



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS CRATEÚS
CURSO DE CIÊNCIA DA COMPUTAÇÃO

ÍCARO DE SENA MOURÃO

**PREDIÇÃO DE SÉRIES TEMPORAIS CLIMÁTICAS COM APRENDIZAGEM
PROFUNDA**

CRATEÚS
2019

ÍCARO DE SENA MOURÃO

PREDIÇÃO DE SÉRIES TEMPORAIS CLIMÁTICAS COM APRENDIZAGEM PROFUNDA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Orientador: Prof. Me. Lívio Antônio Melo Freire

Coorientador: Prof. Me. Marciel Barros Pereira

CRATEÚS

2019

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)

M891p Mourão, Ícaro de Sena.
Predição de Series Temporais Climáticas com Aprendizagem Profunda / Ícaro de Sena Mourão. – 2019.
60 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Crateús,
Curso de Ciência da Computação, Crateús, 2019.

Orientação: Prof. Me. Lívio Antônio Melho Freire.

Coorientação: Prof. Me. Marciel Pereira Barros.

1. Aprendizagem Profunda. 2. Redes Neurais. 3. Série Temporais. 4. Predição. I. Título.

CDD 004

ÍCARO DE SENA MOURÃO

PREDIÇÃO DE SÉRIES TEMPORAIS CLIMÁTICAS COM APRENDIZAGEM PROFUNDA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Aprovada em: 11 de Dezembro de 2019

BANCA EXAMINADORA

Prof. Me. Lívio Antônio Melo Freire (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Me. Marciel Barros Pereira (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Me. André Meireles de Andrade
Universidade Federal do Ceará (UFC)

Prof. Me. José Wellington Franco da Silva
Universidade Federal do Ceará (UFC)

À minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação foi que deram, em alguns momentos, a esperança para seguir. Pai, sua presença significou segurança e certeza de que não estou sozinho nessa caminhada.

AGRADECIMENTOS

Primeiramente a Deus que permitiu que tudo isso acontecesse ao logo da minha vida, e não somente nestes anos como universitário, mas que em todos os momentos é o maior mestre que alguém pode ter.

Aos meus pais, avós e irmãos, pelo amor, incentivo e apoio incondicional.

Ao meu tio por ter acreditado em mim, por não me deixar desistir e pelo seu apoio financeiro.

Ao Prof. Me. Lívio Antônio Melo Freire por me orientar em meu primeiro trabalho científico. À sua paciência para responder todas as minhas dúvidas, mesmo em horas difíceis e à sua tranquilidade. Por seus ensinamentos que foram essenciais para o desenvolvimento deste trabalho. Principalmente por acreditar em mim.

Ao Prof. Me. Marciel Barros Pereira, por me coorientar em meu primeiro trabalho científico. À sua atenção e paciência para me explicar minuciosamente os conteúdos.

Ao Profa. Me. Lisieux Marie Marinho dos Santos por me orientar na disciplina de Projeto de Pesquisa Científica e Tecnológica.

Ao Prof. Me. Rennan Ferreira Dantas por ter acreditado em mim. Por disponibilizar parte de seu preciosíssimo tempo para conversar comigo e mostrar que estudar sempre vale a pena.

À psicóloga Tamylle Kellen Arruda Prestes, por me ajudar num dos momentos mais difíceis de minha vida acadêmica. Por acreditar em mim e por ter me apresentado uma forma eficiente de escrita.

À Universidade Federal do Ceará, pela oportunidade concedida para realização da graduação.

Agradeço a todos os professores e servidores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender.

Ao Doutorando em Engenharia Elétrica, Ednardo Moreira Rodrigues, e seu assistente, Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

A todos que, direta ou indiretamente, fizeram parte da minha formação, o meu muito obrigado!

“Só há duas maneiras de viver a vida: a primeira é vivê-la como se os milagres não existissem. A segunda é vivê-la como se tudo fosse milagre.”

(Albert Einstein)

RESUMO

A capacidade de prever condições climáticas é fundamental para a tomada de decisão em diferentes áreas, como agricultura, pecuária, construção civil, transporte marítimo, aeronáutica e turismo, além de permitir compreender impactos da ação do homem na natureza. O avanço tecnológico, com o desenvolvimento de sensores, recursos computacionais e abordagens para compreender variáveis climáticas, vem aumentando a confiabilidade sobre previsões de modelos meteorológicos. De modo geral, para construir esses modelos, precisa-se de séries temporais climáticas, que são dados coletados sobre uma ou várias variáveis ao longo do tempo. A partir das séries temporais, os parâmetros do modelo são estimados, para produzir previsões sobre as variáveis com base em dados históricos. Os Modelos de Aprendizagem Profunda (AP) permitem a extração automática de características diretamente sobre dados, permitindo a construção de funções capazes de modelar padrões complexos. Diante disso, neste trabalho, propõe-se a aplicação de modelos de AP para construir preditores sobre dados climáticos. Para validar os resultados, os modelos serão submetidos a dados reais, cujas métricas serão utilizadas para validar a modelagem e a solução.

Palavras-chave: Aprendizagem Profunda. Redes Neurais. Séries Temporais. Predição

ABSTRACT

The ability to predict climatic conditions is fundamental for decision-making in different areas, such as agriculture, livestock, construction, maritime transport, aeronautics and tourism, as well as allowing the execution of human nature actions. Technological advancement, with the development of sensors, computational resources and approaches to understanding climate variations, comes with the application of predetermined weather models. In general, to build these models, you need climate time series, which is data collected about one or more variables over time. From time series, model models are estimated to produce predictions about variables based on historical data. Gls AP Models allow automatic extraction of resources directly from data, allowing the construction of functions capable of modeling patterns. Therefore, in this paper, it offers an application of gls AP models to build climate data editors. To validate the results, models can use real data, these measures are used to validate a modeling and a solution.

Keywords: Deep Learn. Neural Networks. Time series. Prediction

LISTA DE FIGURAS

Figura 1 – Representação gráfica da quantidade de passageiros aéreos ao longo do tempo.	19
Figura 2 – Representação gráfica da pontuação de três empresas na bolsa de valores ao longo do tempo.	19
Figura 3 – Temperatura Máxima Absoluta mensal observada em Lages, SC no período de 1961 a 2011.	21
Figura 4 – Modelo de Neurônio Artificial.	25
Figura 5 – Representação visual do Problema de Regressão (PR) e do Problema de Classificação (PC).	28
Figura 6 – Rede Neural Artificial Multicamada.	29
Figura 7 – Gradiente descendente estocástico.	31
Figura 8 – Rede Neural Recorrente.	33
Figura 9 – Diagrama de todo o processo de predição de série temporal.	34
Figura 10 – Arquitetura da Rede Neural Recorrente.	37
Figura 11 – Exemplo de validação cruzada.	39
Figura 12 – Gráfico que representa a base de dados da cidade de Crateús.	43
Figura 13 – Gráfico de dispersão entre a temperatura máxima real e sua predição para a cidade de Crateús.	44
Figura 14 – Gráfico de dispersão entre a temperatura mínima real e sua predição para a cidade de Crateús.	45
Figura 15 – Representação gráfica do desempenho do modelo quando aplicado a diferentes tipos de normalização de dados para a Cidade de Crateús/CE para a temperatura máxima.	46
Figura 16 – Representação gráfica do desempenho do modelo quando aplicado a diferentes tipos de normalização de dados para a Cidade de Crateús/CE para a temperatura mínima.	46
Figura 17 – Gráfico que representa os erros de predições de temperatura máxima do modelo para a Cidade de Crateús/CE.	47
Figura 18 – Gráfico que representa os erros de predições de temperatura mínima do modelo para a Cidade de Crateús/CE.	47
Figura 19 – Gráfico que representa a base de dados da cidade de Maceió/AL.	53
Figura 20 – Gráfico que representa a base de dados da cidade de Campos Sales/CE.	53

Figura 21 – Gráfico que representa a base de dados da cidade de Turiçu/MA.	54
Figura 22 – Gráfico que representa a base de dados da cidade de Belém/PA.	54
Figura 23 – Gráfico que representa a base de dados da cidade de Monte Alegre/PA. . . .	55
Figura 24 – Gráfico que representa os acertos de predições de temperatura máxima do modelo para a Cidade de Maceió/AL.	56
Figura 25 – Gráfico que representa os acertos de predições de temperatura mínima do modelo para a Cidade de Maceió/AL.	56
Figura 26 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de Campos Sales/CE.	57
Figura 27 – Gráfico que representa a precisão de predições de temperatura mínima do modelo para a Cidade de Campos Sales/CE.	57
Figura 28 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de Turiaçu/MA.	58
Figura 29 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de Turiaçu/MA.	58
Figura 30 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de Bélem/PA.	59
Figura 31 – Gráfico que representa a precisão de predições de temperatura mínima do modelo para a Cidade de Maceió/Alagoas.	59
Figura 32 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de Monte Alegre/PA.	60
Figura 33 – Gráfico que representa a precisão de predições de temperatura mínima do modelo para a Cidade de Monte Alegre/PA.	60

