

Atividade 01 - Definição do tema

Equipe

Nome: $\sin(args)^0$

Repositório GitLab da equipe: <https://gitlab.com/antfer/pdi>

Membros:

- Antonio Carlos Farias Ferreira, 2413868, antfer antfer@alunos.utfpr.edu.br, BSI, UTFPR

Tema

Descobrimiento de temas latentes em reviews de produtos da Amazon para inferência de nichos de consumo. Isso será feito a partir de uma análise dos textos de avaliação dos produtos para determinar os temas relacionados ao produto em si. Em seguida os temas inferidos serão associados ao usuário que enviou a avaliação como um tópico de interesse. A partir disso, deve-se determinar quais usuários são próximos ou têm interesses em comum, assim definindo um nicho de consumo.

Objetivos

Determinar que temas mantém os usuários em compras mais engajadas e verificar como a inserção em um determinado nicho de consumo impacta os padrões de consumo. Por exemplo, que nichos se intersectam e como os usuários migram entre os nichos ao longo do tempo.

Fontes de dados

Dados de avaliações de produtos da Amazon entre 1996 e 2023, fazendo um recorte temporal e de categorias a depender dos desafios encontrados. Por exemplo: a categoria de livros pode oferecer textos de avaliação mais densos, mas por ter muitas avaliações, talvez o processamento seja inviável. A princípio, irei usar a categoria de livros do kindle no período de 1996 à 2013. Os dados disponíveis para cada avaliação são: título do produto, preço do produto, usuário que realizou a avaliação, pontuação da avaliação, nota de satisfação, momento da avaliação e texto.

Desafios

Estabelecer perguntas relevantes, mas possíveis de serem respondidas dentro do escopo da matéria; construir um modelo que bem representam as comunidades a partir de somente os temas latentes das avaliações e da proximidade dos temas consumidos pelos usuários; Inferir os temas relacionados aos produtos a partir de suas avaliações.