



I CONGRESSO PERNAMBUCANO DE RECURSOS HÍDRICOS

Água para o Desenvolvimento
Recife, 24, 25 e 26 de Março de 2026

CLASSIFICAÇÃO DOS MUNICÍPIOS DE PERNAMBUCO COM BASE EM ÍNDICES PLUVIOMÉTRICOS: UMA ANÁLISE DE TÉCNICAS DE CLUSTERING

Clara Aquino¹ ; Waldemar Ferreira²

Palavras-chave: precipitação, clustering, classificação

INTRODUÇÃO

A previsão de precipitação no nordeste brasileiro é desafiadora devido à elevada variabilidade climática e à vulnerabilidade hídrica (Comin 2021). Uma das alternativas para mitigar as dificuldades de analisar os padrões de chuva no Nordeste é fazê-lo segundo a organização de mesorregiões (Kleber 2006). As mesorregiões são uma área individualizada de um Estado, que apresenta formas de organização do espaço geográfico definidas por três dimensões — o processo social, o quadro natural e a rede de comunicação (Guimarães 1971). Contudo, essa delimitação frequentemente não reflete os índices pluviométricos de alguns municípios (Correia 2019).

A literatura científica tem explorado o uso de técnicas de agrupamento (clustering) para mitigar o problema de previsão de precipitação, utilizando apenas esse critério de agrupamento. Dentre os métodos de clustering mais usados está o método hierárquico de Ward. Ele se baseia na distância euclidiana entre os municípios para agrupá-los. Por exemplo, Mutti et al. (2020) aplicaram esta técnica na Bacia do Rio Piranhas-Açu (RN/PB), identificando duas sub-regiões distintas. Da mesma forma, Amorim et al. (2020) adotaram a mesma abordagem para zonear o Rio Grande do Norte, isolando o litoral do interior, enquanto Oliveira-Júnior et al. (2021) mapearam períodos secos e úmidos em Alagoas. No contexto específico de Pernambuco, Silva et al. (2021) usaram Ward para agrupar alguns municípios, contudo, restringiram-se à Zona da Mata e à Região Metropolitana do Recife. Além deles, Jardim et al. (2021) focaram em um gradiente longitudinal específico, identificando apenas dois grandes grupos (Recife/Triunfo vs. Sertão).

Até o limite do nosso conhecimento, nenhum trabalho aplica diferentes técnicas de inteligência artificial (IA), além do método hierárquico de Ward, para analisar os municípios com base em seus índices pluviométricos. Neste trabalho, apresentamos diversas técnicas de IA para clusterização, visando avaliar como elas classificam os municípios, e planejamos verificar se essas classificações estão de acordo com a organização das mesorregiões do estado de Pernambuco. Em particular, usaremos os algoritmos baseados em partição (K-Means) (Tan, et al. 2021), em densidade (Density-

¹Afiliação: Universidade Federal Rural de Pernambuco - Unidade Acadêmica de Belo Jardim, (81) 3320-6930, clara.aquino@ufrpe.br.

²Afiliação: Universidade Federal Rural de Pernambuco - Unidade Acadêmica de Belo Jardim, (81) 3320-6930, waldemar.ferreira@ufrpe.br

Based Spatial Clustering of Applications with Noise - DBSCAN) (Tan, et al. 2021) e o método hierárquico de Ward, com base na distância euclidiana (Tan, et al. 2021). Como base de dados, usaremos o Global Precipitation Climatology Project (GPCP) (Alder et al., 2003), que foi o mesmo utilizado por Anochi et al. (2021) no trabalho de previsão meteorológica na América do Sul. Vale salientar que este trabalho está em andamento e apresentaremos apenas os resultados preliminares.

METODOLOGIA

O estado de Pernambuco é dividido em 185 cidades com 5 grandes mesorregiões (Guimarães 1971):

- Região Metropolitana do Recife - inclui a capital e a área metropolitana, com características litorâneas e manguezais;
- Mata Pernambucana - região litorânea;
- Agreste Pernambucano - transição entre a mata úmida e o sertão seco;
- Sertão Pernambucano - região mais seca e árida do interior;
- São Francisco Pernambucano - abrange o vale do rio São Francisco.

