



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE CRATEÚS
CURSO DE ENGENHARIA CIVIL

RODRIGO MACHADO ALVES

**PREVISÃO CLIMÁTICA SAZONAL PARA A REGIÃO HIDROGRÁFICA DOS
SERTÕES DE CRATEÚS UTILIZANDO REDES NEURAIIS ARTIFICIAIS
RECORRENTES**

CRATEÚS
2022

RODRIGO MACHADO ALVES

PREVISÃO CLIMÁTICA SAZONAL PARA A REGIÃO HIDROGRÁFICA DOS SERTÕES
DE CRATEÚS UTILIZANDO REDES NEURAIAS ARTIFICIAIS RECORRENTES

Trabalho de Conclusão de Curso apresentado ao
Curso de Engenharia Civil da Universidade
Federal do Ceará, como requisito parcial à
obtenção do título de bacharel em Engenharia
Civil.

Orientador: Prof. Dr. Alan Michell Barros
Alexandre.

CIDADE

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A482p Alves, Rodrigo.
Previsão climática sazonal para a Região Hidrográfica dos Sertões de Crateús utilizando redes neurais artificiais recorrentes / Rodrigo Alves. – 2022.
44 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Crateús, Curso de Engenharia Civil, Crateús, 2022.
Orientação: Prof. Dr. Alan Michell Barros Alexandre.

1. Previsão pluviométrica. 2. Índices TSM. 3. Redes neurais artificiais. 4. Long Short Term Memory. I.
Título.

CDD 620

RODRIGO MACHADO ALVES

PREVISÃO CLIMÁTICA SAZONAL PARA A REGIÃO HIDROGRÁFICA DOS SERTÕES
DE CRATEÚS UTILIZANDO REDES NEURAIIS ARTIFICIAIS RECORRENTES

Trabalho de Conclusão de Curso apresentado ao
Curso de Engenharia Civil da Universidade
Federal do Ceará, como requisito parcial à
obtenção do título de bacharel em Engenharia
Civil.

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

Prof. Dr. Alan Michell Barros Alexandre (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dra. Janine Brandão de Farias Mesquita
Universidade Federal do Ceará (UFC)

Prof. Ms. Taís Maria Nunes Carvalho
Universidade Federal do Ceará (UFC)

À Deus.

Aos meus pais, Neide Maria e Francisco
Moacir.

À minha avó, Raimunda Machado.

Ao meu irmão, Davi Barbosa.

AGRADECIMENTOS

À Deus.

Agradeço aos meus pais, Neide Maria e Francisco Moacir, pelo apoio, amor e segurança que me forneceram durante toda minha trajetória educacional.

Ao meu irmão, Davi Barbosa, pelos momentos de descontração e brincadeiras.

À minha avó, Raimunda, por ter sido presente em todos os momentos na minha vida e me apoiar em todas as decisões que tomei. Além de mostrar a cada dia sobre como é importante valorizar cada minuto que temos.

Ao meu Orientador, Alan Michell, não só pela orientação, mas pelos momentos de conversas e ensinamentos que levarei para a vida.

A todos meus amigos de faculdade, pelos momentos de diversão e companheirismo nos estudos.

“Nenhuma grande descoberta foi feita jamais
sem um palpite ousado.”

(Isaac Newton)

RESUMO

Um dos grandes problemas que o Sertão do Nordeste Brasileiro enfrenta é a escassez hídrica. Isso se dá pelo baixo índice pluviométrico que essa região possui e pela concentração pluviométrica em poucos meses no ano. Desta forma, faz-se necessário um planejamento à curto prazo do uso dos recursos hídricos, que requer informações climáticas precisas para a tomada de decisões. Assim, este trabalho tem como objetivo gerar modelos de previsão de precipitação para a Região Hidrográfica dos Sertões de Crateús, utilizando uma Rede Neural Artificial Recorrente chamada *Long Short Term Memory*. Para isso, foi analisado o quadrimestre chuvoso (FMAM), utilizando como preditor quatro trimestres, SON, DJF, MAM e JJA. Cada preditor possui 11 índices TSM ranqueado por ordem de relevância e foi gerado 11 modelos de previsão para cada preditor através da linguagem Python. O preditor que apresentou maior acurácia, ou seja, com maiores valores de correlação e NASH foi DJF, com uma correlação de 0,808 e NASH de 0,530, ambos na fase de validação. Não obstante isso, em comparação com trabalhos semelhantes também usando redes neurais, foi observado que os valores de NASH e correlações na fase de validação deste trabalho foram melhores que os presentes na literatura, isto é, a rede desenvolvida neste trabalho possui um poder maior de generalização.

Palavras-chave: Previsão Pluviométrica. Índices TSM. Redes Neurais Artificiais. *Long Short Term Memory*

ABSTRACT

One of the great problems that the Sertão of the Brazilian Northeast faces is water scarcity. This is due to the low rainfall index that this region has and the concentration of rainfall in a few months of the year. Thus, a short-term planning of the use of water resources is necessary, highlighting the importance of climate accuracy for decision making. Thus, this work aims to generate precipitation forecast models for the Hydrographic Region of the Sertões de Crateús, using a Recurrent Artificial Neural Network called Long Short Term Memory. For this, the rainy quarter (FMAM) was analyzed, using as predictor four quarters, SON, DJF, MAM and JJA. Each predictor has 11 TSM indices ranked in order of relevance and 11 forecast models were generated for each predictor using the Python language. The predictor that presented the highest accuracy, i.e., with higher correlation and NASH values was DJF, reaching 0.808 in correlation and 0.530 in NASH, both in the validation phase. Nevertheless, in comparison with similar work also using neural networks, it was observed that the NASH and correlation values in the validation phase of this work were better than the comparison phase, in other words, the network developed in this work has a greater power of generalization.

Keywords: Rainfall Forecasting. TSM Indexes. Artificial Neural Networks. Long Short Term Memory

LISTA DE FIGURAS

Figura 1 - Divisão do Nordeste Brasileiro em três regiões.	18
Figura 2 – Regiões correspondentes a cada índice El Niño.	21
Figura 3 - Neurônio artificial.....	23
Figura 4 – Rede Neural Recorrente.	24
Figura 5 - Camadas de uma Rede Neural Recorrente.	25
Figura 6 - Célula LSTM.	26
Figura 7 - Funções de ativação	27
Figura 8 - Regiões Hidrográficas do Ceará	28
Figura 9 - Estratégia metodológica utilizada para determinação da precipitação na RHSC....	31
Figura 10 – Correlação e NASH da precipitação observada com a obtida com o preditor SON	35
Figura 11 - Correlação e NASH da precipitação observada com a obtida com o preditor DJF	35
Figura 12 - Correlação e NASH da precipitação observada com a obtida com o preditor MAM	36
Figura 13 - Correlação e NASH da precipitação observada com a obtida com o preditor JJA	36
Figura 14 - Precipitação acumulada para a validação do preditor DJF Modelo 8.....	39

LISTA DE TABELAS

Tabela 1 - Reservatórios da RH dos Sertões de Crateús	29
Tabela 2 - Índices TSM utilizados.....	29
Tabela 3 - Estações Pluviométricas utilizadas para geração do Thiessen	30
Tabela 4 - Ranking de variáveis por Evangelista	32
Tabela 5 - Hiperparâmetros utilizados por Vasco (2020).....	33
Tabela 6 - Três maiores correlações da precipitação de FMAM.....	37
Tabela 7 - Três maiores valores de NASH da precipitação de FMAM.....	37
Tabela 8 – Comparação do melhor resultado de correlação para cada preditor.....	38
Tabela 9 - Comparação do melhor resultado de NASH para cada preditor	38

LISTA DE ABREVIATURAS E SIGLAS

AD	Árvore de Decisão
AMO	Oscilação Multidecadal do Atlântico (<i>Atlantic Multidecadal Oscillation</i>)
DJF	Período de dezembro a fevereiro
ENOS	El Niño – Oscilação Sul
FMAM	Período de fevereiro a maio
FUNCEME	Fundação Cearense de Meteorologia e Recursos Hídricos
IBGE	Instituto Brasileiro de Geografia e Estatística
LSTM	<i>Long Short Term Memory</i>
MAM	Período de março a maio
NAO	Oscilação do Atlântico Norte (<i>North Atlantic Oscillation</i>)
NEB	Nordeste Brasileiro
NINO 1.2	Niño 1+2
NINO 3	Niño 3
NINO 3.4	Niño 3.4
NINO 4	Niño 4
NOAA	<i>National Oceanic and Atmospheric Administration</i>
ONI	Niño Oceânico (<i>Oceanic Niño Index</i>)
PDO	Oscilação Decadal do Pacífico (<i>Pacific Decadal Oscillation</i>)
RH	Região Hidrográfica
RHSC	Região Hidrográfica dos Sertões de Crateús
RN	Rede Neural
RNR	Rede Neural Recorrente
TAD	Dipolo do Atlântico Tropical (<i>Tropical Atlantic Dipole</i>)
TNA	Atlântico Norte Tropical (<i>Tropical Northern Atlantic</i>)
TSA	Atlântico Sul Tropical (<i>Tropical Southern Atlantic</i>)
TSM	Temperatura de Superfície do Mar
VCAN	Vórtices Ciclônicos de Alto Nível
ZCIT	Zona de Convergência Intertropical

