SELEÇÃO CÓD. 1512 CIENTISTA DE DADOS I (OBSERVATÓRIO)

Todos os arquivos de resposta devem ser organizados em pastas, compactados em somente um e enviado com assunto "Respostas Seleção Cientista de Dados" para o e-mail: fjfilho@sfiec.org.br e cevalente@sfiec.org.br. A pasta também deve ser disponibilizada em sites como Github e/ou Gitlab, compartilhe o link com o repositório no git com o email: fjfilho@sfiec.org.br, seguindo as recomendações:

- **Questão1 – arquivo jupyter notebook elaborado e comentado**
- **Questão bônus – projeto no MLflow elaborado e comentado**

Este desafio usa um conjunto de dados de séries temporais de clima registrado pelo Instituto Max Planck de Biogeoquímica. Este conjunto de dados contém 14 features diferentes, como temperatura do ar, pressão atmosférica e umidade. Estes dados foram coletados a cada 10 minutos, a partir de 2003. Para este desafio, você usará apenas os dados coletados entre 2009 e 2016.

Download da base de dados.

Recomenda-se lidar apenas com previsões por hora, então é interessante que você utilize uma sub amostra dos dados de intervalos de 10 minutos para intervalos de uma hora, por exemplo:

```
df = pd.read_csv(csv_path)
# Slice [start:stop:step], starting from index 5 take every 6th record.
df = df[5::6]
```

Com a base de dados em mãos, o objetivo deste desafio é de que você faça previsões sobre a variável **'T(degC)'** e por se tratar de um desafio de séries temporais, o modelo deve prever para ***pelo menos*** 2 passos a frante, isto é, dados com o *timestamp* '`01.01.2009 01:00:00`' só podem ser usados para prever os valores a partir de '`01.01.2009 03:00:00`'.

Recomenda-se o uso de um jupyter notebook com os comentários sobre as suas análises sobre os dados, tratamentos das features, modelos escolhidos, métricas utilizadas e o motivo das suas decisões tomadas.
Um arquivo de README com as instruções de como reproduzir o seu experimento é bem vindo.

Bônus:

Como atividade bônus, além do seu experimento em um jupyter notebook, encorajamos que todo o pipeline do seu projeto de machine learning seja convertido para scripts python orquestrados pelo framework MLflow.