

Climate Weather Surface of Brazil

BIG DATA

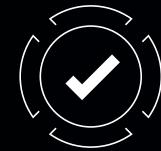
Professor: Fabrício Torquato

Matheus Arizono RA:235912

Felipe Roberto de Souza Silva RA:226752

Caue Henrique Ricardo RA:235873

Sumário



Introdução e Objetivos



Arquitetura do Ambiente Big Data



Dataset Escolhido



Pré-processamento & EDA



Modelos Preditivos



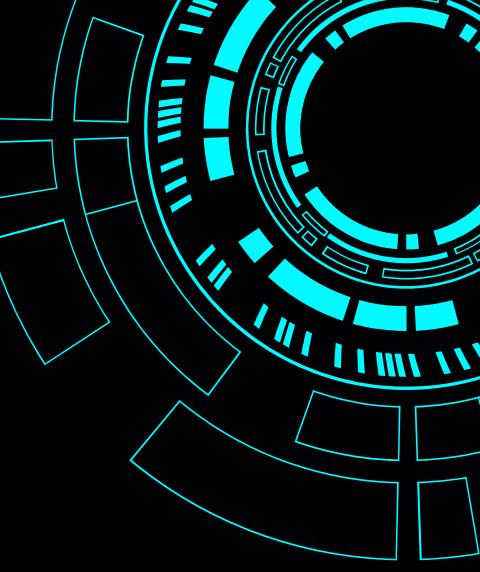
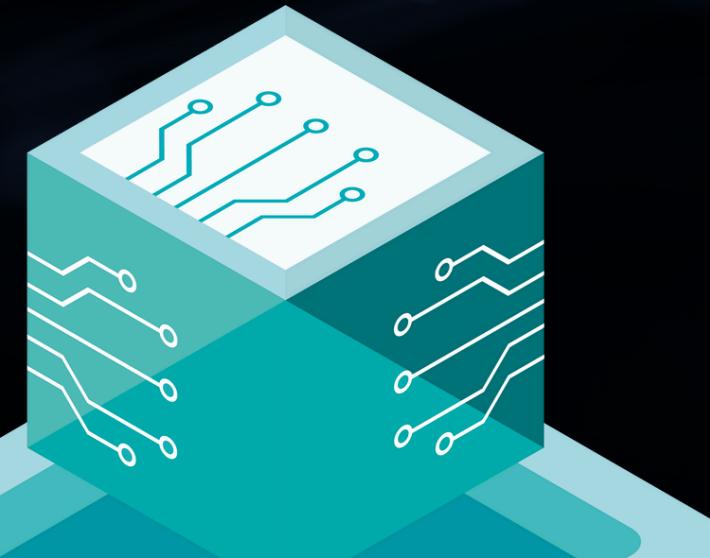
Resultados & Avaliação