LISTA DE TABELAS

Tabela 1 – Dados da estação automática de Crateús/CE no formato CSV.	23
Tabela 2 – Dados da estação automática de Crateús/CE.	23
Tabela 3 – Métricas de resultado do treinamento da cidade de Maceió no estado de Alagoas para as temperaturas mínimas.	48
Tabela 4 – Métricas de resultado do treinamento da cidade de Maceió no estado de Alagoas para as temperaturas mínimas.	48

LISTA DE ABREVIATURAS E SIGLAS

AgriTempo	Sistema de Monitoramento Agrometeorológico
AP	Aprendizagem Profunda
CPTEC	Centro de Previsão de Tempo e Estudos Climáticos
CSV	<i>Comma Separated Values</i>
CV	Validação Cruzada
INMET	Instituto Nacional de Meteorologia
LSTM	<i>Long Short Term Memory</i>
MAE	Erro Absoluto Médio
MCG	Modelos Numéricos de Circulação Geral
MLP	<i>Multilayer Perceptron</i>
MSE	Erro Quadrático Médio
NCV	Validação Cruzada Aninhada
NOAA	<i>National Oceanic and Atmospheric Administration</i>
PC	Problema de Classificação
PR	Problema de Regressão
RNA	Rede Neural Artificial
RNR	Rede Neural Recorrente

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Contextualização	16
1.2	Justificativa	16
1.3	Objetivo Geral	17
1.4	Objetivos Específicos	17
2	SÉRIES TEMPORAIS E PREDIÇÃO	18
2.1	Séries Temporais	18
2.2	Séries Temporais Climáticas	21
2.3	Coleta dos Dados	22
3	REDES NEURAIS ARTIFICIAIS	24
3.1	Neurônio Artificial	24
3.2	Funções de Ativação	26
3.3	Problemas de Regressão e Classificação	27
3.4	Redes Neurais Multicamadas	28
3.5	Treinamento e Avaliação	29
3.6	Redes Neurais Recorrentes	32
4	MODELO PARA PREDIÇÃO DE SÉRIES TEMPORAIS	34
4.1	Pré-processamento de Dados	34
4.2	Modelo de Aprendizagem Profunda	36
4.3	Experimentos e Validação	38
5	TRABALHOS RELACIONADOS	40
5.1	<i>Deep forecast: Deep learning-based spatio-temporal forecasting</i>	40
5.2	<i>A deep hybrid model for weather forecasting</i>	40
6	RESULTADOS	42
6.1	Cenário	42
6.2	Análise de Comportamento	43
6.3	Análise de normalização	44
6.4	Análise de precisão	45
7	CONCLUSÕES E TRABALHOS FUTUROS	49
	REFERÊNCIAS	50

APÊNDICES	53
APÊNDICE A – Representação gráfica da bases de dados utilizadas . .	53
APÊNDICE B – Resultados de precisão das predições	56
ANEXOS	56

1 INTRODUÇÃO

1.1 Contextualização

Uma ferramenta essencial no desenvolvimento econômico do país, a previsão de tempo vem ganhando uma nova e inédita dimensão. O motivo principal é o grande avanço tecnológico. A tecnologia é uma grande aliada da meteorologia, ajudando a entender as variáveis do clima, ao longo do tempo. De acordo com Santos e Costa (2013) os instrumentos utilizados pelos meteorologistas foram se desenvolvendo e, com eles, a precisão das previsões do tempo foi melhorada substancialmente.

A inferência e predição de clima sempre esteve presente nas atividades do homem. A capacidade de prever o tempo é essencial para a tomada de decisão em diferentes áreas, como agricultura, pecuária, construção civil, transporte marítimo, aeronáutica e turismo, além de permitir compreender impactos da ação do homem na natureza (FOLHES; FISCH, 2006). De forma geral, a predição do clima busca auxiliar o homem nas mais variadas atividades que compõem seu dia a dia.

Segundo ESCOBAR (2007) ainda existem limitações de processamento computacional pelo fato dos modelos numéricos clássicos ou Modelos Numéricos de Circulação Geral (MCG) serem muito complexos. Os MCG's são equações matemáticas representativas das leis físicas que regem os movimentos da atmosfera e as interações com os componentes do sistema climático (SAMPAIO; DIAS, 2014). Segundo Sampaio e Dias (2014) as equações do MCG são resolvidas por métodos numéricos e necessitam computadores de altíssimo desempenho. Pensando nas limitações e no modo como este problema é abordado mais frequentemente é que levantamos um questionamento: será que não existe algum método que seja mais eficiente e um tanto menos complexo?

1.2 Justificativa

Quaisquer modelos de predição climática são “alimentados” com informações meteorológicas (vento, temperatura, pressão, umidade, etc.), vindas de toda a Terra, provenientes de estações meteorológicas, navios, boias oceânicas, aviões, balões atmosféricos ou dados de satélites. Os dados coletados a partir de então são tratados como uma série temporal. Segundo Wooldridge (2000), uma série temporal consiste em observações sobre uma variável ou várias

variáveis ao longo do tempo. Dessa forma, séries temporais climáticas são dados temporais coletados com respeito ao clima.

Depois dos dados coletados pode-se pensar na forma como vamos processá-los, para extrair algum resultado útil para o trabalho. A primeira e principal ferramenta utilizada para as previsões de tempo e de clima são os modelos numéricos conhecidos como MCG, que são constituídos por um conjunto de equações físicas descritas em forma numérica e resolvidas com o auxílio de computadores (NOAA, 2017).

Nesse contexto, este trabalho teve como principal contribuição aplicar modelos de AP para construir preditores sobre dados climáticos. Embora já existam diversos métodos de predição do clima, segundo Yan-Tak (2015), a performance da AP vem superando abordagens clássicas para o problema, sejam elas estatísticas ou baseadas em aprendizagem. Por esse motivo, pretende-se utilizar um método de predição do clima baseado em AP.

1.3 Objetivo Geral

Treinar modelos de Aprendizagem Profunda a partir de dados coletados para gerar preditores climáticos.

1.4 Objetivos Específicos

- Realizar coleta de dados públicos sobre indicadores climáticos;
- Treinar modelos de Aprendizagem Profunda;
- Realizar previsões climáticas a partir dos modelos treinados a partir dos dados coletados;

2 SÉRIES TEMPORAIS E PREDIÇÃO

Este capítulo abordará conceitos gerais para o estudo das séries temporais. Na seção 2.1, são apresentadas definições e os tipos de séries temporais. Na seção 2.2 é definida a série temporal que utiliza dados climáticos. E por fim na seção 2.3 é apresentado algumas fontes de coletas de dados e o tipo de dados que são encontrados nessas fontes.

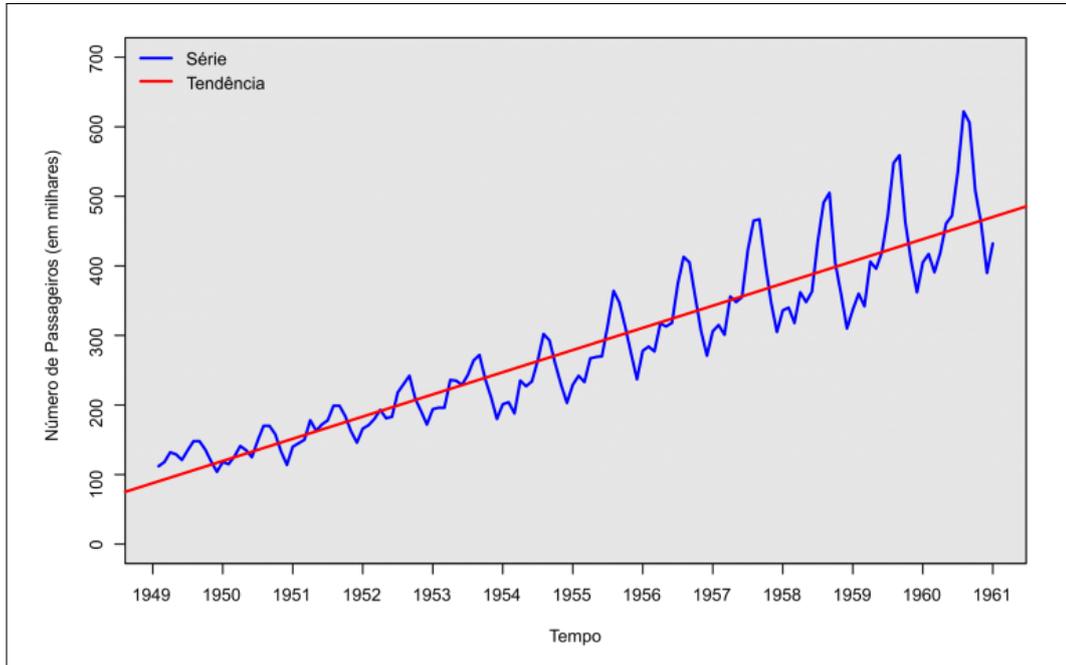
2.1 Séries Temporais

Como mencionado no Capítulo 1, uma série temporal é um conjunto sequencial de pontos de dados (WOOLDRIDGE, 2000). Sendo assim, constituída de observações por uma ou várias variáveis ao longo do tempo, organizados em uma ordem cronológica adequada. Observe a Figura 1 que denota a quantidade de passageiros aéreos ao longo do tempo. Veja que de 1949 até 1961 o número de passageiros cresceu substancialmente. Perceba também que há um tendência de alta, pois por algum motivo as pessoas tendem a confiar em transportes aéreos. Portanto a tendência já é um bom indicio para você apostar que de 1961 em diante a quantidade de passageiros irão aumentar. Tendência, portanto, passa a ser uma característica de uma série temporal. Para que uma série seja necessariamente uma série temporal algumas outras características ainda devem ser levado em consideração, como por exemplo se ela tem sazonalidade. Essas características serão abordadas ainda nesse Capítulo.

