

POSTECH



Tech Challenge

Tech Challenge é o projeto da fase que englobará os conhecimentos obtidos em todas as disciplinas da fase. Esta é uma atividade que, a princípio, deve ser desenvolvida em grupo. Importante atentar-se ao prazo de entrega, pois trata-se de uma atividade obrigatória, uma vez que sua pontuação se refere a 90% da nota final.

O problema

Desafio: Criação de uma API Pública para Consulta de Livros

Você foi contratado(a) como Engenheiro(a) de Machine Learning para um projeto de recomendação de livros. A empresa está em sua fase inicial e ainda não possui uma base de dados estruturada.

Seu primeiro desafio será montar a infraestrutura de extração, transformação e disponibilização de dados via API pública para que cientistas de dados e serviços de recomendação possam usar esses dados com facilidade.

Assim, seu objetivo será desenvolver um pipeline completo de dados e uma API pública para servir esses dados, pensando na escalabilidade e reusabilidade futura em modelos de machine learning.

Entregáveis Obrigatórios

1. Repositório do GitHub Organizado

- Código estruturado em módulos (scripts/, api/, data/, etc.).
- README completo contendo:
 - Descrição do projeto e arquitetura.
 - Instruções de instalação e configuração.
 - Documentação das rotas da API.
 - Exemplos de chamadas com requests/responses.
 - Instruções para execução.

2. Sistema de Web Scraping

- Script automatizado para extrair dados de <https://books.toscrape.com/>
- Dados armazenados localmente em um arquivo CSV.
- Script executável e bem documentado.

3. API RESTful Funcional

- API implementada em Flask ou FastAPI.
- Endpoints obrigatórios (listados a seguir).
- Documentação da API (Swagger).

4. Deploy Público

- API deployada em Heroku, Render, Vercel, Fly.io ou similar.
- Link compartilhável funcional.
- API totalmente operacional no ambiente de produção.

5. Plano Arquitetural

- Diagrama ou documento detalhando:
 - Pipeline desde ingestão → processamento → API → consumo.
 - Arquitetura pensada para escalabilidade futura.
 - Cenário de uso para cientistas de dados/ML.
 - Plano de integração com modelos de ML.

6. Vídeo de Apresentação (3-12 minutos)

- Demonstração técnica (no macro apenas, sem aprofundamento).
- Apresentação da arquitetura e pipeline de dados.
- Execução de chamadas reais à API em produção.
- Comentários sobre boas práticas implementadas.

Objetivos Técnicos Core

Web Scraping Robusto

- Extrair todos os livros disponíveis no site.
- Capturar: título, preço, rating, disponibilidade, categoria, imagem.

Endpoints Obrigatórios da API

Endpoints Core

- GET /api/v1/books: Lista todos os livros disponíveis na base de dados.
- GET /api/v1/books/{id}: Retorna detalhes completos de um livro específico pelo ID.
- GET /api/v1/books/search?title={title}&category={category}: Busca livros por título e/ou categoria.
- GET /api/v1/categories: Lista todas as categorias de livros disponíveis.
- GET /api/v1/health: Verifica status da API e conectividade com os dados.

Endpoints Opcionais da API

Endpoints de Insights

- GET /api/v1/stats/overview: Estatísticas gerais da coleção (total de livros, preço médio, distribuição de ratings).
- GET /api/v1/stats/categories: Estatísticas detalhadas por categoria (quantidade de livros, preços por categoria).
- GET /api/v1/books/top-rated: Lista os livros com melhor avaliação (rating mais alto).
- GET /api/v1/books/price-range?min={min}&max={max}: Filtra livros dentro de uma faixa de preço específica.

Desafios Adicionais (Bônus)

Desafio 1: Sistema de Autenticação

Tech Challenge

Implementar **JWT Authentication** para proteger rotas sensíveis:

- POST /api/v1/auth/login - Obter token.
- POST /api/v1/auth/refresh - Renovar token.
- Proteger endpoints de admin como /api/v1/scraping/trigger.

Desafio 2: Pipeline ML-Ready

Criar endpoints pensados para consumo de modelos ML:

- GET /api/v1/ml/features - Dados formatados para features.
- GET /api/v1/ml/training-data - Dataset para treinamento.
- POST /api/v1/ml/predictions - Endpoint para receber previsões.

Desafio 3: Monitoramento & Analytics

- Logs estruturados de todas as chamadas.
- Métricas de performance da API.
- Dashboard simples de uso (Streamlit recomendado).

Lembre-se de que você poderá apresentar o desenvolvimento do seu projeto durante as lives com professores(as). Essa é uma boa oportunidade para discutir as dificuldades encontradas e pegar dicas valiosas com docentes especialistas e colegas de turma.

E tem um plus para fortalecer seu aprendizado e ainda ganhar pontos na nota!

Para dar um up nas suas certificações e ainda ganhar 10 pontos na nota desse Tech Challenge da Fase 1, desafiamos você a realizar também nessa fase o curso da Google Cloud GenAI skill badge course pathways com o objetivo de desenvolver experiência com IA Generativa, conhecendo desde os fundamentos de grandes modelos de linguagem até os princípios responsáveis da IA.

Os cursos fornecem exemplos de como o Gemini pode ajudar a tornar engenheiros(as) de todos os tipos mais eficientes em suas atividades diárias e como incorporar modelos de IA Generativa pré-treinados no seu fluxo de trabalho.

Tech Challenge

Esse curso tem duração de ~30 horas e inclui 3 skill badges (selos de habilidade). Bora fazer o curso para garantir esses pontos, aprender sobre Gen IA com a Google e postar suas badges no LinkedIn? Não se esqueça de marcar o time acadêmico para celebrar suas conquistas.

⚠ Para ganhar os 10 pontos na nota do Tech Challenge, envie para nós um comprovante de conclusão do curso anexado junto aos entregáveis desse desafio! Além disso, vamos te orientar para se cadastrar no curso e iniciar os estudos.

Não se esqueçam de que isso é um entregável obrigatório! Se atentem para o prazo de entrega até o final da fase.

Boa sorte! 🍀

The background is a dark, abstract network visualization. It features a complex web of glowing nodes and connecting lines. The nodes are represented by small, semi-transparent spheres in various colors, including teal, orange, and grey. The lines are thin, light-colored threads that crisscross the frame, creating a sense of depth and connectivity. Some lines are thicker and more prominent, while others are faint and recede into the background. The overall effect is a futuristic, digital landscape that suggests a global or interconnected network.

POSTECH