

TEMA 7 – HOJA DE EJERCICIOS I

Se plantean diferentes ejercicios para practicar con el aumento de datos en textos.

Ejercicio 1.

Utilizando un subconjunto de datos del dataset "[rotten tomatoes](#)", que contiene comentarios de películas clasificados como positivos o negativos se pide lo siguiente:

- a. Sacar estadísticas del conjunto de datos, para ver cuántos textos tiene de cada clase. Si se elige un subconjunto de este dataset, hacer una división para entrenamiento y test y desbalancear el conjunto de entrenamiento.
- b. Utilizar un enfoque clásico para representar los textos con TF-IDF y utilizar un algoritmo SVM para entrenar y luego clasificar los textos del conjunto de test.
- c. Probar a hacer aumento de datos sobre la clase minoritaria del conjunto de entrenamiento, entrenar de nuevo y ver si los resultados mejoran o no.
 - Hay que probar diferentes tipos de aumento de datos.
 - Utilizar alguna de las librerías vistas en clase.