

MATHEUS PAVANI

POSTECH

MACHINE LEARNING ENGINEERING

TECH CHALLENGE

FACTORY

---

## TECH CHALLENGE

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

### O problema

O transporte aéreo é uma parte vital da infraestrutura global, mas os atrasos de voos impactam milhões de passageiros todos os anos. Neste projeto, você utilizará o conjunto de dados público que contém informações detalhadas sobre voos nos EUA para desenvolver análises e modelos preditivos e/ou exploratórios aplicando técnicas de Machine Learning supervisionado e não supervisionado.

 **Base de dados:** [Base de dados - MLET Fase 3](#)

### Objetivos do Projeto

Aplicar o conhecimento adquirido sobre modelagem supervisionada e não supervisionada e desenvolver um pipeline completo de ciência de dados, desde a exploração dos dados até a interpretação dos resultados. Com esse Tech Challenge queremos estimular a criatividade, o questionamento e a exploração autônoma dos dados.

### O que é obrigatório fazer

Exploração dos dados (EDA):

- Investigue os dados com estatísticas descritivas.
- Crie visualizações que trazem insights.
- Trate valores ausentes de forma adequada.

Modelagem supervisionada (mínimo um modelo):

Escolha entre:

- Classificação: prever se um voo vai atrasar ou não.

## Tech Challenge

- Regressão: prever quanto tempo o atraso vai durar.
  - Compare pelo menos dois algoritmos diferentes e avalie com métricas adequadas.

Modelagem não supervisionada (mínimo uma abordagem):

- Use clusterização (ex.: agrupar rotas, aeroportos ou companhias aéreas) ou redução de dimensionalidade (ex.: PCA).
- Mostre os resultados em gráficos e faça a interpretação.

Apresentação crítica dos resultados:

- Compartilhe as principais conclusões.
- Discuta as limitações dos modelos.
- Proponha melhorias e próximos passos.

### **Vá além do básico (exploração livre e não obrigatória)**

Os grupos que quiserem se destacar podem experimentar:

- Criar variáveis derivadas (ex.: período do dia, feriados, estações do ano).
- Analisar atrasos por aeroporto, companhia ou estado.
- Criar mapas geográficos de rotas e atrasos.
- Identificar padrões sazonais ou horários críticos.
- Construir um dashboard interativo para visualização.
- Explorar detecção de anomalias ou aprendizado semi-supervisionado.

### **Perguntas que podem guiar sua análise**

- Quais aeroportos são mais críticos em relação a atrasos?
- Que características aumentam a chance de atraso em um voo?
- Os atrasos são mais comuns em certos dias da semana ou horários?
- É possível agrupar aeroportos com perfis semelhantes?

## Tech Challenge

- Até que ponto conseguimos prever atrasos com base no histórico e contexto do voo?

### **Entregáveis**

- Repositório no GitHub ou Google Colab com o código completo.
- Vídeo de apresentação (5 a 10 minutos) explicando o trabalho, os resultados e as conclusões.

Caso tenha qualquer dúvida, não hesite em nos procurar no Discord. 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. The overall effect is a futuristic, digital landscape that suggests a global or interconnected network.

POSTECH