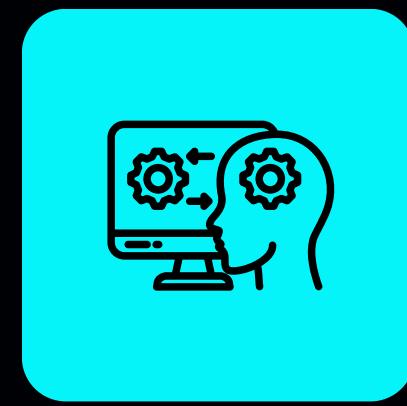
Desafios & Lições Aprendidas



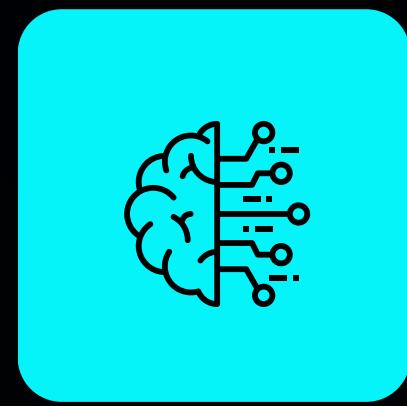
Próximos Passos & Conclusão



Objetivos



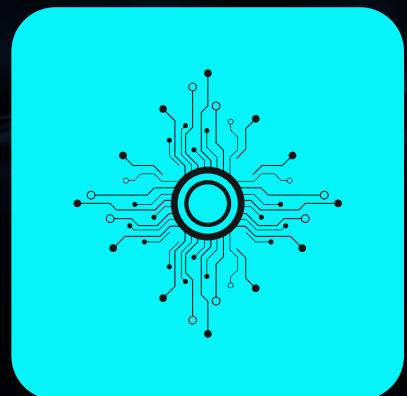
**Montar ambiente Big Data
(Docker + Apache)**



**Limpeza e preparação de
dados (weather dataset)**



Classificação de chuva



**Régressão de
velocidade do vento**

Imports & Leitura de Dados

The image displays a dense network of binary logic gates, primarily AND, OR, and NOT gates, arranged in a grid-like structure. A prominent feature is a large, circular component on the left side, which appears to be a ring oscillator or a similar clock distribution mechanism. The entire circuit is composed of black text on a white background, with various components highlighted in blue. The logic values 0 and 1 are used throughout the diagram.

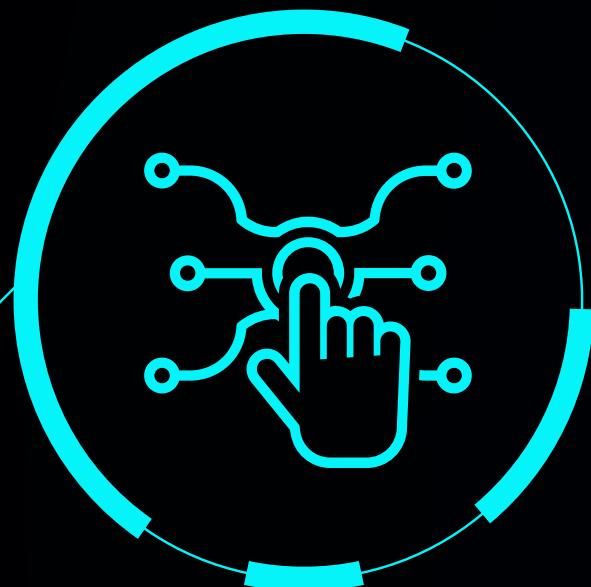
Carregamento do CSV + Tamanho inicial do dataset

Número de linhas no DataFrame de WEATHER: 8.392.320

Após remover as Outlier

Número de linhas no DataFrame de WEATHER: 3.298.717

Tratamento de Dados



Filtrar tipos de dado relevantes

Selecionar apenas colunas que impactam diretamente as análises meteorológicas.

Eliminar valores negativos e placeholders (-999)

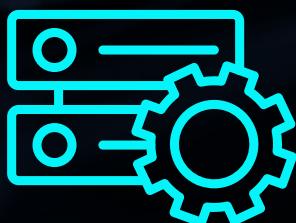
Substituir entradas inválidas por NaN ou valores de referência.

Normalizar nomes de colunas

Padronizar nomenclatura para facilitar scripts e documentação.

Remover outliers simples

Descartar medições muito distantes da média para evitar viés nos modelos.



Remover colunas e campos inúteis

Eliminar variáveis com alta proporção de nulos ou irrelevantes.

Dimensão final do dataset

Apresentar número de registros e colunas após a limpeza completa.