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Contextualização	15
1.2 Justificativa	16
1.3 Área de estudo	17
1.4 Objetivos	17
<i>1.4.1 Objetivo geral</i>	<i>17</i>
<i>1.4.2 Objetivos específicos</i>	<i>17</i>
2 REVISÃO BIBLIOGRÁFICA	18
2.1 Caracterização climática	18
2.2 Temperatura de Superfície do Mar	19
<i>2.2.1 Dipolo do Atlântico</i>	<i>19</i>
<i>2.2.2 ENOS - El Niño e Oscilação Sul</i>	<i>20</i>
<i>2.2.3 Oscilação Decadal do Pacífico</i>	<i>21</i>
<i>2.2.4 Oscilação Multidecadal do Atlântico</i>	<i>21</i>
<i>2.2.5 Oscilação do Atlântico Norte</i>	<i>22</i>
2.3 Random Forest	22
2.4 Redes Neurais Artificiais	22
<i>2.4.1 Introdução às Redes Neurais Artificiais</i>	<i>23</i>
<i>2.4.2 Rede Neural Recorrente (RNR)</i>	<i>24</i>
<i>2.4.3 Long Short Term Memory (LSTM)</i>	<i>25</i>
<i>2.4.4 Ajuste dos modelos e dos hiperparâmetros</i>	<i>26</i>
3 METODOLOGIA	28
3.1 Região de estudo	28
3.2 Levantamento dos dados	29
3.3 Estratégia metodológica	30
3.4 Classificação por ordem de importância dos índices climáticos	31
3.5 Previsão climática utilizando LSTM	32
<i>3.5.1 Tratamento dos dados</i>	<i>32</i>
<i>3.5.2 Construção dos modelos e definição dos hiperparâmetros</i>	<i>32</i>
3.6 Avaliação dos modelos	33
4 RESULTADOS E DISCUSSÃO	35

4.1 Modelos de previsão	35
4.2 Comparação com a literatura	38
5 CONCLUSÃO	40
REFERÊNCIAS	41

1 INTRODUÇÃO

1.1 Contextualização

O Nordeste brasileiro (NEB) é uma região caracterizada por uma climatologia complexa, alta variabilidade interanual e clima predominantemente semiárido. Seu regime de chuvas é concentrado em quatro meses durante o ano e é influenciado por vários movimentos atmosféricos existentes que favorecem ou inibem os processos de formação de nuvens precipitantes (NOBRE, 2012).

Entretanto, ainda que diversos sistemas meteorológicos possam influenciar a precipitação do NEB, existem quatro mecanismos principais: (1) El Niño – Oscilação Sul (ENOS); (2) Temperatura da Superfície do Mar (TSM) do Oceano Atlântico, Ventos Alísios e Pressão ao Nível do Mar (PNM); (3) a Zona de Convergência Intertropical (ZCIT) sobre o Oceano Atlântico; (4) Frentes Frias e Vórtices Ciclônicos de Altos Níveis (VCAN) (UVO, 1989; ALVES; REPELLI, 1992; UVO *et al*, 1995). Araújo Júnior *et al* (2019) adiciona a esta lista mais três sistemas: Linhas de Instabilidade (LI), Sistemas Frontais (SF) e Oscilação Decadal do Pacífico (ODP).

Nobre (2012) reforça que essa gama de mecanismos influenciadores, associados a elevadas taxas de evapotranspiração e solos rasos com pouca capacidade de armazenagem, ao provocarem uma grande variabilidade interanual dos totais pluviométricos, são os responsáveis pela alternância de anos de seca e de cheias na região do NEB e dessa forma, provocam grandes problemas sociais, ambientais e econômicos.

O Ceará, estado situado na porção norte do NEB, é fortemente influenciada pelo El Niño Oscilação Sul, Dipolo do Atlântico e a Oscilação Decadal do Pacífico (NÓBREGA, 2002; SILVA *et al.*, 2011). Seu período chuvoso é dividido em chuvas de pré-estação (dezembro e janeiro), estação chuvosa (fevereiro a maio) e pós-estação (maio e junho) (RIBEIRO FILHO, 2009).

Através de relações entre esse quadro variável da pluviometria do NEB e TSM no pacífico equatorial observadas por Walker (1928), várias pesquisas foram desenvolvidas para entender tal relação de forma mais aprofundada (SILVA *et al.*, 2011).

Como resultado disso, uma interpretação amplamente aceita é que a variabilidade de TSM do Atlântico é a principal causadora das anomalias de precipitação no NEB, enquanto a do Pacífico pode tanto reforçar tais anomalias, como enfraquecê-las (SOUZA *et al.*, 2004;

ANDREOLI; KAYANO, 2007). Entende-se, dessa forma, a possibilidade da realização de previsões pluviométricas a partir de índices TSM.

Para isso, Araújo et al (2019) afirma que é necessária uma ferramenta capaz de modelar fenômenos complexos, não lineares e multivariados. A vista disso, métodos baseados em Redes Neurais Artificiais (RNA) é uma solução viável.

Para Castro *et al.* (2020), RNA são modelos computacionais baseados no sistema nervoso que são capazes de aprender comportamentos e extrair relações de diversas variáveis presentes através de seu paralelismo de atividades, alta conectividade e capacidade elevada em descrever sistemas não-lineares.

1.2 Justificativa

“As secas são fenômenos recorrentes no Nordeste, afetando de forma mais notável a porção semiárida na parte norte do Nordeste” (NOBRE, 2012). Dessa forma, o Ceará, por estar localizado nessa porção semiárida e apresentar altas taxas de evapotranspiração e solos rasos com pouca capacidade de armazenagem, é muito propenso a secas severas.

Conforme FUNCEME (2017), entre 2012 e 2017, o Ceará foi atingido por uma estiagem que provocou uma redução drástica dos principais reservatórios da região, sendo necessário ações como: tarifas de contingências, perfuração de poços e transposições de águas entre reservatórios.

Como exemplo dessas transposições, COGERH (2013) relata que em 2013, com o objetivo de abastecimento do município de Crateús durante tal crise hídrica, foi construída uma adutora emergencial para transferir água do açude Flor do Campo, localizado no município de Novo Oriente, para o açude Carnaubal, em Crateús. Cidade esta, definida como a mais populosa da região hidrográfica dos Sertões de Crateús (IPECE, 2020).

Salienta-se ainda as projeções negativas publicadas no quinto relatório do Painel Intergovernamental de Mudanças Climáticas (IPCC) em 2014. As previsões para o NEB para o século XXI são de aumento de temperaturas e redução de chuvas (IPCC, 2014). Nota-se, dessa forma, a importância da gestão e alocação racional das águas.

Baseado nisso, o presente trabalho tem o potencial de minimizar impactos negativos de eventos climáticos, através do fornecimento de previsões de precipitação mais precisas do que as fornecidas pelos modelos utilizados pelos tomadores de decisão.

1.3 Área de estudo

A área de aplicação deste estudo é a região hidrográfica dos Sertões de Crateús. Local este, situado na macrorregião dos Sertões de Crateús, a oeste do Ceará, com uma área 10821 Km², compreende a rede de drenagem do Rio Poti, é composta por nove municípios e dispõe de dez açudes monitorados pela COGERH (SRH, 2013).

1.4 Objetivos

1.4.1 *Objetivo geral*

Avaliar a previsibilidade da precipitação sazonal através de índices de TSM para a Região Hidrográfica dos Sertões de Crateús utilizando a rede neural *Long Short Term Memory*.

1.4.2 *Objetivos específicos*

Como objetivos específicos, tem-se:

- a) Gerar modelos de previsão de precipitação a partir da associação de índices TSM;
- b) Verificar a acurácia dos diferentes modelos de precipitação;
- c) Definir o modelo preditor de melhor desempenho para o período preditivo adotado;
- d) Comparar os resultados obtidos com a literatura.

2 REVISÃO BIBLIOGRÁFICA

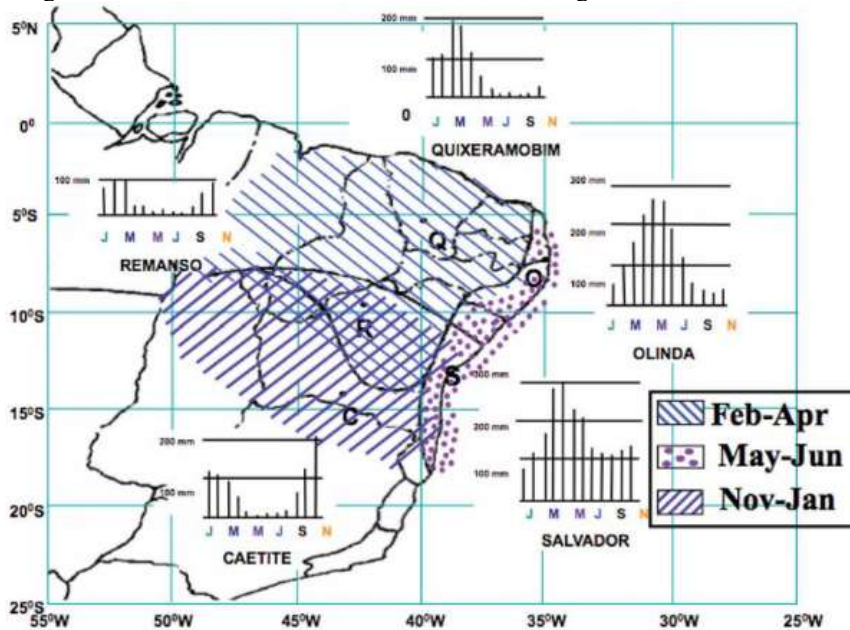
Este capítulo possui como objetivo apresentar uma revisão da literatura acerca da caracterização climática e principais índices TSM atuantes na região de estudo, assim como dos métodos necessários para a realização do trabalho, como o *Random Forest* e por fim uma discussão sobre Redes Neurais Artificiais.

2.1 Caracterização climática

Segundo Uvo e Berndtsson (1996), o NEB de forma geral é influenciado pelos seguintes sistemas atmosféricos: El Niño-Oscilação Sul (ENOS), Temperatura da Superfície do Mar (TSM), Ventos Alísios, Pressão ao Nível do Mar, Zona de Convergência Intertropical (ZCIT), Frentes Frias, Vórtices Ciclônicos, Linhas de Instabilidade e Complexos Convectivos de Mesoescala.

Porém, esses mecanismos não possuem a mesma intensidade de influência para todo o território Nordeste. Nobre (2012) afirma que o regime de chuvas sobre a região Nordeste é dividido em três tipos: porção norte, sul e leste do Nordeste (Figura 1).

Figura 1 - Divisão do Nordeste Brasileiro em três regiões.



Fonte: Kousky et al. (1980).

A porção leste do Nordeste possui um período chuvoso nos meses de abril a julho e seus principais mecanismos causadores de chuvas são a convergência dos ventos alísios de

sudeste, formando uma banda sul da ZCIT e a atuação de distúrbios atmosféricos que se propagam da África para oeste (NOBRE, 2012).