A partir da definição de série temporal é importante observar dois conceitos: séries temporais que possuem uma única variável de observação ou univariada (Figura 1) e séries temporais que possuem mais de uma variável de observação ou multivariada (Figura 2). Como no mundo real o conjunto de dados terá muitas observações, essas observação podem ser vistas como vizinhas umas das outras criando, assim, uma dependência entre elas.

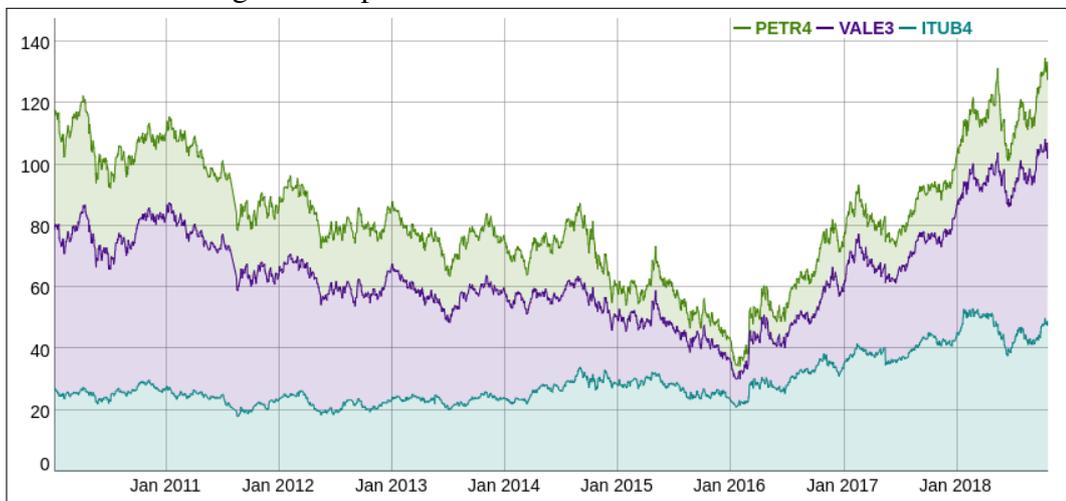
Nota-se que os termos univariada e multivariada referem-se apenas às variáveis observadas. Uma série ainda pode ser continua ou discreta, assim como suas variáveis. Ela é dita continua quando as observações são feitas continuamente no tempo. Por exemplo, leituras de temperatura ou o fluxo de um rio. Formalmente por um certo intervalo de tempo, $T = \{t : t_1 < t < t_2\}$, a série temporal será denotada por $\{X(t) : t \in T\}$, onde X é a série temporal. Ela é dita ser discreta quando as observações são feitas em tempos específicos, geralmente equidistantes. Por exemplo, a população de uma determinada cidade, a produção de uma empresa ou a taxa de câmbio entra duas moedas. Definindo o conjunto $T = \{t_1, \dots, t_n\}$, a série temporal será denotada

Figura 1 – Representação gráfica da quantidade de passageiros aéreos ao longo do tempo.



Fonte: Gonçalves (2018).

Figura 2 – Representação gráfica da pontuação de três empresas na bolsa de valores ao longo do tempo.



Fonte: Costa (2018).

por $\{X_t : t \in T\}$. Por simplicidade, defini-se $T = \{1, 2, \dots, n\}$ (EHLERS, 2012).

A predição com base em séries temporais presume que o histórico de uma variável ao longo do tempo até o presente momento é capaz de auxiliar na predição de um comportamento futuro, aproximadamente do mesmo modo. Existem quatro fatores principais que a previsão com base em séries temporais busca identificar no intuito de realizar previsões, os chamados componentes de uma série temporal. A tendência, o efeito cíclico, a sazonalidade e os efeitos aleatórios.

- **Tendência:** uma tendência é um movimento de longo prazo, normalmente ascendente ou descendente. De acordo com Gonçalves (2018) existem testes específicos para a identificação da tendência, como o Teste de Wald e o de Cox-Stuart. Porém uma das técnicas mais utilizadas é o ajuste de uma Regressão Linear Simples. Na Figura 1, os dados são referentes ao número de passageiros internacionais no transporte aéreo entre os anos de 1949 e 1961. É perceptível uma clara tendência crescente na série ao longo dos anos. Mas tendência não é o único fator que pode influenciar na série.
- **Efeito cíclico:** o efeito cíclico descreve as mudanças de médio prazo na série, causadas por circunstâncias que se repetem em ciclos. Geralmente os efeitos cíclicos em uma série são causados por mudanças na demanda do produto, por ciclos de negócios ou pela inabilidade de se suprir as necessidades do consumidor, por exemplo (MAXWELL, 2018). Pode-se dizer que o efeito cíclico é a repetição de uma tendência ao longo de um médio espaço de tempo.
- **Sazonalidade:** a sazonalidade é uma flutuação dentro de um período específico de tempo que se repete em específicas épocas do ano. Por exemplo, o número de passageiros que utilizam o transporte aéreo geralmente é maior em períodos de férias do que nos demais meses do ano, fato que pode ser verificado na Figura 1. A sazonalidade, muitas vezes, pode ser identificada de maneira visual.
- **Efeito aleatório:** quaisquer dados que não sigam um padrão específico ou que não são regulares são considerados parte das variações irregulares ou efeito aleatório. Apresenta movimentos ascendentes e descendentes da série após a ocorrência de um efeito de tendência, um efeito cíclico, ou de um efeito sazonal. Nas componentes de efeito aleatório aparecem flutuações de período curto, com deslocamento inexplicável e geralmente são causadas, entre outros motivos, por eventos políticos e oscilações climáticas imprevisíveis, por exemplo, o presidente decide extinguir o horário de verão ou um furacão atinge uma determinada região.

Os componentes de uma série temporal podem ser determinísticos e/ou estocásticos. Quando os valores da série podem ser escritos por meio de uma função matemática (um polinômio, por exemplo), diz-se que ela contém apenas componentes determinísticos. Além dos componentes normais de uma função matemática perfeitamente determinada, a representação de uma série temporal pode incluir um componente aleatório, que deverá ser gerado por um processo estocástico. Ser estocástico significa que o estado da variável é indeterminado, com

origem em eventos aleatórios (BUENO, 2008).

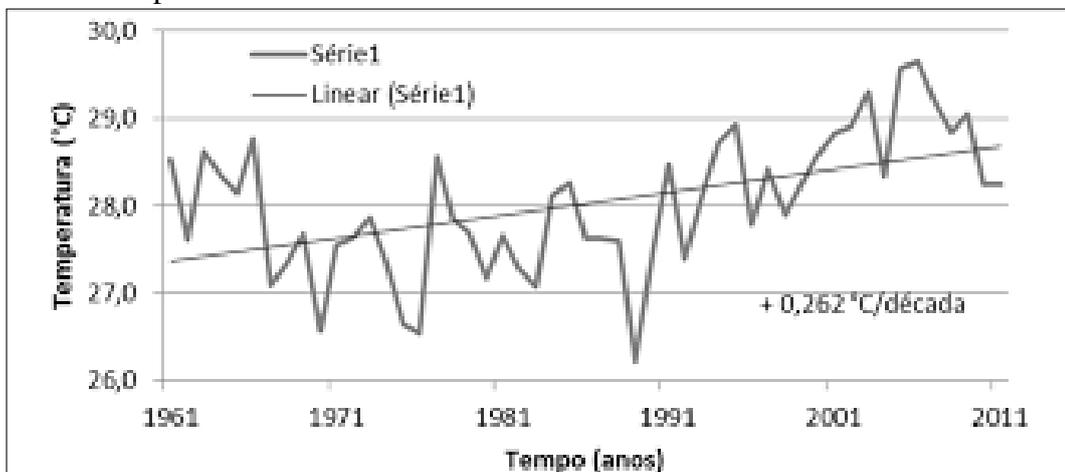
Quanto à variabilidade das observações, podem-se classificar as séries temporais em *estacionárias*, quando as suas estatísticas não são afetadas por variações no tempo e *ergódicas*, se apenas uma realização do processo estocástico é suficiente para se obter todas as estatísticas do mesmo (MORETTIN; TOLOI, 2006). Uma vez série temporal definida é importante ainda defini-la em termos de algum parâmetro conforme será feito na Seção 2.2, a seguir.

2.2 Séries Temporais Climáticas

Séries temporais podem ser definidas pelos seus tipos de dados, por exemplo, existem séries temporais econômicas, que utilizam dados financeiros (GONÇALVES, 2017); séries temporais clínicas (SATO, 2013), etc. Este trabalho utilizará séries temporais climáticas, as quais utilizam dados com respeito ao clima Figura 3. Séries temporais climáticas servem para realizar um diagnóstico sobre os comportamentos climáticos.

