

Predição de níveis de precipitação em Catalão-Goiás com base em séries temporais de múltiplos pontos geográficos

Precipitation forecast in Catalão-Goiás based on time series of multiple geographic points

DOI:10.34117/bjdv7n4-113

Recebimento dos originais: 07/03/2021

Aceitação para publicação: 01/04/2021

Gabriel Santos Resende

Graduando de Ciência da Computação

Departamento de Computação (DC) Instituto de Biotecnologia (IBiotec) – Universidade Federal de Catalão – (UFCAT)

Endereço: Av. Dr. Lamartine Pinto de Avelar, 1120, Setor Universitário - CEP: 75704-020

E-mail: gabriel.s.resende99@gmail.com

Lucas Ávila Oliveira

Graduando de Ciência da Computação

Departamento de Computação (DC) – Instituto de Biotecnologia (IBiotec) – Universidade Federal de Catalão – (UFCAT)

Endereço: Av. Dr. Lamartine Pinto de Avelar, 1120, Setor Universitário - CEP: 75704-020

E-mail: eu.lucasavila@gmail.com

Tércio Alberto dos Santos Filho

Doutor em Engenharia Elétrica

Professor Associado do Departamento de Computação (DC) – Instituto de Biotecnologia (IBiotec) – Universidade Federal de Catalão – (UFCAT)

Endereço: Av. Dr. Lamartine Pinto de Avelar, 1120, Setor Universitário - CEP: 75704-020

E-mail: tercioas@ufg.br

Sérgio Francisco da Silva

Doutor em Ciência da Computação e Matemática Computacional

Professor Associado Departamento de Computação (DC) – Instituto de Biotecnologia (IBiotec) – Universidade Federal de Catalão – (UFCAT)

Endereço: Av. Dr. Lamartine Pinto de Avelar, 1120, Setor Universitário - CEP: 75704-020

E-mail: sergio@ufg.br

RESUMO

Predição de precipitação é importante para planejamento do dia-a-dia e para tomada de decisões de curto à longo prazo. Atualmente, o campo de predição meteorológica conta com o suporte da evolução tecnológica, tal como das redes de sensores, que permitem a coleta automática de grande quantidade de dados com alta precisão, e da comunidade científica que tem desenvolvido várias técnicas de aprendizado de máquina,

especialmente de redes neurais. Este estudo investiga o uso de uma rede neural chamada *long short-term memory* (LSTM) para a previsão da precipitação em Catalão/Goiás com séries temporais de múltiplos pontos geográficos. Considerou-se as tarefas de regressão (previsão de valor real em mm de chuva) e de classificação do nível chuva em cinco classe (sem chuva, pouca chuva, chuva, chuva moderada, muita chuva) usando a base de dados com diferentes periodicidades: horária e diária. Os resultados mostram que o uso de dados com periodicidade horária produz resultados mais precisos tanto de classificação quanto de regressão. Adicionalmente, considerando os dados em periodicidade diária, os resultados de classificação foram mais precisos.

Palavras-chave: Séries temporais, previsão de precipitação, redes neurais, memória de longo a curto prazo.

ABSTRACT

Prediction of precipitation is important for day-to-day planning and making decisions of short to long term. Currently, the field of meteorological prediction has the support of technological evolution, such as sensor networks, which allow the automatic capture of large amounts of data with high precision, and the scientific community that has developed several machine learning techniques, especially neural networks. This study investigates the use of a neural network called long short-term memory (LSTM) for the prediction of precipitation in Catalão / Goiás with time series of multiple geographical points. The regression tasks (prediction of real value in mm of rain) and classification of the rain level into five classes (no rain, little rain, moderate rain, lots of rain) were considered using the database with different periodicities: hourly and daily. The results show that the use of data with an hourly periodicity produces more accurate results for both classification and regression tasks. Additionally, considering the data with a daily periodicity, the classification results were more accurate.

Keywords: Time series, precipitation prediction, neural networks, long-short term memory (LSTM)

1 INTRODUÇÃO

A capacidade de prever com alta precisão as condições meteorológicas futuras, como o nível de precipitação, é de suma importância para sociedade. A partir dessas previsões pode-se planejar de forma adequada as atividades a serem desenvolvidas, principalmente aquelas sensíveis às condições meteorológicas como os plantios, colheitas, aplicações de fertilizantes e herbicidas, entre outros manejos agrícolas (ABREU, 2018), e a segurança pública com relação a problemas infraestruturais deflagrados pelas chuvas como enchentes, inundações, deslizamentos de terra, entre outros (NATIONAL RESEARCH COUNCIL, 2012). Na agricultura o conhecimento das condições meteorológicas futuras é importante para: os plantios (GOMES et al., 2007) - para que não haja problemas de germinação, perda de insumos e mortalidade de plantas; as colheitas (MARCHEZAN, 1995) - principalmente de grãos, que requerem tempo seco

para que não haja perdas futuras devido a humidade; os manejos (MONQUERO et al., 2008) - onde várias atividades exigem condições meteorológicas específicas para uma maior eficiência, como a aplicação de herbicidas e inseticidas, onde muitos têm, por exemplo, a eficiência diminuída com chuvas intensas logo após a aplicação. Na área de segurança pública a previsão de precipitação é importante para que as autoridades possam alertar os cidadãos sobre riscos de problemas infraestruturais como alagamentos, deslizamentos de terra, riscos de queda de árvores, de rompimento em barragens e represas, entre outros. Esses tipos de alertas, além de serem uma oportunidade para evitar perdas materiais, muitos deles, como no caso de deslizamentos de terra, enchentes e rompimentos de barragens são imprescindíveis para evitar a morte de pessoas e animais.

Atualmente há dois tipos básicos de técnicas para previsão do tempo (estado atmosférico): as que se baseiam na simulação da física da atmosfera (MARRAS et al., 2016) e as que se baseiam em modelos matemáticos/computacionais que utilizam de séries temporais de dados meteorológicos passados para a previsão de eventos futuros (DAS & GHOSH, 2018). Enquanto que a simulação da física da atmosfera é complexa devido a sua dinamicidade, os modelos matemáticos/computacionais necessitam de uma grande quantidade de dados. Contudo, com o avanço das tecnologias de sensores e das redes de comunicação, tais dados temporais se tornaram abundantes e de alta precisão, aumentando então a viabilidade de modelos de previsão baseados em dados. Esta pesquisa foca na previsão de precipitação utilizando métodos computacionais de inteligência artificial e séries temporais de múltiplos pontos geográficos. Uma série temporal é um conjunto de dados onde as amostras são atreladas ao tempo, ou seja, cada informação contida em uma série temporal está relacionada a um determinado evento em determinado instante de tempo (CHATFIELD, 2004).