Já na parte sul do Nordeste, o quadrimestre chuvoso está compreendido entre novembro e fevereiro e os principais mecanismos atmosféricos são as frentes frias provenientes de altas latitudes do Hemisfério Sul e a Zona de Convergência do Atlântico Sul (ZCAS) (NOBRE, 2012).

Por fim, Nobre (2012) aborda sobre a porção norte do Nordeste (região foco deste trabalho) que possui um regime pluviométrico anual compreendido no período de janeiro a abril. Seu principal causador de chuvas é a ZCIT, que é fortemente modulada pelos campos de TSM sobre o Atlântico Equatorial.

Conforme Ferreira e Melo (2005), Zona de Convergência Intertropical (ZCIT) é um conjunto de nuvens que cerca a faixa equatorial do globo terrestre e é formada pela confluência dos ventos alísios do hemisfério norte e sul, em baixos níveis, baixas pressões, altas temperaturas da superfície do mar, intensa atividade convectiva e precipitação. Eles ainda complementam que a justificativa dos campos TSM serem um fator determinante para a intensidade e posição da ZCIT está no fato dela ser mais significativa sobre os oceanos.

2.2 Temperatura de Superfície do Mar

Visto que as TSM exercem uma influência significativa na ZCIT, faz-se necessário apresentar quais índices de TSM influenciam a circulação atmosférica na região tropical.

2.2.1 *Dipolo do Atlântico*

O Dipolo do Atlântico é um fenômeno que possui duas fases e é caracterizado pela diferença entre a temperatura superficial das águas do Atlântico Tropical Norte (TNA) e Atlântico Tropical Sul (TSA). Entende-se por fase positiva do dipolo quando há presença de anomalias positivas no Atlântico Norte e negativas no Atlântico Sul e fase negativa do dipolo quando o caso contrário ocorre (MOURA e SHUKLA, 1981; NOBRE e SHUKLA, 1996).

Nobre e Shukla (1996) ainda complementam que esse gradiente meridional de anomalias de TSM influencia a posição latitudinal da ZCIT e conseqüentemente as chuvas no NEB.

Dessa forma, quando o Dipolo do Atlântico está na fase negativa, há um favorecimento de chuvas acima da média sobre o NEB e em sua fase positiva, há um favorecimento de precipitação abaixo da média (ANDREOLI e KAYANO, 2007).

Para descrever esse fenômeno com base em suas duas fases (TNA e TSA) é utilizada a média da anomalia mensal para a área 5°N-23,5°N, 15°W-57,5°W e 0°-20°S, 10°E-30°W, respectivamente (ESRL, 2020).

2.2.2 ENOS - El Niño e Oscilação Sul

Conforme Aragão (1998), El Niño é o aquecimento da água do mar no Pacífico Tropical da costa do Peru/Equador até o oeste do Pacífico, já a Oscilação Sul é uma resposta aérea ao El Niño caracterizada como a variação anômala da pressão atmosférica tropical. O fenômeno resultante da atuação do El Niño e Oscilação do Sul é conhecido como ENOS.

Este é dividido em duas fases: positiva (La Niña) e negativa (El Niño). Enquanto a La Niña representa o resfriamento anômalo das águas do Pacífico Equatorial, o El Niño representa o aquecimento do Pacífico Equatorial, devido a migração de águas quentes a essa região (GRIMM et al., 1998; BERLATO; FONTANA, 2003).

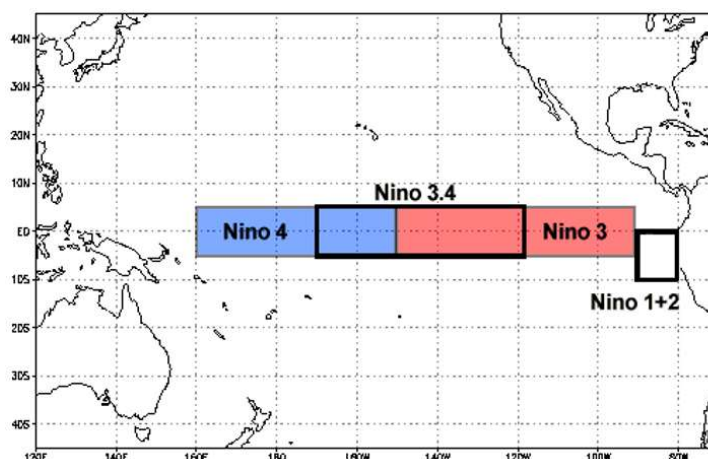
Estudos realizados por Melo e Lima (2021) mostram que o fenômeno ENOS possui influência direta nas secas do semiárido nordestino, uma vez que nos períodos de El Niño, nota-se uma redução na quantidade de chuvas da região.

O monitoramento desse fenômeno é realizado analisando os seguintes índices climáticos:

- a) Niño 1+2: 0°-10°S e 90°W-80°W;
- b) Niño3: 5°N-5°S e 150°W-90°W;
- c) Niño 3.4: 5°N-5°S e 170°W-120°W;
- d) Niño 4: 5°N-5°S e 160°E-150°W.

A Figura 02 mostra a região que corresponde a cada índice El Niño.

Figura 2 – Regiões correspondentes a cada índice El Niño.



Fonte: NOAA (2022).

Outrossim, há o Niño Oceânico (ONI – Oceanic Niño Index), que é obtido usando a média de três meses consecutivos de anomalias da região Niño 3.4 (NOAA, 2022).

2.2.3 Oscilação Decadal do Pacífico

De forma semelhante ao ENOS, a Oscilação Decadal do Pacífico (PDO – Pacific Decadal Oscillation) é caracterizada como a variabilidade na configuração das TSM no Oceano Pacífico, diferenciam-se pela duração climática. Enquanto o prazo do ENOS é de seis a dezoito meses, a PDO dura cerca de 20 a 30 anos. (MANTUA *et al.*, 1997).

O aumento das temperaturas do pacífico representa a fase positiva da PDO, ao passo que a diminuição das temperaturas representa a fase negativa, constituindo assim as fases da PDO. Andreoli e Kayano (2005) ainda complementam que a fase positiva significa uma menor recorrência do La Niña e com menor intensidade, enquanto o El Niño fica mais recorrente e com maior intensidade. Os efeitos se invertem com a fase negativa.

2.2.4 Oscilação Multidecadal do Atlântico

A *Atlantic Multidecadal Oscillation* (AMO) é caracterizada pela alteração da temperatura da superfície do mar (TSM) da região norte do oceano Atlântico, região do Golfo do México e a Baixa da Islândia. Possui uma oscilação de baixa frequência, como o PDO e seu tempo de duração é muito longo, quando comparada com fenômenos como El Niño e La Niña (SIQUEIRA, 2012).

Ademais, as fases frias e quentes da AMO influenciam a posição da ZCIT em escala Multidecadal (KNIGHT; FOLLAND; SCAIFE, 2006). O que mostra Rocha *et al.* (2019), ao detectar que os períodos com fase simultaneamente negativos da PDO e AMO causam precipitações mais fortes no Alto do Jaguaribe.

2.2.5 Oscilação do Atlântico Norte

A Oscilação do Atlântico Norte consiste em um dipolo com orientação aproximada Norte/Sul com um dos centros sobre a Islândia e o outro aos 35°N nos Açores (IPMA, 2014). Siqueira 2010 ainda complementa que na fase positiva, há uma maior pressão sobre Açores e uma baixa pressão na Islândia. Já na fase negativa ocorre o oposto.

Marengo *et al.* (2007) notou que a variabilidade da NAO, associado a outros fatores, podem explicar a variabilidade na escala de tempo interdecadal do NEB.

2.3 Random Forest

O *Random Forest* (RF) é um algoritmo classificador estruturado em árvores $\{h(x, \Theta_k), k = 1, \dots\}$ onde Θ_k são vetores aleatórios independentes e identicamente distribuídos em que cada árvore lança um voto unitário para a variável mais popular da entrada x (BREIMAN, 2022).

O RF é constituído de três componentes: nó interno, ramo e nós folha. O nó interno representa o teste de um atributo, o ramo a um resultado do teste e os nós folha são as classes ou distribuições de classes (RIQUETI; RIBEIRO; ZÁRATE, 2018).

Genuer *et al.* (2010) ainda complementa que o bom desempenho do RF está relacionado à sua aleatoriedade em criar as árvores, onde para cada nó o modelo escolhe a melhor divisão de um subconjunto aleatório de variáveis de entrada.

2.4 Redes Neurais Artificiais

Este tópico possui como foco a apresentação de Redes Neurais Artificiais (RNA), e em seguida, apresentar as principais características da rede *Long Short Term Memory*, modelo utilizado neste trabalho.

2.4.1 Introdução às Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) são algoritmos com uma arquitetura inspirada no funcionamento do cérebro humano a fim de criar sistemas inteligentes capazes de realizar tarefas como classificação, reconhecimento de padrões, processamento de imagens e até previsões (HAYKIN, 2001).

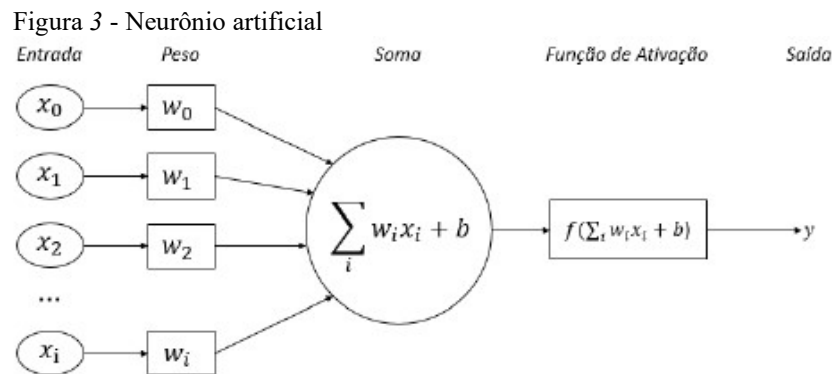
Nesse sentido, as RNAs são construídas por meio de uma rede de unidades de processamento chamadas neurônios. Cada neurônio recebe informação de outro como entrada através das sinapses. Essa informação é multiplicada por um peso definido pelo criador da rede (VASCO, 2020).

