

Datasets RecSys

Solicito que as informações sejam retiradas do mesmo período de análise, algo como 20 semanas de dados em diferentes cidades.

Dataset de Informações do Estabelecimento

Para algoritmos baseado em conteúdo, ter informações do conteúdo é fundamental para definir a similaridade entre os registros. Dessa forma é fundamental ter acesso às informações abaixo:

- **Nome** - O nome do estabelecimento as vezes faz referência a sua especialidade, franquia ou ambientes semelhantes (Texto)
- **Tipos de Categorias:** Categorias pré-definidas no sistema, como : Sushi, Doces & Bolos, Pizza... etc (Category)
- **Estimativa de Valor:** A estimativa de valor, definida no ifood como \$\$\$, contém informações importantes para definir categorias específicas: (Category)
- **Rating:** Rating médio do estabelecimento: (Decimal)
- **Tempo de Entrega:** Estimativa de Tempo de entrega (Decimal)
- **Cardápio:** Nome, descrição e valor dos pratos oferecidos no estabelecimento (Texto e Decimal)
- **Categoria do Cardápio:** No Ifood o cardápio é organizado por categorias. Ex: Bolo de Pote (Texto)
- **Horário de Funcionamento**
- **Localização:** Localização, pode ser o lat/Ing com pouca precisão (Decimal)
- **Fotos:** Fotos disponível na descrição do estabelecimento e dos pratos ofertados (Imagens)



Dataset de Pratos

Para algoritmos baseado em conteúdo, ter informações do pratos é fundamental para definir a similaridade entre os registros.

- **dish_id:** Identificar do prato (essa informação tem que bater com os dados de sessão para identificar qual pra foi comprado)
- **merchant_id:** ID do estabelecimento (essa informação tem que bater com o dataset de estabelecimentos)
- **nome:** Nome do prato
- **descrição:** Descrição em texto
- **preço:** Preço
- **tempo:** tempo de entrega
- **opções:** Opções de adicionais
- **url_image:** Se possível o link acessível da imagem do prato.

Dataset de Estabelecimento Review

Eventos de preferências explícitas com relação ao estabelecimento, contendo os dados cadastrados das avaliações dos restaurantes. Essa informação é interessante por ser uma preferência explícita e por conter dados categóricos (tags), rating, e texto livre.

Um dataset interessante seria algo como:

- **account_id**: ID do usuário
- **merchant_id**: ID do estabelecimento que ele acessou
- **rating**: Nota (k estrelas)
- **comentário**: Texto livre escrito pelo usuário
- **event_timestamp**: Timestamp do evento
- **tags**: Tags adicionadas associadas a pergunta “Quais itens você acha que podem melhorar?” como “Tempo de Entrega”, “Embalagem”



The screenshot shows a mobile app interface for rating a delivery order. At the top, there's a status bar with icons for signal, Wi-Fi, battery (23%), and time (17:05). Below that, a red back arrow and the word "AVALIAÇÃO" are visible. The main heading is "Avalie seu pedido" followed by the date and location: "29/07/2019 • Brásileirinho Delivery - Jardim América". A horizontal line separates this from the order details: "1x Box picanha suína" and "R\$ 18,99". Another horizontal line follows. The next section is titled "Como foi a sua experiência?" with the instruction "Escolha de 1 a 5 estrelas para classificar." Below this are five yellow stars, all of which are filled. The following section is titled "O pedido chegou no tempo informado?" with the instruction "Previsão de entrega: 60-70min". Below this are two radio button options: "Sim, chegou." and "Não, atrasou.". At the bottom, there is a large, light orange button labeled "Avaliar".

Dataset de Sessão/Interação

A sessão é definida como um tempo T em que o usuário entrou no aplicativo, buscou (acessou diversos restaurantes) e comprou (nem sempre termina na compra). Ter acesso à sessão do usuário é importante para modelar com base no comportamento a curto prazo. Um dataset interessante seria algo como:

- **session_id**: ID único da sessão do usuário
- **account_id**: ID do usuário
- **merchant_id**: ID do estabelecimento que ele acessou
- **dish_id**: Qual foi o prato comprado ou visualizado
- **click_timestamp**: Timestamp do acesso
- **buy**: Binário, se nessa ação ele efetivou a compra
- **informações adicionais**: Informações adicionais no momento do pedido que podem variar, como preço, frete(grátis), tempo de preparo.

Dessa forma modelar os passos do usuário até ele realizar a compra. Dependendo do dinamismo do problema, essa será uma solução bem interessante no sentido de prever o próximo restaurante nessa série temporal de busca.

