**TP INTEGRADOR GRUPO 8**

**Participantes: Eliel Berra ; Lucas Choconi ; Federico Mina ; German Scillato.**

● Tema de investigación

Identificación de veracidad en las noticias.

● Antecedentes sobre el tema

Los portales no verifican a los usuarios ni sus publicaciones, por lo que circulan noticias falsas con el fin de generar visitas o influir en la opinión pública. Estas noticias falsas pueden ser una propaganda contra un individuo, sociedad, organización o partido político.

Un ser humano es incapaz de detectar todas estas noticias falsas, por lo tanto, se necesitan clasificadores de aprendizaje automático que puedan detectar estas noticias falsas.

Fuente: <https://arxiv.org/pdf/2102.04458.pdf>

● Aporte esperado

Modelo predictor que permita discernir entre noticias falsas y verdaderas, con una probabilidad asociada, y marcando las palabras con mayor tendencia en cada grupo.

● Disponibilidad de datos e infraestructura

Dataset de Kaggle ya procesado.

[https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset](https://st1.zoom.us/web_client/1q1nf58/html/externalLinkPage.html?ref=https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset)

Fuentes adicionales:

[https://www.kaggle.com/datasets/hassanamin/textdb3](https://st1.zoom.us/web_client/1q1nf58/html/externalLinkPage.html?ref=https://www.kaggle.com/datasets/hassanamin/textdb3) [https://www.kaggle.com/datasets/jruvika/fake-news-detection](https://st1.zoom.us/web_client/1q1nf58/html/externalLinkPage.html?ref=https://www.kaggle.com/datasets/jruvika/fake-news-detection) [https://www.kaggle.com/datasets/ruchi798/source-based-news-classification](https://st1.zoom.us/web_client/1q1nf58/html/externalLinkPage.html?ref=https://www.kaggle.com/datasets/ruchi798/source-based-news-classification)

● Plan de trabajo y cronograma tentativo

- Tiempo disponible: 5 semanas construcción , 6 semana de finalización/consultas/informe.

|  |  |
| --- | --- |
| Semana | Hito / Objetivo |
| 16 - 23/5 | Investigación Text Mining. EDA dataset. Definición de lineamientos. |
| 23 - 30/5 | Evaluar Modelos – Features importance – CV / Gridsearch – Análisis de performance (Ensamble/ Métrica). |
| 30 – 6/6 | NLP – Análisis extra |
| 6 - 13/6 | Pipeline - Formato Notebook - Modelo entregable. |
| 13 – 20/6 | Otras fuentes de información (Web / Dataset alternativos / Noticias actuales). **Reporte Tecnico** y modelo en pickle. |
| 20 – 27/6 | Conclusiones - Armado y preparación de **presentación** |