

## Interações entre Medicamentos, Alimentos e Bebidas: Dados e Abordagem do Modelo

Construir um modelo para prever ou explicar interações adversas entre **medicamentos** e **alimentos/bebidas** (considerando fatores como ingestão em **jejum ou alimentado** e horário **manhã/tarde/noite**) requer um conjunto de dados abrangente e uma estratégia bem definida. A seguir apresentamos recomendações de **datasets**, técnicas de **Análise Exploratória de Dados (EDA)**, e um **roadmap de pipeline** (do preparo dos dados até a produção) para esse projeto.

### Fontes de Dados (Datasets) para Interações Fármaco-Alimento/Bebida

- **Bulário Eletrônico da ANVISA** (Brasil): Contém bulas oficiais (textos completos) de ~7 mil medicamentos registrados no Brasil <sup>1</sup>. Essas bulas incluem seções importantes (ex: **composição, modo de uso, advertências, contraindicações, interações medicamentosas** etc). Por exemplo, é comum a bula alertar sobre ingestão com álcool ou orientações de tomar em jejum. **Vantagem:** dados em português e focados nos medicamentos disponíveis localmente (ANVISA). **Desafio:** As bulas estão em PDF e **não estruturadas** <sup>1</sup>, exigindo extração via processamento de linguagem natural. Já houve iniciativas para estruturar essas bulas – *e.g.*, o projeto *Fácil Bula* mapeou informações das bulas ANVISA para bases como o DrugBank, armazenando-as em um banco de dados grafo (Neo4j) <sup>2</sup>. Isso indica ser viável extrair interações e composições das bulas e estruturá-las.
- **DrugBank** (base de dados internacional): Reúne informações farmacológicas de milhares de fármacos, incluindo uma seção específica de **“Food Interactions”** em cada registro de fármaco <sup>3</sup>. Esse campo descreve como certos alimentos ou bebidas afetam o medicamento – por exemplo, a entrada do *paracetamol* (acetaminofeno) especifica **evitar álcool** devido ao risco de hepatotoxicidade <sup>4</sup>. O DrugBank categorizou *68 tipos de interações* droga-alimento diferentes e inclui descrições textuais para cada interação <sup>5</sup>. Um levantamento mostrou ~2.388 registros de interações alimentares distintas envolvendo cerca de 498 descrições únicas nessa base <sup>5</sup>. **Vantagem:** dados estruturados (disponíveis via licença acadêmica ou versão aberta limitada) e em inglês, cobrindo interações conhecidas globalmente. **Uso:** Podemos extrair tabelas de *food interactions* do DrugBank para obter pares *fármaco - alimento/bebida - descrição da interação*. Como estão em inglês, seria necessário traduzir para PT ou treinar o modelo multilinguisticamente, caso queiramos respostas em português (ver seção de Pipeline).
- **FooDrugs** (Food-Drug Interaction Database): Um dataset recente, aberto, compilado por pesquisadores, que unifica informações textuais de interações alimento-fármaco de múltiplas fontes. A versão 2 dessa base registrou **6.422 interações alimento/fármaco** extraídas de 2.849 documentos (incluindo ~2.312 artigos PubMed, 285 entradas do DrugBank e 252 da base Drugs.com) <sup>6</sup>. Essas interações envolvem **1.464 tipos de alimentos/substâncias** e **3.009 medicamentos** <sup>6</sup>, cobrindo uma vasta gama de combinações possíveis. **Vantagem:** disponível publicamente (via Zenodo), já consolidado de fontes diversas, incluindo literatura científica e bases online. **Observação:** O FooDrugs inclui interações *potenciais* (mapeadas por mineração de texto e dados moleculares) <sup>7</sup>. Isso amplia a cobertura, mas é importante filtrar as interações

bem estabelecidas clinicamente para treinar o modelo (evitando viés de interações apenas teóricas).

