

$$\text{purity} = \frac{1}{N} \sum_k \max_j |d_k \cap c_j|$$

$$\text{entropy} = \sum_k \frac{n_k}{N} \left(-\frac{1}{\log c} \sum_j \frac{n_k^j}{n_k} \log \left(\frac{n_k^j}{n_k} \right) \right)$$

$$\text{Rand index} = \frac{tp + tn}{tp + fp + fn + tn}$$

$$F_1 = \frac{2PR}{P + R}, P = \frac{tp}{tp + fp}, R = \frac{tp}{tp + fn}$$

	Euclidean	Cosine	Jaccard
Bag of Words	37.0708	37.1215	35.8215
Bigrams	27.9799	27.8124	27.2025
tf-idf BoW	51.7989	51.7211	
tf-idf Bigrams	55.0646	54.9467	
LSA	55.5582	55.6255	
LDA	68.3555	68.3549	

	Euclidean	Cosine	Jaccard
Bag of Words	0.43	0.45	0.43
Bigrams	0.41	0.35	0.38
tf-idf BoW	0.42	0.47	
tf-idf Bigrams	0.45	0.40	
LSA	0.55	0.49	
LDA	0.37	0.39	

	Euclidean	Cosine	Jaccard
Bag of Words	0.6669	0.6329	0.6512
Bigrams	0.6836	0.6999	0.6965
tf-idf BoW	0.6567	0.6236	
tf-idf Bigrams	0.6051	0.6726	
LSA	0.4782	0.5530	
LDA	0.7281	0.7031	

	Euclidean	Cosine	Jaccard
Bag of Words	0.7913	0.7840	0.8055
Bigrams	0.7055	0.7125	0.6492
tf-idf BoW	0.8070	0.7861	
tf-idf Bigrams	0.7941	0.7467	
LSA	0.8451	0.8255	
LDA	0.7721	0.7697	

	Euclidean	Cosine	Jaccard
Bag of Words	0.2108	0.2337	0.2113
Bigrams	0.2310	0.2304	0.2250
tf-idf BoW	0.2035	0.2309	
tf-idf Bigrams	0.2611	0.2221	
LSA	0.3494	0.2800	
LDA	0.1742	0.1822	