

APLICAÇÃO DE MODELOS DE INTELIGÊNCIA ARTIFICIAL PARA PREVISÃO DE PERÍODOS DE ESTIAGEM NO SERIDÓ PARAIBANO

Emerson da Silva Freitas (IFPB, Campus Picuí), Wamberto Raimundo da Siva Júnior (IFPB, Campus Picuí), Júlia Souto Lima (IFPB, Campus Picuí).

E-mails: emerson.freitas@ifpb.edu.br, wamberto.silva@ifpb.edu.br, julia.souto@academico.ifpb.edu.br.

Área de conhecimento (Tabela CNPq): 3.01.04.00-9 Engenharia Hidráulica.

Palavras-chave: Machine Learning; Florestas Aleatórias; Seca.

1. Introdução

A escassez de água é um desafio complexo que afeta diversas regiões do mundo. O semiárido brasileiro é uma dessas regiões que sofre constantemente com longos períodos sem precipitação. Além disso, a crescente demanda por água, impulsionada pelo crescimento populacional, pela necessidade de produção agrícola e pela urbanização, combinada com os efeitos das mudanças climáticas, tem aumentado a vulnerabilidade de comunidades à escassez hídrica. Na Paraíba, principalmente em algumas microrregiões como Seridó, Cariri e Sertão, a escassez hídrica prolongada e as recentes secas trazem consequências devastadoras para a economia, o meio ambiente e a qualidade de vida das pessoas.

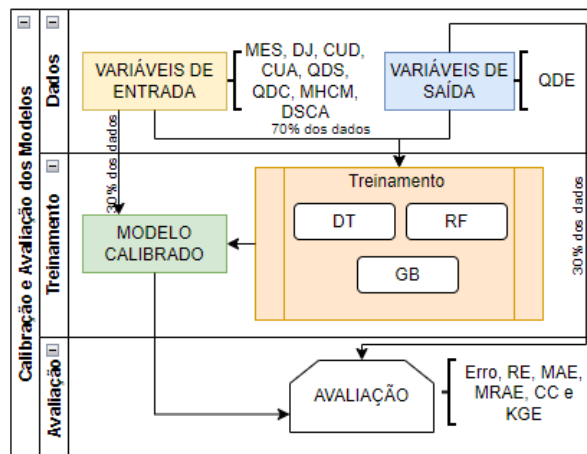
A previsão de secas é imensamente útil no alerta precoce e na preparação das comunidades mais vulneráveis para os seus impactos adversos (KHAN et al., 2020). A previsão precisa da seca e a avaliação de tendências são essenciais para a gestão da água para reduzir os efeitos prejudiciais da seca (HAMEED et al., 2023). Vários algoritmos de inteligência artificial, como machine learning (ML) e deep learning (DL), desenvolveram técnicas eficazes e eficientes para modelagem de sistemas hidrológicos complexos lineares e não lineares, ciência climática e produção agrícola amplamente utilizadas para previsões (ELBELTAGI et al., 2023). Já no Brasil, ainda há poucos trabalhos nessa linha de previsões de seca com uso de técnicas de machine learning, CARVALHO; SOUZA FILHO; LOPES (2021) apresentaram um modelo de classificação para seca na região nordeste do Brasil baseados em modelos de Support Vector machine (SVM) e OLIVEIRA et al. (2023) utilizaram artificial neural network (ANN) para calibração de um modelo de previsão de estiagens na região Sul do Brasil.

Diante disso, este estudo propõe a aplicação de modelos de inteligência artificial com base no machine learning, utilizando dados pluviométricos de longo prazo, na previsão de secas a fim de se obter os padrões dos períodos de estiagem. Com isso, espera-se desenvolver modelos de previsão, robustos e eficazes, capazes de fornecer informações valiosas para apoiar a tomada de decisão e a gestão adaptativa dos recursos hídricos na região.

2. Materiais e métodos

A metodologia de calibração está resumida na Figura 1.