- **Outras fontes internacionais:**

- *Bases de dados de rótulos/bulas internacionais:* por exemplo, o **DailyMed** (FDA/EUA) fornece bulas em XML estruturado, que podem ser consultadas para campos como “*Drug-Food Interactions*” ou recomendações de administração (com ou sem alimentos). Também a EMA (agência europeia) publica relatórios (EPARs) com seções sobre interações.
- *Sites farmacêuticos e referências clínicas:* **Drugs.com**, **MedlinePlus**, **Micromedex**, etc., frequentemente listam interações alimentares importantes por medicamento. Embora não ofereçam um *dataset* para download, podem ser usados para validação cruzada ou extração específica de alguns exemplos (via *web scraping* se necessário, respeitando termos de uso).
- *Publicações e literatura:* Revisões ou artigos (ex: manuais de farmacologia, diretrizes) que listam interações fármaco-nutriente conhecidas podem complementar o dataset, especialmente para consolidar os mecanismos envolvidos.

**Resumo da estratégia de dados:** Idealmente, iniciar pela **base ANVISA** para garantir relevância local (mesmos medicamentos e idioma do usuário). Se a extração das bulas mostrar-se incompleta ou trabalhosa, complementar com dados estruturados do **DrugBank** e recursos como **FooDrugs** para cobrir o maior número de interações. Isso nos daria uma lista robusta de trios (*fármaco, agente externo, efeito*) para alimentar o modelo. A fusão de fontes exige padronizar nomes (por exemplo, mapear nomes comerciais brasileiros ao nome do princípio ativo usado no DrugBank). Uma vez compilados os dados, podemos criar um dataframe com colunas como: **Medicamento**, **Substância ou Alimento/Bebida**, **Condição (jejum ou alimentado)**, **Horário recomendado** (se houver), **Descrição do efeito/risco** e **Severidade** (se disponível).

## Agentes Externos (Alimentos, Bebidas) que Interferem nos Medicamentos

Para “abranjer todos”, devemos definir claramente quais **categorias de alimentos e bebidas** o modelo considerará. A seguir, listamos os principais *agentes externos* conhecidos por interagir com fármacos, conforme literatura farmacêutica:

- **Álcool (bebidas alcoólicas):** Interação muito comum e relevante. O álcool pode potencializar a sedação de remédios depressores do SNC (ex.: benzodiazepínicos como o Rivotril/clonazepam) e sobrecarregar o metabolismo hepático de vários fármacos. Muitas bulas alertam para **não ingerir álcool** durante o tratamento – por exemplo, a bula do paracetamol (acetaminofeno) adverte evitar álcool para prevenir danos ao fígado <sup>4</sup>. Com antibióticos, o álcool pode reduzir a eficácia e aumentar efeitos adversos <sup>8</sup>. **Status de jejum:** ingerir álcool de estômago vazio intensifica sua absorção, podendo agravar essas interações (vale incluir essa distinção no modelo).
- **Cafeína (cafés, chás, energéticos):** Embora não tão frequentemente proibida quanto o álcool, a cafeína pode interagir com certos medicamentos. Por exemplo, antibióticos como ciprofloxacina e alguns antidepressivos podem prolongar os efeitos da cafeína. Já combinada a estimulantes (ex.: alguns descongestionantes ou anfetaminas), pode causar nervosismo excessivo. O modelo deve reconhecer que cafeína pode **potencializar efeitos estimulantes** ou **interferir no sono** quando combinada com medicações que já afetem o sistema nervoso.

