



Reporte de Desempeño de Clasificadores (TC3007C)

Clasificación de Sentimientos

Este informe resume el desempeño de varios modelos de clasificación entrenados con el archivo “tlfornlp-chapters2-3-imdb-traditional.ipynb”.

Se probaron cuatro tipos de clasificadores de acuerdo con su complejidad, cuyos resultados se describen a continuación:

Complejidad de Modelos (de menor a mayor):

1. Regresión Logística: Modelo lineal básico para clasificación binaria.
2. Support Vector Machine: Modelo que optimiza márgenes, más sofisticado y no necesariamente lineal.
3. Random Forest: Ensamblado de árboles de decisión, más robusto y preciso.
4. Gradient Boosting Machine: Modelo secuencial de árboles, que mejora iterativamente.

Rendimiento de Modelos con Features BOW (BagOfWords), se presenta el “accuracy score”:

	Regresión Logística	Support Vector Machine	Random Forest	Gradient Boosting Machine
maxtokens = 50 maxtokenlen = 20	0.6966	0.6716	0.6766	0.6983
maxtokens = 100 maxtokenlen = 100	0.7483	0.7216	0.6683	0.695
maxtokens = 200 maxtokenlen = 100	0.7833	0.7683	0.7016	0.7533

Rendimiento de Modelos con Features TF-IDF, se presenta el “accuracy score”:

	Regresión Logística	Support Vector Machine	Random Forest	Gradient Boosting Machine
maxtokens = 50 maxtokenlen = 20	0.7055	0.6766	0.6416	0.665
maxtokens = 100 maxtokenlen = 100	0.7716	0.7533	0.6666	0.6883
maxtokens = 200 maxtokenlen = 100	0.785	0.8166	0.6416	0.6346