

Desafio Técnico para Cientista de Dados Pleno - MT2 Data

Objetivo

Avaliar a capacidade do candidato de:

1. Desenvolver algoritmos em **R** para previsão de uma série histórica de emplacamento de caminhões da ANFAVEA, utilizando variáveis macroeconômicas como preditores.
 - a. <https://anfavea.com.br/site/edicoes-em-excel/>
 2. Criar um aplicativo interativo em **Python/Streamlit** para visualização dos resultados e simulações "What If" de variáveis macroeconômicas.
-

Tarefas

1. Previsão da Série Temporal (em R)

- Use a base de dados fornecida Anfavea e coleta as séries temporais das variáveis macro com os seguintes campos:
 - **ano_mes**: Data no formato "YYYY-MM".
 - **emplacamentos**: Número de emplacamentos de caminhões (variável dependente).
 - **PIB**: Produto Interno Bruto (em valores ajustados pela inflação).
 - **inflação**: Taxa de inflação mensal.
 - **juros**: Taxa básica de juros (% ao mês).
 - **cambio**: Cotação média do dólar no mês.
 - **produção_industrial**: Índice de produção industrial.
- Desenvolva um modelo de **forecasting** utilizando métodos estatísticos (ARIMA, ETS) e aprendizado de máquina (XGBoost, Random Forest, etc.).
- Avalie a performance do modelo com métricas como **MAE**, **RMSE**, ou **MAPE** em um conjunto de teste.
- Gere previsões para os próximos 12 meses com base nas variáveis fornecidas.

2. Desenvolvimento do Aplicativo em Python/Streamlit

- Desenvolva um aplicativo com as seguintes funcionalidades:
 - **Visualização Interativa**:
 - Gráfico da série temporal de emplacamentos reais e previstos.
 - Visualização das variáveis macroeconômicas ao longo do tempo.
 - **Simulação "What If"**:
 - Permitir que o usuário ajuste manualmente os valores das variáveis macroeconômicas (PIB, inflação, juros, câmbio, produção industrial).

- Gerar previsões atualizadas de emplacamentos com base nos novos valores inseridos.
 - **Exportação de Resultados:**
 - Opção para baixar as previsões atualizadas em formato CSV.
-

Critérios de Avaliação

1. Técnicos (60%)

- Qualidade do modelo preditivo (explicabilidade e performance).
- Escolha e justificativa das técnicas utilizadas.
- Organização e legibilidade do código em R e Python.
- Precisão e robustez das previsões.

2. Interface e Usabilidade (30%)

- Facilidade de uso do aplicativo Streamlit.
- Clareza na visualização dos dados e previsões.
- Funcionalidade correta das simulações "What If".

3. Documentação e Comunicação (10%)

- Entrega de um **README** explicando as decisões técnicas, estrutura do código, e instruções para rodar o projeto.
 - Clareza na comunicação dos resultados e limitações do modelo.
-

Instruções de Entrega

1. Submeta o código completo em um repositório público no GitHub, incluindo:
 - Script em R para o modelo preditivo.
 - Código Python para o aplicativo Streamlit.
 - Arquivo README detalhado.
 2. Certifique-se de que o repositório contenha instruções claras para instalação e execução.
 3. O prazo para entrega do desafio é de **7 dias corridos**.
-

Base de Dados

Base de dados Anfavea com 10 anos de dados + séries históricas das variáveis macro.

Extras (Não Obrigatório)

- Implementação de métodos avançados de análise de sensibilidade na simulação "What If".
 - Uso de técnicas de explainability para descrever o impacto das variáveis preditoras no modelo.
-