**Cenário Fictício para apresentação de trabalho de criação de api em Python para machine learning com dados obtidos do site http://vitibrasil.cnpuv.embrapa.br/  
  
Diagrama em andamento em: https://app.mural.co/t/albani7053/m/albani7053/1716660061141/d6594a4e452e806fcaf10853022011bdb385ae7a?sender=uf7449aba5082740e08ea2944**

**Contexto**

A Embrapa está interessada em oferecer acesso público a seus dados de viticultura por meio de uma API. Os dados fornecidos são sobre produção de vinhos, produção de uva, processamento, comercialização, importação e exportação de vinho e uvas do Brasil com o objetivo de fornecê-los a pesquisadores, viticultores e ao público em geral e também servirá como base para futuros modelos de Machine Learning que podem utilizar esses dados para análises e previsões.

A API deve acessar sempre que solicitado os dados disponíveis no site da Embrapa, armazená-los em um banco de dados MongoDB. Deve possui endpoints para que os usuários possam consultar esses dados de forma performática e organizada.

Além disso, a API deverá utilizar OAuth2/JWT para autenticar e permitir acesos de usuários registrados.

**Cenário Atual (AS-IS)**

Atualmente, os dados de vitivinicultura da Embrapa estão disponíveis apenas por meio do seu site oficial, o que pode dificultar o acesso e a análise sistemática desses dados por parte de pesquisadores e outros interessados.

Além disso, a falta de uma API pública limita a integração desses dados com outros sistemas, como modelos de Machine Learning, que poderiam extrair insights valiosos e previsões baseadas em dados.

**Escopo**

O que a solução deverá oferecer?

Proposição de uma solução que permita a ingestão, armazenamento e disponibilização de dados da Vitibrasil da Embrapa de forma eficiente e segura. A solução deverá incluir:

* **Desenvolvimento de uma REST API em Python (FastAPI):**
  + Endpoints para consulta de dados de produção, processamento, comercialização, importação e exportação.
  + Endpoints para importação e atualização de dados conforme solicitado pelo usuário.
  + Documentação completa utilizando Swagger.
* **Sistema de Aquisição:**
  + Coleta automatizada de dados do site da Embrapa.
  + Atualização automática do banco de dados com os novos dados coletados.
  + Padronização dos dados obtidos.
* **Banco de Dados (MongoDB):**
  + Armazenamento dos dados coletados de forma estruturada e eficiente.
* **Serviço de Autenticação (OAuth2/JWT):**
  + Gerenciamento de autenticação e autorização dos usuários da API.
* **Modelo de Machine Learning (Futuro):**
  + Endpoint para consumo dos dados armazenados para treinamento e predições.