Um neurônio artificial possui “i” entradas (x_i), e cada entrada é conectada ao neurônio através de sinapses com pesos (w_i). O neurônio soma as entradas multiplicadas pelos pesos conforme a seguinte equação:

$$A = \sum_i x_i w_i + b \quad (2.1)$$

Onde b é o valor do viés (bias). Já para o cálculo da saída, à soma é aplicada uma função de ativação (HIRANSHA *et al.*, 2018).

$$y = f(A) \quad (2.2)$$



Fonte: Vasco (2020).

Acharya *et al.* (2003) afirma que as redes neurais derivam seu poder devido sua estrutura massiva e paralela, além da capacidade de aprender por experiência. Mas que para isso, é necessário tomar três decisões importantes no processo de construção de sua arquitetura, são elas: a topologia da rede, algoritmo de aprendizagem e a função de ativação.

Quanto a topologia da rede, há três principais tipos: *feed-forward*, *feed-backward* e redes competitivas. Já para o algoritmo de aprendizagem, ou seja, como a rede deve aprender

ajustando seus pesos sinápticos, há quatro formas: aprendizagem por correção de erro, aprendizagem Hebbiana, aprendizagem de Boltzmann e aprendizagem competitiva (VASCO, 2020).

Para que uma RNA funcione corretamente, é necessário que cada neurônio processe a informação, gere um resultado e repasse para o neurônio subsequente. Como essa informação será passada é definida pela função de ativação (ROCHA *et al.*, 2012).

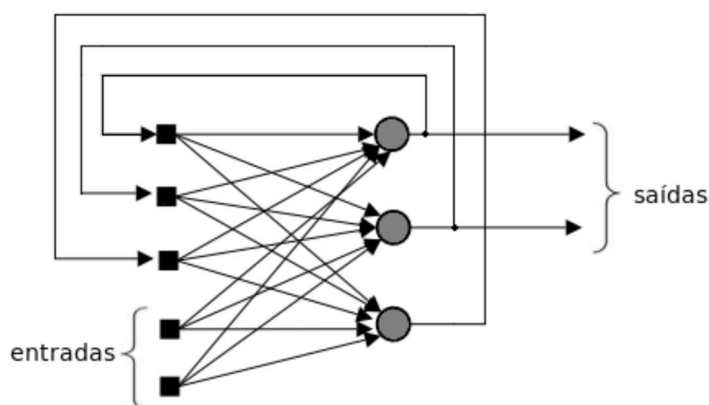
2.4.2 Rede Neural Recorrente (RNR)

Rede Neural Recorrente (*feed-backward*) possui uma arquitetura em que promove uma realimentação da seguinte forma: a saída de um neurônio é aplicada como entrada no próprio neurônio e/ou em outros de camadas anteriores (GONÇALVES, 2008).

Ou seja, RNRs recebem entradas de duas fontes: uma do presente e outra do passado. Portanto, consideram que esse tipo de rede possui memória e se assemelha mais a forma de como os humanos processam informação (ZAREMBA; SUTSKEVER; VINYALS, 2014).

A Figura 04 demonstra a arquitetura de uma rede recorrente genérica.

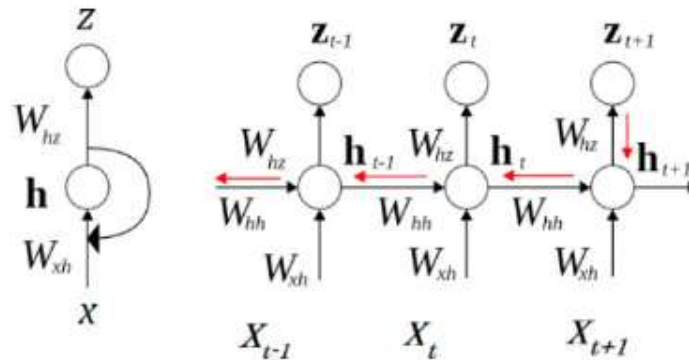
Figura 4 – Rede Neural Recorrente.



Fonte: Gonçalves (2008).

Diferente de outras redes, as RNRs possuem mais de uma camada intermediária. A Figura 05 demonstra mais detalhadamente como funciona essa rede.

Figura 5 - Camadas de uma Rede Neural Recorrente.



Fonte: Hiransha et al. (2018).

Da RNR representada na Figura 05, a equação de entrada das camadas intermediárias é dada por:

$$h_t = g_n(W_{xh}X_t + W_{hh}h_{t-1} + b_h) \quad (2.3)$$

Onde h_t é a camada intermediária no instante t , g_n é a função de ativação, W_{xh} é a matriz de peso de entrada, X_t é a entrada no instante t , W_{hh} é a matriz de peso do neurônio recorrente, h_{t-1} é a camada intermediária no instante $t - 1$ e b_h é o valor de viés.

A equação de saída da camada intermediária é dada por:

$$Z_t = g_n(W_{hz}h_t + b_z) \quad (2.4)$$

Onde Z_t é o vetor de saída, W_{hz} é a matriz de peso para a camada de saída e b_z é o viés (HIRANSHA *et al.*, 2018).

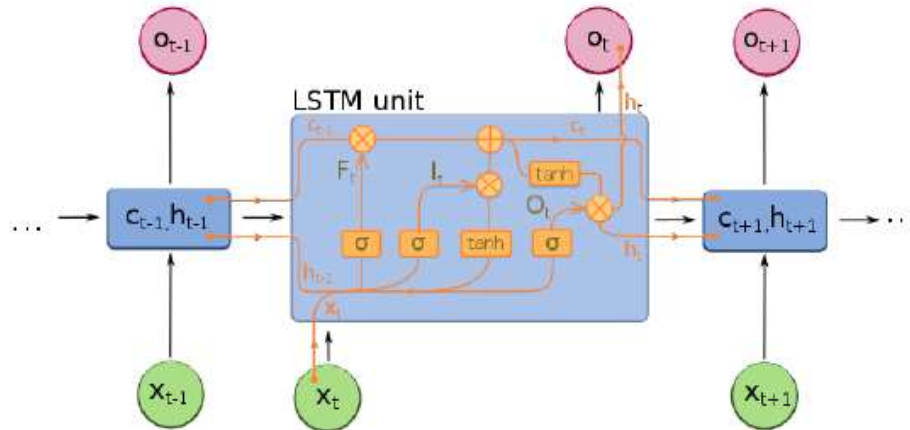
Haykin (2008) destaca que um problema que se encontra nas RNRs é que o gradiente da função custo decai exponencialmente com o tempo. Por isso, a rede para de aprender e pode dissipar ou explodir o gradiente.

2.4.3 Long Short Term Memory (LSTM)

A *Long Short Term Memory* foi criada para contornar esse problema da função custo decair exponencialmente com o tempo. Sua arquitetura impõe um erro constante garantindo que não haja *overfitting* ou *underfitting*. Dessa forma, ela pode aprender séries grandes sem perder eficiência mesmo que algumas sequências de entradas contenham dados incompreensíveis (HOCHREITER; SCHMIDHUBER, 1997).

A Figura 06 mostra a representação gráfica da LSTM.

Figura 6 - Célula LSTM.



Fonte: Vasco (2020).

A arquitetura de uma LSTM é composta por um estado da célula (c_t), saída chamada de *hidden state* (h_t), um *input gate* (I_t), um *output gate* (O_t) e um *forget gate* (F_t). As equações que representam uma célula LSTM são dadas por:

$$I_t = \sigma(W_{xi}X_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2.5)$$

$$F_t = \sigma(W_{xf}X_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2.6)$$

$$c_t = F_t \times c_{t-1} + I_t \times \tanh(W_{xc}X_t + W_{hc}h_{t-1} + b_c) \quad (2.7)$$

$$O_t = \sigma(W_{xo}X_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (2.8)$$

$$h_t = O_t \times \tanh(c_t) \quad (2.9)$$

Hiransha *et al.* (2018) explica que o estado da célula carrega informação da célula anterior para a atual. O *forget gate* decide se guarda a informação no estado da célula. A saída do *forget gate* é adicionada ao estado da célula através da multiplicação de elemento a elemento. O *input gate* junta a camada sigmoide e a tangente hiperbólica no estado da célula. Já a saída é formada pela multiplicação de elemento a elemento do *output gate* e a tangente hiperbólica.

2.4.4 Ajuste dos modelos e dos hiperparâmetros

Conforme Vasco (2020), a função de ativação é responsável por restringir a saída de acordo com o problema proposto. Para séries temporais, a função Sigmoide e Tangente Hiperbólica são as mais comuns.

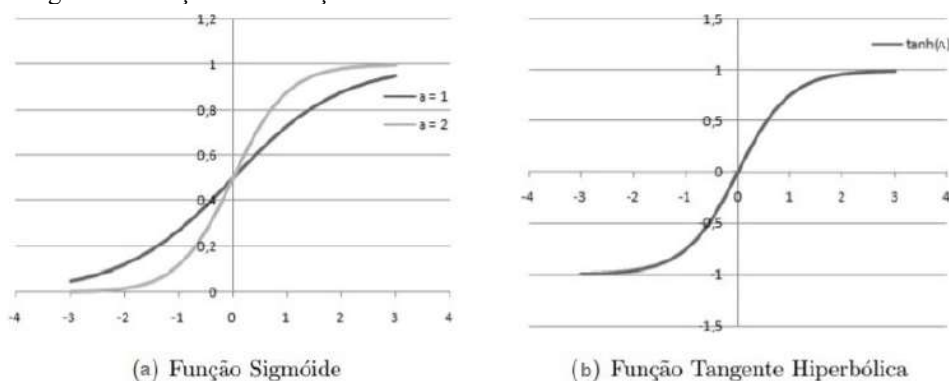
As equações 2.10 e 2.11 demonstram a função Sigmoide e Tangente Hiperbólica, respectivamente.

$$f(A) = \frac{1}{1 + e^{-\alpha A}} \quad (2.10)$$

$$f(A) = \tanh(A) \quad (2.11)$$

A Figura 07 ilustra graficamente as duas funções expostas.