Figura 1- Fluxograma de calibração e avaliação dos modelos



Fonte: Autor, 2024 (MES- mês do ano, DJ- dia juliano, CUD-quantidade de chuva em mm nos dias antecedentes, CUA- quantidade de chuva em mm nos dias antecedentes, QDS- quantidade de dias antecedentes sem chuva, QDC-quantidade de dias antecedentes, MHCM- média histórica de chuva mensal, DSCA- quantidade de dias sem chuva antecedentes e QDE- quantidade de dias de estiagem)

2.1 Área de Estudo

A área a ser estudada consiste nas microrregiões do estado da Paraíba: Seridó oriental e Seridó Ocidental. A primeira região abrange 9 municípios paraibanos (Picuí, Juazeirinho, Seridó, Pedra Lavrada, Cubati, Nova Palmeira, Baraúna, Tenório e Frei Martinho) com aproximadamente 80 mil e uma pluviosidade média anual de cerca de 400 mm. Já a segunda região é composta por 6 municípios (Santa Luzia, São Mamede, Junco do Seridó, São José do Sabugi,

Salgadinho e Várzea) com aproximadamente 40 mil habitantes e uma pluviosidade média de cerca de 500 mm.

2.2 Fonte de dados e controle de qualidade

Os dados pluviométricos utilizados foram obtidos da Agência Executiva de Gestão das Águas (AESAs), com duas estações que possuam dados diários de 1952 a 2017, localizadas na cidade de Picuí e Santa Luzia. Os dados foram analisados quanto à verificação de falhas e essas foram preenchidas com o método do método do Vetor Regional (Hiez, 1978) e posteriormente homogêneos com o Método da Dupla Massa com estações vizinhas.

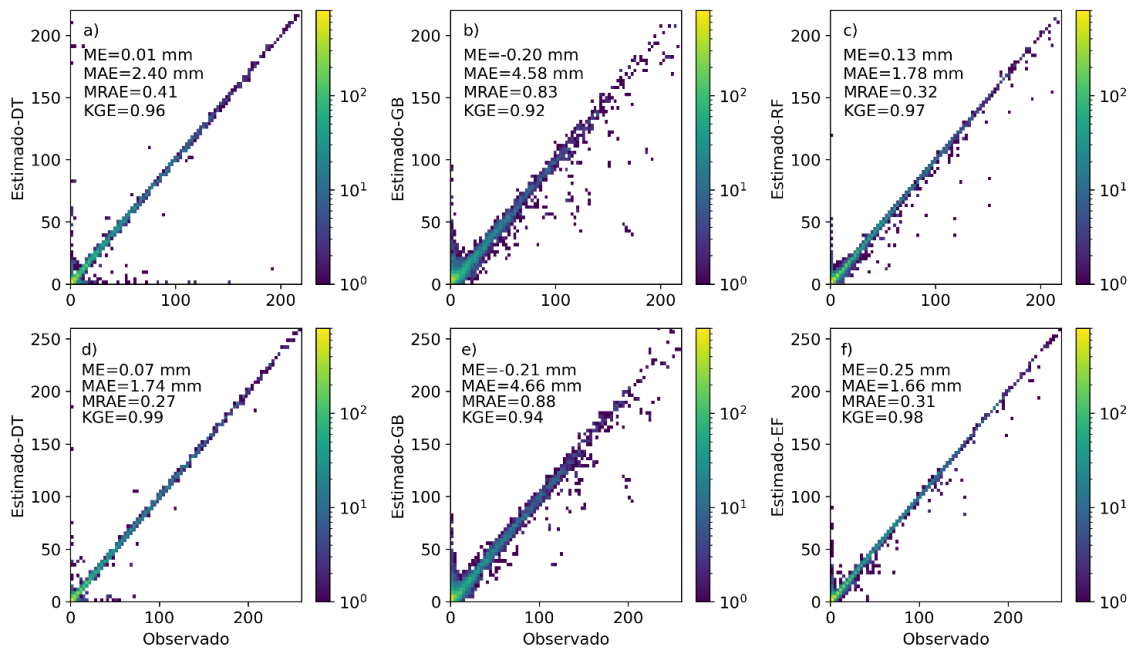
2.3 Modelos de Machine Learning e Calibração

Os modelos de regressão que foram utilizados são aqueles baseados em árvores de decisão: i) Decision Tree (DT); ii) Random forest (RF) e iii) Gradient Boosting (GB). Os modelos de regressão foram calibrados com 8 variáveis de entrada e com 1 variável de saída. Para o treinamento dos modelos os dados serão subdivididos em dois conjuntos de dados: treinamento (70%) e teste (30%), para calibrar e validar o modelo, respectivamente.

3. Resultados e discussão

A Figura 2 mostra os gráficos de dispersão considerando as duas regiões estudadas (Seridó Leste e Oeste). No geral, os modelos utilizados apresentam uma boa concordância entre os dados observado e os dados previstos pelos modelos, principalmente para os maiores valores de QDE. Os valores de KGE, que variam de 0,92 a 0,99 para os diferentes modelos e regiões reforça a boa concordância entre esses dados. Os valores de ME e MAE dos modelos apresentam variações de - 0,20 a 0,13 e 1,78 a 4,58 respectivamente para a região do Seridó Leste e de -0,21 a 0,25 e 1,66 a 4,66 respectivamente para a região do Seridó Oeste. Nota-se que dentre os modelos, o GB é o que apresenta valores de subestimação dos dados (ME negativo), o que pode ser um problema, visto que pode levar a falsas expectativas do tomador de decisão. O ideal é um modelo que possua baixos erros, mas nesse caso errar para menos é um problema, pois a previsão de estiagem seria menor que a que que de fato vai acontecer.

Figura 2- Gráficos de dispersão para o QDE observados e estimados pelos modelos (a) Decision Tree (DT), (b) Gradient Boosting (GB), (c) Randon Forest (RF) para a região Seridó Leste, e (d) Decision Tree (DT), (e) Gradient Boosting (GB), (f) Randon Forest (RF) pa



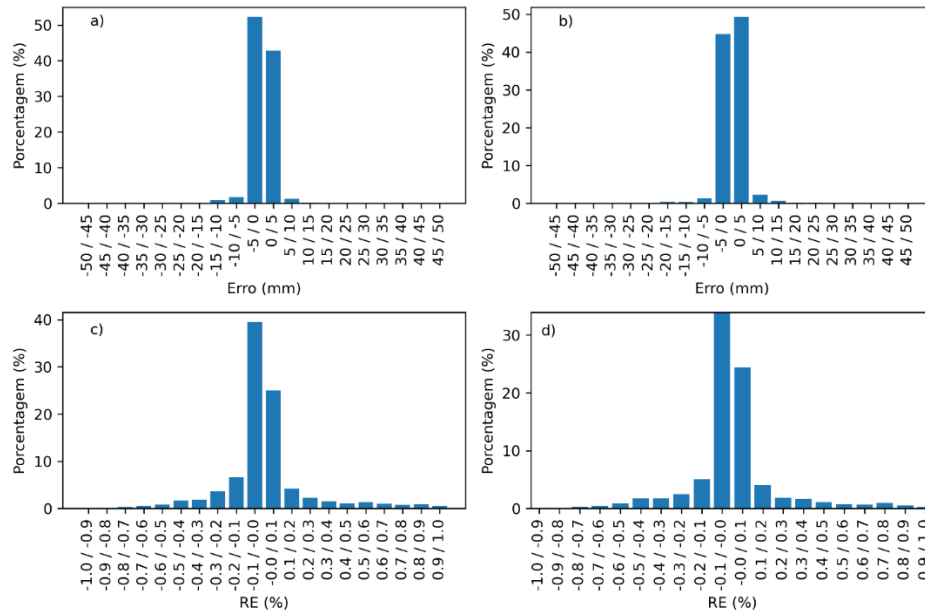
Fonte: Autor (2024)