Os dados meteorológicos a serem utilizados no estudo foram obtidos do Global Precipitation Climatology Project (GPCP). Ele fornece uma análise consistente da precipitação global, a partir da integração de diversos conjuntos de dados de satélite, além de análises de pluviômetros em terra. O conjunto abrange dados mensais de 1979 até agosto de 2025. Além disso, o Banco de Dados do Estado forneceu as informações geográficas (altitude, latitude, longitude) e a divisão mesorregional das cidades.

Como etapa preliminar ao agrupamento das cidades de acordo com o índice pluviométrico, será estruturada uma análise descritiva. Nesta fase, será verificada a distribuição mensal das chuvas em cada mesorregião, permitindo a visualização dos ciclos hidrológicos locais. Em seguida, serão calculadas a média e a mediana de precipitação para cada mesorregião. Além disso, os dados serão agrupados para analisar a frequência trimestral da precipitação. O uso dessa análise indicará quais modelos são plausíveis para a próxima etapa (Bussab, Morettin, 2010). A metodologia exige a confirmação de pressupostos básicos para testes paramétricos (normalidade - Shapiro-Wilk; homogeneidade de variâncias - Levene), a fim de garantir a validade estatística das comparações mesorregionais (Bussab, Morettin, 2010).

Uma vez que os dados forem avaliados, será aplicada a análise de variância para testar a hipótese de que a localização geográfica (mesorregião) influencia significativamente o volume de precipitação. Formalmente, definem-se as seguintes hipóteses: H_0 , na qual, as médias das mesorregiões são iguais e H_1 , na qual, pelo menos uma das médias é estatisticamente diferente.

Caso a hipótese nula (H_0) seja rejeitada, será aplicado um teste de comparação múltipla (Teste de Tukey - Tukey 1949). Este procedimento é fundamental para determinar exatamente onde e em quais períodos (trimestres) ocorrem as divergências, bem como identificar quais regiões apresentam comportamentos pluviométricos semelhantes (Bussab, Morettin, 2010).

Visando identificar a classificação de municípios mais eficiente em função da precipitação, a última etapa será a implementação e a comparação de diferentes paradigmas de aprendizado de máquina. No escopo não supervisionado, serão aplicados os seguintes algoritmos: (i) K-Means, baseado em partição; (ii) o DBSCAN, baseado em densidade; e (iii) o clustering hierárquico de Ward. Usaremos Silhouette, Calinski-

Harabasz Index e Davies-Bouldin Index para avaliar o resultado de cada algoritmo (Tan, et al. 2021).

RESULTADOS PRELIMINARES

Na análise descritiva inicial, a separação dos dados por mesorregiões e a distribuição mensal mostraram precipitações semelhantes na maioria dos meses, exceto entre dezembro e março, quando as diferenças foram maiores. Ao analisar a mediana de cada mesorregião, verifica-se que os valores são bastante próximos (variando de 1.55 no Sertão do São Francisco, a menor, a 1.70 na região metropolitana do Recife, a maior). Essa proximidade das medianas sugere que os grupos não apresentam distinção estatística significativa (Bussab, Morettin 2010; World Meteorological Organization 2008). Examinando a frequência das regiões, o Agreste e a Mata Pernambucana têm características muito parecidas e algumas vezes também semelhantes à Região Metropolitana do Recife. O Sertão e o São Francisco têm características parecidas quando se trata da frequência de chuvas também.

Em seguida, espera-se que os testes de normalidade e homogeneidade revelem a natureza complexa e possivelmente não-paramétrica das séries temporais de precipitação em certas mesorregiões. Com a aplicação da ANOVA, projeta-se a rejeição da Hipótese Nula (H_0), confirmando estatisticamente a existência de variância significativa entre as médias de precipitação das atuais mesorregiões administrativas. Contudo, por meio do teste post hoc, espera-se identificar redundâncias climáticas (World Meteorological Organization, 2008), ou seja, mesorregiões geograficamente distintas (como partes do Agreste e do Sertão) que não apresentam diferença estatisticamente significativa em seus padrões pluviométricos. Este resultado será crucial para demonstrar que a divisão política atual é subótima para fins de modelagem meteorológica.

Na etapa de modelagem não supervisionada, espera-se que os algoritmos K-Means, DBSCAN e Hierárquico de Ward gerem um novo mapeamento climático de Pernambuco. Como resultado visual, prevê-se a produção de um heatmap que reclassifica os 185 municípios em novas zonas homoclimáticas.

