

Análise de Comentários em E-commerces

Jailson Dias (jcd2) e Ramom Pereira (rpss)

1. Introdução

Com a popularização das vendas online, tanto em e-commerces como a Amazon, Magazine Luiza, quanto nas redes sociais como Instagram, Facebook e WhatsApp, surge uma quantidade imensa de avaliações de produtos que muitas vezes são difíceis de identificar o que é nessas avaliações, pois a quantidade é grande e existem muitas avaliações que não são boas para extrair alguma informação sobre o produto. Assim surge um grande problema que é identificar quais os problemas ocorridos durante o processo de compra ou até mesmo depois da compra para assim conseguir dar uma melhor experiência de compra para o cliente.

2. Coleta de Dados

Para fazer esse projeto vai ser utilizado uma base de dados da Amazon sobre avaliações e comentários dos produtos vendidos no seu e-commerce. Como a base dados é muito grande, vamos utilizar apenas a parte de roupas, sapatos e joias que somam 278.677 avaliações. Essa base tem informações do ID da avaliação, nome do usuário na Amazon, texto da avaliação em inglês, resumo feito pelos algoritmos inteligentes da Amazon e a data que o comentário aconteceu.

3. Pré-processamento da base

Nesta base vamos executar um algoritmo de processamento de linguagem natural para identificar algumas características no comentário, que são elas, identificar se o comentário é relevante, ou seja, se ele está falando alguma coisa sobre o produto ou a sobre a loja, e se ele está falando positivamente ou negativamente. O objetivo dessa etapa é retirar todos os comentários de apenas uma palavra ou comentários que não estão falando do produto ou da loja e saber sobre a experiência do usuário no processo de compra e depois da compra quando estiver utilizando o produto.

4. Análise de Dados

A partir desta nova base de dados vai ser feito uma visualização separada por tipo de comentário organizando em, comentários sobre a loja, sobre o produto e que não está falando especificamente de nenhum dos dois. Depois dessa visualização vai ter outra visualização para mostrar se o comentário está falando positivamente ou negativamente, neste caso vai ser analisado apenas as categorias de loja e produto. No caso da categoria que não está falando especificamente de nenhum dos dois, não vai ser feita essa segunda análise.

Abaixo estão os links para acesso ao dataset da Amazon e outro relacionado:

1. <http://jmcauley.ucsd.edu/data/amazon/>

2. <https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews>