

Procesamiento del Lenguaje Natural

Luis Gabriel Moreno Sandoval

Taller NLP 4:

1. Desarrollar un componente que permita la **“clasificación de textos”** de rasgos demográficos del autor como la edad a través de la base de entrenamiento que se provee “age_tweets_v2.rar” en formato CSV, se debe generar un reporte con el accuracy y F1-Score y se ranqueara de acuerdo con estos dos resultados. Se deben entregar, los informes de la construcción y resultados del modelo, código fuente o Jupyter y el compilado del modelo generado.
2. Desarrollar un componente que permita la **“clasificación de textos”** dos rasgos de estos uno asociados a la procedencia del autor teniendo textos de “Ecuador”, “Perú” y “Colombia” y otro asociado al contenido tipificado solo de las noticias de Colombia con las clases de “Política”, “Deporte”, “Economía”, “Tecnología”, “Ocio” a través de la base de entrenamiento que se provee “Corpus_News_LatinoAmerica.rar” en formato JSON, se debe generar un reporte con el accuracy y F1-Score y se ranqueara de acuerdo con estos dos resultados. Se deben entregar, los informes de la construcción y resultados del modelo, código fuente o Jupyter y el compilado del modelo generado.