Figura 7 - Funções de ativação



Fonte: Vasco (2020)

O *dropout* é o parâmetro responsável por impedir que o modelo se ajuste de modo excessivo aos dados do treinamento e não consiga generalizar. Ele escolhe de forma aleatória uma porcentagem das células dentro de cada camada e substitui o valor do *output* por 0. (SRIVASTAVA *et al.*, 2014).

Segundo Vasco (2020), os otimizadores identificam os valores de peso ótimos a serem colocados na rede neural com o objetivo de minimizar o erro. Os mais comuns são o Adam, Stochastic Gradient Descent (SGD) e o RMSProp.

Batch Size ou lote representa o número de dados de treinamento que serão utilizados para o aprendizado em uma iteração, ou seja, é o número de amostras que o modelo analisa antes dele ir para a próxima iteração.

Outro parâmetro importante na construção da rede neural é o número de épocas, ele representa o número de vezes que todo o conjunto de dados passa pela rede. Vasco (2020) salienta que é necessária uma cautela sobre o valor desse parâmetro, haja visto que se for escolhido um número pequeno, a rede pode não se ajustar adequadamente, e se for muito grande, pode haver um excesso de ajuste aos dados de treinamento.

Por fim, há o *lookback*, parâmetro usado para informar o número de dados que a rede usará para fazer a próxima previsão.

3 METODOLOGIA

Richardson (2011) afirma que uma pesquisa quantitativa é aquela que se baseia em dados mensuráveis para verificar e explicar sua existência, relação ou influência. Dessa forma, este estudo se caracteriza como quantitativo, já que se almeja verificar a relação entre Temperatura de Superfície do Mar (TSM) e a precipitação da Região Hidrográfica (RH) dos Sertões de Crateús através de dados mensuráveis.

Não obstante isso, como se pretende obter modelos de previsão pluviométrica para a RH dos Sertões de Crateús a partir dessa verificação dos dados, pode-se caracterizar este trabalho ainda como um estudo de caso com lógica indutiva.

Conforme Yin (2010), estudo de caso é uma investigação empírica que analisa um fenômeno contemporâneo dentro de seu contexto. Godoy (2012) ainda complementa que um estudo de caso indutivo busca compreender padrões e desenvolver conceitos que emergem dos dados coletados.

3.1 Região de estudo

A Região Hidrográfica dos Sertões de Crateús possui uma área de 10.821 km², compreende a rede de drenagem do Rio Poti, é composta por 9 municípios e apresenta uma capacidade de acumulação de águas superficiais de 446.685.647 m³ (SHR, 2013).

Figura 8 - Regiões Hidrográficas do Ceará



Fonte: SRH (2013).

SHR (2013) ainda ressalta que esta região possui um total de 10 açudes monitorados pela COGERH.

Tabela 1 - Reservatórios da RH dos Sertões de Crateús

Reservatório	Capacidade (hm ³)	Reservatório	Capacidade (hm ³)
Barra Velha	99,560	Flor do Campo	105,000
Barragem do Batalhão	1,638	Jaburu II	101,641
Carnaubal	73,200	Realejo	31,550
Colina	4,290	São José III	7,960
Cupim	4,600	Sucesso	6,600

Fonte: Portal Hidrológico do Ceara – FUNCEME (2021)

3.2 Levantamento dos dados

Para gerar os modelos de previsão foram necessários obter a média das precipitações de estações pluviométricas que compõem a Região Hidrográfica dos Sertões de Crateús. Assim como Pereira (2022), calculou-se a média das precipitações através do Thiessen e utilizando um código em linguagem R.

Os dados das anomalias de TSM utilizados foram retirados do sítio do Earth System Research Laboratory (SERL) do National Oceanic and Atmospheric Administration (NOAA) disponível em <http://www.esrl.noaa.gov/psd/data/climateindices/list>. Dos diversos índices disponibilizados, escolheu-se os mesmos utilizados por Evangelista (2019) e Pereira (2022). A Tabela 02 a seguir apresenta esses índices.

Tabela 2 - Índices TSM utilizados

Índices	Sigla
<i>Tropical Atlantic Dipole</i>	TAD
<i>Tropical Northern Atlantic</i>	TNA
<i>Tropical Southern Atlantic</i>	TSA
<i>Niño 1 + 2</i>	NINO1.2
<i>Niño 3</i>	NINO3
<i>Niño 3.4</i>	NINO3.4
<i>Niño 4</i>	NINO4
<i>North Atlantic Oscillation</i>	NÃO
<i>Pacific Decadal Oscillation</i>	PDO
<i>Atlantic Multidecadal Oscillation</i>	AMO
<i>Oceanic Niño Index</i>	ONI

Fonte: Pereira (2022).

Os dados de precipitação da RH dos Sertões de Crateús foram disponibilizados pela Agência Nacional de Águas (ANA), no sítio <http://www.snirh.gov.br/hidroweb>. A Tabela 03 mostra todas as estações pluviométricas utilizadas.

Tabela 3 - Estações Pluviométricas utilizadas para geração do Thiessen

Estação	Nome do Posto	Estação	Nome do Posto	Estação	Nome do Posto
640037	Baixio	540031	Jaburu	440020	Sucesso
640041	Algodões	540027	Vila Santo Antônio	440055	Ipaporanga
540052	Quiterianópolis	540019	Independência	440007	Tamboril
540002	Coutinho	540016	Independência	440080	Açude Carão
540042	Cruz	540035	Recife	440060	Canindezinho
540041	Monteiro	540111	Fazenda Gracinha	440044	Ararendá
540030	Iapi	540028	Tucuns	440017	Nova Russas
540033	Várzea Grande	540034	Ematuba	440045	Santo Antônio
540115	Muquem	540020	Aeroporto		
540113	Fazenda Realeza	540000	Crateús		
540007	Iapi	540026	Irapuã		
540005	Tranqueiras	540025	Morro Alegre		
540038	Cachoeira do Fogo	540039	Barreiros		
540003	Novo Oriente	540044	Oliveira		
540054	Caminho da Cachoeira	540043	Logradouro		
540040	Emaus	540107	Poti		
540036	Jandrangoeira	540023	Ibiapaba		
540108	Desejo	540029	Santa Terezinha		
540032	Monte Sinai	440003	Curatis		
540024	Monte Nebo	440068	Curatis		

Fonte: Pereira (2022).

Vale ressaltar que, como os índices climáticos estão compreendidos entre 1950 e 2020, os dados de precipitação também foram calculados para o mesmo espaço de tempo.

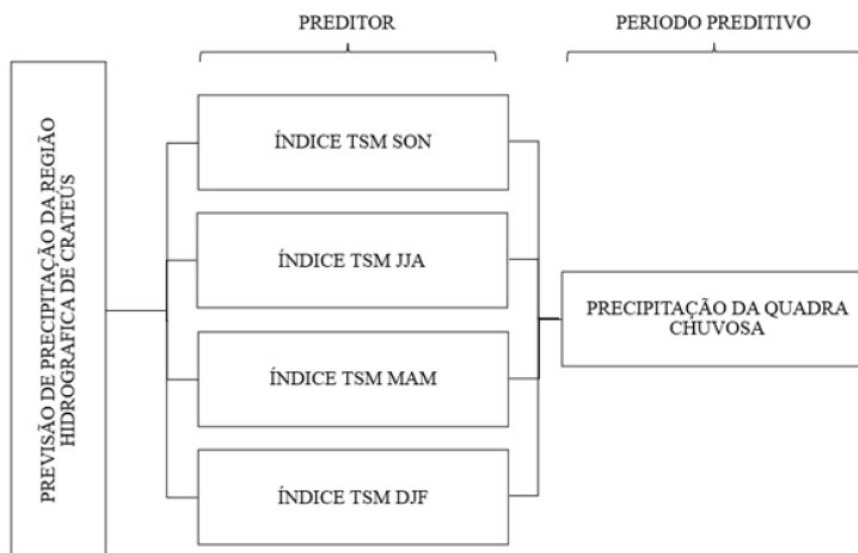
3.3 Estratégia metodológica

Para os modelos de previsão, utilizou-se como período preditivo a quadra chuvosa, que é composta pelos meses de fevereiro, março, abril e maio (FMAM), já que representam as maiores precipitações ao longo do ano na RHSC. Já como preditores, utilizou-se os seguintes trimestres de dados TSM: setembro, outubro e novembro (SON); junho, julho e agosto (JJA); março, abril e maio (MAM); dezembro, janeiro e fevereiro (DJF).

Dessa forma, os dados TSM empregados para cada trimestre foram definidos como a média dos valores dos índices nos meses que constitui o trimestre. Outrossim, para os valores de precipitação da quadra chuvosa, aplicou-se a soma das precipitações dos meses que a compõem, que foram obtidos a partir do Thiessen. Como exemplo de aplicação dessa estratégia

metodológica, tem-se o trabalho de Pereira (2022), que gerou modelos de previsão para a RHSC com regressão linear múltipla. A Figura 9 mostra a estratégia utilizada por ele.

Figura 9 - Estratégia metodológica utilizada para determinação da precipitação na RHSC.



Fonte: Pereira (2022).

Salienta-se que para cada trimestre, foram gerados 11 modelos (M1-M11) em que M1 é o modelo apenas com a primeira variável explicativa, M2 é o modelo com a primeira e segunda variável explicativa e assim por diante até o M11.

3.4 Classificação por ordem de importância dos índices climáticos

Para a classificação dos índices climáticos por ordem de importância, utilizou-se o ranqueamento gerado por Evangelista (2019), que também produziu modelos de previsão para a quadra chuvosa usando a rede neural *Multi Layer Perceptron*.

O ranqueamento produzido por Evangelista (2019) foi implementado no software R-Studio utilizando o pacote “randomForest” e compilando o modelo com 50 *sets* de treino/teste. Além disso, o mesmo adotou o valor de 400 para o número de árvores de decisão e o número de variáveis explicativas dividido por três como valor de entrada escolhidas aleatoriamente em cada divisão. A Tabela 04 mostra o resultado obtido por ele.

