

Miniproyecto clasificación de texto.

Aplicar un modelo de clasificación multiclase para texto utilizando las clases:

**Clasicismo.** Cualquier manifestación que promueva una actitud o tendencia a discriminar o minimizar a alguien basándose en la diferencia de *estatus social*.

- **Racismo.** Cualquier manifestación que promueva una actitud o tendencia a discriminar o minimizar a alguien basándose en *características étnicas* o que promueva la *superioridad de un grupo*.
- **Sexismo.** Cualquier manifestación que promueva una actitud o tendencia a discriminar o minimizar a alguien por sus características de *género*. Esto incluye contenido misógino, misandrista y relacionado con LGBTQ+.
- **Otros.** Cualquier manifestación que promueva una actitud o tendencia a discriminar o minimizar a alguien en base a características que *no pertenezcan a las anteriormente definidas*, tales como *profesión, religión, lugar de origen, afiliación política e intereses*.
- **Contenido inapropiado.** El meme presenta *cualquier tipo de manifestación de contenido ofensivo, vulgar* (profano, obsceno, con contenido sexual) y de *humor morboso*.
- **Ninguna de las anteriores**

*El procedimiento a seguir el es siguiente:*

- 1.- Adquisición de los datos: generar frases con modelos como chatgpt, gemini, claud, etc y etiquetarlas manualmente o buscar un dataset que ya tenga frases etiquetadas. Recuerda que en la competencia DIMEMEX ya hay datos, pero se necesita generar un conjunto más grande.
- 2.- Preprocesamiento de los datos y extracción de las características: decidir si es necesario aplicar limpieza de datos como normalización de texto, lematización, stemming, etc. Proponer una forma de representación de texto a partir de características definidas.
- 3.- Seleccionar un modelo de clasificación justificando la decisión por medio de un diagrama de dispersión, curvas de distribución, u alguna otra técnica.
- 4.- Evaluación del modelo: mostrar métricas de evaluación Precisión, recall , f-measure y Accuracy, en las etapas de entrenamiento y prueba. Las oraciones de prueba serán liberadas un día antes de la presentación.

La presentación deberá reflejar todo el proceso de análisis y desarrollo del proyecto, además de incluir código que sea relevante para ilustrar la toma de decisiones., también pueden mencionar intentos fallidos.

Proyecto 70% y presentación 30%