

## **XV SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE**

# **APLICAÇÃO DE *DEEP LEARNING* PARA A PREDIÇÃO DA PRECIPITAÇÃO PLUVIOMÉTRICA MENSAL NO MUNICÍPIO DE CARUARU-PE**

*Jean Firmino Cardoso*<sup>1</sup>; *Lucas Ravellys Pyrrho de Alcântara*<sup>2</sup>; *Severino Martins dos Santos Neto*<sup>3</sup>;  
*Artur Paiva Coutinho*<sup>4</sup> & *Antonio Celso Dantas Antonino*<sup>5</sup>

**RESUMO** – Estudos sobre a chuva e clima ganharam grande relevância em diversos setores econômicos e sociais, tendo em vista que a previsão do tempo é um fator determinante do potencial hídrico de uma região. Nesse sentido, o objetivo desse trabalho construir um modelo de Deep Learning para a predição da precipitação mensal utilizando apenas a série de histórica de dados. Para isto foram utilizados dados das precipitações mensais de 1960 até 2018 do município de Caruaru-PE. Esses dados foram implementados em um modelo de redes neurais artificiais (RNA) no manejo e análise das informações. Os dados foram analisados em conjunto com diferentes redes neurais utilizadas, a melhor aproximação obtida foi determinada. É possível utilizar as redes neurais artificiais para a problemática, prevendo os volumes de precipitações mensais tendo como base o banco de dados históricos. Mesmo obtendo resultados razoáveis, a quantidade de dados não foi suficiente para o processamento eficaz almejado.

**ABSTRACT**– Studies on rain and climate have gained great relevance in several economic and social sectors, considering that the weather forecast is a determining factor of a region's water potential. In this sense, the objective of this work is to build a Deep Learning model for the prediction of monthly precipitation using only the series of historical data. For this, data from monthly rainfall from 1960 to 2018 in the municipality of Caruaru-PE were used. These data were implemented in a model of artificial neural networks (ANN) in the management and analysis of information. The data were analyzed together with different neural networks used, the best approximation obtained was determined. It is possible to use artificial neural networks for the problem, predicting the monthly precipitation volumes based on the historical database. Even with reasonable results, the amount of data was not sufficient for the effective processing desired.

**PALAVRAS-CHAVE** – Machine Learning, Hidrologia estatística, Recursos Hídricos.

---

1) Graduando em Engenharia Civil, NT/CAA/UFPE: Av. Marielle Franco, s/n - Km 59 - Nova Caruaru, Caruaru - PE, CEP: 55014-900, [jean.firmino@ufpe.br](mailto:jean.firmino@ufpe.br)

2) Doutorando em Tecnologias Energéticas e Nucleares, DEN/CT/UFPE, Av. Prof. Luiz Freire, 1000, Cidade Universitária, Recife, CEP – 50740-540, [lucas.ravellys@ufpe.br](mailto:lucas.ravellys@ufpe.br)

3) Doutor em Tecnologias Energéticas e Nucleares, DEN/CT/UFPE, Av. Prof. Luiz Freire, 1000, Cidade Universitária, Recife, CEP – 50740540, [martinsdsn@gmail.com](mailto:martinsdsn@gmail.com)

4) Professor Doutor em Engenharia Civil, NT/CAA/UFPE: Av. Marielle Franco, s/n - Km 59 - Nova Caruaru, Caruaru - PE, CEP: 55014-900, [Arthur.coutinho@ufpe.br](mailto:Arthur.coutinho@ufpe.br)

5) Professor Doutor em Tecnologias Energéticas e Nucleares, DEN/CT/UFPE, Av. Prof. Luiz Freire, 1000, Cidade Universitária, Recife, CEP – 50740-540, [antonio.antonino@ufpe.br](mailto:antonio.antonino@ufpe.br)

## INTRODUÇÃO

Existe duas principais abordagens capazes de obter a previsão dos volumes de precipitações pluviométricas: a modelagem determinística dos fenômenos hidrometeorológicos e a análise empírica como os dados das séries históricas (dos Santos Sousa et al., 2017). Entretanto, a primeira abordagem é pouco viável, visto que, a precipitação é o resultado de procedimentos com grande complexidade. Por outro lado, pela segunda alternativa, a previsão é trabalhada a partir do reconhecimento de padrões no qual as variáveis históricas são usadas para buscar padrões periódicos anuais dos quais podem ser relevantes para a previsão dos próximos eventos de precipitação (dos Santos Sousa et al. 2017).

Dentre dos métodos empíricos, as técnicas de Machine Learning (ML) vem ganhando grande destaque na predição de series temporal. As MLs são uma forma de algoritmos de inteligência artificial, que diverge com a programação tradicional. Nos algoritmos tradicionais o programador fornece à lógica e a entrada almejando a saída dos dados. Já na ML o programador fornece a entrada e a saída, criando meios com que o programa retorne a lógica contida para obter a relação entre eles (Valente, 2019).

Dentro das abordagens de Machine Learning, as Redes Neurais Artificiais (RNAs) consistem em um modelo matemático complexo, no qual trabalha com a técnica semelhante à do comportamento do cérebro humano. São consideradas algoritmos poderosos, bastante utilizadas para resolver problemas lineares ou não-lineares (Carvalho, 2014). De acordo com Zaher et al. (2016), a RNA se baseia em uma estrutura com um único elemento chamado nó, onde a cada processamento de informação terá um novo neurônio associado, no qual respeitará o seu funcionamento com base em uma função de ativação em cada saída dos nós. De forma geral pode-se dizer que modelo do RNA consiste três níveis: uma entrada, no qual possui um ou mais números de nós de entrada que varia conforme a quantidade de parâmetros para o modelo; uma camada para a função de ativação, da qual pode ser linear ou não linear; e uma camada de saída, que levará a informação para o próximo nó ou ao output do modelo.

Desse mesmo modo, a técnica de Deep Learning (aprendizado profundo de máquina) é um dos ramos do Machine Learning baseado em um conjunto de algoritmos que detenham da capacidade de modelar abstrações de alto nível de dados usando um grafo profundo com várias

camadas de processamento. Nesse sentido, este trabalho teve como objetivo construir um modelo de Deep Learning para a predição da precipitação mensal utilizando apenas a série histórica de dados precipitação. Para tanto, foi utilizado como estudo de caso o município de Caruaru-PE.

## MATERIAIS E MÉTODOS

### Área de estudo

O município de Caruaru (Figura 1), localizado no agreste Pernambucano aos 8° 16' 58'' S de latitude e 35° 58' 33'' W de longitude, tem o clima o semiárido e com grandes variações de temperatura, entre de 20°C no inverno e 38°C no verão. Na Figura 1 são apresentados os postos pluviométricos presentes no município de Caruaru. Foram utilizadas séries históricas das precipitações obtidas pela Agência Pernambucana de Águas e Climas (APAC, 2019). Neste trabalho, por apresentar a maior e mais continua série de dados, foram utilizados os dados presentes no posto pluviométrico do IPA (Instituto Agrônômico de Pesquisa).

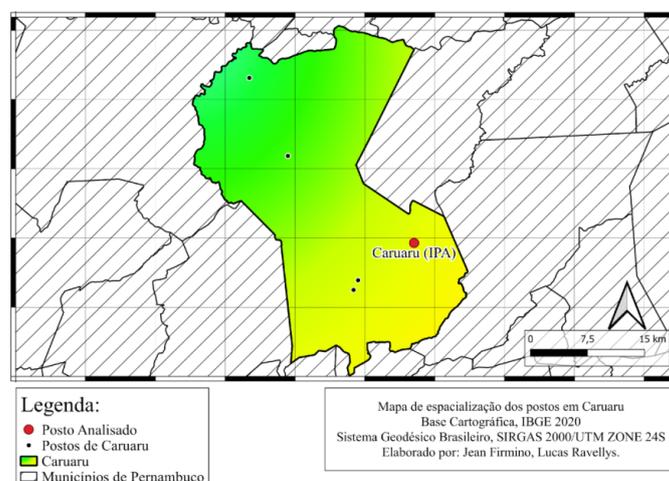


Figura 1 - Localização dos postos de dados pluviométricos no município de Caruaru-PE.