No geral, o modelo de RF teve um desempenho melhor do que os seus concorrentes com os erros absolutos mais baixos (MAE = 1,78 mm e 1,66 mm para as regiões Seridó Leste e Oeste respectivamente) e com os relativos absolutos (MRAE = 0,32 e 0,31 para as regiões Seridó Leste e Oeste respectivamente). As pontuações do KGE confirmam que o modelo RF (KGE = 0,97 e 0,98) fornece uma melhor estimativa de QDE do que os modelos GB (KGE = 0,92 e 0,94) e DT (KGE = 0,96 e 0,99), para as regiões Seridó Leste e Oeste respectivamente. A exceção é Seridó Oeste que o DT possui KGE igual a 0,99 na região Oeste. Ao se analisar visualmente o gráfico, nota-se também que o RF é o que possui menos dados abaixo da reta 1:1 indicando menos dados de subestimativa.

A Figura 3 mostra que aproximadamente 96% e 98% dos valores de QDE do modelo RF apresentam erros inferiores a 5 dias de subestimação e superestimação para as regiões Seridó Leste e Oeste respectivamente. Para os valores de QDE entre 5 e 10 dias são apenas 2% para ambas. Já com relação ao RE, o panorama é praticamente o mesmo, com 80

e 71% dos dados com erros relativos menores que 10% de subestimação e superestimação para as regiões Seridó Leste e Oeste respectivamente. Para os intervalos de 10 a 20% e 20 a 30% tem-se 10 e 5% dos dados respectivamente para ambas as regiões. Isso mostra que o modelo RF erra em sua grande parte (mais que 95% dos casos) em até 5 dias e esse erro em grande parte dos dados (95% dos casos) representam um erro relativo menor que 30%.

Figura 3- Histograma do (a) e (b) Erro (dia) e (c) e (d) RE para as regiões do Seridó Leste e Seridó Oeste, respectivamente, dos valores de QDE estimados pelo modelo RF.



Fonte: Autor (2024)

5. Considerações finais

Esse estudo propôs a aplicação de modelos de inteligência artificial com base no machine learning, utilizando dados pluviométricos de longo prazo para as regiões Seridó Leste e Oeste do estado da Paraíba, para a previsão de períodos de estiagem. As conclusões obtidas por essa pesquisa são:

- Modelos baseados em árvores de decisão são satisfatórios em previsão de dias de estiagem;
- O modelo baseado em florestas aleatórias (RF) mostrou-se mais eficiente que o DT e GB;
- O modelo de previsão de estiagem e seca mostrou-se eficiente e robusto, podendo ser usado como ferramenta de auxílio ao tomador de decisão no gerenciamento dos recursos hídricos.

Agradecimentos

Os autores agradecem a Chamada Interconecta IFPB - Edital Nº 03/2024 - Apoio a projetos de pesquisa, inovação, desenvolvimento tecnológico e social pela verba para concessão de bolsas para os alunos de IC.

Referências

CARVALHO, T. M. N.; SOUZA FILHO, F. DE A. DE; LOPES, T. M. X. DE M. DETECÇÃO DE SECAS E VISUALIZAÇÃO DE PADRÕES CLIMÁTICOS COM APRENDIZADO DE MÁQUINA. XXIV SIMPÓSIO BRASILEIRO DE RECURSOS HIDRÍCOS. Anais...Belo Horizonte. MG: 2021.

ELBELTAGI, A. et al. Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. Environmental Science and Pollution Research, v. 30, n. 15, p. 43183–43202, 1 mar. 2023.

HAMEED, M. M. et al. Machine learning models development for accurate multi-months ahead drought forecasting: Case study of the Great Lakes, North America. PLoS ONE, v. 18, n. 10 October, 1 out. 2023.

KHAN, N. et al. Prediction of droughts over Pakistan using machine learning algorithms. Advances in Water Resources, v. 139, 1 maio 2020.

OLIVEIRA, G. G. DE et al. MODELO BASEADO EM REDES NEURAIAS ARTIFICIAIS E DADOS DE CHUVA POR SATÉLITES PARA ALERTA DE ESTIAGENS E SECAS. Simpósio Brasileiro de Sensoriamento Remoto. Anais...Florianópolis, SC: 2023.