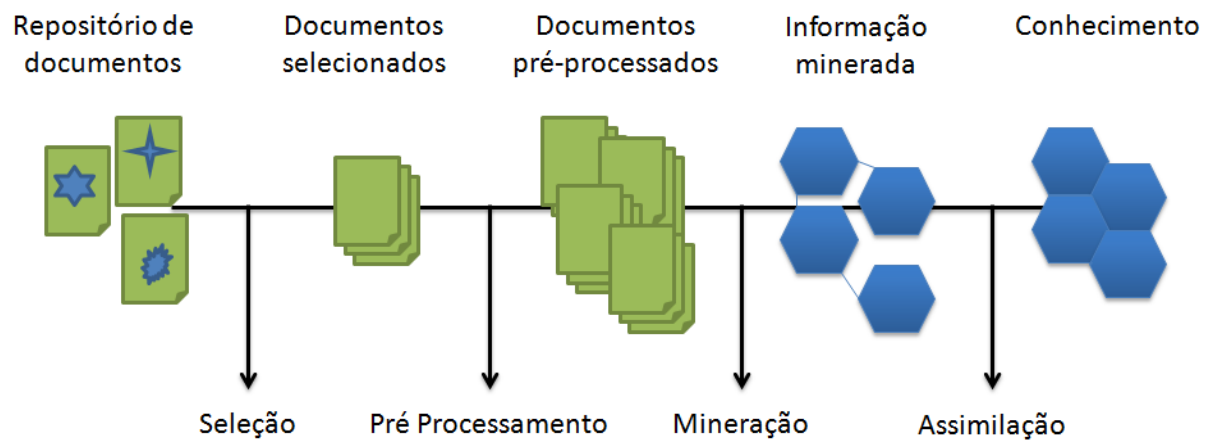


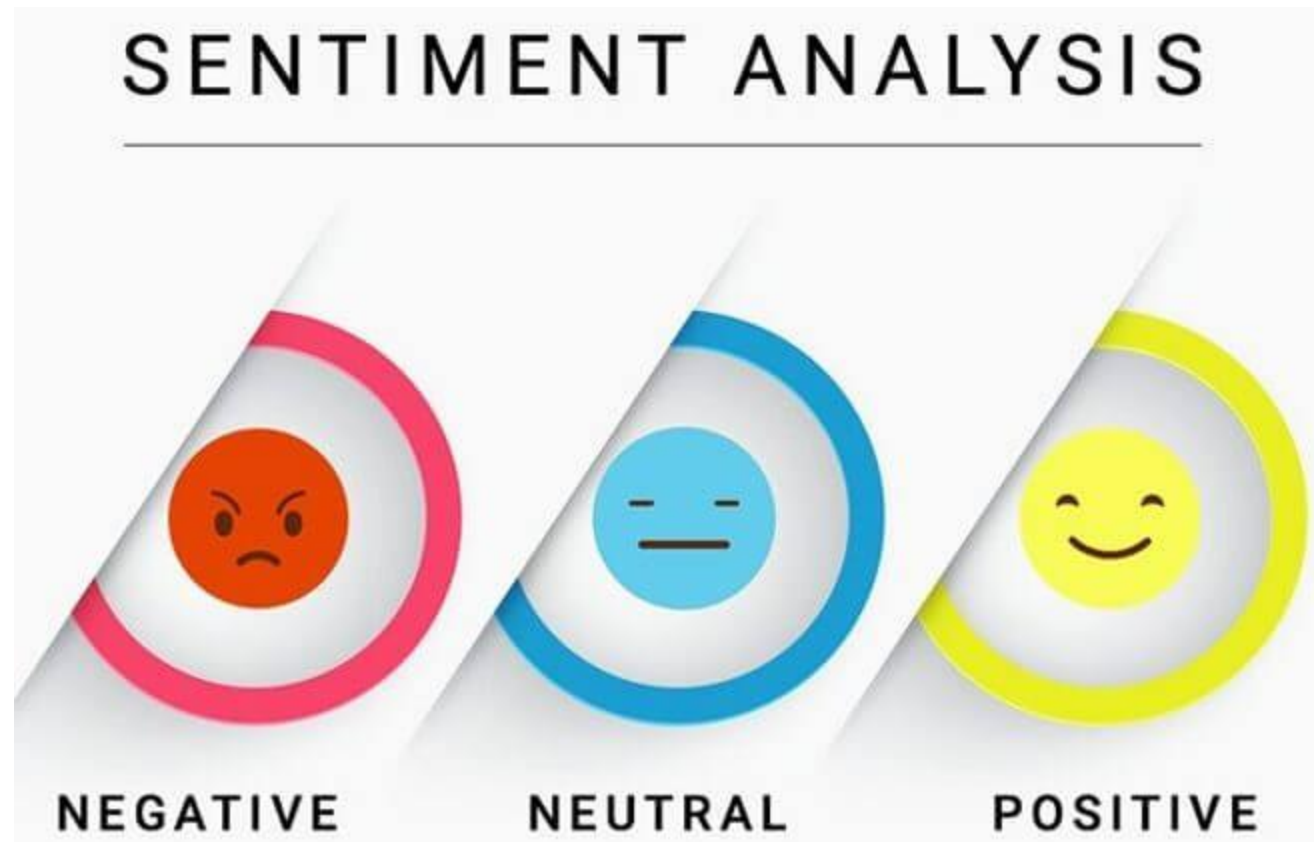
Analise De Sentimento Polarizada

Raissa Camelo

Mineração de Texto:



Mineração de
Texto: Analise
de sentimento.