É importante que se definam as variáveis que serão utilizadas e sua terminologia, isto é, se ela é contínua ou discreta. É comum que dados climáticos sejam representados, dentre outras, na forma de *temperatura*, *pressão atmosférica* e *velocidade do vento*. E que além disso a observação destas sejam feitas em tempos específicos, ou seja equidistantes, o que caracteriza uma série discreta. As três variáveis anteriormente citadas serão utilizadas para predizer o clima. Por essa série possuir mais de uma variável de observação, ela recebe o nome de multivariada, como visto na seção anterior.

Figura 3 – Temperatura Máxima Absoluta mensal observada em Lages, SC no período de 1961 a 2011.



Fonte: Costa *et al.* (2015).

Por se tratar de uma série temporal, a série apresentada na Figura 3 também possui os quatro fatores principais apresentados anteriormente. A tendência, que pode ser observada de acordo com a variável temperatura. Com o aquecimento global a temperatura se mostra em tendência ascendente, pois ela tende a aumentar a medida que o tempo passa (CASAGRANDE *et al.*, 2011). Segundo Parreira (2011), o clima na Terra tem se comportado de forma cíclica, com períodos frios e períodos quentes, isso caracteriza o efeito cíclico climático. A sazonalidade é caracteriza pelas estações do ano, por exemplo. Já a variação irregular de uma série temporal climática pode ser observada quando qualquer alteração climática acontece sem uma explicação ou padrão. Onde e como pode-se extrair os dados que caracterizam essa série? É isso que será abordado na próxima seção.

2.3 Coleta dos Dados

As bases de dados meteorológicas são construídas para receber, armazenar, processar e disponibilizar os dados das variáveis meteorológicas, tais como temperatura, velocidade do vento, umidade do ar, etc (VIANNA; MASSIGNAN,). Esses dados são gerados a partir de sensores que estão presente nas estações meteorológicas. Instituições especializadas em previsão do tempo e climatologia no Brasil, como o Instituto Nacional de Meteorologia (INMET), o Centro de Previsão de Tempo e Estudos Climáticos Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) e o Sistema de Monitoramento Agrometeorológico Sistema de Monitoramento Agrometeorológico (Agritempo) disponibilizam os dados publicamente em seus respectivos sítios.

O INMET disponibiliza gráficos, tabelas e arquivos com dados climáticos, os quais podem ser gerados da forma que melhor atenda a necessidade do problema. No INMET encontra-se dados com respeito a diversas variáveis que podem ser gerados ano a ano por quem quiser. Os dados são gerados automaticamente de hora em hora pelas estações meteorológicas. Portanto encontra-se 24 pontos de dados com respeito à temperatura, pressão, velocidade do vento, etc, por dia. Os quais podem ser consultados em um intervalo máximo de um ano. O Agritempo oferece gráficos, estatísticas, mapas e tabelas que representam dados meteorológicos relacionadas ao meio rural. O CPTEC disponibiliza, assim como os outros, dados meteorológicos nos mesmos formatos. Serão utilizados, pelo menos, 5 anos de dados para treinamento do modelo.

Um dos formato de dados que merece destaque é o formato *Comma Separated Values* (CSV) (em português *Valores separados por Vírgula*), pois a maioria das organizações

disponibilizam seus dados nesse formato. O CSV pode ser descrito da seguinte forma. Cada linha possui diversos campos, com uma vírgula indicando onde um campo termina e o outro começa (Tabela 2). Esse formato pode ser visualizado mais facilmente na forma de uma tabela, isto é, são representados por uma matriz, em que suas linhas podem representar o momento em que os dados foram observados e as colunas podem representar as variáveis ou vice-versa (Tabela 1). Desta forma, esse formato é adequado para descrever séries temporais, pois facilita o processamento e análise, etapa anterior à construção do modelo preditivo. Para este trabalho, dados serão coletados das fontes mencionadas.

Tabela 1 – Dados da estação automática de Crateús/CE no formato CSV.

codigoEstacao,	data,	hora,	tempInst,	tempMax,	tempMin,
A342,	22/05/2019,	23,	25.1,	26.1,	25.0,
A342,	22/05/2019,	22,	26.1,	27.8,	26.1,
A342,	22/05/2019,	21,	27.6,	29.2,	27.6,
A342,	22/05/2019,	20,	29.2,	30.6,	29.2,
A342,	22/05/2019,	19,	30.5,	30.8,	29.8,
A342,	22/05/2019,	18,	30.6,	31.9,	29.9,
A342,	22/05/2019,	17,	30.8,	32.1,	29.9,

Fonte: INMET (2019).

Tabela 2 – Dados da estação automática de Crateús/CE.

Data	Hora UTC	Temperatura (oC)			Umidade (%)			Vento (m/s)			Chuva (kJ/m2)
		Inst.	Máx.	Mín.	Inst.	Máx.	Mín.	Vel.	Dir. (o)	Raj.	
22/05/2019	00	25.4	26.0	25.3	85	85	80	0.9	168	3.6	0.0
22/05/2019	01	25.3	25.4	25.1	84	86	84	1.1	203	2.3	0.0
22/05/2019	02	25.0	25.3	24.7	87	89	84	0.8	284	2.4	0.0
22/05/2019	03	25.2	25.3	24.9	85	87	85	1.0	357	2.2	0.0
22/05/2019	04	24.5	25.3	24.5	95	95	85	0.7	271	2.2	0.2

Fonte: INMET (2019).

A coleta será feita considerando todos as características dos conjuntos de dados de cada uma das organizações. Serão utilizados dados coletados de acordo com a necessidade. Após esta etapa, os arquivos serão pré-processados, a fim de obter apenas as informações necessárias para a definição dos parâmetros do modelo preditivo em rede neural, a qual será apresentada no Capítulo 3.

3 REDES NEURAIS ARTIFICIAIS

Este capítulo abordará conceitos importantes para o estudo de redes neurais como um todo, em todas as seções. Cada seção aborda um ponto específico de uma rede neural, bem como algumas de suas variações. Na Seção, 3.1 será abordado os conceitos sobre neurônio artificial. Na Seção 3.2, serão tratadas as principais funções de ativações. Na Seção 3.4 será abordado uma variação da Rede Neural Artificial (RNA). Na Seção 3.5 apresenta-se toda a parte de treinamento e avaliação de um RNA. E por fim, na Seção 3.6 será apresentada outra variação de RNA, a Rede Neural Recorrente.

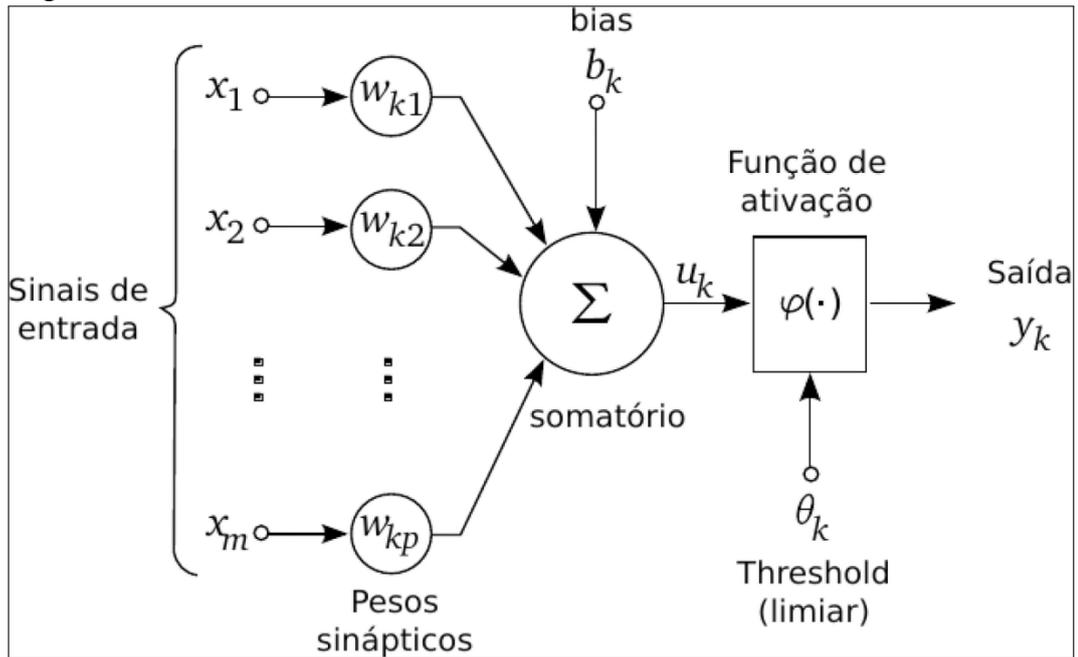
3.1 Neurônio Artificial

Com o trabalho pioneiro de Santiago e Cajal (1911), foi possível entender mais facilmente o cérebro humano. Eles introduziram a ideia dos neurônios como estruturas do cérebro. Estima-se que existam aproximadamente 10 bilhões de neurônios no córtex humano e 60 trilhões de conexões entre eles através de sinapses, formando, juntos, uma grande rede, chamada rede neural biológica (SHEPHERD; KOCH, 1990). As RNA são sistemas que se baseiam na rede neural biológica, contudo, possuindo um conjunto muito limitado de neurônios. De acordo com Zambiasi (2011), o neurônio artificial é um modelo simplificado e simulado do neurônio real e suas características básicas são a adaptação e a representação de conhecimentos baseada em conexões.