Classificação de Chuva

- Definição: chuva se precipitação > 0
- Balanceamento (undersampling / oversampling)
- Pré-processamento (encoding, normalização)

Validação - Classificação

- Split treino/teste
- Métricas: Acurácia, F1-Score
- Resultados principais



Modelo de Regressão

Modelo

- Objetivo: prever velocidade do vento na próxima hora
- Algoritmos: (ex: Linear Regression, Random Forest)
- Métricas: RMSE, R²

Resultados

- Comparação de métricas
- Gráfico de dispersão real vs previsto

Gráfico de Métricas de Modelo

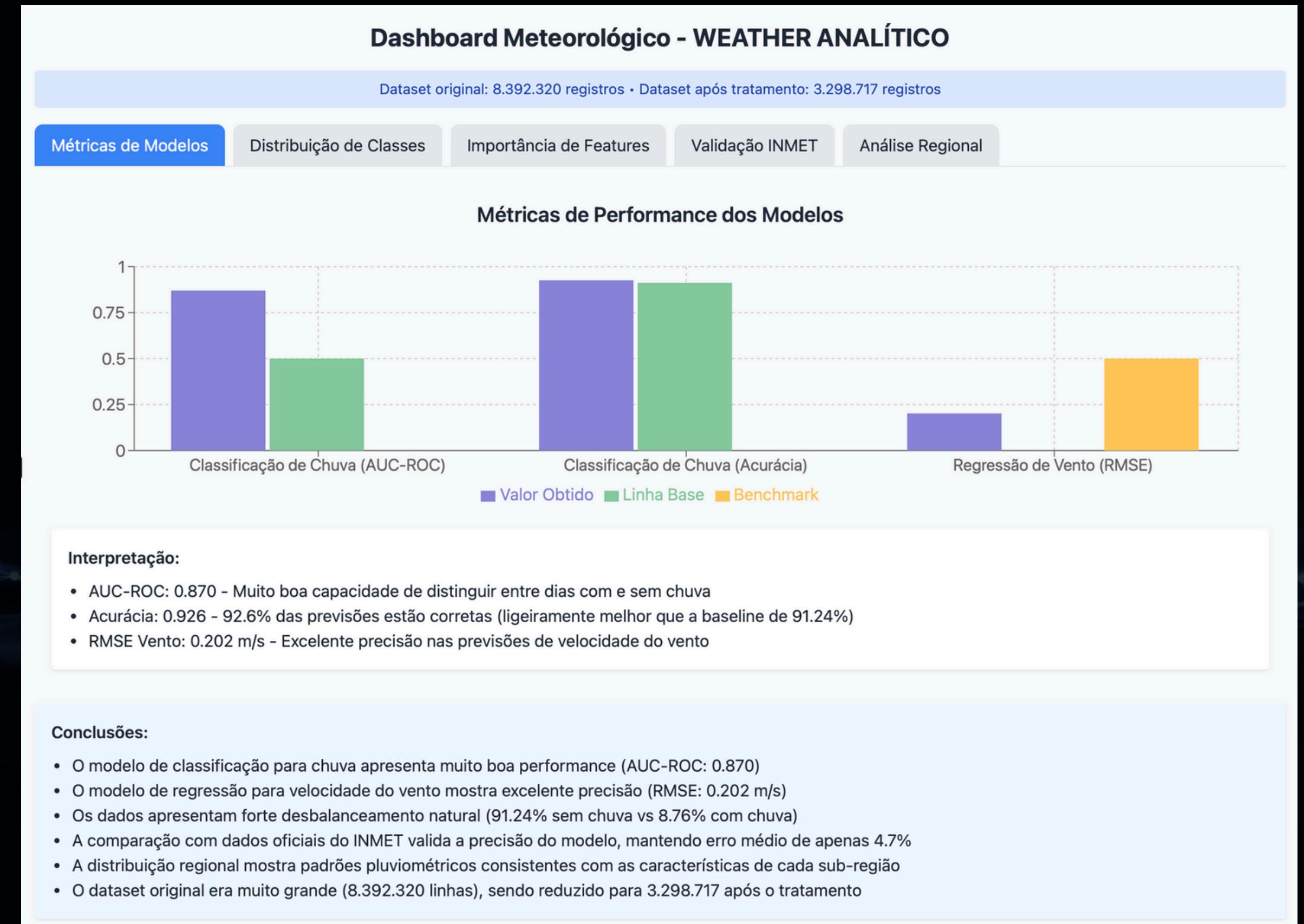


Gráfico de Distribuição de Classes

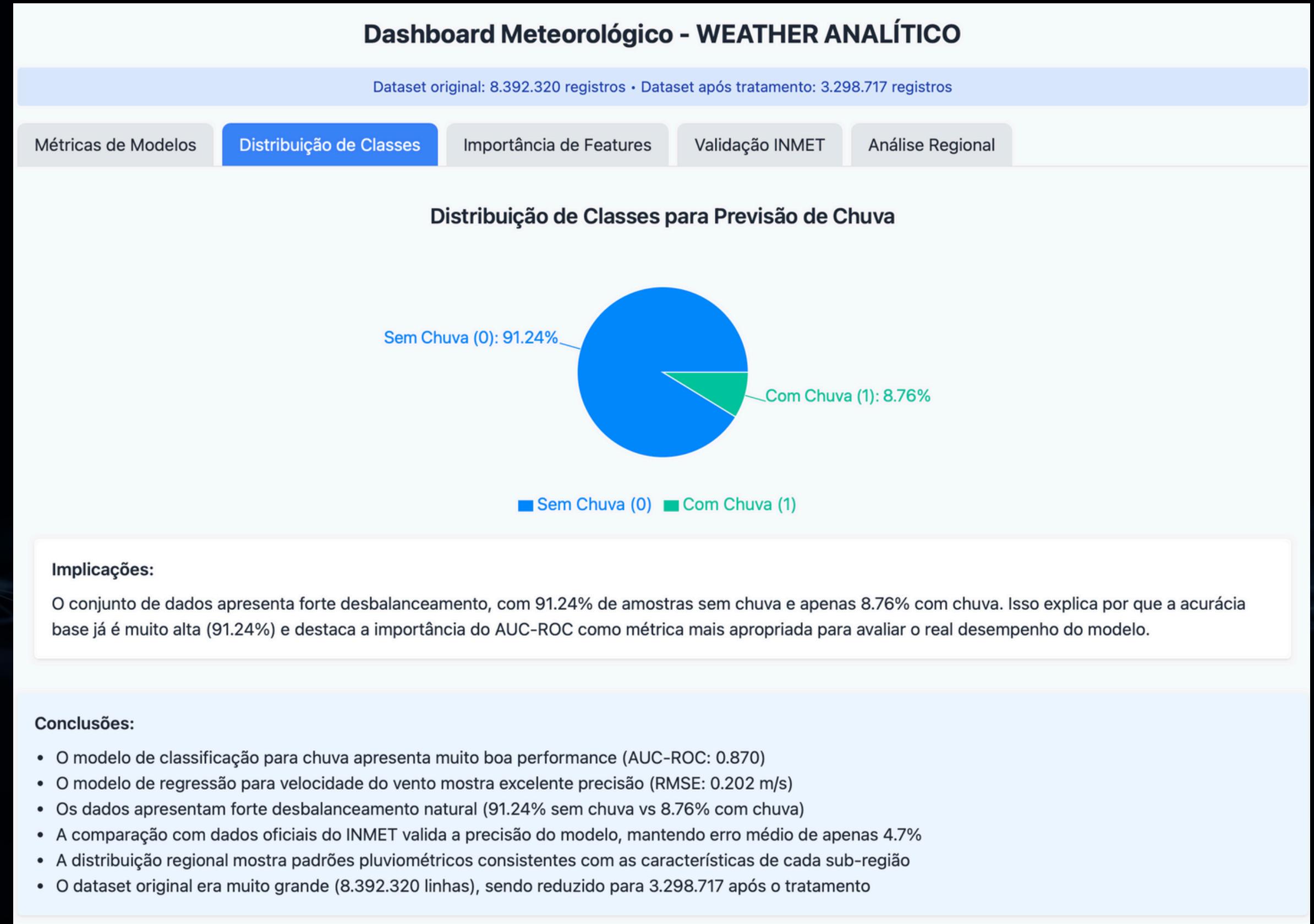


Gráfico de Importância de Features

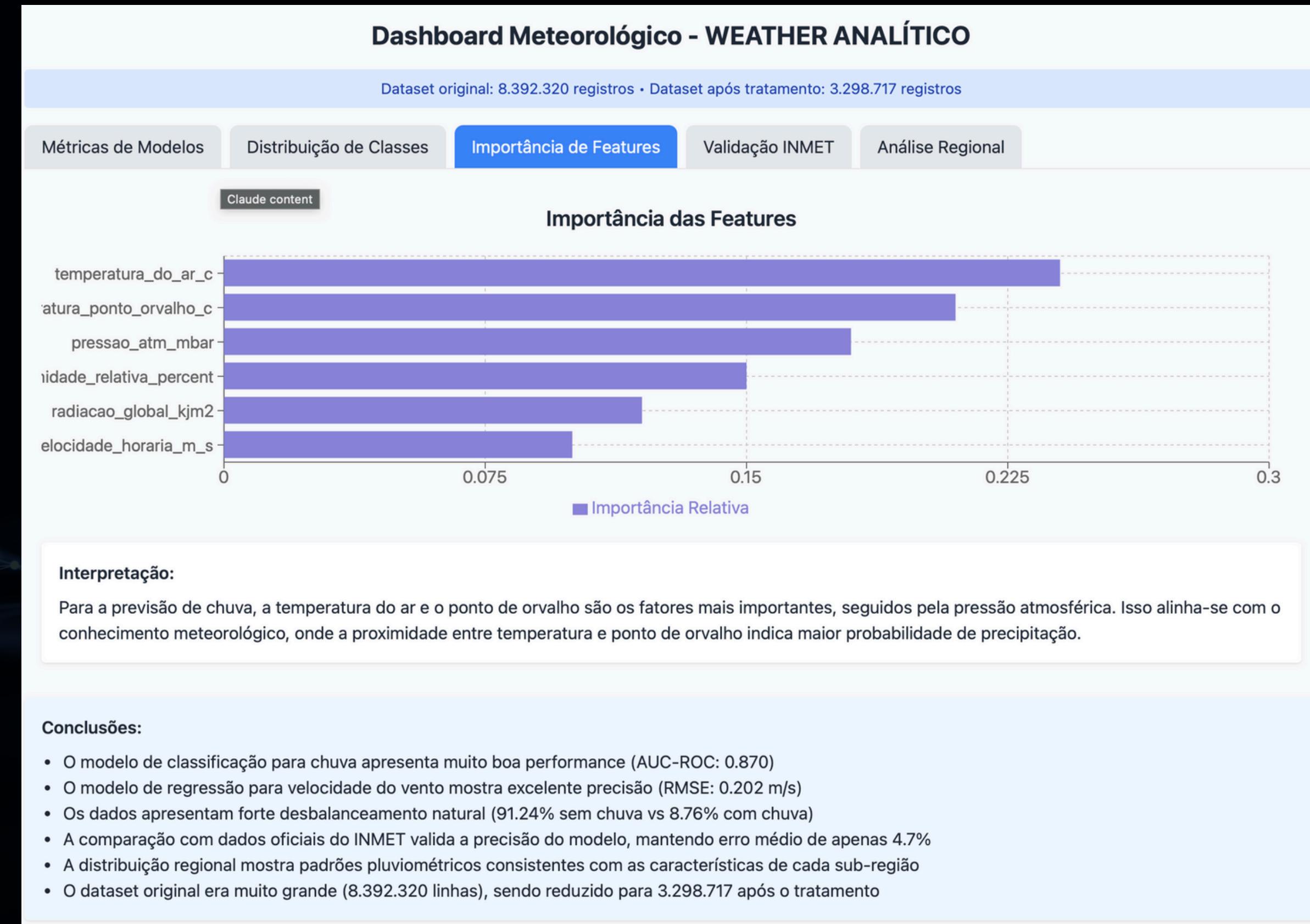