Tabela 4 - Ranking de variáveis por Evangelista

Ordem de Classificação	Índices			
	SON	DJF	MAM	JJA
1°	TAD	TAD	NINO3	AMO
2°	AMO	AMO	TSA	TAD
3°	NAO	NINO4	TNA	TSA
4°	TNA	PDO	NINO3.4	TNA
5°	NINO3	TNA	NINO1.2	NINO3
6°	TSA	TSA	TAD	ONI
7°	PDO	NINO1.2	ONI	NINO1.2
8°	ONI	NINO3	NINO4	NINO3.4
9°	NINO4	ONI	NAO	NAO
10°	NINO3.4	NINO3.4	AMO	NINO4
11°	NINO1.2	NAO	PDO	PDO

Fonte: Autor (2020)

3.5 Previsão climática utilizando LSTM

Este tópico se destina a expor todos os passos e decisões realizadas na construção dos modelos de previsão usando a rede LSTM.

3.5.1 Tratamento dos dados

Dos 70 anos de dados (1950-2020), 80% (56 anos) foi usado para o treinamento da rede e 20% (14 anos) para a validação.

James *et al.* (2018) afirma que para aumentar o desempenho de aprendizado da rede neural, é importante que se normalize os dados em valores entre 0 e 1. Dessa forma, normalizou-se todos os dados de TSM e precipitações, conforme a instrução. Para isso, aplicou a seguinte equação nos dados:

$$x_{norm} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (3.1)$$

3.5.2 Construção dos modelos e definição dos hiperparâmetros

Com os dados normalizados, foi implementada a rede neural LSTM no *software* Python e utilizou-se a biblioteca de aprendizado de máquina TensorFlow.

Vasco (2020) produziu um trabalho para avaliar a precisão de predições da rede LSTM aplicada ao mercado de ações variando os hiperparâmetros e obteve os melhores resultados com os seguintes parâmetros:

Tabela 5 - Hiperparâmetros utilizados por Vasco (2020)

Hiperparâmetro	Valor
Número de camadas	4 (1 de entrada, 2 intermediárias e 1 de saída)
Neurônios para cada camada	128
<i>Dropout</i>	30%
Otimizador	Adam
Lote (<i>Batch Size</i>)	16
Função de ativação	Tangente Hiperbólica
Número de épocas (<i>Epochs</i>)	400
Janela (<i>Lookback</i>)	2

Fonte: Autor (2020)

Inicialmente, foi utilizado os parâmetros demonstrados na Tabela 5 para construir a rede neural LSTM. Porém, após alguns testes se fez necessário alterar os valores do lote para 10, número de épocas para 100, janela para 1 e neurônios para 64.

Tal alteração se deu pelo fato de o modelo ter se ajustado demais aos dados de treinamento, principalmente pelo número de épocas anterior.

3.6 Avaliação dos modelos

A acurácia dos modelos foi analisada a partir do coeficiente de correlação Pearson e o valor de Nash-Sutcliffe, parâmetros esses utilizados nos trabalhos de Pereira (2022) e Evangelista (2019).

Com os valores de precipitação real e dos obtidos pelas previsões, calculou-se o coeficiente de correlação Pearson da seguinte forma:

$$R = \frac{1}{N-1} \sum_{i=1}^n \left(\frac{P_i - P_{méd}}{S_p} \right) \left(\frac{O_i - O_{méd}}{S_o} \right) \quad (3.2)$$

Onde P_i e P_i são os valores previstos de precipitação; O_i e $O_{méd}$ são os valores observados de precipitação; S_p e S_o são os valores do desvio padrão dos valores de precipitação previstos e observados.

O coeficiente de correlação Pearson varia de -1 a 1 e mensura a direção e o grau de relação linear entre duas variáveis. Valores iguais a -1 e 1 significam uma correlação perfeita negativa e positiva, respectivamente. Já para valores próximos a zero representam uma relação baixa entre as variáveis. (FIGUEIREDO FILHO; JÚNIOR, 2009).

Para Cohen (1988), valores entre 0,10 e 0,29 são considerados pequenos; escores entre 0,30 e 0,49 são considerados como médios e valores entre 0,50 e 1,00 como grandes. Já para Dancey e Reidy (2005), valores entre 0,10 e 0,30 são considerados fracos; 0,40 até 0,60 são moderados e 0,70 até 1,00 são fortes. Em todo caso, quanto mais próximo de 1,00, melhor a correlação entre as variáveis.

O Nash-Sutcliffe é um coeficiente que varia entre $-\infty$ a 1,00 para demonstrar o quão eficiente o modelo de previsão é. Nesse sentido, valores negativos e positivos indicam uma menor ou maior eficiência do modelo com relação à média e valores iguais a zero indicam que a previsão gerada apresenta a mesma eficiência da média dos dados observados como modelo de previsão (MORIASI *et al.*, 2007). A seguir é demonstrada a equação utilizada para obter esse coeficiente.

$$Nash = 1 - \frac{(P_i - O_i)^2}{(O_i - O_{méd})^2} \quad (3.3)$$

Vale ressaltar que o coeficiente de correlação Pearson e Nash-Sutcliffe foram implementados no Excel.

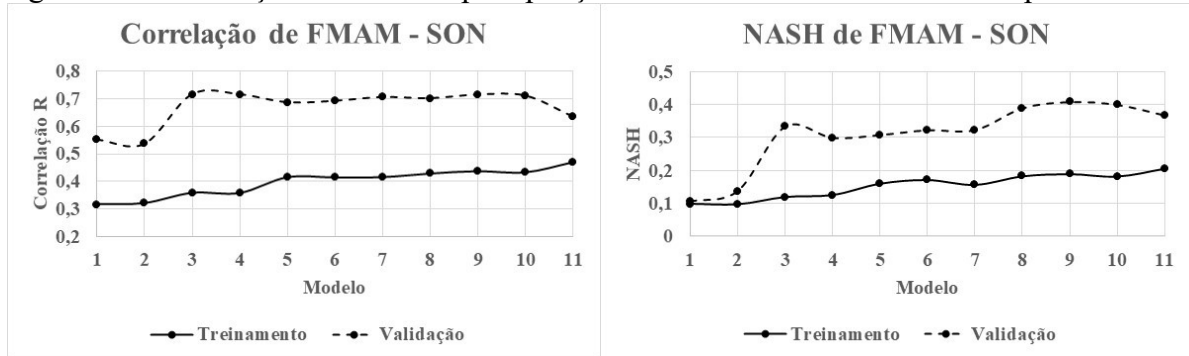
4 RESULTADOS E DISCUSSÃO

Esta seção tem como objetivo apresentar a discussão dos resultados obtidos. De forma mais específica, a geração dos modelos de previsão de precipitação para cada período preditor e sua acurácia.

4.1 Modelos de previsão

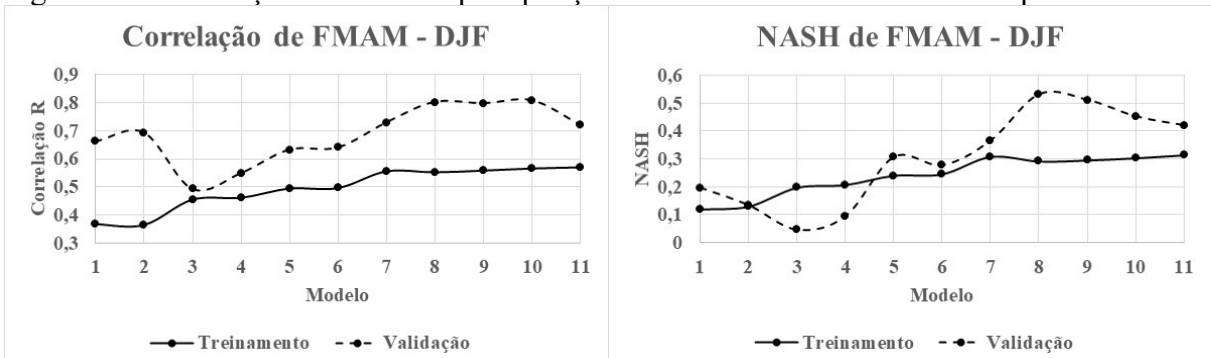
Os resultados dos 11 modelos gerados na rede LSTM para cada um dos quatro preditores são mostrados nas figuras a seguir. Salienta-se que os resultados são referentes aos valores de NASH e coeficiente de correlação Pearson para a etapa de treinamento, assim como para a de validação.

Figura 10 – Correlação e NASH da precipitação observada com a obtida com o preditor SON



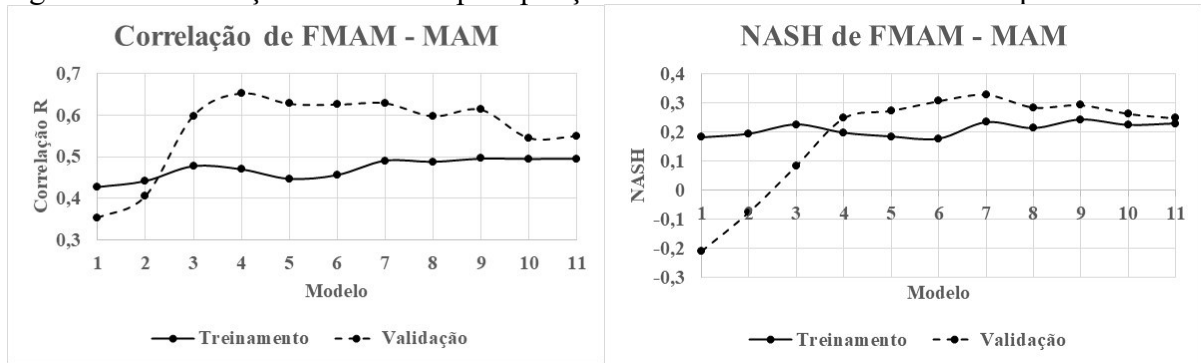
Fonte: Autoria própria (2022)

Figura 11 - Correlação e NASH da precipitação observada com a obtida com o preditor DJF