Modelos de inteligência artificial, sobretudo redes neurais recorrentes, ganharam destaque e têm produzido resultados expressivos de previsão com base em séries temporais (MA et al., 2015). As redes neurais recorrentes são um tipo de rede neural especializada para dados temporais, como por exemplo, dados meteorológicos, sendo capazes de aprender informações temporais. Porém, o tipo mais simples de redes neurais recorrentes possui problemas especialmente quanto ao aprendizado de dependências temporais de longo prazo, o que normalmente resulta em precisão insatisfatória nas previsões. Na tentativa de superar estas limitações, outros tipos de redes neurais recorrentes foram derivadas deste modelo simples, dentre essas, a rede *Long-short Term Memory* (LSTM). LSTM introduz o conceito de blocos de memória para superar o

gargalo da aprendizagem de dependências temporais de longo à curto prazo (HOCHREITER & SCHMIDHUBER, 1997; MA et al., 2015) .

Diversos trabalhos utilizam LSTM por suas capacidades de previsão de dados com dependências temporais de longo prazo, como em Ma et al. (2015) na predição de fluxo de tráfego; Ke et al. (2017) na predição de demanda de passageiros em serviços de transporte; Kim et al. (2020) na predição e gerenciamento/equilíbrio de cotas para serviços sobre demanda; Werner et al. (2020) na predição de volume diário de negociações das ações da Petrobrás; Kao et al. (2020) na previsão de enchentes; Dikshit et al. (2021) na previsão de longas estiagens (secas); Liu et al. (2018) na predição da velocidade do vento; Duan et al. (2021) e Xiaoyun et al (2016) na predição da produção de energia eólica; e por fim, Zaytar & Amrani (2016) e Hewage et al. (2020) na predição do tempo.

Neste trabalho apresenta-se um estudo de LSTM para a predição de valores de precipitação na cidade de Catalão-GO utilizando dados de múltiplos pontos geográficos, com base em quatro casos: 1) predição por classificação de níveis de chuva com os dados em periodicidade horária. 2) predição por classificação de níveis de chuva com os dados em periodicidade diária. 3) predição por regressão (valor estimado de precipitação em mm.) com os dados em periodicidade diária. 4) predição por regressão com os dados divididos em periodicidade horária. Na predição por classificação foram considerados cinco níveis de precipitação: sem chuva, pouca chuva, chuva, chuva moderada, muita chuva. Com os experimentos foram extraídas várias conclusões. As principais são: 1) com dados em periodicidade horária tem-se maior precisão dos resultados de predição, o que indica ser promissor o uso dos dados meteorológicos providos pelas estações automáticas que fazem medições de hora em hora; 2) com os dados em periodicidade diária, que resulta em um menor número de amostras, considerando um intervalo fixo de tempo, obteve resultados de classificação mais precisos do que de regressão, indicando que tarefas de regressão exigem uma maior quantidade de dados de treinamento do que a tarefa de classificação.

Este trabalho é estruturado da seguinte forma. A Seção 1 discutiu sobre a importância da predição meteorológica, principalmente da precipitação, e contextualizou o problema de pesquisa e abordagem de investigação utilizada neste trabalho. A Seção 2 apresenta os conceitos fundamentais para a compreensão do trabalho, incluindo os conceitos básicos sobre tempo, clima e os fatores que regem estes, além de conceitos fundamentais sobre redes neurais e LSTM, que constitui a técnica de predição utilizada

neste estudo. A Seção 3 apresenta a base de dados utilizada para a predição de precipitação, o pré-processamento, normalização, janelamento e particionamento (em treinamento e teste) dos dados, seguido do detalhamento da arquitetura e do método de treinamento da rede neural LSTM empregada. A Seção 4 apresenta e discute os resultados, os quais foram analisados com base na matriz de confusão da classificação do nível de chuva, acurácia (*accuracy*), precisão (*precision*), revocação (*recall*), e erro médio absoluto. Concluindo, a Seção 5 lista as principais contribuições do estudo.

2 FUNDAMENTAÇÃO TEÓRICA

Essa seção descreve de forma sucinta os conceitos básicos empregados nesta pesquisa. Primeiramente são descritos os conceitos básicos de tempo e clima, seguido dos fatores que influenciam ambos. Tal descrição visa prover um entendimento sobre os principais fatores atuantes nos estados atmosféricos no Brasil e especialmente na cidade alvo da predição de precipitação. Posteriormente são apresentados os conceitos elementares de redes neurais artificiais destacando o seu elemento primordial, o modelo de neurônio. Por fim, é apresentado o conceito de célula de memória LSTM, que forma a base da rede neural recorrente utilizada para a predição de precipitação.

2.1 TEMPO, CLIMA E FATORES INFLUENCIAIS

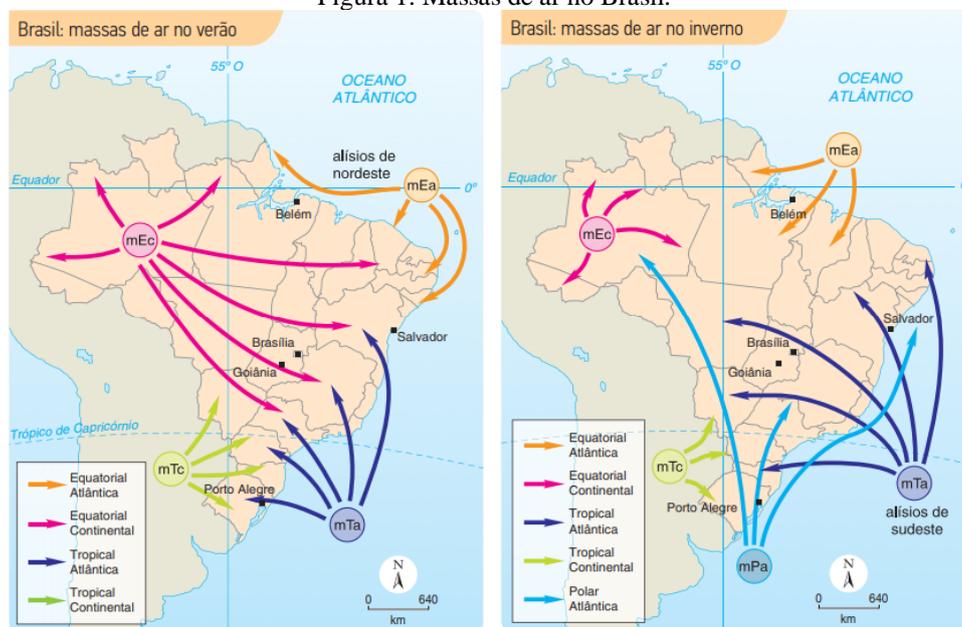
Primeiramente, é importante diferenciar os termos tempo e clima. Tempo se refere a um estado momentâneo da atmosfera em uma determinada área da superfície terrestre, expresso por meio da combinação fenômenos como temperatura, umidade, velocidade do vento, precipitação, radiação, entre outros (MOREIRA E SENE, 2004). Por exemplo, quando falamos "hoje, em Catalão, o dia está quente e úmido", nos referimos ao tempo. Já o clima, corresponde ao comportamento do tempo em uma determinada área durante um período longo, de pelo menos 30 anos (MOREIRA E SENE, 2004). Por exemplo, "o clima em Goiás é tropical semi úmido, com estação seca e chuvosa".

