

 06

## Caracaterizando o TF-IDF e Ngrams

Júlia quer otimizar ainda mais seus resultados, então ela resolve pesquisar mais sobre as técnicas **TF-IDF** e **Ngrams**. Para isso ela precisa da sua ajuda para descobrir:

Qual a principal característica do **TF-IDF**? Qual cuidado Júlia deve ter ao utilizar o **Ngrams**?

*Seleciona uma alternativa*

**A** A principal característica do TF-IDF é ser uma pontuação proporcional à frequência da palavra no texto e equilibrada pela frequência no corpus, ou seja, palavras que se repetem muito em todo o corpus tendem a ter pontuações menores. Embora o Ngrams seja um método que pode otimizar o seu resultado devemos ter cuidado, já que ao utilizá-lo o nosso vetor Bag of Words aumenta consideravelmente de tamanho, o que também faz com que o custo computacional aumente.

**B** A principal característica do TF-IDF é ser uma pontuação proporcional à frequência da palavra no texto e equilibrada pela frequência no corpus, ou seja, palavras que se repetem muito em todo o corpus tendem a ter pontuações menores. Embora o Ngrams seja um método que pode otimizar o seu resultado, devemos ter cuidado, já que ao utilizá-lo o nosso vetor Bag of Words diminui consideravelmente de tamanho, o que pode impactar negativamente a classificação.

**C** A principal característica do TF-IDF é ser uma pontuação proporcional à frequência da palavra no texto e equilibrada pela frequência no corpus, ou seja, palavras que se repetem muito em todo o corpus tendem a ter pontuações maiores. Embora o Ngrams seja um método que pode otimizar o seu resultado devemos ter cuidado, já que ao utilizá-lo o nosso vetor Bag of Words aumenta consideravelmente de tamanho, o que também faz com que o custo computacional aumente.