- **Sucos cítricos, especialmente Toranja (Grapefruit):** O suco de toranja é clássico em interações – ele **inibe enzimas hepáticas (CYP3A4)**, podendo elevar os níveis de diversos medicamentos no sangue <sup>9</sup>. Consequências: efeitos exacerbados e mais efeitos colaterais. Interage notoriamente com:
  - *Antihipertensivos e imunossupressores:* a toranja pode aumentar a concentração, causando hipotensão excessiva ou toxicidade <sup>10</sup>.
  - *Estatinas* (atorvastatina, sinvastatina, etc.): suco de grapefruit aumenta os níveis e risco de efeitos adversos (miopatia) <sup>11</sup>.
  - *Digoxina:* toranja pode elevar níveis séricos, risco de intoxicação digitalis <sup>12</sup>.
- **Recomendação:** “Evitar suco de toranja/grapefruit durante o tratamento” – o dataset deve conter tais avisos <sup>10</sup>.
- **Leite e Laticínios (cálcio):** Produtos ricos em **cálcio** podem **quelar** certos fármacos, reduzindo sua absorção. Exemplos clássicos:
  - *Antibióticos tetraciclina e quinolonas* (ciprofloxacino, etc.): devem ser tomados longe de laticínios, pois o cálcio **precipita o fármaco** no GI e diminui a absorção em até 50% ou mais <sup>13</sup>. As bulas recomendam administração **1h antes ou 2h depois das refeições/leite** <sup>14</sup>.
  - *Suplementos de ferro e Levotiroxina:* também têm absorção reduzida por lácteos e outros alimentos minerais – usualmente indicam tomar em jejum com água.
- O modelo precisa capturar instruções de “**tomar em jejum**” ou “**não tomar com leite/antiácidos**” como sinal de interação alimento-medicamento.
- **Alimentos Ricos em Vitamina K:** *Vegetais verdes folhosos* (brócolis, espinafre, couve, etc.) contêm vitamina K, que **antagoniza anticoagulantes cumarínicos** (ex.: Varfarina). Ingestão alta ou variável de vitamina K reduz o efeito do anticoagulante e pode predispor à coagulação <sup>15</sup>. A recomendação não é necessariamente cortar esses alimentos, mas manter consumo consistente <sup>16</sup>. Para nosso modelo, o importante é saber que **grandes quantidades de vegetais verdes diminuem eficácia da varfarina** – uma interação alimentar crítica a sinalizar.
- **Alimentos Ricos em Tiramina:** Certos alimentos fermentados ou envelhecidos (queijos maturados, embutidos, fígado, extrato de levedura, vinho tinto, algumas cervejas, etc.) possuem tiramina. Com medicamentos **inibidores da MAO** (monoaminooxidase), como antidepressivos antigos (fenelzina, tranilcipromina), ocorre acúmulo de tiramina no corpo, levando a uma **crise hipertensiva potencialmente fatal** <sup>17</sup>. As bulas desses medicamentos destacam **dieta restrita em tiramina**. Nosso dataset deve listar esses alimentos disparadores e associá-los aos fármacos MAOIs, com descrição do risco (p.ex. “**pode causar aumento perigoso da pressão**” <sup>17</sup>).
- **Fibras e Alimentos Abundantes em Fibras:** Uma dieta muito rica em fibras pode **atrapalhar a absorção** de determinados medicamentos. Exemplo: grandes quantidades de farelo (fibra solúvel, como em *farinha de aveia*) reduzem a absorção da **digoxina**, diminuindo seu efeito <sup>18</sup>. Assim, orienta-se não tomar digoxina junto com comidas ricas em fibra. O modelo deve aprender que **fibra em excesso pode diminuir a eficácia de certos fármacos**.
- **Refeições Ricas em Gordura:** Alimentos gordurosos podem tanto aumentar a absorção de fármacos lipofílicos (às vezes levando a picos mais altos) quanto retardar o esvaziamento gástrico. Ex.: a absorção de medicamentos como itraconazol ou certos antivirais é **melhorada**

**com refeição gordurosa** (recomendado tomar com alimento para melhorar biodisponibilidade), enquanto outras vezes refeições pesadas podem atrasar o efeito de analgésicos de ação rápida. O dataset deve codificar casos onde **“tomar com alimento” aumenta eficácia ou tolerabilidade** e casos onde **“evitar refeições pesadas”** é indicado.

- **Outros:** Diversos outros itens podem constar:
- **Suplementos herbais** (erva de São João, ginseng, etc.) – embora não exatamente “alimento”, podem induzir ou inibir metabolismo de drogas.
- **Sucos e frutas específicas:** p.ex. suco de **laranja ou maçã** pode reduzir absorção de alguns antialérgicos (fexofenadina) por afetar transportadores.
- **Chocolate** (teobromina/cafeína) pode interagir com antidepressivos inibidores de MAO semelhante à tiramina.
- **Tabaco** (nicotina) e até **cigarro** – não é alimento, mas citado pois substâncias do fumo induzem enzimas que reduzem efeito de drogas como teofilina <sup>19</sup>. (É um “agente químico externo” relevante, embora fora do eixo alimento/bebida estrito, podemos incluir pelo menos como nota no conhecimento do modelo.)

