la extracción de características relevantes.
3. **Entrenamiento del Modelo:** Se entrenarán modelos de machine learning utilizando características extraídas de las noticias preprocesadas para clasificarlas como verdaderas o falsas.
4. **Despliegue del Modelo:** Una vez entrenado, el modelo se desplegará como un servicio web para que las noticias puedan ser clasificadas en tiempo real.
5. **Monitoreo y Mantenimiento:** Se implementará un sistema de monitoreo continuo para evaluar el rendimiento del modelo en producción y realizar ajustes según sea necesario.



En conclusión, este proyecto aborda un problema importante y actual, utilizando técnicas de machine learning para combatir la propagación de noticias falsas durante la pandemia de COVID-19, pudiéndose extrapolar a muchos otros ámbitos como la política, las noticias amarillistas y en general todo lo que engloba las fake news.

Con un enfoque colaborativo y una adecuada asignación de roles, buscamos desarrollar un sistema efectivo y confiable para identificar potenciales noticias falsas y proteger así la integridad de la información sobre la salud pública.