Conforme Moreira e Sene (2004), os principais fatores que influenciam no clima, e conseqüentemente, nas sucessões de tempo de uma região, são: a latitude (distância em relação a linha do Equador), a altitude (altura em relação ao nível do mar), o albedo (que caracteriza a reflexão da superfície), as massas de ar (a circulação do ar ao longo do planeta ou território), a continentalidade e maritimidade (que se refere a proximidade da região de grandes volumes de água), correntes marítimas (os grandes volumes de água

que se deslocam pelo oceano), a vegetação (que se refere a cobertura vegetal da região) e o relevo (que além de estar associado à altitude, influência na circulação das massas de ar).

No Estado de Goiás, mais especificamente na cidade de Catalão (Latitude 18° 09' 57" S, Longitude 47° 56' 47" W, Altitude 835m) o fator mais determinante para o clima e a sucessão de estados meteorológicos são as massas de ar [PRECISO DE REFERÊNCIAS AQUI. VOU PESQUISAR DEPOIS]. A Figura 1 ilustra as massas de ar que atuam no Brasil. As massas de ar Equatorial Atlântica (mEa), Equatorial Continental (mEc) e Tropical Atlântica (mTa) são quentes e úmidas, a Tropical Continental (mTc) é quente e seca, e a Polar Atlântica (mPa) é fria e úmida.

Figura 1: Massas de ar no Brasil.



Fonte: SENE e MOREIRA (2013, 153).

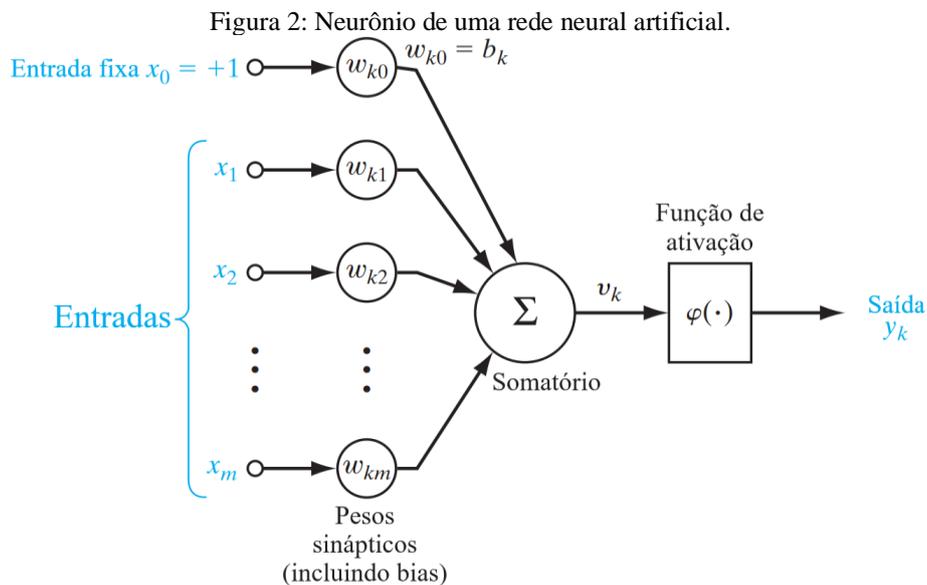
2.2 REDES NEURAIS ARTIFICIAIS

Redes neurais artificiais, segundo Aggarwal (2018) e Haykin (2009), são mecanismos que modelam a estrutura básica do cérebro humano, sendo formadas por neurônios e suas conexões. No cérebro humano as forças das conexões (sinapses) mudam de acordo com os estímulos recebidos, e isso caracteriza o aprendizado. Nas redes neurais artificiais, as sinapses são modeladas por pesos de conexões que são ajustados conforme os exemplos de treinamento. A forma de organização das conexões varia bastante nas redes neurais artificiais, contudo, o elemento básico de formação dessas é o neurônio, que segue uma especialização padronizada, sendo está representada na Figura 2. Um

neurônio k recebe $m+1$ sinais de entrada, e para cada sinal j aplica um peso w_{kj} , produzido o que é denominado de campo local induzido v_k dado por:

$$v_k = \sum_{j=0}^m w_{kj}x_j \quad (1)$$

onde x_1, \dots, x_m são os estímulos de entrada, w_{k1}, \dots, w_{km} são os respectivos pesos sinápticos, $x_0 = 1$ é uma entrada constante que é associada ao peso w_{k0} que aprende a tendência dos dados, também chamada de bias.



Fonte: Adaptada de Haykin (2009).

Em seguida, o campo local induzido v_k passa por uma função de ativação φ , para limitar a saída em um dado intervalo (tipicamente valores entre 0 e 1 ou -1 e 1), resultando na saída y_k , isto é:

