



**Data Science
Academy**

www.datascienceacademy.com.br

Processamento de Linguagem Natural

Modelos N-grama x Bag of Words



Até aqui já analisamos os conceitos de n-grams e BOW. Então, vamos ver agora como n-grams e BOW são diferentes ou relacionados uns aos outros.

Vamos primeiro discutir as diferenças. Aqui, a diferença está em termos de seu uso em aplicativos de PLN. Em n-gramas, a ordem das palavras é importante, enquanto que na BOW não é importante manter a ordem das palavras. Durante a aplicação de PLN, n-gram é usado para considerar as palavras em sua ordem real, para que possamos ter uma ideia sobre o contexto da palavra em particular. BOW é usado para construir vocabulário para o seu conjunto de dados de texto.

Agora vamos ver alguns relacionamentos significativos entre n-grams e BOW, que lhe darão uma ideia de como n-grams e BOW estão relacionados entre si. Se você está considerando n-gram como um recurso, então BOW é a representação de texto derivada usando um unigrama. Portanto, nesse caso, um n-grama é igual a um recurso e o BOW é igual a uma representação de texto usando um unigrama (um grama) contido nele.

Vejamos um exemplo de aplicação do BOW.