Gráfico de Análise Temporal

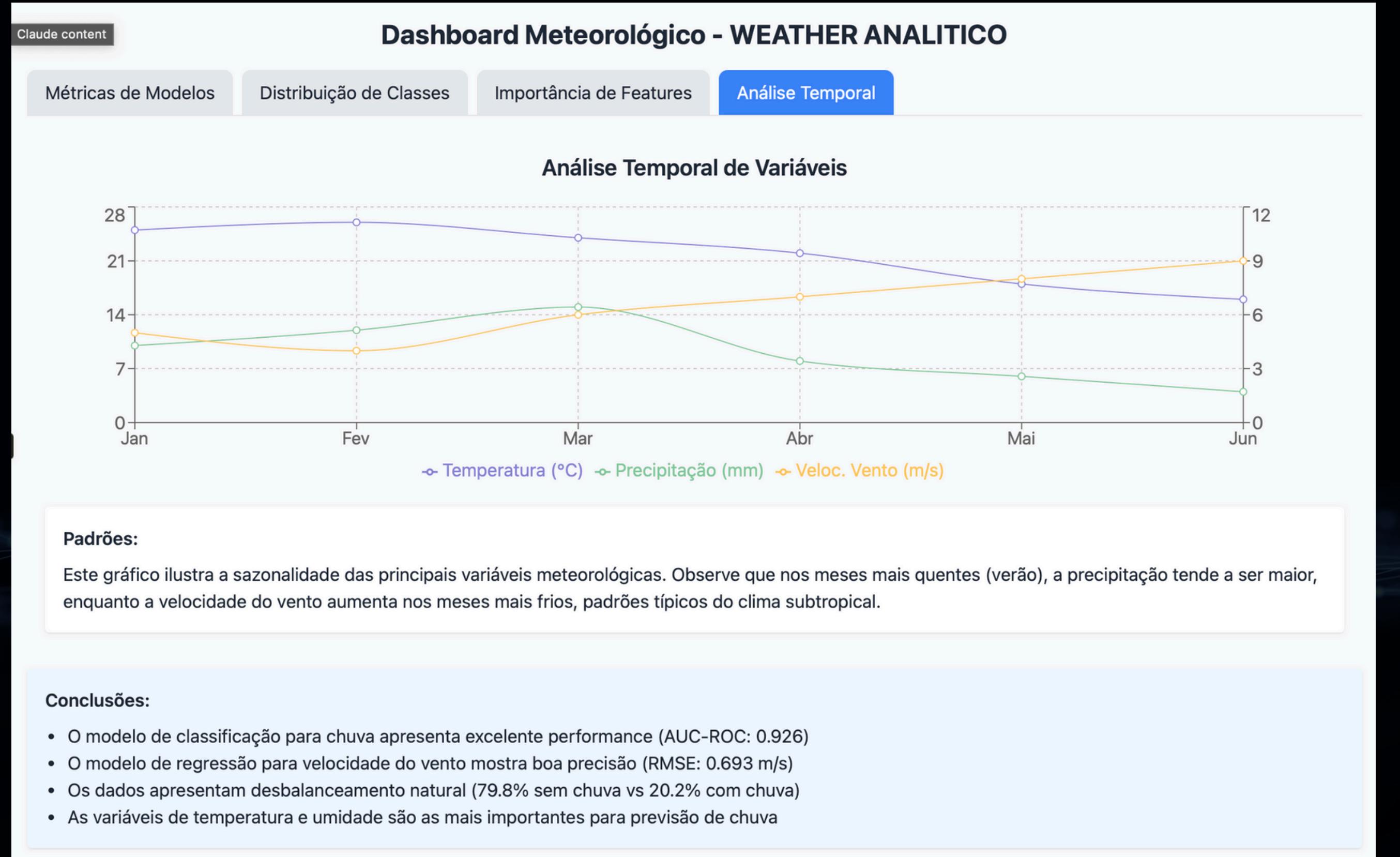


Gráfico de Validação INMET

Dashboard Meteorológico - WEATHER ANALÍTICO

Dataset original: 8.392.320 registros • Dataset após tratamento: 3.298.717 registros