$$y_k = \varphi(v_k) \quad (2)$$

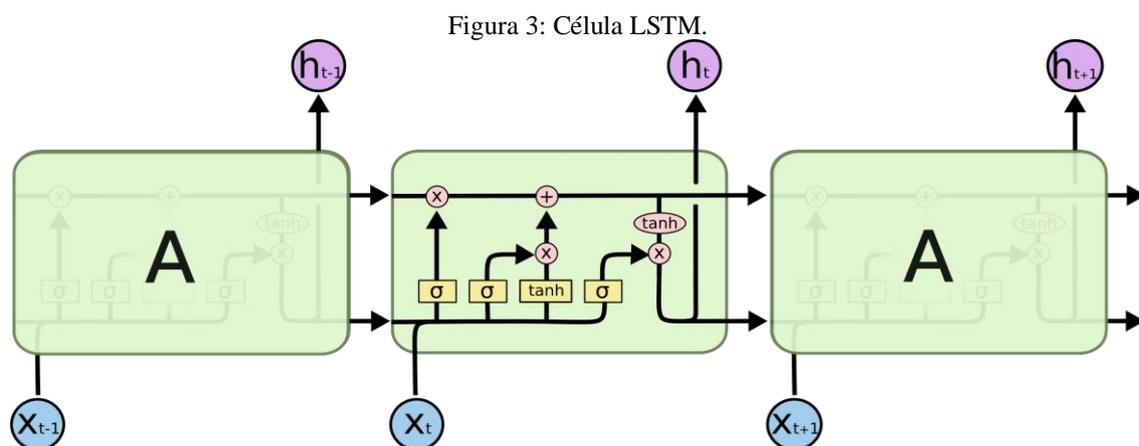
Existem várias funções de ativação e vários modos de organização das conexões entre os neurônios para formar as arquiteturas de redes neurais. As arquiteturas elementares de redes neurais são as de alimentação adiante e as recorrentes, podendo se definir várias ramificações dentro de cada uma dessas categorias. Este trabalho usa uma rede neural recorrente denominada *long-short term memory* (LSTM), que é uma das redes neurais recorrentes promissoras quando é necessário que conhecimento temporal seja

aprendido pela rede, tendo várias aplicações com sucesso na área de séries temporais e processamento de linguagem natural.

2.3 LONG-SHORT TERM MEMORY (LSTM)

A ampla maioria das redes neurais existentes utiliza o cálculo de gradiente, mais especificamente o método de gradiente descendente, para atualizar os pesos da rede neural, de forma a produzir o aprendizado com base nos dados de treinamento. Contudo, quando se necessita aprender dependências temporais de grande a curto prazo incorre-se nos problemas de desaparecimento e explosão do gradiente. Assim, as redes neurais LSTM surgiram como uma tentativa de lidar com esses problemas. Para isso, no lugar de simples neurônios ligados uns aos outros em uma arquitetura recorrente, ou seja, com retroalimentação de sinais, foram desenvolvidas células de memória que retêm informações temporais de curto à longo prazo, utilizando na sua construção, da formulação básica de neurônios e dos conceitos de portões.

A Figura 3 ilustra uma célula LSTM. Cada célula possui três portões: o portão de entrada, o de esquecimento e o de saída. As funções destes portões são respectivamente definir o quanto de informação do estado atual entrará em memória (portão de entrada), o quanto de informação da memória será esquecida (portão de esquecimento) e combinar a memória atual com as novas informações de memória (portão de saída). Com esses portões, cada célula possui um ciclo interno de dados que mantém a memória ou estado da célula.



Fonte: Olah (2015).

O fluxo interno de dados em uma célula LSTM começa com o portão de esquecimento. Este observa a entrada X_t e a saída anterior da célula h_{t-1} e aplica uma

função sigmóide (σ , na Figura 3) para encontrar um valor entre 0 e 1, onde 0 significa esquecer completamente e 1 lembrar de tudo, para cada informação da memória, e então atualiza o estado da célula. O portão de entrada analisa também a entrada \mathbf{X}_t e a saída do cálculo anterior \mathbf{h}_{t-1} , aplica uma função sigmóide (σ), e uma função tangente hiperbólica (tanh) para normalizar os valores de entrada. O resultado da função sigmóide (σ) é multiplicado pelo resultado da função tanh. Dessa forma, o cálculo resulta nas informações que devem ser guardadas na memória. O resultado então é somado ao estado da célula para adicionar as informações na memória. No final, o portão de saída analisa também a entrada \mathbf{X}_t e a saída do cálculo anterior \mathbf{h}_{t-1} e aplica a função sigmóide (σ) multiplicando o resultado pelo estado da célula aplicado uma função tanh, resultando então no processamento final da célula de memória.

As redes neurais LSTM normalmente contém múltiplas camadas de células LSTM, uma camada final completamente conectada de neurônios tradicionais e são treinadas para uma versão do tradicional método de retropropagação de erro conhecido como conhecida como retropropagação de erro através do tempo (backpropagation through time - BPTT). Detalhes sobre as equações que constituem o algoritmo de treinamento das redes LSTM são dados em (COLOCAR REF.)

3 MATERIAL E MÉTODO

3.1 CONJUNTO DE DADOS METEOROLÓGICOS

3.1.1 Fonte

Os dados utilizados no trabalho foram coletados do site do Instituto Nacional de Meteorologia (INMET)¹, no qual foram escolhidas as estações meteorológicas A003 - Morrinhos/GO, A015 - Itapaci/GO, A034 - Catalão/GO, A037 - Silvânia/GO, A520 - Conceição Das Alagoas/MG, A535 - Florestal/MG, A737 - Ibatinga/SP, A738 - Casa Branca/SP, A912 - Campo Verde/MT, A920 - Juína/MT e A932 - Guiratinga/MT. A Figura 4 ilustra a localização das estações meteorológicas, designadas pelos nomes das respectivas cidades. As variáveis usadas foram: hora da medição; precipitação total (mm); pressão atmosférica ao nível da estação (mb); radiação global (kJ/m²); temperatura do ar em bulbo seco (°C); temperatura máxima (°C); temperatura mínima (°C); temperatura do ponto de orvalho (°C); umidade relativa do ar (%); umidade relativa máxima (%);

¹ <https://portal.inmet.gov.br/>

umidade relativa mínima (%); direção do vento ($^{\circ}$ (grau)); rajada de vento máxima (m/s) e velocidade média do vento (m/s).

Figura 4. Localização das cidades de onde foram obtidos os dados meteorológicos.