CONCLUSÕES

O trabalho conclui que a otimização da previsão meteorológica em Pernambuco exige uma segmentação territorial data-driven que supere as divisões político-administrativas, propondo uma metodologia robusta que combina inferência estatística e aprendizado de máquina não supervisionado para mitigar generalizações em zonas extensas como Sertão e Agreste; a validação cruzada com algoritmos supervisionados assegura que os novos agrupamentos representem fielmente a realidade fenomênica das chuvas, e não meros artefatos matemáticos. Futuramente, pretende-se concluir a análise comparativa dos algoritmos de agrupamento para definir a segmentação espacial definitiva do estado, cujos clusters — representativos de zonas climaticamente homogêneas — servirão de base para o desenvolvimento de modelos de previsão com redes neurais, com a expectativa de que o treinamento nessas novas zonas otimizadas possibilite criar um software de previsão para o interior de Pernambuco com acurácia superior à obtida pelas divisões mesorregionais tradicionais.

AGRADECIMENTOS

Agradecemos à Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) pelo apoio financeiro por meio do processo nº BIC-1463-1.03/25.

REFERÊNCIAS

- Adler, R.F., G.J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, and P. Arkin, 2003: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *J. Hydrometeor.*, 4, 1147-1167.
- ALVARES, Clayton Alcarde et al. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, p. 711–728, 1 dez. 2013.
- AMORIM, Ana Cleide Bezerra et al. Regionalização e Análise da Tendência da Precipitação do Rio Grande do Norte Associados a Padrões de TSM. *Revista Brasileira de Meteorologia*, v. 35, n. 2, p. 269–280, jun. 2020.
- ANOCHI, J. A.; DE ALMEIDA, V. A.; DE CAMPOS VELHO, H. F. Machine learning for climate precipitation prediction modeling over South America. *Remote Sens* 13: 2468 [em linha]. 2021.
- BDE. Disponível em: <http://www.bde.pe.gov.br/estruturacaogeral/institucional_o_que_e_a_bde.aspx>. Acesso em: 14 dez. 2025.
- BUSSAB, Wilton de O.; MORETTIN, Pedro A. Estatística básica. In: *Estatística básica*. 2010. p. xvi, 540-xvi, 540.
- COMIN, Alcimoni Nelci et al. Extreme rainfall event in the Northeast coast of Brazil: a numerical sensitivity study. *Meteorology and Atmospheric Physics*, v. 133, n. 2, p. 141–162, 1 abr. 2021.
- CORREIA FILHO, Washington Luiz Félix *et al.* Rainfall variability in the Brazilian northeast biomes and their interactions with meteorological systems and ENSO via CHELSA product. **Big Earth Data**, v. 3, n. 4, p. 315–337, 2 out. 2019.
- GUIMARÃES, Augusto. FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA PRESIDENTE. 1971.
- JARDIM, Alexandre Maniçoba da Rosa Ferraz et al. Spatiotemporal climatic analysis in Pernambuco State, Northeast Brazil. *Journal of Atmospheric and Solar-Terrestrial Physics*, v. 223, p. 105733, 15 out. 2021.
- KELLER FILHO, Thadeu; ZULLO JUNIOR, Jurandir; LIMA, Paulo Roberto Schubnell De Rezende. Análise da transição entre dias secos e chuvosos por meio da cadeia de Markov de terceira ordem. *Pesquisa Agropecuária Brasileira*, v. 41, n. 9, p. 1341–1349, set. 2006.
- MUTTI, Pedro R. et al. A detailed framework for the characterization of rainfall climatology in semiarid watersheds. *Theoretical and Applied Climatology*, v. 139, n. 1–2, p. 109–125, jan. 2020.
- OLIVEIRA-JÚNIOR, José Francisco de et al. Wet and dry periods in the state of Alagoas (Northeast Brazil) via Standardized Precipitation Index. *Journal of Atmospheric and Solar-Terrestrial Physics*, v. 224, p. 105746, 1 nov. 2021.
- SILVA, Marcos Vinícius da et al. Spatial modeling of rainfall patterns and groundwater on the coast of northeastern Brazil. *Urban Climate*, v. 38, p. 100911, 1 jul. 2021.
- TAN, PANG-NING et al. *Introduction to Data Mining*. 2. ed. NY, NY: Pearson, 2021.
- TUKEY, John W. Comparing individual means in the analysis of variance. *Biometrics*, p. 99-114, 1949.
- World Meteorological Organization. *Guide to meteorological instruments and methods of observation*. Secretariat of the World Meteorological Organization, 2008.