

Data Science do ZERO

Capítulo 06 - Machine Learning

Naive Bayes

Naive Bayes e Bag of Words

- Por que texto??



Naive Bayes e Bag of Words

- Modelagem usada para processamento de textos
- Mapeia a quantidade de cada termo no documento.
- É gerado um vetor de termos e suas frequências.
- Naive Bayes costuma ter bons resultados com essa abordagem.
- Alta dimensionalidade.

The Bag of Words Representation

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!

15



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...

Representação Bag of Words

- Exemplo de representação da frase “The dog is on the table”

the dog is on the table

0	0	1	1	0	1	1	1
are	cat	dog	is	now	on	table	the

Representação Bag of Words

- A frequência das palavras é usada para treinar um modelo.
- Por exemplo, imagine as seguintes sentenças:
Sentença 1: “O smartphone é bom, porém não gosto de smartphone grandes” , Negativo
Sentença 2: “Gostei bastante desse smartphone, mesmo sendo caro” , Positivo
- Veja como seria as features:

{o, smartphone,é, bom, porém, não, gosto, de, grandes,gostei,bastante,desse,mesmo,sendo,caro}

Sentença 1: {1, 2,1,1,1,1,1,1,1,0,0,0,0,0,0, Negativo}

Sentença 2: {0,1,0,0,0,0,0,0,0,1,1,1,1,1,1, Positivo}

Hands on!