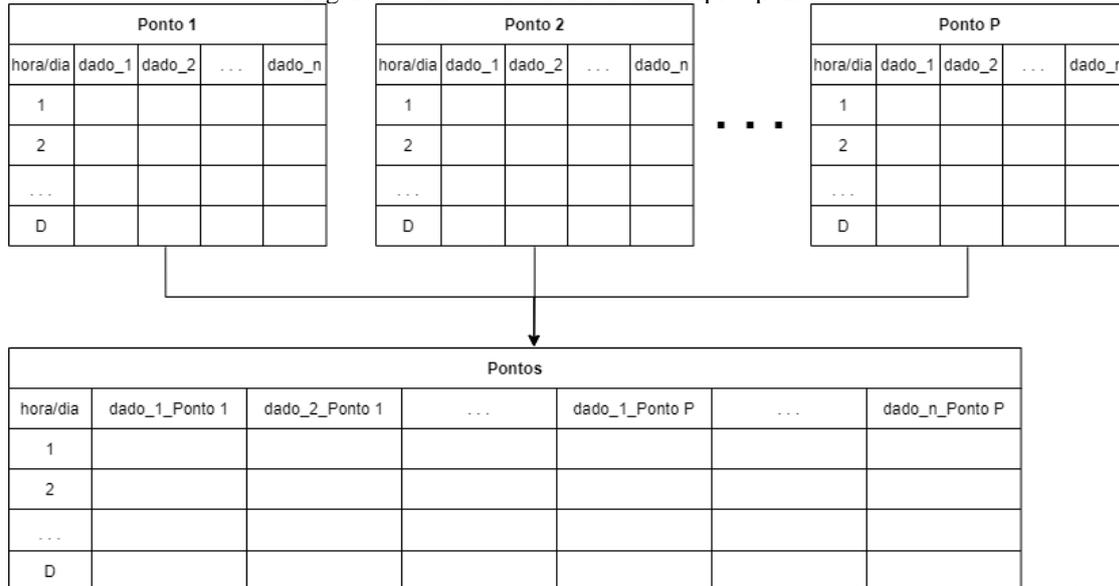


3.1.2 Pré Processamento

As tabelas de dados obtidas do INMET possuíam dados ausentes, os quais foram preenchidos utilizando a média dos valores presentes de cada variável. Como os dados originais tem periodicidade horário, para as tarefas de classificação e regressão em periodicidade diária, é feita uma sumarização dos dados das 24 horas de cada dia pelos seguintes cálculos: Precipitação - Soma; Pressão atmosférica à nível da estação, Radiação, Temperatura, Temperatura ponto de orvalho, Umidade relativa e Velocidade do vento - Média simples; Temperatura máxima, Umidade relativa máxima, Vento rajada máxima e Direção do vento - Maior valor; Temperatura mínima e Umidade relativa mínima - Menor valor.

Como o INMET fornece uma tabela de dados separada para cada estação, então essas foram unidas em uma só tabela (Figura 5) para melhor representação dos dados de entrada, e os dados de precipitação foram extraídos da tabela da estação A034 - Catalão/GO como os dados de saída a serem preditos.

Figura 5: União das tabelas dos múltiplos pontos.



Para a predição por classificação os níveis de precipitação foram discretizados em cinco classes definidas por: ‘Sem chuva’ para precipitação igual a 0 mm; ‘Pouca chuva’ para maior que 0 mm e menor que 3 mm; ‘Chuva’ para maior ou igual a 3 mm e menor que 5 mm, ‘Chuva moderada’ para maior ou igual a 5 mm e menor que 20 mm e ‘Muita chuva’ para maior ou igual a 20 mm . Por boa parte do ano não ocorrer chuvas as classes ficam desbalanceadas e portanto foi utilizado uma técnica chamada SMOTE (N. V. Chawla et. al., 2002) para o balanceamento das classes.

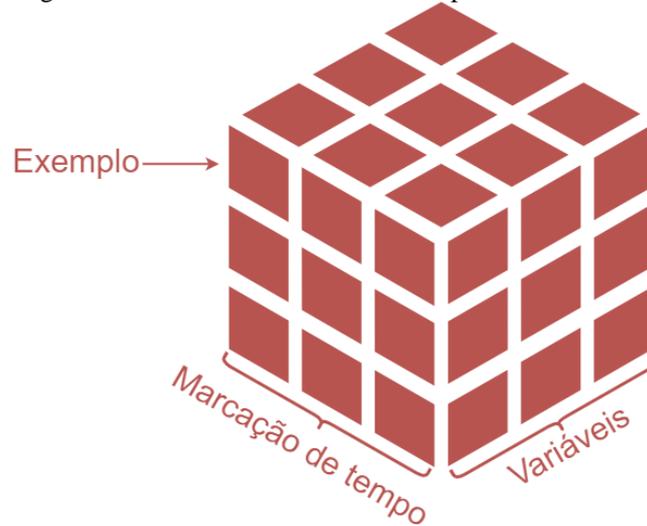
3.1.3 Normalização

Visando uma melhor predição por redes neurais, é usual normalizar os dados, geralmente entre 0 e 1 ou entre -1 e 1; isto faz com que a rede consiga lidar de maneira mais efetiva com os dados. Nesta pesquisa os dados foram normalizados entre -1 e 1 com exceção dos dados de saída na tarefa de classificação, pois cada classe será representada por um número.

3.1.4 Janelamento

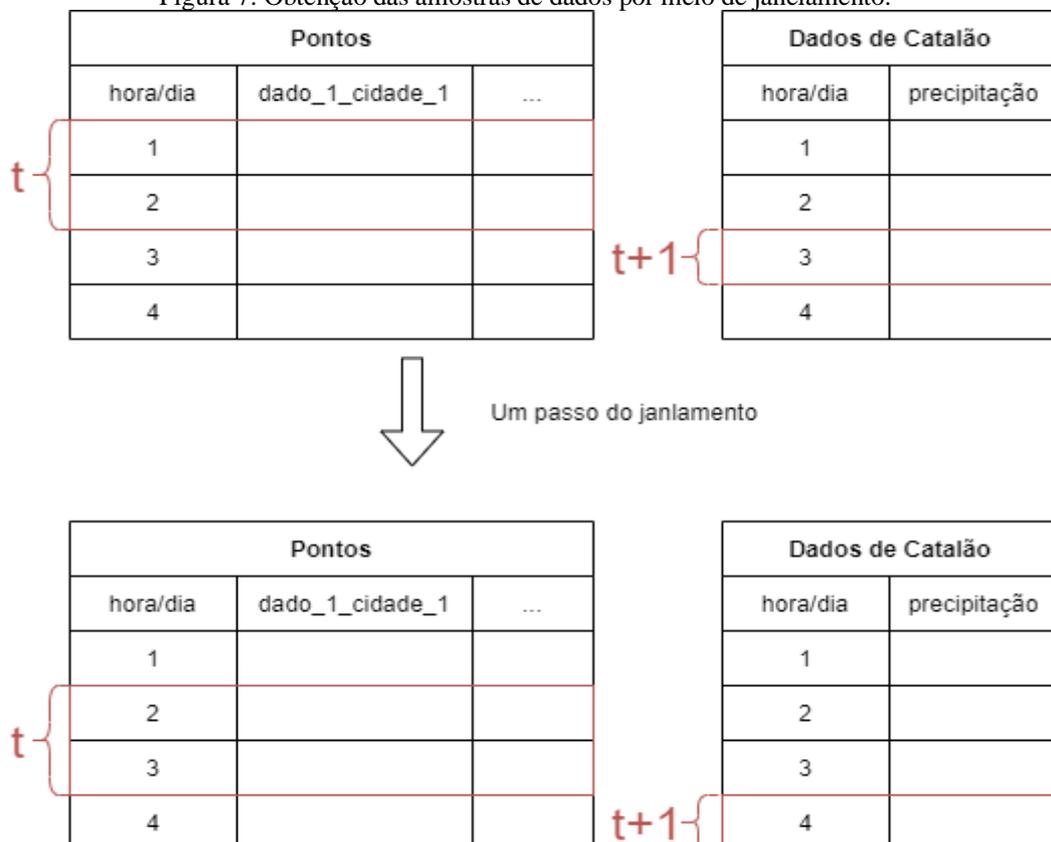
Uma camada LSTM tradicionalmente recebe os dados de entrada com três dimensões, sendo elas [exemplo, marcação de tempo, variáveis] (Figura 6). Para isso as amostras de dados resultantes do pré-processamento e normalização que estão nas dimensões [exemplo, variáveis] precisam ser remodeladas para a adição da dimensão do tempo, que é definida com base no janelamento usado.

