

Previsão de Precipitação Acumulada Utilizando Redes Neurais MLP e Dados de Municípios Vizinhos

Ryan Alves¹, Fabrício Galende Marques de Carvalho¹, Juliana Anochi¹

¹Instituto Nacional de Pesquisas Espaciais (INPE)
São José dos Campos – SP – Brazil

ryan246.contato@gmail.com, {fabricio.galende, juliana.anochi}@inpe.br

Abstract. *This study contemplates the implementation and application of a machine learning model of the Multi-Layer Perceptron (MLP) type for forecasting the daily accumulation of precipitation in the municipality of Araraquara, state of São Paulo. The forecast is based on data from four neighboring municipalities within a 90-kilometer radius: Dois Córregos, Luís Antônio, Jaboticabal, and Cândido Rodrigues, using data from January, February and March 2023 to predict the months of April, May, and June 2023. The results demonstrate the potential of artificial neural networks for environmental estimates based on regional data.*

Resumo. *Este estudo contempla a implementação e aplicação de um modelo de aprendizagem de máquina do tipo Multi-Layer Perceptron (MLP) para previsão do acúmulo diário de precipitação no município de Araraquara, estado de São Paulo. A previsão é realizada com base em dados provenientes dos quatro municípios vizinhos em até 90 quilômetros de distância: Dois Córregos, Luís Antônio, Jaboticabal e Cândido Rodrigues, utilizando dados dos meses de janeiro, fevereiro e março de 2023 para prever os dados do mês de abril, maio e junho de 2023. Os resultados demonstram o potencial de redes neurais artificiais para estimativas ambientais baseadas em dados regionais.*

1. Introdução

A precipitação constitui um dos principais desafios ambientais em regiões urbanas e rurais no século 21. Entre os fenômenos meteorológicos mais estudados está a precipitação diária (ou mensal), devido à sua influência direta no abastecimento hídrico, agricultura, gestão de recursos hídricos e riscos de enchentes e secas [Gonçalves 2021]. Estimar suas concentrações, ou volumes, com precisão espacial e temporal é essencial para o monitoramento climático, além de ajudar significativamente na mitigação de desastres naturais e planejamento urbano.

A previsão da precipitação ainda representa um desafio, devido à complexidade dos processos atmosféricos envolvidos, à variabilidade sazonal e à forte interdependência entre diversos parâmetros meteorológicos, o que impacta diretamente o desempenho dos modelos.

Este trabalho investiga a previsão do acúmulo de precipitação diária no município de Araraquara a partir de dados históricos de quatro municípios vizinhos, utilizando redes neurais do tipo Multi-Layer Perceptron (MLP), uma das arquiteturas mais utilizadas em problemas de regressão não linear.

2. Metodologia

2.1. Dados Utilizados

Foi utilizado um arquivo CSV contendo o acúmulo diário de precipitação dos meses de janeiro, fevereiro e março de 2023 para todos os municípios do Brasil. Através de um filtro de municípios, foi feita uma análise de:

- Araraquara (município-alvo da previsão)
- Dois Córregos (86 km de distância do município-alvo da previsão)
- Luís Antônio (72 km de distância do município-alvo da previsão)
- Jaboticabal (72 km de distância do município-alvo da previsão)
- Cândido Rodrigues (80 km de distância do município-alvo da previsão)

Cada linha do arquivo contém o código do município (IBGE), nome do município, estado em que se localiza, a data e o valor da precipitação acumulada.