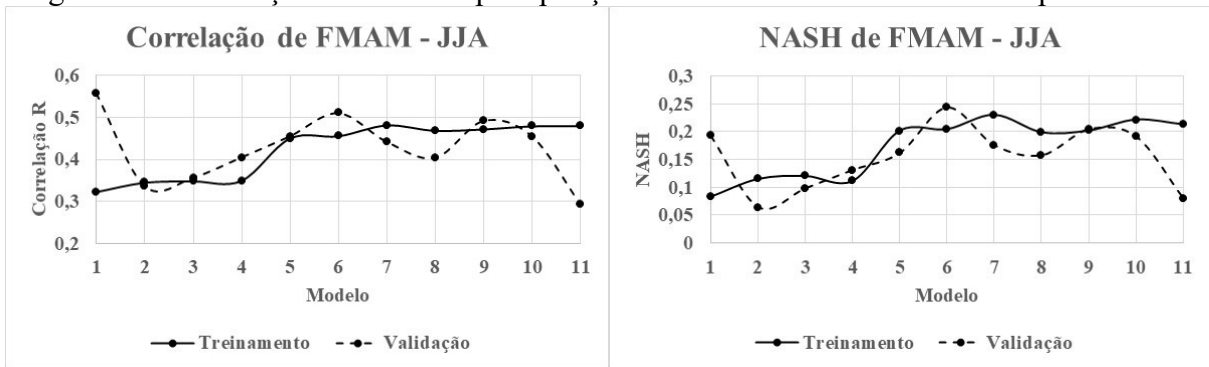
Fonte: Autoria própria (2022)

Figura 12 - Correlação e NASH da precipitação observada com a obtida com o preditor MAM



Fonte: Autoria própria (2022)

Figura 13 - Correlação e NASH da precipitação observada com a obtida com o preditor JJA



Fonte: Autoria própria (2022)

Ao analisar a Figura 10, nota-se que na validação para o preditor SON, obteve-se os melhores resultados de correlação no Modelo 3 (M3) ao Modelo 10 (M10) com valor quase constante em 0,7. O maior valor obtido foi no Modelo 4 (M4) com 0,716. Já no treinamento, percebe-se que a correlação se manteve praticamente crescente à medida que se aumentava o número de variáveis, no qual o maior valor obtido foi de 0,470 no Modelo 11 (M11).

Já para o NASH, houve um padrão de crescimento e decaimento parecido no treinamento e validação, exceto no Modelo 3 (M3) em que houve um crescimento mais acentuado na validação do que no treinamento. Isso fez a curva de validação se distanciar da curva de treinamento. Para a validação, obteve-se o melhor resultado com 0,408 no Modelo 9 (M9) e para o treinamento o valor de 0,206 no Modelo 11 (M11).

Na Figura 11, tem-se os valores para o preditor DJF. Nesse caso, a correlação no treinamento seguiu um crescimento com o aumento dos modelos, atingindo o maior valor no Modelo 11 (M11) com 0,569, já na validação também seguiu um crescimento, exceto no Modelo 3 e 11, em que reduziu. O melhor valor encontrado na validação foi de 0,807 no Modelo 10 (M10).

O NASH do treinamento apresentou uma curva semelhante à correlação e seu melhor resultado também foi para o Modelo 11 (M11) com 0,314. Na validação, o maior NASH foi obtido no Modelo 8 (M8) com 0,530.

Para o preditor MAM (Figura 12), percebe-se que para os Modelos 1 e 2 as correlações para a validação foram menores que a do treinamento. Porém, a partir do Modelo 3, a correlação de validação ultrapassou o treinamento e se permaneceu próximo ao valor de 0,6 até o Modelo 10, enquanto a correlação de treinamento apresentou valores próximos a 0,5 em todos os Modelos. Os maiores resultados obtidos para a validação e treinamento foram 0,653 (M4) e 0,495 (M9).

No NASH para validação do preditor MAM (Figura 12), identifica-se um comportamento linear entre os Modelos 1 ao 4 e após isso, uma tendência ao valor de 0,3 com o melhor resultado sendo no Modelo 7 com 0,328. Já para o treinamento, é notório a tendência ao valor de 0,2 em todos os Modelos. Para este caso, o melhor resultado foi de 0,243 no Modelo 9.

A Figura 13 demonstra que não só a curva de treinamento, mas como a da validação tiveram comportamentos semelhantes no NASH e na correlação. A melhor correlação foi obtida no Modelo 1 (0,559) e Modelo 7 (0,482) e o melhor NASH no Modelo 6 (0,244) e Modelo 7 (0,229) para validação e treinamento, respectivamente.

Tabela 6 - Três maiores correlações da precipitação de FMAM

Preditor	Treinamento						Validação					
	1°		2°		3°		1°		2°		3°	
	R	M	R	M	R	M	R	M	R	M	R	M
SON	0,470	11	0,437	9	0,433	10	0,716	4	0,716	3	0,715	9
DJF	0,569	11	0,566	10	0,558	9	0,808	10	0,802	8	0,797	9
MAM	0,495	9	0,495	11	0,495	10	0,653	4	0,629	7	0,628	5
JJA	0,482	7	0,480	11	0,480	10	0,559	1	0,511	6	0,492	9

Fonte: Autor (2022)

Tabela 7 - Três maiores valores de NASH da precipitação de FMAM

Preditor	Treinamento						Validação					
	1°		2°		3°		1°		2°		3°	
	NASH	M	NASH	M	NASH	M	NASH	M	NASH	M	NASH	M
SON	0,206	11	0,189	9	0,183	8	0,409	9	0,400	10	0,390	8
DJF	0,314	11	0,308	7	0,304	10	0,530	8	0,511	9	0,452	10
MAM	0,243	9	0,233	7	0,229	11	0,328	7	0,306	6	0,293	9
JJA	0,229	7	0,221	10	0,213	11	0,244	6	0,205	9	0,193	1

Fonte: Autor (2022)

As Tabelas 6 e 7 mostram os três melhores modelos para cada preditor referentes ao valor de correlação e NASH, respectivamente. É perceptível que, de forma geral, os melhores resultados foram obtidos com os modelos com mais variáveis, só na validação do preditor JJA que o Modelo 1 foi obtido como um bom resultado e nos preditores SON e MAM que o Modelo 4 apareceu.

Outro ponto a se destacar é que na maioria dos modelos, obteve-se melhores resultados de NASH e correlação para a etapa de validação do que para a etapa de treinamento. Além disso, dentre todos os preditores, DJF apresentou o melhor resultado de correlação e de NASH no treinamento e validação.

4.2 Comparação com a literatura

Evangelista (2019) também produziu modelos de previsão para a quadra chuvosa (FMAM) e realizou um *ranking* com os três melhores resultados para cada preditor. As Tabelas 8 e 9 apresentam uma comparação entre o melhor modelo obtido para cada preditor neste trabalho com o apresentado por Evangelista (2019).

Tabela 8 – Comparação do melhor resultado de correlação para cada preditor

Preditor	Treinamento		Validação	
	Autor	Evangelista (2019)	Autor	Evangelista (2019)
SON	0,470	0,603	0,716	0,627
DJF	0,569	0,447	0,808	0,261
MAM	0,495	0,631	0,653	0,227
JJA	0,482	0,558	0,559	0,724

Fonte: Autor (2022)

Tabela 9 - Comparação do melhor resultado de NASH para cada preditor

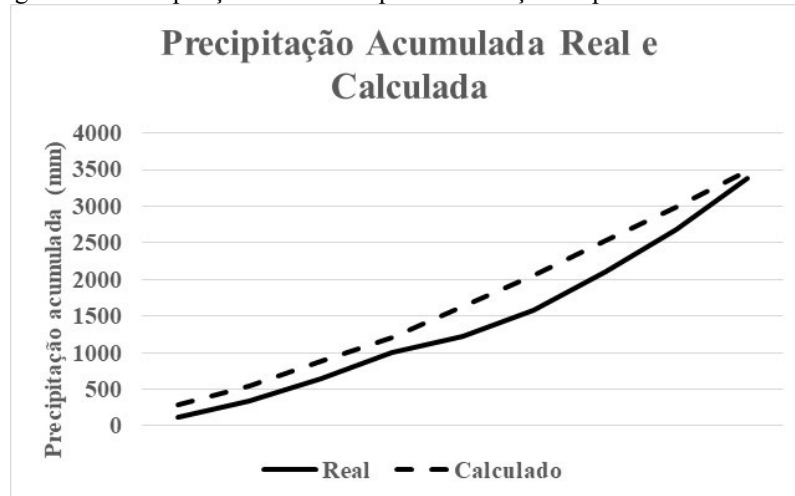
Preditor	Treinamento		Validação	
	Autor	Evangelista (2019)	Autor	Evangelista (2019)
SON	0,206	0,361	0,409	0,300
DJF	0,314	0,224	0,530	-0,218
MAM	0,243	0,393	0,328	0,297
JJA	0,229	0,302	0,244	0,240

Fonte: Autor (2022)

De modo geral, os resultados de correlação e NASH de treinamento obtidos neste trabalho foram menores do que os apresentados por Evangelista (2019). No entanto, na validação e para todos os preditores, os valores de correlação e NASH adquiridos se mostraram melhor do que o trabalho em comparação.

Evidencia-se que o melhor preditor adquirido por Evangelista (2019) foi SON com 0,300 de NASH na validação. Enquanto neste trabalho obteve-se NASH de 0,530 com o preditor DJF Modelo 8.

Figura 14 - Precipitação acumulada para a validação do preditor DJF Modelo 8



Fonte: Autor (2022)

Pela Figura 14 se nota a proximidade entre a precipitação acumulada dos 14 anos de validação real e a calculada usando a rede neural LSTM com o preditor DJF e Modelo 8. O erro percentual para o último ano de precipitação acumulada comparado com o calculado chegou a ser 3%.

5 CONCLUSÃO

Este trabalho teve como objetivo gerar modelos de previsão de precipitação para a quadra chuvosa (FMAM) da Região Hidrográfica dos Sertões de Crateús, além de verificar a eficácia deles. Como período preditor, utilizou-se os trimestres SON, DJF, MAM e JJA, gerando no total 44 modelos de previsão.

