



FMAP

alura

POSTECH

MACHINE LEARNING ENGINEERING

Tech Challenge 01

Grupo F-61

Cenário: Painel Inteligente para Análise e Projeção de Tendências na Produção e Venda de Vinhos e Sucos

1. Contextualização

1.1 Embrapa Uva e Vinho

Embrapa Uva e Vinho é um centro de pesquisa da Embrapa (Empresa Brasileira de Pesquisa Agropecuária). Este centro mantém um site que disponibiliza dados sobre a Viticultura no Brasil, como produção de uvas, vinhos e sucos, além de estatísticas do setor desde 1970 - URL: http://vitibrasil.cnpuv.embrapa.br/index.php?opcao=opt_01.

1.2 API Restfull

Como escopo do projeto do Tech Challenge 1, foi construída uma API Restfull que, por meio de técnicas de webscraping, permite o acesso a todos os dados históricos de uvas e vinhos de diversas fontes: Produção, Processamento, Comercialização, Importação e Exportação. Foi preocupação deste projeto a disponibilidade de dados (por meio de redundância) e segurança da informação (por meio de autorização). A documentação da API também seguiu padrões técnicos de mercado.

Algumas funcionalidades da API são:

- Coleta automática e atualizada de dados do site Vitibrasil da Embrapa Uva e Vinho;
- Armazenamento dos dados extraídos do site Vitibrasil para garantia de disponibilidade;
- Fornecimento de endpoints para consulta de dados de diversas fontes tais como Produção, Processamento, Comercialização, Importação e Exportação, podendo filtrar por ano;
- Cenários futuros: são propostos modelos de Machine Learning (ML) para implementação futura, para as diversas fontes de dados.

Potenciais usuários desta API são:

- Produtores rurais e viticultores – buscam informações sobre cultivo e produtividade;
- Pesquisadores e acadêmicos – utilizam os dados da Embrapa para estudos, publicações científicas e desenvolvimento de novas tecnologias para a viticultura;
- Empresas e cooperativas do setor – produtoras de vinhos, sucos e uvas de mesa que acompanham estatísticas de produção e consumo;
- Órgãos governamentais e formuladores de políticas – utilizam os dados para subsidiar políticas públicas voltadas para a agricultura e o setor vitivinícola;
- Consumidores e entusiastas do vinho – Alguns podem acessar o site para entender melhor a produção nacional e suas tendências.

A partir dos dados colhidos pela API, tornou-se possível sugerir cenários para o desenvolvimento de Modelos de Machine Learning - para este intento, é proposto um **Painel Inteligente para Análise e Projeção de Tendências na Produção e Venda de Vinhos e Sucos.**

2. Arquitetura da Solução proposta

(desenvolver)

[Usuário ou Dashboard Web]



[API Flask]



[Módulo de Webscraping Agendado]



[Base de Dados (PostgreSQL ou MongoDB)]



[Camada de Análise/BI]



[Futuro: Módulo de Machine Learning para Previsão de Vendas]

3. Pipeline

3.1 Coleta de Dados

Webscraping é feito a partir das fontes de dados descritas anteriormente, com rotas disponíveis para cada fonte em particular.

Tecnologias empregadas: Python, Flask, Pandas e BeautifulSoup. A autenticação foi feita com JWT. A documentação foi feita com Swagger (biblioteca Flasgger). A aplicação foi organizada e modularizada usando-se blueprints.

3.2 Pré-processamento

Os dados coletados são tratados de forma a garantir um nível de qualidade mínima - por exemplo, quando um valor vem com “nd” é substituído por “0.00”.

Tecnologias empregadas: Python e Pandas.

3.3 Cenários de Modelos para Machine Learning

Para a proposta de cenários, serão abordadas as fontes de dados com potenciais tarefas de Machine Learning (ML) correlatas para possíveis setores de interesse. Também serão propostos potenciais algoritmos para endereçar solução para cada tarefa de ML proposta.

3.3.1 Fonte de dados: Processamento

Tarefa de ML: Previsão de processamento, em Kg, por tipo de Cultivar (Uva)

Setor Interessado: Planejamento da Produção

Algoritmos possíveis: Regressão Linear ou Séries Temporais

3.3.2 Fonte de dados: Produção

Tarefa de ML: Previsão de produção, em Litros, por Produto categorizado

Setor Interessado: Planejamento de Vendas

Algoritmos possíveis: Regressão Linear ou Séries Temporais

Tarefa de ML: Agrupamento de Variedades de Produtos com melhor desempenho

Setor Interessado: Planejamento de Produção, em toda sua cadeia, desde o plantio

Algoritmos possíveis: K-Means ou DBSCAN

3.3.3 Fonte de dados: Exportação

Tarefa de ML: Previsão de preço, em US\$, por País categorizado

Setor Interessado: Planejamento de Vendas, identificando sazonalidades de mercado

3.3.4 Fonte de Dados: Importação

Tarefa de ML: Previsão de preço, em US\$, por País categorizado

Setor Interessado: Planejamento de Compras e Vendas

3.3.5 Fonte de Dados: Importação

Tarefa de ML: Previsão de processamento, em Litros, por tipo de Cultivar (Uva)

Setor Interessado: Planejamento de Vendas, identificando sazonalidades de mercado

4. Modelo de Deploy

Para o atendimento dos cenários de Machine Learning sugeridos, a seguir são apresentados as etapas e tecnologias passíveis de serem aplicadas para o Deploy.

Etapas	Descrição
Containerização	Uso de Docker para isolar API e Webscraper
Hospedagem da API	Render, Heroku ou AWS EC2
Banco de Dados	PostgreSQL (recomendado para dados relacionais) ou MongoDB (para flexibilidade)
Agendamento do Scraping	Cron, Celery + Redis ou GitHub Actions programado
Visualização dos Dados	Gráficos com Plotly/Dash, Streamlit ou integração com BI (ex: Metabase)
Monitoramento	Logs com Flask-Logging + observabilidade via Grafana ou serviços da nuvem