2.2. Obtenção dos dados

Foi realizado um levantamento preliminar de possíveis bases de dados para servirem de fonte principal da aquisição dos dados brutos de precipitação e a escolha foi a base de dados do Centro de Previsão de Tempo e Estudos Climáticos (CPTEC), do produto MERGE. O MERGE [Rozante et al. 2010] combina precipitação observada com estimativa de precipitação por satélite, que visa minimizar incertezas em dados de precipitação associados a interpolações sobre regiões com baixa densidade de pluviômetros. Na etapa de download, o dado que é disponibilizado pelo CPTEC de forma diária é do tipo GRIB2. Ele é convertido para o tipo NetCDF após o download para posteriormente ser processado. Já na etapa de processamento, o dado é extraído do arquivo NetCDF e processado, assegurando que as coordenadas geográficas estejam corretas de acordo com o recorte de cada estado e município a ser analisado. Dados disponíveis em <https://ftp.cptec.inpe.br/modelos/tempo/MERGE/GPM/DAILY> [CPTEC/INPE 2020]

Dado	MERGE
Nome da variável	prec
Unidade	km/m ²
Período	2010 – presente
Resolução temporal	Diário
Resolução espacial	10 km
Cobertura vertical	superfície

Tabela 1. Características do conjunto de dados MERGE

2.3. Processamento dos dados de precipitação

O sistema calcula a precipitação acumulada considerando pontos de amostragem localizados no centro da área urbana de cada município e fazendo o cálculo por interpolação bilinear, considerando os quatro pixels mais próximos para gerar um valor ponderado. A precipitação acumulada é obtida a partir dos dados de entrada e ajustada para milímetros com atributos descritivos e unidades correspondentes.

Após o processamento, os dados são colocados no banco de dados para posteriormente poderem ser transformados em um arquivo Comma Separated Values (CSV).

2.4. Pré-processamento

Os dados foram filtrados para conter apenas os cinco municípios de interesse, fazendo a verificação e exclusão de valores nulos. Posteriormente, o DataFrame foi transformado para o formato *wide*, em que cada município corresponde a uma coluna, e as linhas representam os dias do mês, além de ter sido ordenado para que os dados mantenham a ordem cronológica. Foi criada uma máscara temporal para selecionar os dados do período de primeiro de janeiro a trinta de junho de 2023, garantindo que o conjunto de dados utilizado abrangesse apenas os meses desejados para análise.

2.5. Modelo MLP

O modelo utilizado foi o **MLPRegressor** pertencente a biblioteca `scikit-learn` [Pedregosa et al. 2011], com as seguintes configurações obtidas através da busca por hiperparâmetros realizada com a utilização da biblioteca `Optuna` [Shekhar et al. 2022]:

- Camadas ocultas: 17, 23, 43, 26, 26)
- Função de ativação: ReLu
- Taxa de aprendizagem inicial: 0.019
- Regularização: L2 (0.035)
- Iterações máximas: 350
- solver: adam
- early stopping: true
- Aleatoriedade controlada com `random_state=190`

2.6. Validação

A validação utilizada foi a `TimeSeriesSplit`, que mantém a ordem temporal dos dados, sendo a opção mais adequada neste caso de previsão. Como os dados não podem acessar informações futuras, cada fold utiliza dados passados para treinamento e dados futuros para teste, simulando cenários reais de previsão (número de splits utilizados: 5).

3. Resultados

Após o treinamento, o modelo foi avaliado utilizando os dados de abril, maio e junho de 2023. O gráfico na Figura 2 abaixo mostram a comparação entre os valores reais de precipitação acumulada em Araraquara e as previsões realizadas pelo modelo MLP, com base nos 4 municípios próximos. Também é demonstrado as métricas de Mean Absolute Error (MAE), que foi escolhida para este modelo já que é uma métrica robusta a outliers pois evita que o modelo seja penalizado de forma exagerada por alguns dias extremos de chuvas intensas.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Figura 1. Fórmula do Mean Absolute Error

O MAE obtido foi de 2.55 mm, indicando que, em média, a previsão diária de precipitação acumulada do modelo difere do valor observado em 2.55 mm. Esse indicador reflete a precisão média do modelo para cada dia do período de teste.