A partir do *ranking* de TSM da literatura, gerou-se os 44 modelos na rede neural recorrente *Long Short Term Memory* (LSTM) implementada no Python. Diante dos resultados, é válido constatar que foram satisfatórios. Já que se obteve valores de correlações médias em todos os preditores na fase de treinamento e de valores considerados elevados em todos preditores na fase de validação, chegando ao máximo de 0,808.

Ademais, exceto pela validação no Modelo 1 e 2 do preditor MAM, todos os modelos apresentaram valores de NASH positivos, o que mostra que as previsões geradas possuem um resultado melhor que a média dos valores observados, tanto no treinamento como na validação. Para o melhor modelo de cada preditor, obteve-se apenas valores positivos de NASH na validação, variando entre 0,244 e 0,530.

Após a comparação entre o melhor modelo obtido para cada preditor neste trabalho com o retirado da literatura, nota-se que a rede atingiu um poder de maior generalização, já que em todos os valores de NASH para validação, os adquiridos com a rede LSTM foram maiores que os apresentados para comparação. O que demonstra que a rede não se ajustou de forma excessiva aos dados de treinamento apresentados.

Evidencia-se que também para os valores de correlação na validação, a maioria obtida com o LSTM atingiu melhores resultados que o de comparação.

Ao analisar a curva de precipitação acumulada para os 14 anos de teste real e a calculada, nota-se uma boa proximidade com um erro percentual de 3% para o último ano.

Por fim, para estudos futuros se recomenda:

- a) Utilizar a própria variável de precipitação como dado de entrada na rede neural;
- b) Verificar a eficiência dos modelos para diferentes períodos preditivos;
- c) Verificar a eficiência da rede neural LSTM com outras arquiteturas e até mesmo com outra rede neural recorrente;
- d) Analisar a influência de outros índices TSM na precipitação da região.

REFERÊNCIAS

- ACHARYA, U. R. *et al.* Classification of heart rate data using artificial neural network and fuzzy equivalence relation. *Pattern Recognition*, v. 36, 2003.
- ANDREOLI, R.V.; KAYANO, M.T. A importância relativa do Atlântico Tropical Sul e Pacífico Leste na variabilidade de precipitação do Nordeste do Brasil. *Revista Brasileira de Meteorologia*, v. 22, n.1, p.63-74, 2007.
- ARAGAO, JOR de. O Impacto do ENSO e do Dipolo do Atlântico no Nordeste do Brasil. **Bull. Inst. fr. études andines**, v. 27, n. 3, p. 839-844, 1998.
- BERLATO, M. A.; FARENZENA, H.; FONTANA, D. C. **Associação entre El Niño Oscilação Sul e a produtividade do milho no estado do Rio Grande do Sul.** *Pesquisa Agropecuária Brasileira*, Brasília, v.40, n. 5, maio/2005. P. 423-432.
- BREIMAN, L. **Random Forests** Kluwer Academic Publishers, 2022.
- CARVALHO, T. M. N. Water demand modelling using machine learning techniques. Dissertação (Mestre) em Recursos Hídricos, Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará, Fortaleza, 2019.
- COGERH. **Sertões de Crateús.** Disponível em: <<https://portal.cogerh.com.br/sertoes-decrateus/>>.
- COHEN, Jacob. *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ, Erlbaum, 1988.
- DANCEY, Christine & REIDY, John. *Estatística Sem Matemática para Psicologia: Usando SPSS para Windows.* Porto Alegre, Artmed, 2006.
- EVANGELISTA, Francisco Igor Siqueira. **Previsão climática sazonal para a região hidrográfica dos sertões de Crateús utilizando redes neurais artificiais.** Curso de Engenharia Civil, Universidade Federal do Ceará, Crateús, 2019.

FERREIRA, A. G.; MELLO, N. G. da S. Principais Sistemas Atmosfericos Atuantes Sobre a Regiao Nordeste Do Brasil E a Influencia Dos Oceanos Pacifico E Atlantico No Clima Da Regiao. **Revista Brasileira de Climatologia**, v. 1, n. 1, p. 15–28, 2005.

FIGUEIREDO FILHO, Dalson Britto; SILVA JÚNIOR, José Alexandre da. Desvendando os Mistérios dDesvendando os Mistérios do Coeficiente de Correlação de Pearson (r)o Coeficiente de Correlação de Pearson (r). **Revista Política Hoje**, Pernambuco, v. 18, n. 1, p. 115-146, 2009.

GENUER, R.; POGGI, J.; TULEAU-MALOT, C. Variable selection using Random forests. *Pattern Recognition Letters*, v. 31, n. 14, p. 2225-2236, 2010.

GONÇALVES, André Ricardo. *Redes Neurais Artificiais*. Unicamp, 2008.

GRIMM, A.; FERRAZ, S. E. T.; GOMES, J. Precipitation anomalies in Southern Brazil Associated with El Niño and La Niña events, **Journal of Climate**, Boston, v.11, p.2863-2880, NOV., 1998.

HAYKIN, S. *Redes Neurais Artificiais: Princípios e práticas*. 2ª edição, Editora Bookman, Porto Alegre, 2001.

HIRANSHA, M. et al. Nse stock market prediction using deep-learning models. *Procedia Computer Science*, v. 132, p. 1351 – 1362, 2018. ISSN 1877-0509. International Conference on Computational Intelligence and Data Science. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1877050918307828>>.

INSTITUTO DE PESQUISA E ESTRATEGIA ECONOMICA DO CEARA - IPECE. **Perfil das Regiões de Planemento**. Fortaleza: IPECE 2020.

IPMA, Instituto Português do Mar e da Atmosfera. Área educativa (2014) NAO (North Atlantic Oscillation - Oscilação do Atlântico Norte).

JAMES, G. et al. *An Introduction to Statistical Learning: with Applications in R*. 2018.

KNIGHT, J. R.; FOLLAND, C. K.; SCAIFE, A. A. Climate impacts of the Atlantic multidecadal oscillation. **Geophysical Research Letters**, v. 33, n. 17, p. 2–5, 2006.

MARENGO, J. A., VALVERDE M. C. Caracterização do clima no Século XX e Cenário de Mudanças de clima para o Brasil no Século XXI usando os modelos do IPCC-AR4. *Revista Multiciência*, Campinas, n.8, 2007.

MCCULLOCH, W. S.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 115-133, 1943.

MELO, V. S.; LIMA, L. M. Caracterização da Chuva da Microrregião de Catolé do Rocha no Estado da Paraíba Baseada em Estatística Aplicada. *Revista Brasileira de Meteorologia*, v. 36, n. 01, p. 97-106, 2021.

MORIASI, Daniel N. *et al.* Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. **Transactions of the ASABE**, v. 50, n. 3, p. 885-900, 2007.

MOURA, A.D.; SHUKLA, J. On the dynamics of droughts in northeast Brazil: Observations, theory and numerical experiments with a general circulation model. *J. Atmos. Sci.*, 38, p.2653-2675, 1981.

NOAA. Índice ENSO multivariado version 2 (MEI.v2). v. 2, p. 1-5, 2022. Disponível em: <<https://psl.noaa.gov/enso/mei/>>.

NOBRE, P.; SHUKLA, J. Variations of sea surface temperature, wind stress, and rainfall over the tropical Atlantic and South America. *Journal of Climate*, v. 9, n.10, p.2464-2479, 1996.

NOBREGA, J. N. DA *et al.* RELACOES COM A TSM DOS OCEANOS TROPICAIS Universidade Federal de Campina Grande (UFCG), Unidade Academica de Ciencias Atmosfericas ,. **Revista Brasileira de Meteorologia**, p. 197-208, 2014.

PEREIRA, Samuel Martins. **Previsão climática sazonal para a região hidrográfica dos sertões de Crateús utilizando regressão linear múltipla**. Curso de Engenharia Civil, Universidade Federal do Ceará, Crateús, 2022.

ROCHA, R. V.; SOUZA FILHO, F. A. S.; SILVA, S. M. O. Análise da Relação entre a Precipitação Média do Reservatório Orós, Brasil - Ceará, e os Índices PDO e AMO Através da

Análise de Changepoints e Transformada de Ondeletas. *Revista Brasileira de Meteorologia*, v. 34, n. 1, p. 139-149, 2019.

RIQUETI, G. A.; RIBEIRO, C. E.; ZARATE, L. E. Classificando perfis de longevidade de bases de dados longitudinais usando floresta aleatoria. *In: Symposium on Knowledge Discovery; mining and Learning – KMMILE*, 2018.

SILVA, Vicente de Paulo Rodrigues da; PEREIRA, Emerson Ricardo Rodrigues; ALMEIDA, Rafaela Silveira Rodrigues. Estudo da variabilidade anual e intra-anual da precipitacao na regioa nordeste do Brasil. *Revista Brasileira de Meteorologia*, v. 27, n. 2, p. 163-172, 2012.

SIQUEIRA, A. H. B. VARIABILIDADE DO CLIMA DA AMÉRICA DO SUL E SUA RELAÇÃO COM OS ÍNDICES OCEÂNICOS E ATMOSFÉRICOS. Dissertação (Mestrado em Meteorologia) – Instituto de Ciências Atmosféricas, Universidade Federal de Alagoas. 77 f. Maceió, 2011.

SRIVASTAVA, N. et al. Dropout: A simple way to prevent neurais networks from overfitting. *Journal of Machine Learning Research*, v. 15, n. 56, p. 1929–1958, 2014.

UVO, C. B.; BERNDTSSON, R. Regionalization and spatial properties of Ceara State rainfall in northeast Brazil. *Journal of Geophysical Research Atmospheres*, v. 101, n. D2, p. 4221–4233, 1996.

VASCO, Lucas Pimenta. **Um Estudo de Redes Neurais Recorrentes no Contexto de Previsões no Mercado Financeiro**. Curso de Engenharia da Computação, Universidade Federal de São Carlos, São Carlos, 2020.

Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent Neural Network Regularization. *arXiv e-Prints*, p. arXiv:1409.2329, set. 2014.