### Tratamento dos dados

Foram utilizados dados mensais de precipitação pluviométrica entre os anos de 1960 até 2018, resultando em um total de 708 dados. Os dados de precipitação pluviométrica acumulada mensal no período de 1960 até 2018 são apresentados na Figura 4. Em vermelho é demonstrada a precipitação média, da qual teve uma magnitude de cerca de 50 mm/mês.

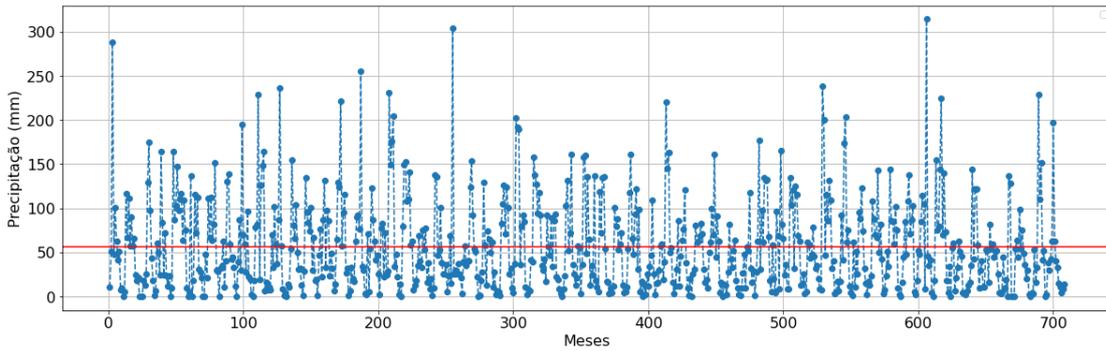


Figura 2. Dados de Precipitação acumulada mensal no período de 1960 até 2018.

Para o ajuste do modelo, foi utilizada a normalização do logaritmo dos dados, utilizando a Equação (1):

$$P'_i = \frac{\ln(P_i + 1) - \overline{\ln P}}{S_{\ln}} \quad (1)$$

Sendo:  $P'_i$  o dado normalizado no tempo  $i$ ;  $P_i$  a precipitação em mm no tempo  $i$ ;  $\overline{\ln P}$  a média do  $\ln(P+1)$  de treino; e  $S_{\ln}$  o desvio padrão do  $\ln(P)$  de treino. Na Figura 3 é apresentado o histograma da precipitação pluviométrica acumulada mensal com os dados brutos e normalizados.

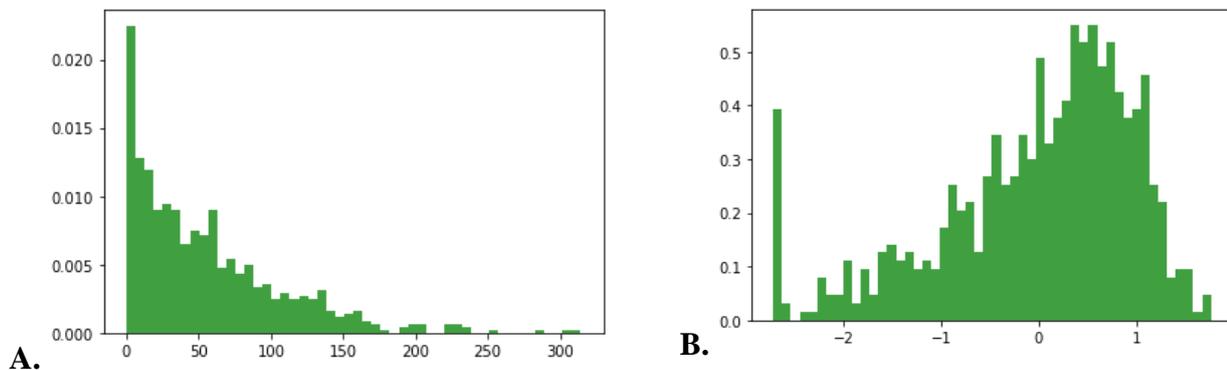


Figura 3. Histograma da precipitação pluviométrica acumulada mensal com os dados brutos (A) e normalizados (B)

### Construção do modelo de *Deep Learning*

Para a arquitetura do modelo de Machine learning, foi utilizado modelo de Redes Neurais Artificiais (RNA). Para isso, foi utilizado a biblioteca Keras do python em conjunto com o Backend do TensorFlow 2.0. Como entrada, foi selecionada uma janela com 12 meses de dados, ou seja, para a predição de um mês futuro é necessário o input de um ano de dado anterior a este mês. Para a

otimização foram ainda utilizados tamanhos de batch de 20 e um critério de parada de 1000 iterações.

Como função objetivo foi utilizada a função de Huber (1964), na qual, por definição é uma função de perda, geralmente usada quando a análise estatística trata de valores discrepantes entre si. A função descreve a perda incorrida por um procedimento estimativo. A função de perdas de Huber é dada por:

$$L(a) = \begin{cases} \frac{1}{2} a^2 & \text{para } |a| \leq \delta \\ \delta \left( |a| - \frac{1}{2} \delta \right) & \text{para } |a| > \delta \end{cases} \quad (2)$$

Onde ‘a’ refere-se ao resíduo, ou seja, à diferença entre os valores observados e os previstos sendo igual a  $a = y - f(x)$ .

Para a otimização da função de Huber foi utilizada o método do Gradiente Descendente Estocástico com taxa de aprendizado de  $10^{-5}$  e momento de 0,3. Este é um método de otimização utilizado em aprendizado de máquina, no qual objetivo é minimizar uma dada perda  $L(a)$  (Pinto, 2018), encontrando para isso um valor  $\theta^* \in \mathfrak{R}^k$  ( $k \geq 1$ ), começando com um valor inicial  $\theta_0$ .

$$\theta_{n+1} = \theta_n - \varepsilon_n L'(\theta_n) \quad (3)$$

Onde  $n$  é o número do passo da iteração e  $\varepsilon_n$  é a sequência decrescente do tamanho do passo, assume-se que  $\varepsilon_n$  é não negativo e finito (Gemulla et al., 2011).

Uma arquitetura de Deep Learning Recursiva foi utilizada para ajustar o modelo. Para o treinamento e validação do modelo foram utilizados 70% (496 meses) e 30% (212 meses), respectivamente. As métricas utilizadas para avaliar a qualidade do ajuste foram o Coeficiente de eficiência do modelo de Nash (NSE), Equação ( 4 ), e o erro absoluto médio (MAE), Equação ( 5 ).

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |(O_i - P_i)| \quad (5)$$

No qual:  $O_i$  é o valor da precipitação observada em  $i$ ,  $P_i$  é o valor da precipitação predita em  $i$ ,  $n$  é o número de registros, e  $\bar{O}$  é a precipitação média da série analisada.

## RESULTADOS E DISCUSSÃO

Na Figura 4 são apresentados os Box-Plots das precipitações pluviométricas mensais durante o período de 1960 até 2018. Nota-se que as maiores precipitações se concentram entre os meses de março a julho, enquanto que as menores precipitações estiveram concentradas de setembro a dezembro. Observa-se também que a maior precipitação mediana ocorreu no mês de julho. O mesmo foi constatado na Bacia do Tapacurá por Silva et al. (2010), entretanto, em outras regiões do Agreste a maior precipitação mensal pode ocorrer no mês de junho (Rodrigues et al., 2017).

Nota-se ainda que os limites superiores dos históricos de chuvas possuem uma maior amplitude entre os meses de fevereiro a abril, gradativamente vão diminuindo até o mês de novembro onde encontra o menor limite superior. Resultados semelhantes foram observados Santos e Montenegro (2012). Contrariamente a isso, os limites inferiores foram mais uniformes ao longo dos meses. Uma das variáveis hidrológicas que podem influenciar para a ocorrência das outliers ao longo da série histórica vem da anomalia ocorrida na temperatura do atlântico e pacífico causada pelo fenômeno global chamado La Niña. Segundo Rodrigues et al. (2007), entre os anos de 1964 e 2016, 56% dos registros chuvas intensas foram provocadas por este evento. Essas informações mostram a heterogeneidade temporal na distribuição da precipitação no município de Caruaru, como constatado por Santos et al. (2010) para a região do agreste Pernambucano.

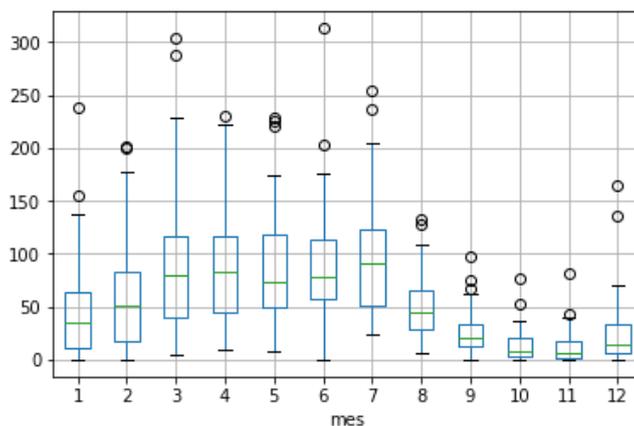
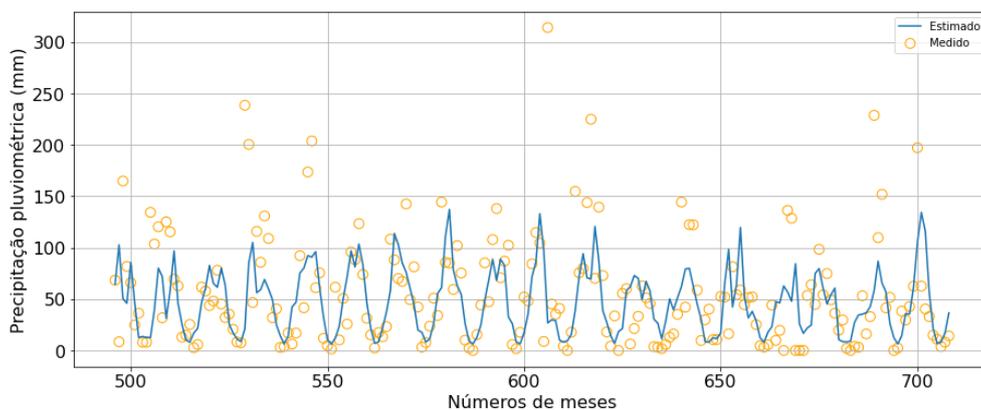


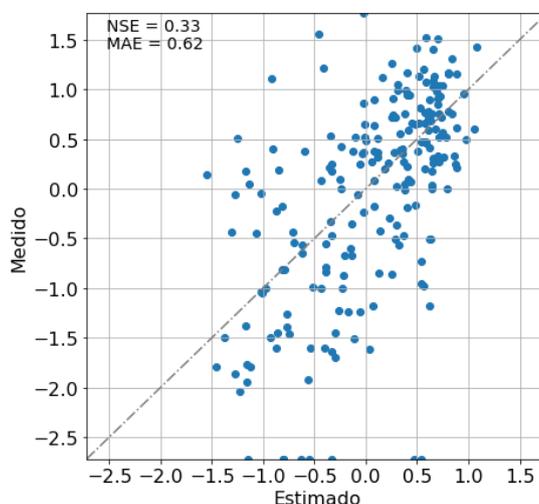
Figura 4. Boxplot dos meses no intervalo entre 1960 até 2018