Base de
Dados:

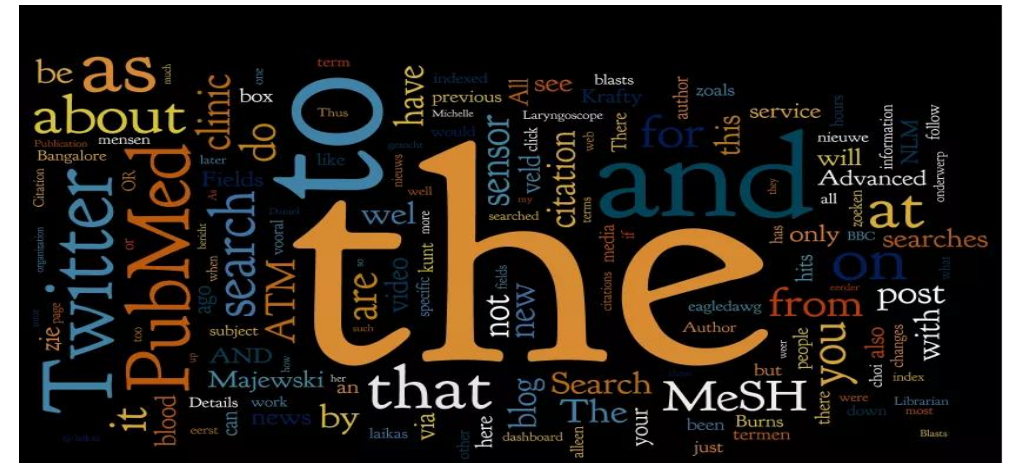


- 12.500 textos de review positivos
- 12.500 textos de review negativos
- Textos de aproximadamente 5 frases.

Pré - processamento

- Tokenização
- Stop Words removal

'strangers', 'come', 'along', '.', 'What', 'makes',
'changes', 'outlooks', 'life', '.', 'Unfortunately'
'Ann', 'Harding', 'plays', 'important', 'role', 'fi
'really', 'make', 'wonder', "n't", 'zombie', 'movie
'bombastic', 'I', 'close', 'turning', 'film', '--',
'understand', '.', 'Sure', ',', 'film', 'great', 'a
'.', 'Sorry', ',', 'film', 'aching', 're-make', '!'



TF - IDF

$$\log(N/df)$$

$$TF = nt/n$$

21	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	1
22	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	1
23	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	1
24	0.0	0.0	0.0	0.010995	0.000000	...	0.0	0.0	0.0	0.0	1
25	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	1
26	0.0	0.0	0.0	0.003765	0.000000	...	0.0	0.0	0.0	0.0	1
27	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	1
28	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	1
29	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	1

- Cálculo do TF-IDF
- Geração de Tabela TF-IDF em dataframe

Resultados

Média, Desvio Padrão e Mediana			
Classificador	Média	Desvio Padrão	Mediana
Árvore	0.6536	0.0388	0.6545
Naive Bayes	0.8727	0.0353	0.8636
KNN (n=1)	0.7254	0.0283	0.7363
KNN (n=3)	0.6927	0.0412	0.7090
KNN (n=5)	0.7136	0.0579	0.6954
WKNN (n=1)	0.7254	0.0283	0.7363
WKNN (n=3)	0.6927	0.0412	0.7090
WKNN (n=5)	0.7136	0.0579	0.6954
SVM-LINEAR	0.4763	0.0179	0.4863
SVM-RBF	0.4763	0.0179	0.4863
REG-LOGI	0.8263	0.0622	0.8363

Resultados

KNN e SVM variações		
Comparação	t-value	p-value
KNN-1 e KNN-3	35.345	0.00062
KNN-1 e KNN-5	-0.1146	0.9089
KNN-3 e KNN-5	-30.908	0.0025
WKNN-1 e WKNN-3	35.345	0.00062
WKNN-1 e WKNN-5	-0.1146	0.9089
WKNN-3 e WKNN-5	-3.090	0.0025
SVM-LIN e SVM-RBF	0.0	1.0

Melhores Algoritmos		
Comparação	t-value	p-value
KNN-1 e NBAYES	-28.155	9.170
KNN-1 e REG - LOG	-18.904	18.445
NBAYES e REG - LOG	57.047	12.378