Figura 6: Formato dos dados de entrada para a rede LSTM.



A **Figura 7** ilustra o janelamento dos dados, no qual é escolhido um tamanho t da janela de tempo, e partindo do início da tabela de dados normalizada, as amostras são extraídas; a precipitação no tempo dia/hora($t+1$) da cidade de Catalão é adicionada como saída da amostra. A janela é deslizada de uma em uma unidade ao longo da tabela de dados até $p - (t+1)$, onde p é o índice da última linha da tabela de dados.

Figura 7: Obtenção das amostras de dados por meio de janelamento.



3.1.5 Separação treino e teste

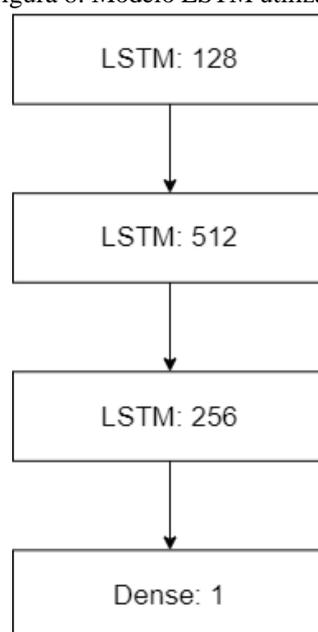
Os dados são então separados aleatoriamente em treino e teste, sendo que 80% para treino e 20% para teste. No caso da tarefa de classificação os dados são ainda estratificados para que ambos dados de treino e teste possuam a mesma proporção de amostras para cada classe.

3.2 MÉTODOS PROPOSTOS PARA PREDIÇÃO DE PRECIPITAÇÃO

Para realização das previsões foi utilizado uma rede LSTM de múltiplas camadas, representada pela Figura 8, sendo três camadas LSTM, de 128, 512 e 256 neurônios, respectivamente, e uma camada com um neurônio tradicional completamente conectado, para gerar as previsões no formato desejado. Cada camada LSTM é sucedida por uma camada de Dropout para remover aleatoriamente 20% das unidades para prevenir o ‘overfitting’. Foi utilizada a função ‘rmsprop’ como otimizador, a função de erro médio quadrático como função de ‘loss’, a quantidade de épocas em 200 e tamanho do mini lote 128. Para evitar excesso de treinamento, a rede foi configurada para parar o treinamento quando o valor de ‘loss’ não diminuir durante três épocas.

No caso da classificação a saída da rede continua sendo um valor decimal, portanto após ser realizada a previsão os valores são mapeados com uma função de arredondamento para o inteiro mais próximo assim correspondendo às classes pré definidas.

Figura 8: Modelo LSTM utilizado



4 RESULTADOS

Para a tarefa de classificação foram utilizadas as métricas acurácia global (que mede a taxa de classificação correta) e *precision*, *recall*, aplicados por classe. Precision (Eq. 3) mede a proporção das instâncias preditas em uma dada classe que realmente pertencem àquela classe. Ex.: a proporção das predições de chuva forte em que realmente ocorreu chuva forte. Na Eq. 3, TP mede a quantidade de amostras da classe que foram preditas corretamente e FP mede a quantidade de amostras preditas erroneamente como pertencente a classe (também chamado de número de falsos positivos)

$$precision = \frac{TP}{TP + FP} \quad (3)$$

Recall (Eq. 4) mede a proporção das instância de uma classe que foram preditas naquela classe. Ex.: a proporção das chuvas fortes que foram preditas como chuvas fortes. Na Eq. 4, TP mede a quantidade de amostras da classe que foram preditas corretamente e FN mede a quantidade da classe que não foram detectadas (ou seja, que pertencem a classe mas não foram preditas corretamente).

$$recall = \frac{TP}{TP + FN} \quad (4)$$

A acurácia (Eq. 5) mede a proporção das instâncias em que a classe foi predita corretamente.

$$acurácia = \frac{\text{quantidade de amostras classificações corretamente}}{\text{quantidade de amostras classificadas}} \quad (5)$$

Já para a tarefa de regressão foi utilizado o erro médio absoluto como métrica de desempenho e para comparação entre as tarefas de classificação e regressão, que possuem métricas diferentes, foi utilizado um método para definir a acurácia da predição por regressão. Este método considera uma variação positiva e negativa para definir se a predição foi correta ou não, e dessa forma se torna possível a comparação entre os dois tipos de predição, por regressão e classificação.

A tarefa de classificação utilizando os dados divididos em horas obteve uma acurácia de 98%, sua matriz de confusão está representada na Figura 9, e as métricas precision e recall na tabela 1. Já a tarefa de classificação utilizando os dados divididos em

dias obteve uma acurácia de 89%, sua matriz de confusão está representada na Figura 10, e as métricas precision e recall na tabela 2.

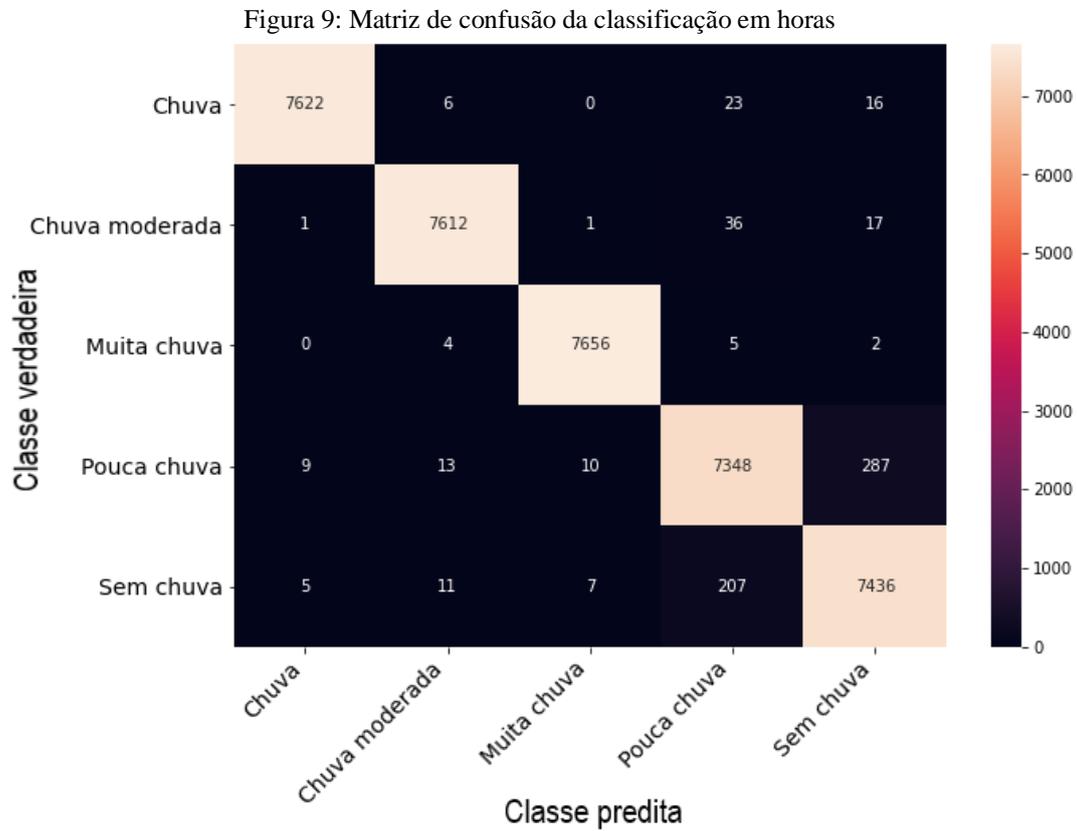


Tabela 1: Precision e Recall para tarefa de classificação em horas

Classe	Precision	Recall
Sem chuva	0.96	0.97
Pouca chuva	0.96	0.96
Chuva	1.00	0.99
Chuva moderada	1.00	0.99
Muita chuva	1.00	1.00

Figura 10: Matriz de confusão da classificação em Dias.

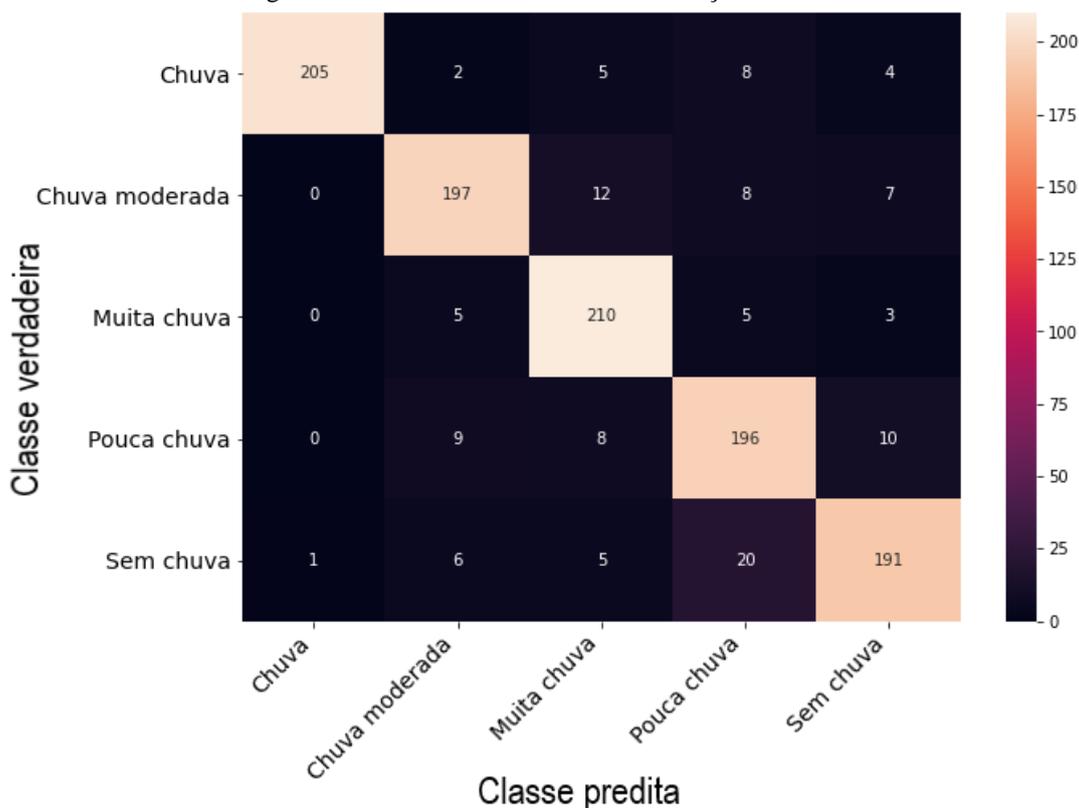


Tabela 2: Precision e Recall para tarefa de classificação em dias.

Classe	Precision	Recall
Sem chuva	0.89	0.86
Pouca chuva	0.83	0.88
Chuva	1.00	0.92
Chuva moderada	0.90	0.88
Muita chuva	0.88	0.94

Para a tarefa de regressão por hora o erro absoluto médio foi de 0.18 e acurácia de 98% com taxa de variação de 2mm, já a tarefa de regressão por dia o erro absoluto médio foi de 5,07 e acurácia de 67% com taxa de variação de 2mm. Devido à baixa acurácia, foi realizado um experimento com a tarefa de regressão com dados divididos em dias

utilizando uma variação maior de 6mm, porém a acurácia aumentou somente para 74,7%, portanto esse experimento não será levado em consideração.

5 CONCLUSÕES E TRABALHOS FUTUROS

Tanto a tarefa de classificação como a de regressão obtiveram melhores resultados utilizando dados em periodicidade horária em comparação com o uso de dados em periodicidade diária. Vale ressaltar que para um intervalo fixo de tempo, com os dados em periodicidade horária tem-se em uma maior quantidade de dados, e conseqüentemente amostras, do que se coletasse os dados com periodicidade diária. Também, é praticamente um consenso de que as redes neurais, entre elas a LSTM usada neste estudo, tem melhor desempenho de predição quando treinadas com uma grande quantidade de amostras de dados. Dados meteorológicos de múltiplos pontos com periodicidade diária têm sido viabilizados pelas estações meteorológicas automáticas, que tendem a substituir as estações não-automáticas. Utilizando dados em periodicidade horária obteve-se 98% de acurácia tanto para a tarefa de classificação quanto para a tarefa de regressão (na regressão a acurácia foi medida considerando uma variação de ± 2 mm). É importante salientar que a tarefa de regressão entrega uma predição mais informativa, que é a estimativa da quantidade de chuva (em mm), em uma determinada hora.