Na Figura 5A é apresentada precipitação pluviométrica mensal e sua predição pelo modelo no período de validação. Nota-se que o modelo foi capaz de prever as tendências dos dados medidos, apresentando uma boa predição na ausência de chuva, contudo, subestimando os picos de chuvas. Gorodetskaya et al. (2018) utilizaram os métodos de Redes Neurais Artificiais e Random Forest e também obtiveram dificuldade em prever a vazão média diária de uma estação fluviométrica do rio Paraíba do Sul com um horizonte de previsão de até 7 dias. Assim como no presente trabalho, eles ainda observaram que durante períodos com vazões médias e baixas, a previsão foi satisfatória. Além disso, os autores ainda observaram que os modelos de Redes Neurais mostraram melhor desempenho na previsão dos picos de vazão em comparação aos modelos Random Forest.

Em relação às métricas de quantificação do ajuste adotadas (Figura 5B) é possível identificar várias discrepâncias entre os dados inicialmente processados. Nota-se também tendência positiva entre os dados medidos e estimados no gráfico de comparação, entretanto, também é observado uma grande dispersão dos valores. O coeficiente NSE foi de 0,33 e o seu MAE foi de 0,62, mesmo assim, ainda não foi possível prever os dados com uma eficiência satisfatória para o objetivo, ou seja, foi almejado o NSE para mais próximo de 1. Também é possível supor um dos motivos para o MAE calculado, provavelmente foi influenciado pelas precipitações diárias anômalas a média anual, por exemplo, precipitações acima de 50mm (Figura 2) podem alterar o MAE para distante do MAE ideal(zero).



A.



B.

Figura 5. (A) Predição da Precipitação Mensal pelo modelo no período de Validação, (B) Scatter plot dos dados normalizados medidos e estimados pelo modelo no período de validação.

## CONCLUSÃO

Com o desenvolvimento do trabalho foi possível utilizar as Redes Neurais Artificiais para a predição de precipitações mensais no município de Caruaru, a partir o banco de dados da série histórica pluviométrica. Mesmo obtendo resultados interessantes, a quantidade de dados não foi suficiente para o processamento eficaz almejado, com base nos resultados das métricas de quantificação, ou seja, os coeficientes MAE e NSE.

Mesmo assim, ainda foi possível prever as tendências de ausência de chuvas, de acordo com os dados, foi possível identificar padrões cíclicos para os meses com baixo volume de chuva, especialmente entre o mês de novembro. Também foi possível identificar os meses entre março e

julho com as maiores probabilidades de ocorrer eventos de grandes volumes de precipitações pluviométricas ao ano, coincidindo com o período das festas anuais tradicionalmente populares, com comemorações pela colheita feita entre o final do período chuvoso identificado.

Nesse sentido, uma análise multivariada da precipitação, levando em consideração a temperatura do oceano atlântico e o efeito do el niño/la niña podem melhorar significativamente a predição do modelo.

## **AGRADECIMENTOS**

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo N° 465764/20142, do processo N° 117603/20201, da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES), processo N° 88887.136369/2017-00 e do projeto “Observatório Nacional da Dinâmica da Água e de Carbono no Bioma Caatinga (ONDACBC)”.

## **REFERÊNCIAS**

AGÊNCIA Pernambucana de Águas e Clima - APAC. [S. l.], 31 dez. 2018. Disponível em: <http://www.apac.pe.gov.br/>. Acesso em: 10 ago. 2019.

CARVALHO, H. M. (2014). Aprendizado de máquina voltado para mineração de dados: árvores de decisão.

