

## Desafio\_CESAR\_DATASCIENCE:

Precisamos que você crie um repositório no github ou gitlab e inicie um projeto para o problema que vai ser apresentado a seguir. Pode colocar nesse repositório seu código em Python ou um jupyter notebook. É importante que adicione em seu README os passos necessários para que possamos executar e verificar sua implementação. Tendo em mente a base de dados Bias correction of numerical prediction model temperature forecast Data Set(<https://archive.ics.uci.edu/ml/datasets/Bias+correction+of+numerical+prediction+model+temperature+forecast>) da UCI, queremos que você faça uma análise descritiva dos dados explicando aspectos fundamentais, tais como a distribuição dos dados e correlação entre as variáveis. Fique a vontade para fazer sugestões, alterações e transformações que achar necessárias a fim de facilitar o entendimento do problema.

Além da análise descritiva, queremos que você treine e avalie um modelo de Machine Learning para fazer a estimação da variável Next\_Tmax, explicando a qualidade do modelo a partir das métricas de avaliação utilizadas. Como métricas de avaliação, sugerimos como obrigatórias a *accuracy*, *precision* e *recall*. Mas você fica livre de avaliar outras métricas que sejam oportunas de acordo com seu julgamento e necessidade na avaliação do modelo. Queremos que você nos mostre o que fez no código, mas também explique suas decisões e conclusões, pode gerar gráficos ou outras visualizações para auxiliar em sua explicação. Queremos ver também sua linha de raciocínio.

Por fim, queremos ouvir de você o que poderia ser feito de diferente se tivesse mais tempo e mais recursos. Quais outros testes e análises poderiam ser feitas e qual a importância destas.

Não é necessário pensar nesse momento em eficiência ou otimização de código e modelos, mas se considerar oportuno, pode adicionar esse tópico em sua análise. Fique à vontade para determinar o que é relevante.

Extra: Como atividade complementar e, portanto, não obrigatória, você pode treinar e avaliar o modelo de Machine Learning considerando além da variável Next Tmax, também a variável Next Tmin. Novamente explicando a qualidade do modelo e trazendo pelo menos as métricas obrigatórias de accuracy, precision e recall.