

Desafio



Processo Seletivo
Instituto Tecgraf

Desafio da previsão meteorológica

O Instituto XPTO de Pesquisas Climáticas deseja criar uma previsão meteorológica para uma cidade. Espera-se prever a temperatura e a chuva para uma janela de um ano a partir de dados dos últimos dez anos. Para isso utilize uma combinação de modelos, ferramentas, abordagens e algoritmos conhecidos nas áreas de ciência de dados, visualização, aprendizado de máquina e inteligência artificial. É muito importante documentar o processo de análise e geração dos dados de previsão, inclusive listando os caminhos descartados. Evidentemente a acurácia das previsões também será analisada. Inclua as fontes de consulta utilizadas na pesquisa.

Como entrada de dados é fornecido um arquivo histórico com dados meteorológicos de um local ao longo de uma década, com informações de hora em hora. No final espera-se a geração de um arquivo de saída CSV com uma linha de cabeçalho e as mais de 8000 linhas de dados, sendo uma para cada hora do dia, ao longo de um ano, contendo temperatura (em graus Celsius) e chuva (em milímetros). Este arquivo deve ser enviado juntamente com a documentação da solução e do processo utilizado.

Descrição do arquivo de entrada HISTORICO.CSV fornecido:

- Uma linha de cabeçalho contendo o nome das colunas
- Uma linha de dados para cada hora de cada dia ao longo de dez anos
- As seguintes colunas:
 - **time**: referente à série temporal no formato “yyyy-MM-ddTHH:mm”
 - **temperature_2m**: temperatura do ar a 2 metros acima do solo.
 - **relative_humidity_2m**: umidade relativa a 2 metros acima do solo.
 - **rain**: apenas precipitação líquida da hora anterior.
 - **pressure_msl**: pressão do ar atmosférico reduzida ao nível médio do mar.
 - **surface_pressure**: pressão do ar atmosférico na superfície.
 - **cloud_cover**: cobertura total de nuvens como uma fração da área.
 - **wind_speed_10m**: velocidade do vento a 10 metros acima do solo.
 - **wind_speed_100m**: velocidade do vento a 100 metros acima do solo.
 - **wind_direction_10m**: direção do vento a 10 metros acima do solo.
 - **wind_direction_100m**: direção do vento a 100 metros acima do solo.
 - **wind_gusts_10m**: rajadas a 10 metros acima do solo na hora indicada.
 - **soil_temperature_0_to_7cm**: temperatura média de 0 a 7 cm abaixo do solo.
 - **soil_moisture_0_to_7cm**: conteúdo médio de água no solo como razão de mistura volumétrica na profundidade de 0 a 7 cm.

As unidades de medida de cada coluna aparecem no cabeçalho do arquivo.

Exemplo de um recorte da tabela gerada a partir do arquivo saída PREVISOA.CSV resultante:

time	temperature	rain
...
2010-11-23T12:00	0.2	1.3
2010-11-23T13:00	-0.1	0.0
2010-11-23T14:00	1.5	0.1
...

Pede-se:

1. Geração de um arquivo **PREVISOA.CSV** contendo com as seguintes colunas:
 - a. **time**: referente à série temporal no formato “yyyy-MM-ddTHH:mm”
 - b. **temperature**: número com uma casa decimal
 - c. **rain**: é um número maior ou igual a zero com uma casa decimal
2. Documentação de todo o processo (análise exploratória dos dados, tratamento dos dados, criação e treinamento de modelos, análise dos resultados, etc). Para esta documentação utilize um **Jupyter Notebook que deve ser compatilhado**.
<https://jupyter.org/try-jupyter/notebooks/?path=notebooks/Intro.ipynb>

Em caso de dúvidas, documente as premissas, os caminhos e decisões tomadas em seu notebook.

Boa sorte!