

Previsão de consumo de energia elétrica:

Vinícius Person de Oliveira – Ra:221022503

O meu trabalho teve como objetivo prever o consumo anual de energia elétrica das cinco regiões do Brasil. A ideia surgiu da necessidade de entender como a demanda de energia evolui ao longo do tempo e como isso pode ajudar no planejamento da infraestrutura elétrica, na tomada de decisão e até na formulação de políticas públicas. Trabalhar com séries anuais também traz desafios, porque os dados são limitados e apresentam comportamentos diferentes entre as regiões.

O conjunto de dados utilizado foi derivado de informações públicas de consumo de energia, consolidadas por ano e por região. No final do processo de limpeza e preparação, ficamos com séries temporais anuais para cada região, com aproximadamente vinte anos de histórico. Cada série contém o total de energia consumida naquele ano, em megawatt-hora, sempre identificado pela região correspondente: Norte, Nordeste, Centro-Oeste, Sudeste e Sul.

Ao iniciar a modelagem, tentamos alguns métodos mais comuns, como regressão linear, Gradient Boosting e Random Forest considerando defasagens. Esses modelos funcionam bem em muitos cenários, mas percebemos que não eram adequados para esse tipo de série temporal. As técnicas de machine learning tradicionais não conseguem lidar bem com dependência temporal e tendência de longo prazo, o que fez com que os erros fossem maiores do que o esperado, principalmente para regiões que apresentavam mudanças mais abruptas no consumo.

Diante desses resultados, o método que escolhemos para o modelo final foi o SARIMA, que é uma variação do ARIMA. Esse método se adapta melhor porque considera autocorrelação e tendência, aspectos muito presentes no comportamento do consumo de energia ao longo dos anos. No nosso caso, como os dados eram anuais, não havia sazonalidade evidente, então a parte sazonal do modelo não teve grande impacto, mas o ARIMA em si já foi suficiente para capturar a dinâmica das séries.

Para avaliar o desempenho, reservamos os três últimos anos de cada série como conjunto de teste e compararamos as previsões do modelo com os valores reais. Utilizamos como principais métricas o MAE e o RMSE. Os resultados mostraram que o SARIMA foi significativamente mais preciso que o Random Forest com lags em todas as regiões. No Norte e no Sul, por exemplo, o erro ficou abaixo de dois milhões de megawatt-hora, o que é considerado muito bom para séries desse tipo. O Centro-Oeste e o Nordeste também apresentaram erros relativamente baixos. A região mais difícil foi o Sudeste, pois é a que

possui maior variação e o maior consumo absoluto. Mesmo assim, o SARIMA reduziu bastante o erro em comparação ao modelo inicial.

Região	RMSE – RF	RMSE – SARIMA	MAE – RF	MAE – SARIMA	RF (% do real)	SARIMA (% do real)
Norte	5,75 milhões	0,70 milhões	5,30 milhões	0,62 milhões	14,02%	1,72%
Nordeste	9,90 milhões	2,58 milhões	8,99 milhões	2,30 milhões	10,47%	2,73%
Centro-Oeste	3,29 milhões	0,58 milhões	2,83 milhões	0,50 milhões	7,68%	1,34%
Sudeste	17,64 milhões	6,63 milhões	15,50 milhões	5,78 milhões	6,85%	2,58%
Sul	7,47 milhões	1,67 milhões	6,62 milhões	1,49 milhões	7,56%	1,69%

No geral, os resultados indicam que o SARIMA foi a melhor escolha entre as técnicas testadas. Ele conseguiu capturar a tendência histórica de forma mais consistente e apresentou previsões mais confiáveis. Esse tipo de abordagem se mostrou adequado para séries anuais e pode ser útil no planejamento energético e em análises de evolução da demanda no longo prazo.