Um neurônio é uma unidade de processamento de informação e é fundamental para a operação de uma rede neural. A Figura 4 representa um modelo de um neurônio artificial. Nela, pode se identificar três elementos básicos de um neurônio:

1. Entrada, que recebem sinais e possuem pesos. O peso representa o quão significativo determinado objeto é para o neurônio. Um sinal x_j , na entrada j conectada ao neurônio k , é multiplicado pelo peso w_{kj} . É importante notar que o primeiro índice do peso w_{kj} , refere-se ao neurônio em questão e o segundo refere-se ao terminal de entrada. Ao contrário de um neurônio biológico, o neurônio artificial pode conter tanto pesos negativos quanto positivos.
2. Um somador: que possui a funcionalidade de somar os sinais de entrada, ponderados pelos seus respectivos pesos. Essa operação é basicamente uma combinação linear.
3. Função de ativação: para limitar a amplitude de saída de um neurônio. Tipicamente, o

Figura 4 – Modelo de Neurônio Artificial.



Fonte: Zambiasi (2011).

intervalo de saída de um neurônio é escrito como um intervalo unitário fechado $[0,1]$ ou $[-1,1]$.

O modelo da Figura 4, também inclui um *bias*, aplicado externamente, representado por b_k . Ele tem o efeito de aumentar ou diminuir a entrada da função de ativação. Pode-se descrever um neurônio artificial k da seguinte maneira:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (3.1)$$

e

$$y_k = \varphi(u_k + b_k) \quad (3.2)$$

onde x_1, x_2, \dots, x_m são os dados de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos do neurônio k ; u_k é a saída da combinação linear de acordo com os dados de entrada; b_k é o *bias*, já falado anteriormente; φ é a função de ativação; e y_k é a saída do neurônio. O *bias* é somado à saída u_k para que assim se obtenha o processamento da informação pelo neurônio, conforme a seguir:

$$v_k = u_k + b_k \quad (3.3)$$

Essa é a abstração mais usual de um neurônio biológico. A depender da função não linear utilizada definem-se vários tipos de neurônios artificiais, como por exemplo, o linear, sigmoide, *rectifier* linear ou *relu*, ou limiar, demonstrados na Seção 3.2.

3.2 Funções de Ativação

A função de ativação, representada anteriormente por $\varphi(v)$, define a magnitude de saída de um neurônio em termos do valor v . Vale ressaltar que as funções de ativação limitam o sinal propagado pelo neurônio. Existem diversos tipos de funções de ativação. Segundo Haykin (2001), existem três tipos básicos de funções de ativação:

1. **Função linear:** onde a saída do neurônio é igual a zero se seu valor for negativo e 1 se seu valor for positivo. Sua definição pode ser escrita da seguinte forma:

$$\varphi(v) = \begin{cases} 0 & \text{se } v \geq 0 \\ 1 & \text{se } v < 0 \end{cases} \quad (3.4)$$

2. **Função linear por partes:** esse tipo de função pode ser visto como uma aproximação de uma operação não linear. Que pode ser definida como:

$$\varphi(v) = \begin{cases} 0 & \text{se } v \geq 0 \\ \frac{v}{2} & \text{se } -0,5 < v < 0,5 \\ 1 & \text{se } v \leq -0,5 \end{cases} \quad (3.5)$$

3. **Função Sigmoid:** é o tipo de função de ativação mais utilizado em redes neurais artificiais, por serem biologicamente mais plausíveis. Como neurônios biológicos funcionam de forma binária, a função sigmoide é uma boa forma de modelar esse comportamento, já que ela assume valores apenas entre 0 e 1. É definida como uma função crescente, que apresenta um balanço entre o comportamento linear e não-linear. Um exemplo de função sigmoide é a função tangente hiperbólica ou função logística, definidas respectivamente da seguinte forma:

$$\varphi(v) = \frac{1 - e^{-av}}{1 + e^{+av}} \quad (3.6)$$

Onde a é o parâmetro de inclinação da função.

$$\varphi(v) = \frac{1}{1 + e^{-v}} \quad (3.7)$$

A função sigmoide tem algumas características que a difere das demais funções. Ao calcular suas derivadas, nota-se que ela satura para valores muito altos ou muito baixos. Além de que ela é uma função diferenciável, o que a torna bem mais interessante para a aplicação do gradiente descendente, já que o gradiente precisa calcular derivadas para encontrar o ponto de mínimo da função.

Mas, e como se faz para escolher qual a melhor entre elas? O fato é que não existe nenhuma regra que defina qual a melhor, mas essa escolha depende do problema que se está trabalhando. Outro fato que deve-se atentar é que existem dois tipos de problemas de aprendizagem supervisionada: problema de regressão e problema de classificação, os quais serão descritos nas Seções a seguir.

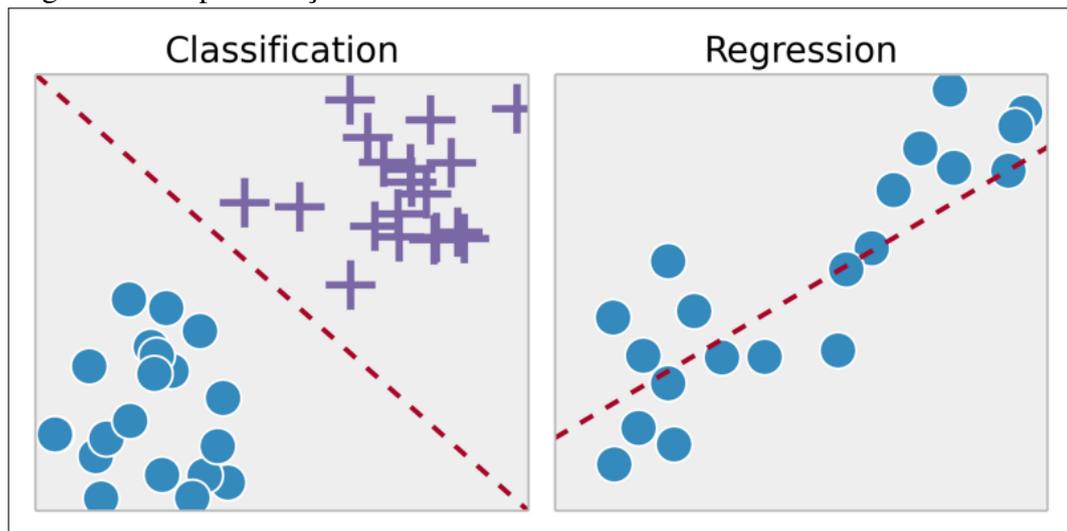
3.3 Problemas de Regressão e Classificação

O PR é uma tarefa de aproximação de uma função de mapeamento das variáveis de entrada para uma variável de saída contínua (BERNARDO; BERGER, 1998). Uma variável contínua é um valor no mundo dos números reais, como um valor inteiro ou ponto flutuante. Em outras palavras, o PR busca uma relação entre uma ou mais variáveis de entrada para explicar a saída. Por exemplo, dado um conjunto de dados sobre o tamanho de casas no mercado imobiliário, tentar prever o seu preço. Preço em função do tamanho retorna como saída um resultado contínuo caracterizando, assim, um PR. Para se estimar a habilidade de um modelo preditivo do PR, o mais comum é calcular o Erro Quadrático Médio (MSE) cujo procedimento será apresentado em maior detalhe na Seção 3.5.

Da mesma forma que o PR, segundo Bernardo e Berger (1998) o PC é uma tarefa de aproximação de uma de função de mapeamento das variáveis de entrada para variáveis de saída, só que discretas e não mais contínuas. Uma variável discreta é um conjunto de resultados que pode ser finito ou enumerável como por exemplo o número de filhos de um casal ou a quantidade de alunos numa escola. As variáveis de saída geralmente são chamadas de rótulos ou categorias. A função de mapeamento prevê a classe ou categoria para uma determinada observação. Em outras palavras, o PC tenta mapear variáveis de entrada em categorias distintas. Por exemplo, dado um tumor cancerígeno, temos de prever se ele é benigno ou maligno através do seu tamanho e idade do paciente (BARROS, 2016). A precisão da classificação é a porcentagem de exemplos corretamente classificados de todas as previsões feitas. Por exemplo, se um modelo preditivo do PC fizer 5 previsões e 3 delas forem corretas e 2 incorretas a taxa de acerta será de 60%.

Em alguns casos é possível converter um PR para um PC e vice-versa. O primeiro caso é conhecido como discretização. Como exemplo, o primeiro caso poderia ser transformado em um problema de classificação da seguinte maneira: ao invés de prever o preço da casa, o objetivo é saber se a casa foi vendida por mais ou por menos do que o preço que foi pedido. Dessa forma, têm-se uma classificação em duas categorias distintas. Para complementar esses

Figura 5 – Representação visual do PR e do PC.



Fonte: Barros (2016)

problemas a Seção 3.4 trata de um assunto que é muito importante para sua resolução.

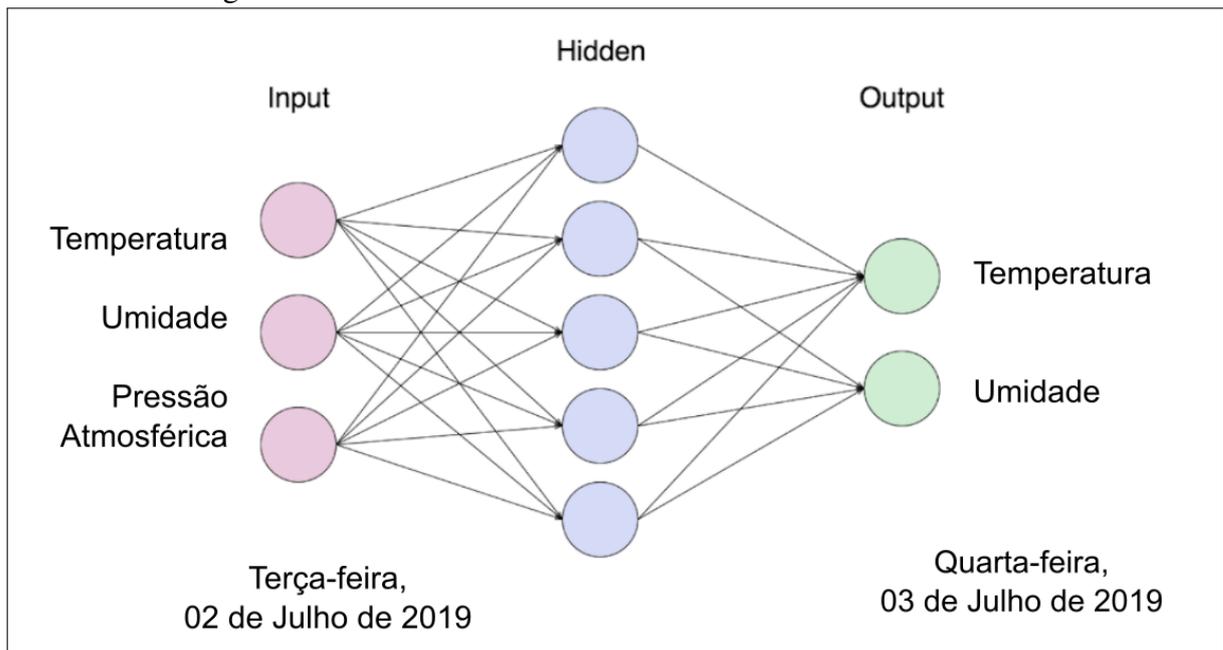
3.4 Redes Neurais Multicamadas

Também conhecidas como *Multilayer Perceptron* (MLP), as Redes Neurais Multicamadas são formadas basicamente por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída, como pode ser visto na Figura 6. Cada neurônio de uma camada recebe a saída de todos os neurônios da camada anterior e propaga sua saída para todos os neurônio da camada posterior.

A camada de entrada é formada por neurônios denominados especiais por não fazerem nenhum tipo de processamento. Eles servem exclusivamente para distribuir cada uma das entradas da rede para todos os neurônios da camada seguinte. A camada de saída tem a função de armazenar as respostas obtidas pela rede. Entre a camada de entrada e a camada de saída existem uma ou mais camadas ocultas, que são responsáveis por todo o processamento da rede. Quanto mais camadas escondidas na RNA, mais poder de processamento possui a rede.

Para melhor representar uma RNA observe o exemplo da Figura 6 no contexto deste trabalho. Imagine que deseja-se obter a previsão de temperatura e umidade do ar de quarta-feira, dia 03 de julho de 2019 tendo em vista que foram coletados dados com respeito a temperatura, umidade e pressão atmosférica da terça-feira. A RNA construída para esse modelo de dados tem a responsabilidade de receber os dados pela camada de entrada, processá-los pela camada oculta e por fim gerar a previsão para o dia 03 de julho.

Figura 6 – Rede Neural Artificial Multicamada.



Fonte: Oliveira *et al.* (2010)

Note que pode-se construir um modelo com RNA de diversas formas e uma delas é utilizar mais de uma camada oculta. Quanto maior for o número de camadas, maior a capacidade de processamento da Rede. Em problemas não-linearmente separáveis, as RNA multicamadas são essenciais, pois uma RNA simples com apenas uma camada consegue apenas separar o conjunto de dados de forma linear, isto é, por meio de uma reta. Assim como qualquer rede neural, esta também precisa ser treinada. Um conceito que será abordado na próxima Seção.

3.5 Treinamento e Avaliação

De acordo com Bocanegra (2002), a propriedade das redes neurais que mais se assemelha às características humanas é a habilidade de aprender. Ao invés de especificar todos os detalhes de uma tarefa, tem-se a possibilidade de treinar uma rede para isso. O objetivo principal do treinamento de uma RNA é fazer com que a aplicação de um conjunto de entradas produza um conjunto de saídas desejadas ou, no mínimo, um conjunto de saídas consistentes (FERREIRA; GALO, 2010).

Métodos iterativos de otimização são usados para realizar o treinamento de uma rede neural, geralmente fazendo o uso do método de gradiente descendente estocástico. A seguir está descrito a atualização dos parâmetros de uma RNA em cinco passos:

1. Apresenta-se uma amostra de exemplos de treinamento à RNA;

2. Calcula-se o erro com base na função objetivo, tendo em vista que a classe correta é conhecida nessa etapa;
3. Calcula-se a derivada parcial de cada parâmetro da rede, utilizando o valor do erro;
4. Com base no gradiente, atualiza-se os parâmetros no sentido contrário ao informado por esse vetor;
5. Volta-se ao passo 1 ou encerra o procedimento caso se atinja um critério de parada.

Com todo esse processo realizado, o esperado é que os valores de parâmetros objetivos sejam satisfatórios ou no mínimo consistentes, como já mencionado anteriormente. Esse processo é chamado de treinamento, ao qual no final do processo a função de custo deve estar minimizada. Para exemplificar o processo do gradiente descendente, observe o seguinte pseudo-código:

Algoritmo 1: Gradiente descendente estocástico

```

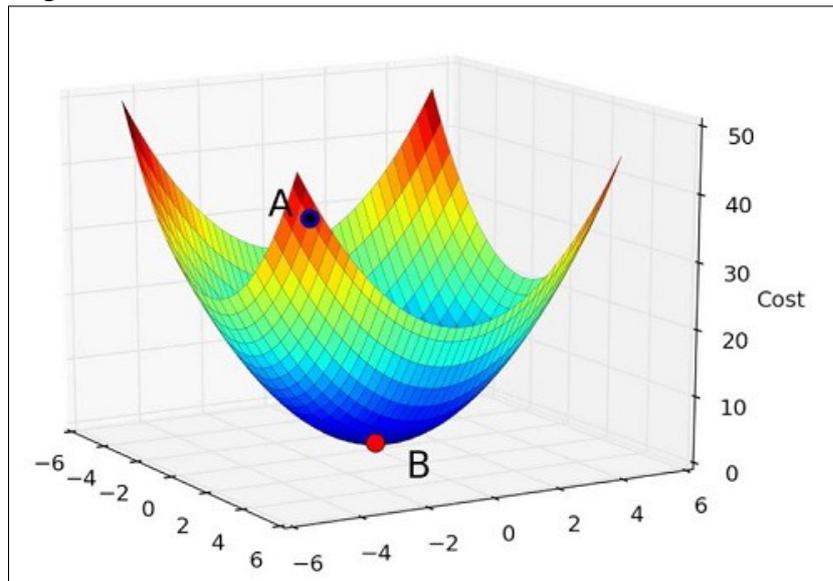
 $w_0 \leftarrow$  inicialização;
while True do
    |  $w_t = w_{t-1} - n_t \nabla L(w_{t-1});$ 
    | if  $\|w_t - w_{t-1}\| < \epsilon$  then
    | | break;
    | end
end

```

Como mostra o Algoritmo 1, inicializa-se o vetor de peso inicial w_0 . Em seguida, tem-se um laço iterativo *while* que executa enquanto a norma da diferença entre o vetor de pesos atual e anterior for maior do que um erro pré-definido, que significa que a aprendizagem atingiu o critério de aceitação. A terceira linha representa o processo principal do gradiente descendente. O w_t é atualizado sempre recebendo o valor do w anterior diminuído da taxa de aprendizagem n_t multiplicado pelo resultado de uma aplicação de derivadas parciais em relação ao w anterior. O processo principal é responsável por encontrar o ponto de mínimo do gradiente utilizando-se da operação de derivada, observe a Figura 7.

O Gradiente Descendente Estocástico utiliza apenas uma observação dos dados. Além dele, existem outras variações desse método. Eles diferem apenas na quantidade de dados da amostra para atualizar os parâmetros. Existe o Gradiente Descendente que utiliza toda a amostra de dados e o Gradiente Descendente em *Batch* que utiliza uma quantidade pré-determinada e fixa de observações. Algo em comum entre eles é que todos buscam uma forma

Figura 7 – Gradiente descendente estocástico.



Fonte: Imram (2017).

de minimizar a função de custo do modelo.

A função de custo é um método utilizado para avaliar o quão bem treinado está o modelo. Existem diversas funções que podem ser utilizadas como funções de custo. Vale ressaltar que a escolha da função de custo deve ser feita de acordo com o objetivo a ser alcançado, tendo em vista que cada função têm suas características. São exemplos de funções de custo: *minimax*, função L, função de erro quadrático médio etc (BERGER, 2013). Abaixo o exemplo da função de Erro Quadrático Médio de um Rede Neural é denotado por

$$C(w, b) = \frac{1}{2n} \sum_x \|c(x) - h(x)\|^2, \quad (3.8)$$

Em que w são as matrizes de pesos dos neurônios, b são os *bias*, n é o número total de amostras de treino, x representa cada exemplo de treino, $c(x)$ é a saída esperada da rede, e $h(x)$ é o vetor de ativações da saída da rede quando x é a entrada.

Em seu artigo Stewart (2019) cita a interpretabilidade do modelo como sendo uma das principais limitações da Aprendizagem de Máquina. Ele dá como exemplo o fato de que se quem criou modelo não consegue convencer seu público de que entende como o algoritmo gerou os resultados, a probabilidade de confiarem nos resultados gerados pelo algoritmo é muito baixa. É com base nesse fato que a maioria dos cientistas costumam não aceitar resultados provindos de algoritmos de Aprendizagem de Máquina.

3.6 Redes Neurais Recorrentes

Em um contexto prático, suponha que se possua dados diários de temperatura. Seria interessante relacionar as duas variáveis – temperatura e dia – passados 2 dias para prever a temperatura no dia seguinte. Nesse caso, a temperatura do dia em que se deseja prever seria uma variável dependente, isto é, ele depende das temperaturas nos dois dias que se passaram, estas sendo variáveis independentes. Formalmente, $y = temperatura_{t+1}$ é a variável dependente e $x = [temperatura_t, temperatura_{t-1}, dia_t, dia_{t-1}]$ são as variáveis independentes. Porém, existem alguns problemas com essa representação, por exemplo, como saber quantos dias defasados são necessários para prever bem o dia seguinte? Mesmo que com isso definido, o que fazer quando algumas das observações não tem sequer o número de dias necessários para prever o próximo dia? E o que fazer quando o tamanho das sequências de dados é variável? Com a finalidade de suprir essa necessidade, foi que surgiu o conceito de Rede Neural Recorrente (RNR).

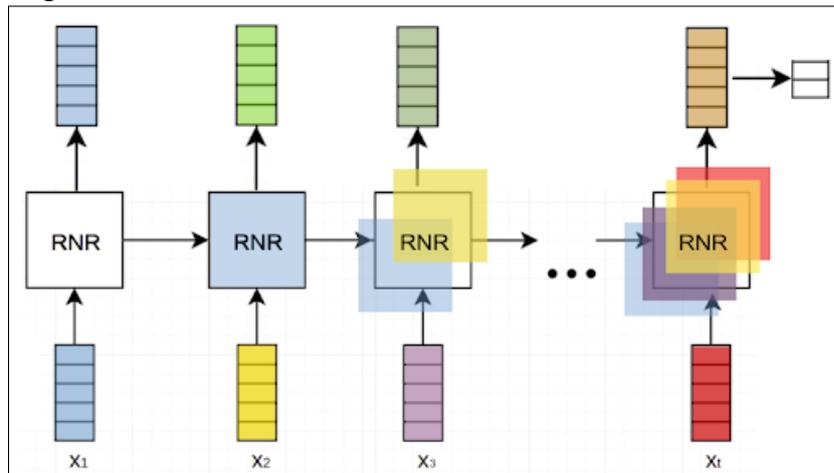
RNR é uma extensão de RNA clássica, que é capaz de receber como entrada uma sequência de dados de comprimento variável (GOODFELLOW *et al.*, 2016). A principal diferença entre RNR e RNA é que a camada oculta da segunda recebe sinais apenas da camada de entrada, os processa e transmite para a camada de saída. Já a camada oculta da RNR, recebe sinais tanto da camada de entrada quanto da camada oculta da iteração de tempo anterior. Em outras palavras, é como se a camada oculta da RNR armazenasse no seu estado oculto informações dos dados observados naquele período de tempo e recuperasse apenas informações relevantes do estado oculto anterior. Formalmente, dada uma sequência $x = x_1, x_2, \dots, x_t$, a RNR atualiza o estado oculto h_t da seguinte forma:

$$h_t = \begin{cases} 0, & t = 0 \\ \varphi(h_{t-1}, x_t), & \text{caso contrário} \end{cases} \quad (3.9)$$

Na Figura 8, cada vetor x_t representa o conjunto de valores das sequências de dados ou sinais para cada período de tempo e cada RNR representa um estado oculto. A rede processa toda a informação dos sinais de entrada e propaga sua saída para o próximo estado oculto, até que se tenha passado por toda a série.

RNRs não são uma única classe de rede, mas sim uma coleção de topologias que se aplicam a diferentes problemas. Elas foram construídas para processar sequências de dados - os sinais de entrada - com a capacidade de memorizar entradas de estados ocultos anteriores, dessa forma abrindo a porta para novas possibilidades. Agora, no Capítulo 4, será definido todo

Figura 8 – Rede Neural Recorrente.



Fonte: Próprio autor.

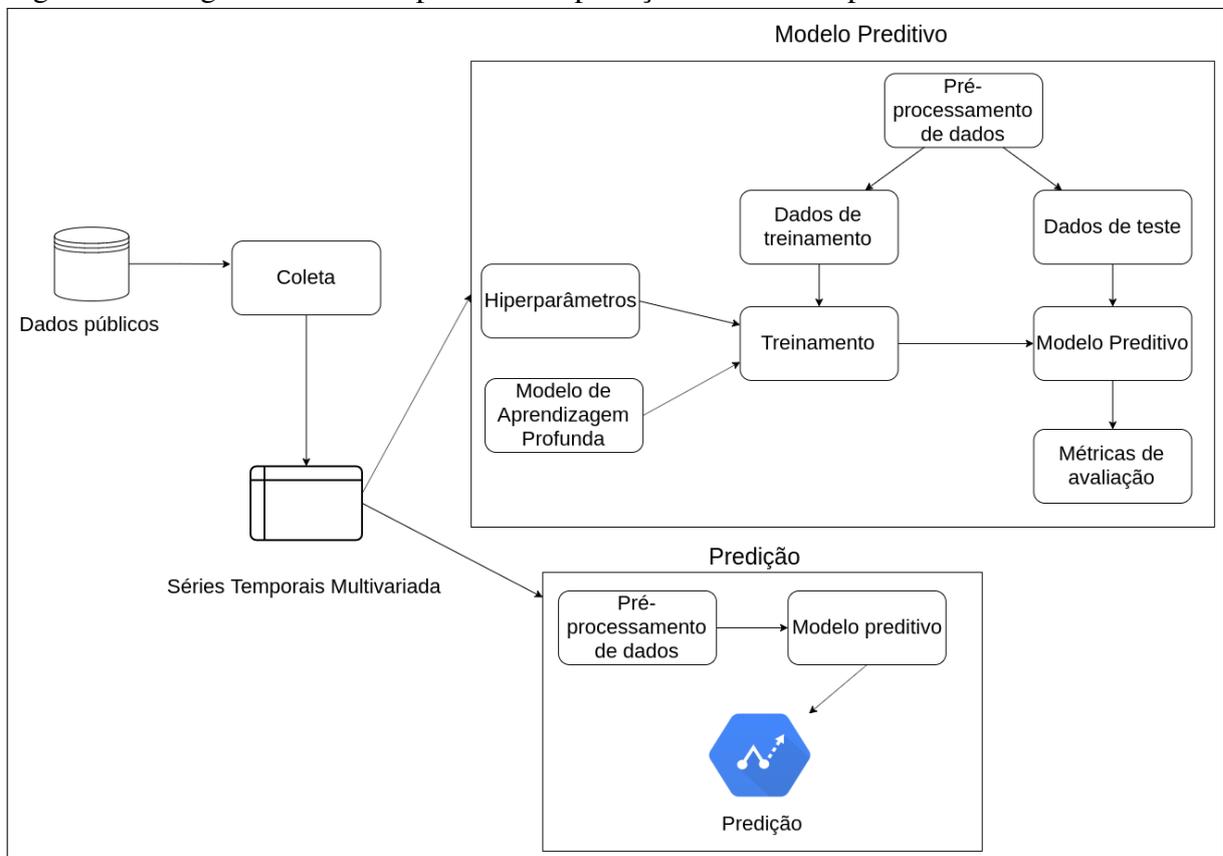
o processo de predição proposto por esse trabalho, utilizando todos os conceitos anteriormente descritos.

4 MODELO PARA PREDIÇÃO DE SÉRIES TEMPORAIS

Esse capítulo abordará, de forma geral, os procedimentos para construção do modelo proposto. Na Seção 4.1 será tratado como se dará a fase de pré-processamento de dados, a Seção 4.2 se trata da arquitetura da rede neural utilizada no trabalho e por fim a Seção 4.3 tem por finalidade mostrar como se dará o processo de validação do modelo.

A Figura 9 representa, de forma geral, todos os passos que serão utilizados no processo de predição proposto por este trabalho.

Figura 9 – Diagrama de todo o processo de predição de série temporal.



