Procesamiento de Lenguaje Natural (NLP)

Especialización y Maestría en Analítica Estratégica de Datos

2021-I

Plan de Proyecto Final

1. Estudiante(s) — Máximo 3:
   * Leonardo Hoyos
   * Diego Barreto
   * Diana Marcela Gómez
2. Título provisional del Proyecto

*“Clasificación automática por categoría de noticias”*

1. ¿De qué se trata el proyecto?

Con este proyecto buscamos realizar una clasificación automática de noticias por categorías (política, entretenimiento, salud, ciencia, dinero, negocios, comedia, deporte, entre otros), de acuerdo al texto que contiene cada una de estas.

1. ¿Cuál es la motivación?

Pone en práctica los conocimientos adquiridos en la clase de Lenguaje de procesamiento natural.

1. ¿Cómo se relaciona con NLP?

El proyecto está relacionado directamente con NLP, puesto que realizaremos un análisis de texto para hacer una clasificación por categorías.

1. ¿Dónde va conseguir los datos? ¿Son etiquetados?

Los datos los obtendremos del siguiente link:

<https://www.kaggle.com/rmisra/news-category-dataset>

Son datos etiquetados.

1. De las cosas que hemos visto (o vamos a ver) en clase, ¿qué piensa usar?

Adquisición de datos, limpieza de texto, pre - procesamiento, Feature engineering y modelado.

1. ¿Cómo piensa presentar los datos y sus resultados?

Pensamos presentar los datos y resultados en un Jupyter Notebook.

1. Otros comentarios: