

75.06/95.58 Organización de Datos

Finger 1: Real or Not?

Tiempo Estimado: 3hs.

El objetivo de este ejercicio es realizar un muy rápido análisis del set de datos del TP, es una buena oportunidad para instalar las herramientas que van a usar para la primera entrega del TP que consiste en realizar un análisis exploratorio y visualizaciones.

Herramientas recomendadas:

- Python con Pandas, Matplotlib y otros.
- R y R-Studio con Ggplot2 y otros.

Para realizar el finger, deberán bajar el set de datos del TP:

<https://www.kaggle.com/c/nlp-getting-started>

Se pide:

1. Sobre el set de datos del TP, analizar la relación que existe entre la longitud de los tweets, y la veracidad de los mismos. Para esto se espera que analicen las longitudes de los distintos tweets, long máxima, mínima, promedio, y si existe alguna relación entre estos valores y el target (que indica si la información provista en el tweet es real o no).
2. Realizar una visualización que permita comunicar los resultados del punto 1.

Entrega:

Para entregar el resultado, generar un pdf que contenga el código y la visualización, y subirlo a Gradescope.

Sobre Gradescope:

Ingresar a <https://www.gradescope.com/> y registrarse.

Una vez registrados, buscar el curso de la materia. El código del curso es : **MJJB4R**
Ahi encontraran la opción para subir el pdf con el resultado.

El finger aparecerá publicado en gradescope el día 23-4, y habrá tiempo para realizar la entrega hasta el día 7-5. La presentación es de carácter obligatoria para los alumnos regulares, y optativa para los alumnos oyentes.