Essa **lista de agentes externos** deve ser refinada e usada para anotar/criar features no dataset. Por exemplo, podemos ter categorias binárias ou tags para: *álcool, cafeína, dairy (cálcio), fibra, alta gordura, toranja, vitamina K, tiramina, etc.*. Assim, ao treinar o modelo, ele reconhece padrões de interação associados a cada categoria.

## Análise Exploratória de Dados (EDA) sobre Interações

Antes de treinarmos qualquer modelo, é fundamental realizar uma EDA minuciosa no dataset compilado. Alguns passos e técnicas recomendadas:

- **Distribuição de Interações por Categoria:** Contar quantas interações de cada tipo de alimento/bebida existem. Por exemplo, quantos medicamentos têm interação com álcool versus quantos com toranja ou leite. Isso revela quais agentes são mais recorrentes. Estudos prévios mostram que **álcool** e **toranja** tendem a aparecer com alta frequência nas descrições <sup>20</sup> (álcool muitas vezes em “evitar álcool”, e toranja presente em várias classes). Uma visualização de barras ou um *histograma* de frequências por agente externo pode ser útil.
- **Interações por Classe de Medicamento:** Usando informações de classe/função dos medicamentos (disponíveis no DrugBank ou anotadas nas bulas), podemos agrupar os fármacos e ver padrões. Exemplo: *estatinas* e *anti-hipertensivos* compartilhando a interação com suco de toranja <sup>11</sup> <sup>21</sup>; *antibióticos* vários indicando evitar leite (tetraciclina, quinolona) e álcool (metronidazol e outros). Isso ajuda a validar se o dataset está consistente com o esperado: p.ex., **todos os MAOIs no conjunto devem mostrar interação com tiramina** – se algum faltar, pode indicar dado faltante.
- **Análise de Texto das Descrições:** As descrições de efeitos adversos/interações geralmente são texto livre (ou semi-estruturado). Aplicar técnicas de NLP simples pode trazer insights:
- Faça uma nuvem de palavras (*word cloud*) ou contagem de termos mais frequentes nas descrições. Provavelmente veremos termos como “evitar”, “aumenta efeito”, “reduz absorção”, “risco”, “efeitos colaterais”, etc. No projeto de análise do DrugBank, os termos alimentares mais frequentes incluíram “álcool”, “food” (refeições), “grapefruit”, “milk”\*, etc <sup>20</sup> <sup>22</sup>.

- Extrair bigramas/trigramas comuns: ex.: “evitar o consumo de álcool”, “tomar em jejum”, “tomar junto com alimentos” – isso mostra recomendações padrão que o modelo deverá aprender a reproduzir.
- Identificar descrições duplicadas ou muito similares. Por exemplo, várias drogas podem ter a mesma frase “Não ingerir com suco de toranja” – confirmações assim indicam que podemos generalizar essa interação.
- **Deteção de Outliers ou Lacunas:** Procurar medicamentos que tenham  *muitas* interações listadas versus os que têm nenhuma. Um fármaco polimedicado (ex.: varfarina) possui múltiplas interações (vitamina K, muitos medicamentos, álcool) – se nosso dataset só focou em alimento/bebida, varfarina deve aparecer pelo menos com vitamina K e álcool. Se um medicamento conhecido por interações não tiver nenhuma listada, pode ser que a fonte (bula) não menciona ou houve falha na extração – vale checar manualmente alguns casos. A EDA também pode revelar interações inesperadas (outliers) citadas raramente, e podemos decidir se incluí-las (se relevantes clinicamente) ou filtrar (se forem obscurecidas ou sem importância prática).
- **Visualizações adicionais:**
  - Uma matriz (heatmap) de medicamentos vs. alimentos marcando onde há interação pode ser informativa, mas seria enorme se incluirmos todos. Alternativamente, um  *grafo* simples ilustrando conexões: nós representando fármacos ligados a nós de alimentos. A densidade de conexões de um nó indica grau de interações. Essa análise de grafo pode mostrar, por exemplo, que “álcool” conecta-se a muitos medicamentos (nó altamente conectado), enquanto “toranja” conecta-se principalmente a subgrupos específicos (estatinas, etc.).
  - Plotar distribuições de quantas interações por medicamento e quantos medicamentos por interação alimentícia.
- **Consistência e Qualidade dos Dados:** Durante a EDA, também limpamos dados:
  - Padronizar nomes (remover duplicatas de um mesmo fármaco com nome levemente diferente, corrigir grafias de alimentos).
  - Remover entradas irrelevantes: o DrugBank, por exemplo, contém descrições como “No food effects” <sup>23</sup> – essas devem ser filtradas, pois não são interações. Já identificamos essas frases redundantes e as eliminaremos do conjunto de treinamento <sup>24</sup> .
  - Verificar se o atributo “tempo de administração” (manhã/noite) aparece: poucas bases mencionam explicitamente horário, mas as bulas às vezes dizem “preferencialmente à noite” (p.ex., anti-histamínicos sedativos) ou “de manhã” (p.ex., diuréticos). Essas informações podem não estar estruturadas, mas vale procurar na mineração de texto por palavras-chave como “pela manhã”, “à noite”. Se obtidas, podemos criar uma feature indicativa (0/1 para manhã/noite recomendada) para ver quantos medicamentos possuem essa recomendação.

Ferramentas recomendadas para EDA incluem **Python (pandas)** para manipulação tabular, **matplotlib/Seaborn** para gráficos, e bibliotecas de NLP como **NLTK** ou **spaCy** para análise textual básica. A EDA deve ser feita em um ambiente interativo (Jupyter Notebook, por exemplo) para iterar rapidamente em perguntas sobre os dados. Documente bem esses achados, pois eles guiarão as próximas etapas (modelagem e ajustes do dataset).

## Pipeline de Modelagem e Produção (Roadmap)

Com os dados preparados e insights da EDA, delineamos um pipeline completo – desde o treinamento do modelo até a disponibilização em produção. Abaixo está um **roadmap** com etapas chave e as **melhores práticas/ferramentas** em cada fase:

1. **Coleta e Integração de Dados:** Utilize a linguagem/ferramenta de sua preferência (Python é bastante adequado pela disponibilidade de bibliotecas).
2. Para as **bulas ANVISA**: Empregue técnicas de *web scraping* ou APIs se houver. Por exemplo, baixar os PDFs do Bulário Eletrônico e usar uma biblioteca (como **PyMuPDF**, **PDFPlumber** ou **OCR** se necessário) para extrair o texto. Em seguida, usar expressões regulares ou algoritmos de NLP para localizar seções de interesse (buscando por palavras como “Interações” ou “Advertências” dentro do texto da bula).
3. Para **DrugBank/FooDrugs**: Se disponível em formato CSV/SQL, importar diretamente (usando `pandas.read_csv` ou conexão SQL). Caso contrário, a API do DrugBank (necessita chave) pode ser usada para obter as interações alimentares por fármaco.
4. **Unificação:** Ter uma estratégia de *merge* entre fontes. Idealmente, trabalhar no nível de **ingrediente ativo** (princípio ativo). Por exemplo, “Rivotril” mapear para “Clonazepam”. Podemos montar um dicionário de equivalência a partir das próprias bulas (que normalmente listam o nome do fármaco). Assim unificamos interações por substância. Se múltiplas fontes falam da mesma interação, consolidar ou verificar consistência das descrições.
5. **Limpeza e Pré-processamento:** Aplique técnicas de **ETL** nos dados brutos:
  6. Normalizar **nomes** (caixas alta/baixa, remover acentuação se necessário), tanto de medicamentos quanto de alimentos.
  7. Criar campos categóricos para *tipo de interação* ou *severidade*, se possível. Por exemplo, no DrugBank cada interação tem um tipo categorizado (68 tipos distintos, como “*Metabolism/PK interaction*”, “*QT prolongation*”, etc.)<sup>25</sup>. Se disponíveis, essas categorias podem ser usadas como **labels** para classificação.
  8. Gerar **features** binárias/indicadoras conforme discutido (álcool, toranja, etc.) e outras contextuais: *jejum* (se descrição contém “tomar em jejum” ou “com alimentos”), *horário* (se contém “à noite/manhã”).
  9. Dividir a coluna de descrição em partes se útil – por exemplo, isolar a consequência física vs. recomendação. Muitas descrições já vêm no formato “Faça X porque Y”. Podemos separar: *Ação recomendada* (evitar/ingerir com ou sem alimento) e *Motivação/efeito* (aumenta absorção, causa sonolência, etc.).
10. Remover duplicatas exatas e linhas irrelevantes (“sem interação conhecida”).
11. **Definição da Tarefa de Modelagem:** Com o dataset estruturado, definimos o objetivo do modelo:
  12. Pode ser **Classificação binária** (há ou não interação adversa para um par medicamento-alimento dado?),
  13. ou **Classificação multiclasse/multirrótulo** (qual tipo de efeito ocorre – e.g., classes: *reduz absorção*, *aumenta toxicidade*, *causa sedação*, etc., podendo ter múltiplas por par),
  14. ou até **Geração de texto** (produzir uma breve frase explicando a interação).

15. Dado o exemplo fornecido ( `rivotril + vodka + noite + jejum -> consequência física/colateral, impacto no sono` ), parece desejar uma **saída textual descritiva**. Isso sugere uso de **modelos de NLP** que possam concatenar o conhecimento e gerar explicações. Uma abordagem segura seria gerar a resposta a partir de **templates preenchidos com os dados** (por exemplo: “Combinar [Medicamento] com [Agente] à [horário] em jejum pode [Efeito]”), usando o dataset como base das frases. Outra abordagem mais avançada é treinar um **modelo de linguagem** para essa tarefa de geração condicionada.

#### 16. Seleção do Modelo e Framework:

17. Para **modelos preditivos estruturados** (classificação/regressão), bibliotecas como **scikit-learn** (Random Forest, XGBoost, etc.) funcionam bem em datasets tabulares. Features poderiam incluir vetores que representam medicamento (p.ex., embeddings aprendidos a partir de descrições das drogas), flags de agentes, etc., para prever a presença/ausência de interação ou severidade.

18. Para **modelos de NLP/geração**: Aproveitar frameworks como **Hugging Face Transformers**. Podemos, por exemplo, fine-tunar um modelo *seq2seq* (como T5 ou BART multi-língua) onde a entrada seria uma string estruturada (“Medicamento: X; Alimento: Y; Contexto: jejum/noite”) e a saída a frase de aconselhamento. Teríamos que fornecer ao modelo vários exemplos do dataset durante o fine-tuning. Dado que o volume de interações identificadas é na ordem de algumas milhares, um modelo pré-treinado pode generalizar bem. *Observação*: Caso os dados de treino estejam majoritariamente em inglês (ex: DrugBank), podemos primeiro traduzir para português ou treinar o modelo multi-língua diretamente com um conjunto bilíngua (fontes em PT e EN) para que ele aprenda termos técnicos nas duas línguas.

19. Uma alternativa leve é usar um **banco de conhecimento** e lógica determinística para este MVP: por exemplo, armazenar as interações conhecidas num banco de dados e simplesmente consultar quando o usuário fizer a pergunta. Isso garantiria 100% precisão em não “alucinar” interações inexistentes. Porém, treinar um modelo tem a vantagem de lidar com combinações novas ou inferir efeitos não explícitos (p.ex., se conhece que remédio causa sonolência e álcool causa sonolência, inferir aumento da sonolência).

#### 20. Avaliação do Modelo:

21. Se for classificação, dividir o dataset em treino/validação/teste e usar métricas como **accuracy**, **F1-score** (especialmente se classes desbalanceadas) ou **ROC-AUC**.

22. Se for geração de texto, avaliar **BLEU score** ou Rouge vs. frases esperadas. Porém, como as descrições podem ter múltiplas formas corretas, também é útil uma avaliação humana ou baseada no conteúdo (por exemplo, checar se as palavras-chave certas aparecem na saída).

23. Garantir que o modelo **não perca interações críticas** (melhor ter falso-positivos que omitir um alerta importante). Uma estratégia é calibrar o modelo para ser conservador nas recomendações de evitar combinações quando incerto.

24. **Considerações de Multilinguagem**: Conforme indicado (item 1 do usuário: “Qualquer linguagem ... + tradução”), a aplicação final deve **detectar o idioma do usuário** e responder naquele idioma.

25. Podemos implementar a detecção de idioma no front-end ou na camada de API (há bibliotecas Python como **langdetect** para isso).

26. Se o modelo for treinado só em português (usando bulas ANVISA), ele naturalmente servirá bem aos usuários PT. Para outras línguas, integrar um serviço de tradução. Por exemplo, usar a API

do Google Translate ou um modelo open source (como **MarianMT** da HuggingFace) para traduzir a pergunta do usuário para português (alimentar o modelo), e depois traduzir a resposta PT de volta ao idioma do usuário. Isso permitiria a abrangência global sem precisar treinar modelos separados por idioma.

27. Em longo prazo, treinar diretamente um modelo multilíngue com dados traduzidos poderia melhorar a fluência das respostas em outros idiomas sem depender de APIs externas.

**28. Armazenamento e Serviço dos Dados/Modelo:**

29. Durante o desenvolvimento, usar arquivos CSV/JSON ou um banco SQL para manipular o dataset está ok. Mas para produção, considere uma solução que facilite consultas rápidas. Um **banco de dados de grafos** (como Neo4j, utilizado no projeto *Fácil Bula* <sup>2</sup>) é interessante, pois permite query do tipo: “*dado fármaco X, quais interações possui e com quais agentes?*” de forma muito eficiente. Poderíamos armazenar nós de tipo **Medicamento** e nós de tipo **SubstânciaAlimento**, com arestas rotuladas como **INTERAGE\_COM** contendo propriedades (descrição, severidade, etc.). Esse grafo pode ser consultado via API para gerar respostas dinâmicas também.

30. Se optarmos por modelo ML puro, então o armazenamento é focado no modelo em si. Use ferramentas como **Pickle/joblib** para salvar modelos scikit, ou se for um Transformer, salvar o *checkpoint* treinado. Documente a versão do modelo e data do dataset (versionamento de dados com **DVC** ou similar pode ser útil para reprodutibilidade).

31. **Implementação do Serviço (API):** Para expor a funcionalidade do modelo, um **web service RESTful** é o caminho. Frameworks Python como **FastAPI** ou **Flask** permitem construir endpoints para receber o input (medicamento + alimento + contexto) e retornar a resposta.

32. Pipeline exemplo no endpoint:

1. Recebe JSON com {medicamento: "X", agente: "Y", horario: "noite", jejum: true, lingua: "pt"}.
2. Normaliza/valida os valores (p. ex., converte "rivotril" em "clonazepam" via nosso dicionário de mapeamento, "vodka" em "álcool").
3. Consulta o modelo ou banco de conhecimento. Se for modelo de geração, formata a entrada e obtém texto gerado. Se for banco de dados, recupera a descrição armazenada.
4. Se a língua solicitada não for português, traduz a resposta.
5. Retorna a resposta formatada (possivelmente incluindo também um indicativo de severidade ou recomendações extras).

33. **Teste:** garantir com alguns casos conhecidos (ex.: Rivotril + álcool, antibiótico + leite, MAOI + queijo) que o sistema responde corretamente.

34. **Desempenho:** para produção, considerar desempenho. Modelos grandes de NLP podem ser lentos; se a opção for um Transformer, talvez utilizá-lo via **PyTorch serve** ou converter para **ONNX** format para otimizações. Como alternativa, gerar respostas previamente para as combinações mais comuns e colocar em cache (embora as combinações possam ser muitas).

35. **Escalabilidade:** Containerizar a aplicação (Docker) facilita o deploy. Orquestrar com Kubernetes ou similar se esperar grande volume de requisições. Monitorar uso e latência.

36. **Manutenção e Atualização:**



37. Novos medicamentos são lançados e novas interações descobertas continuamente. Portanto, planeje uma rotina de atualização de dados (por exemplo, verificar trimestralmente o Bulário da ANVISA por novas bulas ou mudanças).
38. Se houver uma mudança significativa (ex.: um fármaco passa a ter contraindicação com certo alimento), atualizar o dataset e **re-treinar** ou **atualizar a base de conhecimento**. Usar ferramentas de **versionamento** para acompanhar melhorias no modelo.
39. Logar as consultas dos usuários (anonimamente) em produção pode revelar padrões de busca não cobertos – e isso retroalimenta a necessidade de expandir o modelo. Por exemplo, se muitos perguntam sobre interação de medicamento com *café* e não prevemos isso explicitamente, podemos incluir cafeína de forma mais robusta na próxima versão.

