

Trabalho Final – NLP

- 1) Implemente o algoritmo Bayesian Sets segundo orientações dada em aula.
- 2) Utilizando o algoritmo implementado e o dataset movies.csv como base de pesquisa, use as seguintes frases (queries) para obter exemplos semelhantes no dataset.
 - a. Remova o ano do título do filme antes
 - b. Queries: ['toy story', 'the lion king', 'alladin', 'beauty and the best', 'cinderella', 'little mermaid', 'hercules']
 - c. Resultado esperado (aproximado):

	input	score
1997	Little Mermaid, The	10.081729
360	Lion King, The	9.668139
0	Toy Story	8.836643
7960	Cinderella Story, A	8.836643
3027	Toy Story 2	8.836643
15401	Toy Story 3	8.836643
9398	Lion King 1½, The	8.163891
18252	Another Cinderella Story	7.331624
19335	Little Mermaid: Ariel's Beginning, The	7.072975
15368	Mermaid, The	6.931102
26654	Hercules	6.396474
23754	Hercules	6.396474
1515	Hercules	6.396474
13118	Cinderella	6.155449
27069	Cinderella	6.155449

- 3) Utilizando o dataset movie_review1, visto na aula de Naive Bayes, crie um classificador de sentimentos utilizando word2vec.
 - a. Utilize F1-score como métrica
 - b. Compare CBOW com Skip-gram. Qual obteve o melhor resultado?
- 4) Execute o algoritmo LDA no mesmo dataset (movie_review1) usando apenas 2 tópicos. Descreva os resultados indicando se ele conseguiu separar bem reviews positivos de reviews negativos