



**Data Science
Academy**

www.datascienceacademy.com.br

Processamento de Linguagem Natural

Mini-Projeto 2

Aplicação de Classificação de E-mails



No mundo da Internet hoje, uma quantidade enorme de dados é transferida entre computadores na forma de e-mails. Consequentemente, está ficando difícil classificar manualmente os e-mails importantes dos não importantes.

A classificação de e-mail foi extensivamente estudada e pesquisada no passado, mas a maior parte da pesquisa foi no campo de detecção e filtragem de spam.

Este mini-projeto se concentra na classificação de e-mails em pastas personalizadas relevantes ao usuário. Utilizamos duas abordagens diferentes aqui - o classificador Naïve Bayes e o algoritmo k-vizinhos mais próximos (KNN).

O classificador Naïve Bayes é baseado em um modelo probabilístico, enquanto o algoritmo k-vizinhos mais próximos é baseado em uma medida de similaridade com os e-mails de treinamento. Propomos o método de usar essas duas abordagens na classificação de e-mail, analisar o desempenho desses algoritmos e comparar seus resultados. Em seguida, propomos alguns trabalhos futuros para otimização adicional e melhor eficiência.

Em anexo você encontra a aplicação completa com todas as instruções para execução. Esta é uma aplicação Python 100% funcional.

Bons estudos.