

Reporte de comparación de desempeño

Run A

Number of samples to generate in each class - 'spam', 'not spam'

Nsamp = 1000

Maximum number of tokens per document

maxtokens = 50

Maximum length of each token

maxtokenlen = 20

Run de B

Number of samples to generate in each class - 'spam', 'not spam'

Nsamp = 1000

Maximum number of tokens per document

maxtokens = 100

Maximum length of each token

maxtokenlen = 100

Run de C

Number of samples to generate in each class - 'spam', 'not spam'

Nsamp = 1000

Maximum number of tokens per document

maxtokens = 200

Maximum length of each token

maxtokenlen = 100

Bag-of-Words (BOW)

Run	Logistic Regression	SVC Classifier	Random Forest	Gradient Boosting
A	0.66	0.65	0.66	0.62
B	0.73	0.73	0.76	0.67
C	0.77	0.75	0.78	0.76

TD-IDF

Run	Logistic Regression	SVC Classifier	Random Forest	Gradient Boosting
A	0.72	0.73	0.65	0.65
B	0.74	0.77	0.71	0.71
C	0.81	0.78	0.72	0.72