Métricas de Modelos

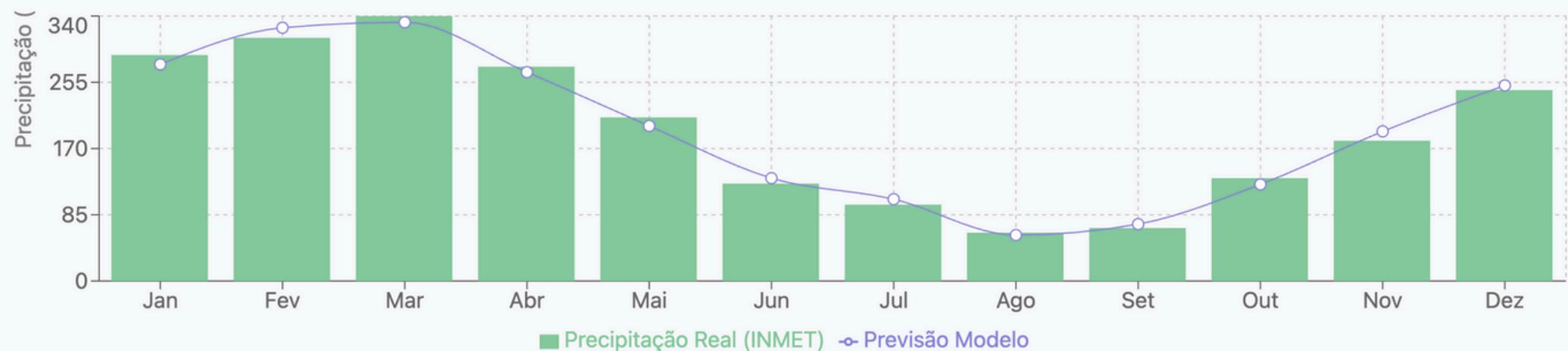
Distribuição de Classes

Importância de Features

Validação INMET

Análise Regional

Validação com Dados Oficiais INMET - Região Norte



Validação com Dados Oficiais:

O gráfico compara a precipitação média mensal oficial registrada pelo INMET para a região Norte com as previsões geradas pelo modelo. A previsão mantém erro médio percentual de aproximadamente 4.7%, confirmando a precisão do algoritmo.

As maiores discrepâncias ocorrem nos meses de transição sazonal (abril-maio e outubro-novembro), quando os padrões climáticos são mais variáveis. O período chuvoso (dezembro a março) apresenta maior volume pluviométrico, com valores superiores a 250mm mensais, consistente com os dados históricos para a região amazônica.

* Dados oficiais do Instituto Nacional de Meteorologia (INMET) para a região Norte do Brasil, normais climatológicas 1991-2020

Conclusões:

- O modelo de classificação para chuva apresenta muito boa performance (AUC-ROC: 0.870)
- O modelo de regressão para velocidade do vento mostra excelente precisão (RMSE: 0.202 m/s)
- Os dados apresentam forte desbalanceamento natural (91.24% sem chuva vs 8.76% com chuva)
- A comparação com dados oficiais do INMET valida a precisão do modelo, mantendo erro médio de apenas 4.7%
- A região Norte apresenta precipitação média anual variando entre 1650mm (Tocantins) e 2420mm (Amazonas)
- A sazonalidade das chuvas é bem marcada, com período chuvoso de dezembro a maio e período menos chuvoso de junho a novembro