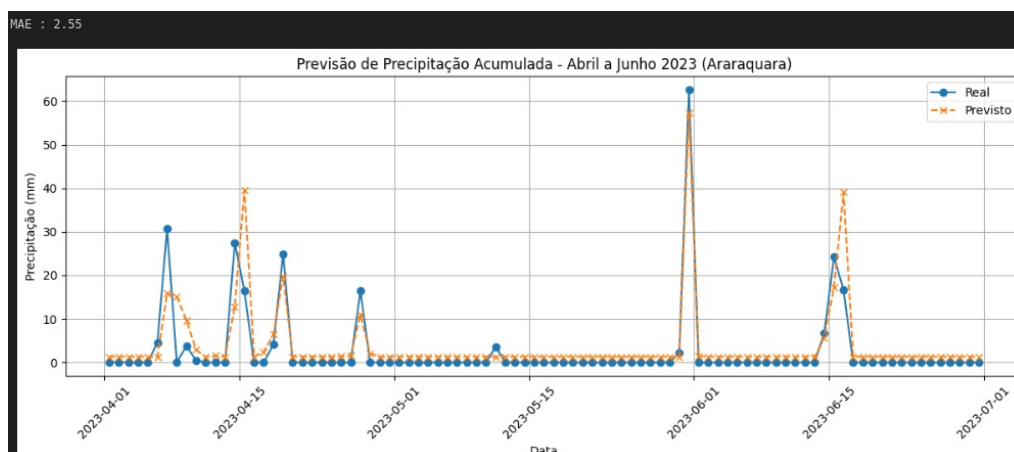


Figura 2. Resultado da previsão do mês de abril de 2023 utilizando os dados históricos dos meses de fevereiro e março de 2023

O modelo embora não preveja exatamente o valor de precipitação acumulada do dia avaliado, consegue captar as tendências de aumento de chuva para o mês avaliado.

4. Conclusão

Este estudo demonstrou que por meio da utilização de redes neurais como MLP é possível capturar tendências da precipitação acumulada em um município a partir de dados de municípios vizinhos, em meses que apresentam comportamento similares. O modelo mostrou-se eficaz na previsão temporal do mês de abril, maio e junho de 2023 quando foi treinado com dados de janeiro, fevereiro e março de 2023. Com isso, foi possível notar também a possibilidade real de se prever a precipitação acumulada de um município com base em seus municípios próximos.

Futuramente, buscando-se melhorar e diversificar os testes realizados, pretende-se também testar com uma maior quantidade de períodos tendo em vista que MLPs tendem a exigir um tamanho de amostra maior em comparação a outros modelos de aprendizado de máquina[Zhang et al. 2021] e diversidade de arquiteturas para maior robustez do modelo.

5. Agradecimentos

O autor agradece ao INPE, ao CNPQ, ao programa Queimadas, ao BiomasBR e ao Ministério da Saúde pelo apoio.

Referências

- CPTEC/INPE (2020). Produto de precipitação merge. Online. Disponível em: <https://ftp.cptec.inpe.br/modelos/tempo/MERGE/>. Acesso em: 15 de agosto de 2025.
- Gonçalves, S. T. N. (2021). Índices e metodologias de monitoramento de secas. *Revista Brasileira de Meteorologia*, 39(1):1–10. Disponível em: <https://www.scielo.br/j/rbmet/a/rGqGprMLXW5F66b6R98D98j/?format=pdf&lang=pt>. Acesso em: 15 ago. 2025.

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830. Available at: <https://jmlr.org/papers/v12/pedregosa11a.html>. Accessed on: August 15, 2025.
- Rozante, J. R., Moreira, D. S., de Goncalves, L. G. G., and Vila, D. A. (2010). Combining trmm and surface observations of precipitation: technique and validation over south america. *Weather and Forecasting*, 25(3):885–894.
- Shekhar, S. et al. (2022). A comparative study of hyper-parameter optimization tools. *arXiv preprint arXiv:2201.06433*. Available at: <https://arxiv.org/abs/2201.06433>. Accessed on: August 15, 2025.
- Zhang, Y. et al. (2021). Machine learning for precipitation forecasts postprocessing. *Journal of Hydrometeorology*, 22(11):2345–2358.