Em resumo, a arquitetura de dados e modelo deve ser robusta e escalável: desde a **engenharia dos dados** (integrando ANVISA + bases globais), passando por uma **EDA cuidadosa** para direcionar o modelo, até a escolha de um **framework** adequado (modelos de linguagem pré-treinados para texto ou abordagem baseada em grafo/consultas). Com esse roadmap, conseguimos ir do desenvolvimento inicial até a implementação em produção de um sistema capaz de responder, por exemplo, que *“Clonazepam (Rivotril) + álcool (vodka) à noite em jejum” pode potencializar a depressão do sistema nervoso central, levando a sedação profunda, prejuízo da coordenação e riscos à respiração e ao sono – devendo ser evitado*, e fazer isso de forma automatizada e consistente.

#### Referências Utilizadas:

- ANVISA Bulário Eletrônico – ~6.961 bulas em PDF, informações não estruturadas <sup>1</sup>. Desafio de extrair automaticamente dados de interações, como feito pelo projeto *Fácil Bula* (mapeando bulas ANVISA a DrugBank e estruturando em Neo4j) <sup>2</sup>.
- DrugBank – Base internacional com campo específico de interações com alimentos. Interações podem alterar absorção, aumentando ou diminuindo efeitos ou causando **efeitos adversos** <sup>26</sup>. Exemplo: entrada do Acetaminofeno indica evitar álcool <sup>4</sup>.
- FooDrugs – Dataset aberto compilado (2022) contendo **6422 interações alimento-fármaco** extraídas de literatura, DrugBank e Drugs.com <sup>6</sup>. Cobre ~1464 alimentos/substâncias e ~3009 fármacos.
- Manual MSD de Farmacologia – Exemplos de interações: **Suco de toranja** inibe metabolismo e intensifica efeito de várias drogas <sup>9</sup>; Alimentos ricos em **tiramina** + IMAO podem causar crise hipertensiva <sup>27</sup>; **Laticínios (cálcio)** reduzem absorção de tetraciclina/quinolona <sup>13</sup>; Vegetais com **vitamina K** diminuem eficácia da varfarina <sup>15</sup>; Fibra em excesso (aveia) prejudica digoxina <sup>18</sup>.
- Artigo informativo TuaSaúde – Recomendações práticas, por ex: **evitar álcool com antibióticos** (risco de toxicidade e perda de efeito) <sup>8</sup>; evitar *grapefruit* com estatinas e anti-hipertensivos <sup>11</sup> <sup>21</sup>, etc., corroborando os pontos considerados no dataset.

---

<sup>1</sup> <sup>2</sup> Portal eduCapes: Fácil bula: sistema que estrutura o bulário eletrônico da Anvisa  
<https://educapes.capes.gov.br/handle/capes/666054?mode=full>

<sup>3</sup> <sup>4</sup> <sup>26</sup> Food interaction | DrugBank Help Center  
<https://dev.drugbank.com/guides/terms/food-interaction>

<sup>5</sup> <sup>20</sup> <sup>22</sup> <sup>23</sup> <sup>24</sup> GitHub - kedith/food-interacton-analysis: An analysis of the food interaction data from Drugbank  
<https://github.com/kedith/food-interacton-analysis>

- 6 7 FooDrugs database: A database with molecular and text information about food - drug interactions  
<https://zenodo.org/records/8192515>
- 8 10 11 12 21 Interações entre remédios e alimentos: quais são e como evitar - Tua Saúde  
<https://www.tuasaude.com/interacao-medicamentosa/>
- 9 13 14 15 16 17 18 19 27 Interações medicamentosas - Medicamentos - Manual MSD Versão Saúde para a Família  
<https://www.msmanuals.com/pt/casa/medicamentos/fatores-que-afetam-a-resposta-a-medicamentos/inter%C3%A7%C3%B5es-medicamentosas>
- 25 DrugBank 6.0: the DrugBank Knowledgebase for 2024  
<https://academic.oup.com/nar/article/52/D1/D1265/7416367>