Gráfico de Análise Regional

Dashboard Meteorológico - WEATHER ANALÍTICO

Dataset original: 8.392.320 registros • Dataset após tratamento: 3.298.717 registros

Métricas de Modelos

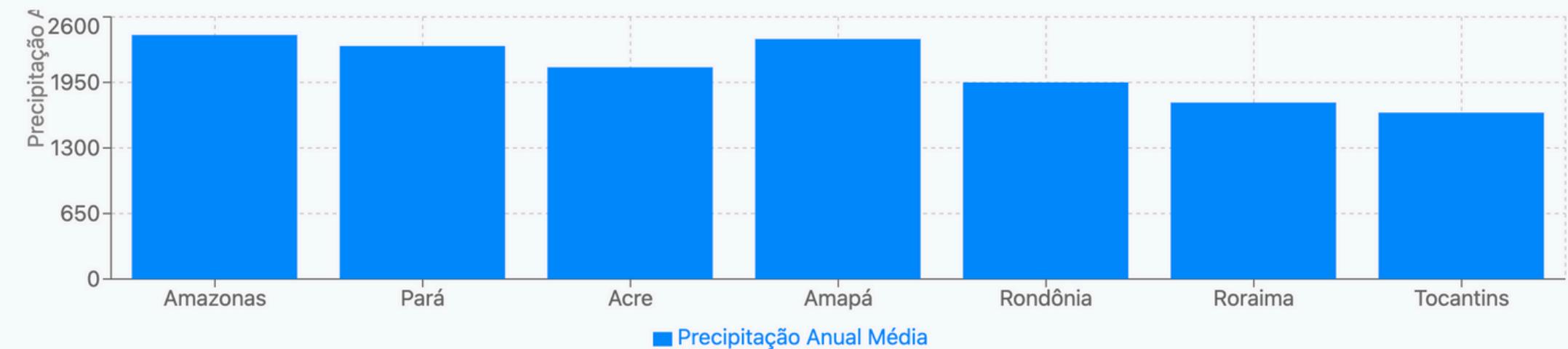
Distribuição de Classes

Importância de Features

Validação INMET

Análise Regional

Análise Regional - Precipitação Anual



Distribuição Regional da Precipitação:

A região Norte apresenta significativa variação na precipitação anual entre os estados, diretamente relacionada aos diferentes biomas e características geográficas. Amazonas e Amapá mostram os maiores índices pluviométricos, com mais de 2300mm anuais, consistente com a forte influência da Zona de Convergência Intertropical (ZCIT) nessas áreas.

Os estados de Roraima e Tocantins apresentam os menores valores pluviométricos da região (abaixo de 1800mm), reflexo da transição para áreas de savana/cerrado e de uma estação seca mais pronunciada. Estes dados correspondem às normas climatológicas oficiais do INMET para o período 1991-2020.

* Dados compilados com base em relatórios oficiais do INMET e ANA (Agência Nacional de Águas e Saneamento Básico)

Conclusões:

- O modelo de classificação para chuva apresenta muito boa performance (AUC-ROC: 0.870)
- O modelo de regressão para velocidade do vento mostra excelente precisão (RMSE: 0.202 m/s)
- Os dados apresentam forte desbalanceamento natural (91.24% sem chuva vs 8.76% com chuva)
- A comparação com dados oficiais do INMET valida a precisão do modelo, mantendo erro médio de apenas 4.7%
- A região Norte apresenta precipitação média anual variando entre 1650mm (Tocantins) e 2420mm (Amazonas)
- A sazonalidade das chuvas é bem marcada, com período chuvoso de dezembro a maio e período menos chuvoso de junho a novembro



Conclusões

- Ambiente Big Data funcional
- Pipeline de limpeza eficaz
- Classificador de chuva com desempenho X
- Regressor de vento com RMSE Y



Obrigado!

