

Exercício Prático

Data Science

Olá! Obrigado pelo interesse e esforço dedicado ao nosso processo seletivo. Esta etapa busca avaliar a sua desenvoltura, criatividade e capricho na solução de um problema técnico, utilizando ferramentas conhecidas. Veja as instruções abaixo.

Desafio: Neste exercício, você deve analisar dados e desenvolver modelos de predição sobre concentrações de CO₂ na atmosfera e a evolução da temperatura global. Para isso, você vai precisar dos seguintes conjuntos de dados:

- Concentração de CO₂ na atmosfera (dados mensais, desde o ano zero):
ftp://data.iac.ethz.ch/CMIP6/input4MIPs/UoM/GHGConc/CMIP/mon/atmos/UoM-CMIP-1-1-0/GHGConc/gr3-GMNHSH/v20160701/mole_fraction_of_carbon_dioxide_in_air_input4MIPs/GHGConcentrations_CMIP_UoM-CMIP-1-1-0_gr3-GMNHSH_000001-201412.csv
- Temperatura global (dados mensais, desde o ano 1850):
<https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data/data#GlobalTemperatures.csv>

Em seguida, desenvolva as seguintes tarefas:

1. **Estatísticas:** Determine estatísticas relevantes para cada uma das séries, como autocorrelação, estacionariedade, tendência, sazonalidade e ruído. Avalie e comente os resultados da forma que achar mais relevante.
2. **Correlação vs. Causalidade:** Uma das principais questões de interesse atual é entender se o aumento da temperatura média global é ou não causado (pelo menos em parte) pelo aumento da concentração de CO₂ na atmosfera. Como você responderia essa questão com base nos dados

fornecidos acima? Determine a correlação entre as duas séries de dados e explique sua abordagem e os resultados.

3. **Predição:** Desenvolva um modelo de predição para cada série temporal. Escolha o modelo que lhe parecer mais adequado para o problema, e justifique sua escolha. Avalie a qualidade e as limitações de cada predição.
4. **Predição com série de suporte:** Desenvolva um modelo de predição para a temperatura média global utilizando a concentração de CO₂ na atmosfera como série de suporte. Escolha o modelo que lhe parecer mais adequado para o problema, e justifique sua escolha. Avalie a qualidade e as limitações da predição.

Como entregar os resultados: Você deve nos enviar por e-mail um link para um notebook Jupyter que deverá conter todo o código-fonte, as análises e comentários para as tarefas acima. Nós executaremos os scripts do lado de cá e avaliaremos os resultados.

O que avaliaremos: O funcionamento correto dos scripts, a organização, estrutura e documentação do código fonte, a qualidade dos resultados e, principalmente, o capricho e a profundidade na apresentação das análises. Não deixe de evidenciar qualquer abordagem que julgar relevante no trabalho (por exemplo, se foi necessário tratar os dados originais de alguma maneira). Lembre-se de utilizar os recursos visuais que achar mais adequados para apresentar suas análises. Sua capacidade de buscar informações e resolver problemas por conta própria também conta!

Prazo: Informado no e-mail. Tente fazer o máximo que puder, mesmo se achar que o prazo não for suficiente. E se tiver dúvidas ou imprevistos, não deixe de entrar em contato conosco!

Bom trabalho!