

Matching de Produtos

Contexto:

Um dos maiores desafios do e-commerce em plataformas de marketplace é a comparação de informações de produtos a partir de descrições textuais. Cada vendedor registrado na plataforma pode cadastrar os dados do produto de maneiras bem diversas, às vezes contendo informação incompleta, e mesmo a descrição mais básica – o nome que aparece na página da plataforma – não possui uma padronização.

A importância do agrupamento de produtos idênticos influencia desde estratégias de precificação e previsão de vendas até estratégias de busca e de recomendação. Ter produtos duplicados na base também dificulta o gerenciamento de estoque e a torna a execução de estratégias de marketing menos eficientes.

Problema:

O Magazine Luiza, sendo um dos maiores varejistas do país, possui um grande inventário de produtos em seu catálogo, muitos deles provenientes de vendedores externos que vendem em sua plataforma de marketplace.

A amostra de produtos disponibilizada no arquivo CSV contém algumas informações que foram fornecidas pelos vendedores no cadastro do produto. Essas informações são: o **nome** do produto, o **tipo**, a **categoria**, a **marca**, a **cor** e o **modelo**. O arquivo está separado por **TABs** (caractere ‘t’) e os campos que não foram preenchidos contém a palavra **None**.

Crie uma solução de matching usando os dados fornecidos. A solução deve responder às questões apresentadas abaixo:

- Dados dois produtos, qual a probabilidade de eles serem o mesmo produto?
- Dado um nome de produto, como recuperar as informações de **tipo**, **categoria**, **marca**, **cor** e **modelo**? Caso o produto não apresente a informação, o sistema deve retornar **None**.
- Como usar a solução de matching para retornar os produtos mais relevantes para uma busca simples de um cliente do site? Exemplo de busca: **notebook sony 4 GB**. Observação: adapte a solução de matching criada para resolver este item; não é necessário criar um sistema ou API.

Critérios de avaliação:

- Compreensão do problema e design de soluções;
- Conhecimento em processamento de linguagem natural e de modelos de machine learning;
- Capacidade de transmitir conhecimentos complexos de forma simples, ou seja: comunicação com pessoas que não possuem o mesmo nível acadêmico e entendimento sobre tecnologia.