Ao comparar as tarefas de classificação e regressão utilizando os dados divididos em dias, a diferença de acurácia foi grande, tendo melhores resultados para a tarefa de classificação. Presume-se portanto que a tarefa de classificação consegue aprender com uma quantidade menor de dados e que portanto se torna uma escolha mais interessante para casos onde a quantidade de dados disponíveis seja limitada.

AGRADECIMENTOS

Os autores agradecem ao Instituto Nacional de Meteorologia (INMET) pela disponibilização dos dados usados nesta pesquisa e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro, Processo (431552/2016-9).

REFERÊNCIAS

ABREU, J. P. M. **Agrometeorologia. Aplicação da Meteorologia para Maximizar a Produção Agrícola**. Editora Agrobook, 2018.

AGGARWAL, C. C. **An Introduction to Neural Networks**. Cham: Springer International Publishing, 2018. 1–52 p. ISBN 978-3-319-94463-0. Disponível em: <https://doi.org/10.1007/978-3-319-94463-0_1>. Acesso em: 14 de Março de 2021.

CHATFIELD, C. (2004). **The analysis of time series: an introduction**. Florida, US, 6th edition.

DAS, M.; GHOSH, S. K. **Data-driven approaches for meteorological time series prediction: A comparative study of the state-of-the-art computational intelligence techniques**. Pattern Recognition Letters, v. 105, p. 155 – 164, 2018.

DIKSHIT, A.; PRADHAN, B.; ALAMRI, A. M. **Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model**. Science of The Total Environment, v. 755, p. 142638, 2021.

DUAN, J.; WANG, P.; MA, W.; TIAN, X.; FANG, S.; CHENG, Y.; CHANG, Y.; LIU, H. **Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and correntropy long short -term memory neural network**. Energy, v. 214, p.118980, 2021.

GOMES, A. C. A.; OLIVEIRA, M. C. F; FERREIRA, A. N. P. **Influência dos elementos meteorológicos na produção do feijão (Phaseolus vulgares L.) no município de Traquateua-PA**. XV Congresso Brasileiro de Agrometeorologia, Aracaju-SE, 2007.

HAYKIN, S. S. **Neural networks and learning machines**. Third. Upper Saddle River, NJ: Pearson Education, 2009.

HOCHREITER, S.; SCHMIDHUBER, J. Long-short term memory. Neural Computation, v. 9, n. 8, p. 1735–1780, 1997.

HEWAGE, P.; BEHERA, A.; TROVATI, M.; PEREIRA, E.; GHAREMANI, M.; PALMIERI, F.; LIU, Y. **Temporal convolutional neural (tcn) network for an effective weather forecasting using time-series data from the local weather station**. Soft Computing, v. 24, 2020

KE, J.; ZHENG, H.; YANG, H.; CHEN, X. M. **Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach**. Transportation Research, Part C: Emerging Technologies, v. 85, p. 591 – 608, 2017.

KIM, T.; SHARDA, S.; ZHOU, X.; PENDYALA, R. M. **A stepwise interpretable machine learning framework using linear regression (lr) and long short-term memory (lstm): City-wide demand-side prediction of yellow taxi and for-hire vehicle (fhv) service**. Transportation Research Part C: Emerging Technologies, v. 120, p. 102786, 2020.

LIU, H.; MI, X. wei; LI, Y. fei. **Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and elman neural network.** Energy Conversion and Management, v. 156, p. 498 – 514, 2018.

NATIONAL RESEARCH COUNCIL. **Urban Meteorology: Forecasting, Monitoring, and Meeting Users' Needs.** Washington, DC: The National Academies, 2012.

MA, X.; TAO, Z.; WANG, Y.; YU, H.; WANG, Y. **Long short-term memory neural network for traffic speed prediction using remote microwave sensor data.** Transportation Research Part C: Emerging Technologies, v. 54, p. 187 – 197, 2015.

MARCHEZAN, E. **Efeito de elementos meteorológicos na época de colheita sobre a quantidade de grãos inteiros de arroz.** Ciência Rural, vol 25, no. 2, Santa Maria, 1995.

MARRAS, S.; KELLY, J. F.; MORAGUES, M.; MÜLLER, A.; KOPERA, M. A.; VÁZQUEZ, M.; GIRALDO, F. X.; HOUZEAUX, G.; JORBA, O. **A review of element-based galerkin methods for numerical weather prediction: Finite elements, spectral elements, and discontinuous galerkin.** Archives of Computational Methods in Engineering, v. 23, n. 4, p. 673–722, 2016.

MONQUERO, P.A., AMARAL, L.R., BINHA, D.P., SILVA, A.C., & SILVA, P.V.. (2008). **Potencial de lixiviação de herbicidas no solo submetidos a diferentes simulações de precipitação.** Planta Daninha, 26(2), 403-409.

N. V. Chawla; K. W. Bowyer; L. O. Hall; W. P. Kegelmeyer. **SMOTE: Synthetic Minority Over-sampling Technique.** Journal of Artificial Intelligence Research, vol. 16 (2002), pág. 321–357.

OLAH, C. **Understanding LSTM Networks.** 2015. Disponível em: <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em: 10 de Março de 2021.

Qu Xiaoyun; Kang Xiaoning; Zhang Chao; Jiang Shuai; Ma Xiuda. **Short-term prediction of wind power based on deep long short-term memory.** In: 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC). [S.l.: s.n.], 2016. p. 1148–1152.

SENE, E. D.; MOREIRA, J. C. **Geografia geral e do Brasil : espaço geográfico e globalização.** 2th. ed. [S.l.]: scipione, 2013. v. 1. ISBN 978 85262 9128 7.

ZAYTAR, M. A.; AMRANI, C. E. **Sequence to sequence weather forecasting with long short-term memory recurrent neural networks.** International Journal of Computer Applications, v. 143, p. 7–11, 06 2016.

WERNER, L.; BISOGNIN, C; ARAÚJO, C. W. **Análise De Técnicas De Previsão: Um Estudo De Caso Para O Volume De Ações Da Petrobras/ Forecasting Techniques Analysis: A Case Study For Petrobras Stock Volume.** Brazilian Journal of Development. V. 16, No. 1 2020, p. 1103-1115. DOI: <<https://doi.org/10.34117/bjdv6n1-078>>.