

Obligatorio

Introducción

Como consecuencia de la aparición del Covid-19, una gran cantidad de noticias se dispersaron diariamente a través de las redes sociales. De esta forma fueron generados muchos tweets fidedignos, pero varios también resultando ser noticias falsas o **fake news**.

Debido a la aparición de **fake news**, muchas personas fueron desinformadas sobre aspectos relacionados al virus o acerca de cómo cuidarse adecuadamente.

Mediante técnicas de Machine Learning, y más específicamente NLP, se solicita diseñar un método que permita detectar las **fake news** con la mayor exactitud posible.

Datos

Los datos se proveen como ***obligatorio.csv*** y ***test.csv***.

Implementación

Se pide

1. Realizar las transformaciones preliminares del dataset que considere necesaria.
2. Separar el dataset ***obligatorio.csv*** en train y validation.
3. Realizar un baseline utilizando bags of words, haciéndole un preprocesamiento básico.
4. Aplicando una técnica clásica de análisis supervisada, analizar el resultado observando errores en validation y test. Mencionar posibles puntos de mejora.
5. Realizar ingeniería de atributos (feature engineering) y luego evaluar el impacto de los atributos agregados sobre las predicciones.
Algunas de las features que se le sugiere extraer son:
 - Largo del documento
 - Cantidad de palabras
 - Cantidad de letras
6. Realizar un modelo con TfIdf, stop word removal y reduce to a root.
7. Aplicando una técnica clásica de análisis supervisada, analizar el resultado observando errores en validation y test, compararlo con el modelo anterior y ajustar hiperparámetros.

Deberá entregar la notebook y todos los archivos de código generados para obtener las predicciones y métricas.

Entrega Online:

La entrega de los obligatorios será en formato digital online. Los principales aspectos a destacar sobre la entrega online de obligatorios son:

1. La entrega se realizará desde gestion.ort.edu.uy
2. Previo a la conformación de grupos cada estudiante deberá estar inscripto a la evaluación. Sugerimos realizarlo con anticipación.
3. Uno de los integrantes del grupo de obligatorio será el administrador del mismo y es quien formará el equipo y subirá la entrega.
4. Cada equipo debe entregar un único archivo en formato zip o rar (los documentos de texto deben ser pdf, y deben ir dentro del zip o rar)
5. El archivo a subir debe tener un tamaño máximo de 40mb.
6. Les sugerimos realicen una 'prueba de subida' al menos un día antes, donde conformarán el 'grupo de obligatorio'.
7. La hora tope para subir el archivo será las 21:00 del día fijado para la entrega.
8. La entrega se podrá realizar desde cualquier lugar (ej. hogar del estudiante, laboratorios de la Universidad, etc).
9. Aquellos que presenten alguna dificultad con su inscripción o tengan inconvenientes técnicos, por favor pasar por la oficina del Coordinador o por Coordinación adjunta antes de las 20:00hs. del día de la entrega.

En caso de tener una situación particular de fuerza mayor, es necesario dirigirse con suficiente antelación al plazo de entrega, al Coordinador de Cursos o Secretario Docente.

| | |
|--------------------------|--------------------------|
| Lectura de obligatorio: | 03-10-2022 |
| Plazo máximo de entrega: | 08-12-2022 |
| Defensa: | A definir por el docente |
| Puntaje mínimo / máximo: | 10 / 40 puntos |