

## Tarea 3: Clasificación de sentimientos base de datos IMBD

Estudiante: Oscar Gutierrez

Profesor: Dr. Alfredo Esquivel Jaramillo

### 1 Introducción

Se clasificarán las reviews de películas con datos obtenidos de IMBD utilizando 4 modelos diferentes, con los cuales se toman en cuenta los siguientes parámetros:

- (a) Nsamp = 1000, maxtokens = 50, maxtokenlen = 20
- (b) Nsamp = 1000, maxtokens = 100, maxtokenlen = 100
- (c) Nsamp = 1000, maxtokens = 200, maxtokenlen = 100

### 2 Resultados utilizando BOW

Modelo	Parámetros	Accuracy
Regresión Logística	a	0.706
Regresión Logística	b	0.738
Regresión Logística	c	0.763
Support Vector Machine	a	0.685
Support Vector Machine	b	0.725
Support Vector Machine	c	0.763
Random Forest	a	0.700
Random Forest	b	0.751
Random Forest	c	0.766
Gradient Boosting	a	0.676
Gradient Boosting	b	0.693
Gradient Boosting	c	0.725

Table 1: Resultados de diferentes modelos con BOW

### 3 Resultados utilizando TD-IDF

Modelo	Parámetros	Accuracy
Regresión Logística	a	0.736
Regresión Logística	b	0.766
Regresión Logística	c	0.805
Support Vector Machine	a	0.718
Support Vector Machine	b	0.745
Support Vector Machine	c	0.800
Random Forest	a	0.687
Random Forest	b	0.712
Random Forest	c	0.775
Gradient Boosting	a	0.648
Gradient Boosting	b	0.723
Gradient Boosting	c	0.765

Table 2: Resultados de diferentes modelos con TD-IDF