DE OLIVEIRA RODRIGUES, L., DE SOUZA, W. M., DE OLIVEIRA COSTA, V. S., & PEREIRA, M. L. T. Influência dos eventos de El Niño e La Niña no regime de precipitação do Agreste de Pernambuco (Influence of El Niño and La Niña events on rainfall of Agreste Pernambuco). *Revista Brasileira de Geografia Física*, 10(6), 1995-2009.

DOS SANTOS SOUSA, R., SILVA, L., GUEDES, E. B., & DE OLIVEIRA, M. B. L. (2017). Redes neurais artificiais aplicadas à previsão antecipada de precipitações na região central de Manaus. *IV Escola Regional de Informática Norte I*, 1-9.

DOS SANTOS, T. E., & MONTENEGRO, A. A. (2012). Erosividade e padrões hidrológicos de precipitação no Agreste Central pernambucano. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 16(8), 871-880.

GEMULLA, R.; NIJKAMP, E.; HAAS, P.J.; SISMANIS, Y. Large-scale matrix factorization with distributed stochastic gradient descent. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 69–77. ACM, 2011.

GORODETSKAYA, Yulia et al. Daily streamflow forecasting for Paraíba do Sul river using machine learning methods with hydrologic inputs. In: **Anais do XV Encontro Nacional de Inteligência Artificial e Computacional**. SBC, 2018. p. 162-173.

KÖPPEN, W.; GEIGER, R. “Klimate der Erde. Gotha: Verlag Justus Perthes”. Wallmap 150cmx200cm. 1928.

MORIASI, D. N., ARNOLD, J. G., VAN LIEW, M. W., BINGNER, R. L., HARMEL, R. D., and VEITH, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, T. ASABE, 50, 885–900, 2007.

PEREIRA, D. R., ULIANA, E. M., MARTINEZ, M. A., & SILVA, D. D. (2016). Desempenho de um modelo hidrológico concentrado e de um semidistribuído na predição de vazões diárias. *Irriga*, 21(2), 409-409.

PINTO, J. C. (2018). Geração de representação distribuída de palavras fatorando a matriz de informação mútua pontual utilizando o gradiente descendente estocástico no modelo de MapReduce.

RODRIGUES, L. O.; SOUZA, W. M.; COSTA, V. S. O.; PEREIRA, M. L. T. Influência dos eventos de El Niño e La Niña no regime de precipitação do Agreste de Pernambuco. **Revista Brasileira de Geografia Física**, v. 10, n. 6, p. 1995-2009, 2017.

SANTOS, T. E. M.; SILVA, D. D.; MONTENEGRO, A. A. A. Temporal variability of soil water content under different surface conditions in the semiarid region of the Pernambuco State. *Revista Brasileira de Ciência do Solo*, v.34, p.1733-1741, 2010.

SCHNITMAN, L. (1998). Controladores Preditivos Baseados em Redes Neurais Artificiais (Doctoral dissertation, Dissertação de Mestrado do curso de Pós Graduação em Eng. Elétrica da-UFBa).

SHORTRIDGE, J. E., GUIKEMA, S. D., & ZAITCHIK, B. F. (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology & Earth System Sciences*, 20(7).

SILVA, R. M.; PEREIRA E SILVA, L.; MONTENEGRO, S. M. G. L.; SANTOS, C. A. G. Análise da variabilidade espaço-temporal e identificação do padrão da Precipitação na Bacia Do Rio Tapacurá, Pernambuco. *Sociedade & Natureza*, v.22, n.2, p. 357-372, ago. 2010.

VALENTE, P. O. L. (2019). PROGRAMACAO FUNCIONAL APLICADA AO GERENCIAMENTO DE SISTEMAS DE APRENDIZADO DE MAQUINA (Doctoral dissertation, Universidade Federal do Rio de Janeiro).

YASEEN, Z. M., ALLAWI, M. F., YOUSIF, A. A., JAAFAR, O., HAMZAH, F. M., & EL-SHAFIE, A. (2018). Non-tuned machine learning approach for hydrological time series forecasting. *Neural Computing and Applications*, 30(5), 1479-1491.