Tech Challenge – Fase 01

Machine Learning Engineering | FIAP

Rodrigo Matheus da Silva (rodrigorizando@gmail.com)

Vitor Efigênio Neto (vitorefigenio@gmail.com)

Descritivo do Projeto

São Paulo 2025

Resumo do Projeto: Consulta à Embrapa via API

Este projeto foi desenvolvido como parte do **Tech Challenge 1** da Pós-Graduação em Engenharia de *Machine Learning* da FIAP.

O objetivo principal foi criar uma solução para consulta automatizada ao site da Embrapa (http://vitibrasil.cnpuv.embrapa.br/), com foco em viabilizar o acesso estruturado aos dados de vitivinicultura para uso em projetos futuros relacionados a modelos de *machine learning* e análise de dados.

Foi utilizado no início (como um “ponta pé inicial”) o auxilio da LLM chatGPT para conseguirmos explorar e entender a linguagem python e a FAST API.

Repositório do github: https://github.com/rodrigorizandobr/fiap-tech-challenge-1/

Objetivo do Projeto

Automatizar a coleta de dados do site da Embrapa, disponibilizando os resultados por meio de uma API RESTful. Essa API será a base para ingestão e análise em sistemas futuros, facilitando o uso de dados históricos e categorizados.

Artefatos Criados

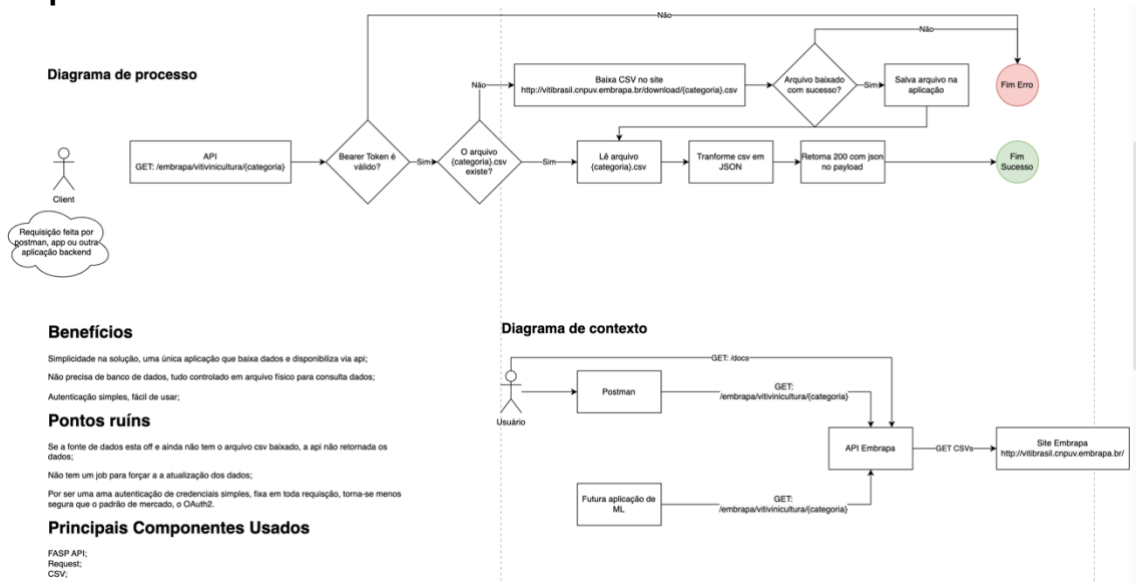
API RESTful com FastAPI:

Desenvolvemos uma API que consulta, processa e retorna os dados estruturados de arquivos CSV disponíveis no site da Embrapa. A API suporta autenticação por *token Bearer*, garantindo a segurança do acesso. *Endpoints* foram criados para diferentes categorias de dados, como produção, comercialização, processamento, importação e exportação.

Documentação:

A documentação automática da API foi gerada com *Swagger UI* e **ReDoc**, integradas ao FastAPI. Isso permite fácil navegação pelos *endpoints* e teste direto na interface.

Arquitetura:



Descrição do Código

O código principal do projeto está no arquivo `main.py`, que contém as principais funcionalidades:

Framework Utilizado:

- **FastAPI**: Utilizado para criar a API devido à sua facilidade de uso, alto desempenho e suporte nativo a documentação.

Bibliotecas Importantes:

- **requests**: Para realizar o download dos arquivos CSV diretamente do site da Embrapa.
- **csv**: Para leitura e processamento dos dados em formato CSV.
- **Pathlib**: Para manipulação segura e organizada de diretórios e arquivos.
- **certifi**: Para garantir segurança em conexões HTTPS.
- **enum**: Para organizar categorias como constantes.
- **uvicorn**: Para executar o servidor FastAPI.

Funcionalidades do Código:

Autenticação:

Implementada com o uso de `HTTPBearer`, permitindo acesso apenas mediante um token específico.

Processamento de CSV:

Transforma arquivos CSV em JSON, com lógica específica para agrupar anos repetidos em `quantidade` e `valor`, pois nas categorias de importação e exportação o site da embrapa mostra dessa forma e no csv os campos vem repetidos. Outros campos não numéricos são mapeados diretamente como atributos do JSON.

Importante destacar que utilizamos a lógica para ler o header do arquivo para gerar o payload de forma automática. Ex.: em processamento o campo CULTIVAR aparece, já em produção ou comercialização é PRODUTO.

Download Automático:

Faz o download de arquivos CSV apenas quando não estão presentes localmente, garantindo eficiência e adicionamos o nome das categorias para facilitar a manipulação se necessário.

Tratamento de Erros:

Exceções e falhas são tratadas com mensagens detalhadas e respostas HTTP apropriadas.

1. **Exemplo de Saída JSON de Importação**:

```
{
    "Id": "3",
    "País": "Argentina",
    "anos": {
        "1970": {"quantidade": "4980", "valor": "3836"},
        "1971": {"quantidade": "8811", "valor": "7543"}
    }
}
```

Destaques Técnicos

Segurança: A API utiliza autenticação via token Bearer.

Flexibilidade: O código é preparado para lidar com múltiplas categorias de dados, cada uma com URLs específicas.

Processamento de Dados: A lógica de transformação garante a estruturação dos dados, agrupando valores de anos repetidos.

Documentação: A integração nativa com Swagger permite fácil exploração da API.