Fonte: Próprio autor

4.1 Pré-processamento de Dados

Inicialmente, houve a fase de coleta dos dados (Etapa 1), que consiste em obter dados disponíveis publicamente em portais de informação climática. Essa coleta é facilitada por necessitar apenas baixá-los e armazenar no local desejado. Com o processo da coleta finalizado, parte-se para a etapa do pré-processamento.

Difícilmente algum conjunto de dados no mundo real é perfeito, livre de inconsistên-

cias. É bastante comum, inclusive, que seja necessário uma etapa exclusivamente para a limpeza dos dados, definida nesse trabalho como pré-processamento. A etapa de pré-processamento dos dados é muito importante para que o modelo seja construído corretamente. Em certas situações, 80% da análise de dados é gasto no processo de limpeza e preparação dos dados, como mencionado em Malik *et al.* (2016). A preparação dos dados servirá para converter os dados para o formato adequado, além da extração de algumas características estatísticas sobre os dados. Por exemplo, pode-se obter os valores da média, desvio padrão, mediana, etc dos indicadores nos últimos k dias, com k assumindo diferentes valores. Esses valores podem capturar representações sobre o histórico e auxiliar na codificação de entrada para o modelo de predição.

Os dados que serão utilizados para treinamento do modelo são matrizes que possuem 11 colunas, em que cada uma dessas colunas representa uma característica específica dos dados. A primeira é o número da estações onde o dado foi gerado, a segunda e a terceira dizem respeito a data e hora em que o dado foi gerado, respectivamente. Em seguida, temos os dados propriamente ditos da série temporal. Das quais precipitação, temperatura máxima, temperatura mínima, insolação, evaporação, temperatura média, umidade relativa média e por fim velocidade do vento. Todos os pontos de dados são gerados duas vezes por dia, às 12 : 00 e às 00 : 00. Na primeira hora são gerados os dados que dizem respeito à temperatura máxima, insolação, temperatura média, umidade relativa e velocidade do vento, já na segunda hora são gerados os dados com respeito à precipitação e temperatura mínima.

O pré-processamento dos dados foi realizado da seguinte forma: excluiu-se as colunas desnecessárias, particularmente a coluna de estação. Em seguida formatou-se a data e comprimiu-se os dados em um único dia com todos os dados de seu respectivo dia resultando, assim, uma série com alguns pontos de dados que possuem 8 características cada. Além disso foi utilizada uma estratégia de exclusão de valores em branco na série, o chamado *forward fill*, que consiste em repetir valores defasados para os que estão em branco. A série resultante é transformada em uma série de aprendizagem supervisionada da seguinte maneira: cada dia da série é agrupado com outros 60 dias precedentes resultando, assim, uma estrutura tridimensional, onde a primeira dimensão diz respeito à quantidade de exemplos, a segunda aos 60 dias precedentes e a última as 8 características do dado.

Finalizada a fase de limpeza e preparação dos dados, os dados serão tratados como séries temporais climáticas, as quais servirão como entrada para a construção do modelo preditivo. O conjunto de dados será dividido em dois conjuntos, da seguinte maneira: a maior parte será

utilizada para treinar e a outra parte será utilizada para realizar testes acerca do modelo, conforme descrito na Seção 4.2.

4.2 Modelo de Aprendizagem Profunda

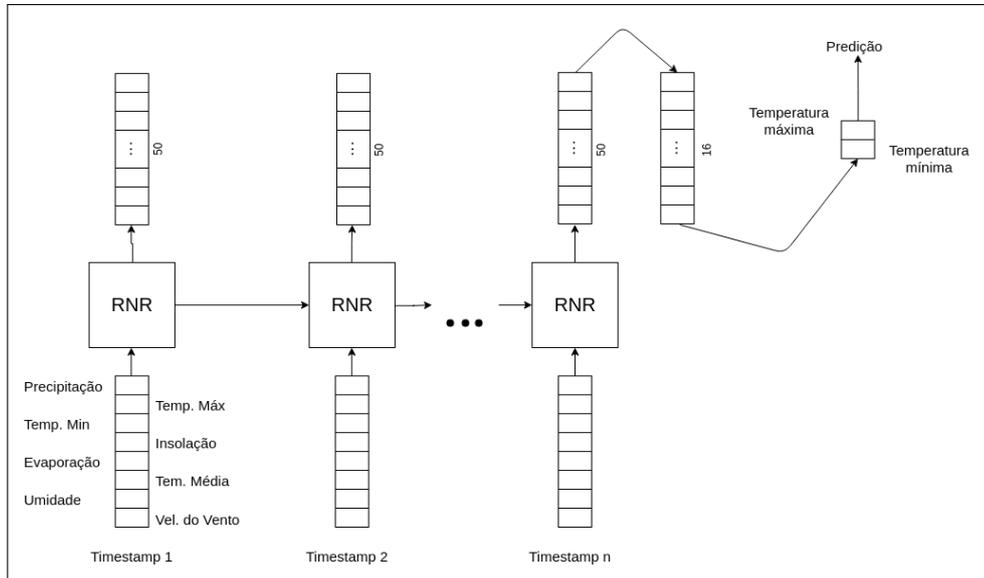
De uma maneira simples, um modelo preditivo consiste em uma função matemática que, quando aplicada a uma massa de dados, é capaz de identificar padrões e oferecer como saída uma previsão do que pode ocorrer. O objetivo principal de uma predição é prever algum evento, baseando-se na análise de condições existentes em eventos passados. Segundo Yan-Tak (2015), a performance da AP vem superando abordagens clássicas para predição, sejam elas estatísticas ou baseadas em aprendizagem. Motivado por essas definições é que esse trabalho se propõe realizar predições climáticas utilizando um modelo de AP.

O modelo de AP será construído com a utilização de uma RNA, mais especificamente RNR. Com a Figura 10 é possível observar uma representação desse modelo. Cada elemento dos *Timestamp* representa uma *feature* ou característica de cada observação do histórico. Essa característica pode ser, por exemplo, o valor de uma série no histórico dos dados. Dessa forma, a quantidade de estados ocultos da rede será definido pelo tamanho n do histórico da observação. O horizonte de predição será sempre um dia a frente. Por exemplo, se o espaço de tempo das observações for dado em dias, para cada dia existirá um estado oculto, obtido pelo processamento das características daquele dia e do estado oculto do dia anterior. Além disso, dispendo-se de séries sobre “temperatura máxima”, “temperatura mínima”, “precipitação”, “insolação”, “evaporação”, “temperatura média”, “Umidade” e “velocidade do vento”, o vetor de característica de cada *Timestamp* será dado pela valor de cada uma dessas variáveis coletadas para o dia.

Levando em consideração o pré-processamento dos dados descrito na na Seção 4.1, foi definida a arquitetura do modelo: para representar o histórico, foi utilizada a Rede Neural *Long Short Term Memory* (LSTM). A LSTM é uma arquitetura de uma RNR que memoriza valores para realizar predições ou classificações mais inteligentes (SUNDERMEYER *et al.*, 2012). Cada ponto de dado é representado na rede como um estado oculto, que fará o processamento de cada *feature*, representando-as como um vetor de tamanho 50, que logo em seguida servirá como entrada do próximo estado até que se chegue no ultimo ponto de dados. Logo em seguida, são aplicadas duas transformações no vetor de saída. Uma que transfere os dados para um vetor de tamanho 16 e outra que transforma em um vetor de tamanho 2, da forma com que esse vetor seja

a predição propriamente dita das temperaturas máxima e mínima. Veja a Figura 10.

Figura 10 – Arquitetura da Rede Neural Recorrente.



Fonte: Próprio autor.

A técnica utilizada para validação e ajuste dos *hyperparâmetros* do modelo foi um tipo de Validação Cruzada (CV), mais especificamente a Validação Cruzada Aninhada (NCV), descrita na Seção 4.3.

De outra forma modelo irá receber todo o histórico e codificar em único vetor de saída, que representará toda a informação. Neste trabalho, o último estado oculto, ou seja, h_T , será usado para representar o histórico. Também é possível concatenar a esse vetor uma representação do histórico dada por atributos modelados sobre os dados, como os estatísticos mencionados na Seção 4.1. O histórico será utilizado como entrada para as duas ultimas camadas do modelo. Elas são responsável por realizar as predições propriamente ditas por meio do uso de transformação linear. Ou seja, a ultima camada irá gerar um estados oculto d_1 , no qual representa a predição. Assim, para cada período k do horizonte de predição, o valor predito será dado por:

$$y_k = \varphi(W_d d_k + b_d), \quad (4.1)$$

em que o W_d realiza uma transformação linear sobre o estado oculto d_k , b_d é o *bias*, colocando-o em uma dimensão n , correspondente ao número de séries que se deseja prever. A função φ realiza uma transformação sobre a saída, podendo ser, inclusive, uma identidade. Abaixo, apresenta-se o cálculo de d_k , de maneira recorrente, como esperado.

$$d_k = \rho(h_t, d_{k-1}), \quad (4.2)$$

Vale ressaltar que, uma vez o tendo o vetor codificado h_T , ele não será mais alterado e será aplicado a mais duas camadas para que assim se tenha a predição. Resumidamente, o modelo processa as informações advindas dos sinais de entrada, os processa e gera a predição. Uma vez definido, o modelo terá que passar por uma fase de experimentos e validação, os quais estão descritos na Seção 4.3.

4.3 Experimentos e Validação

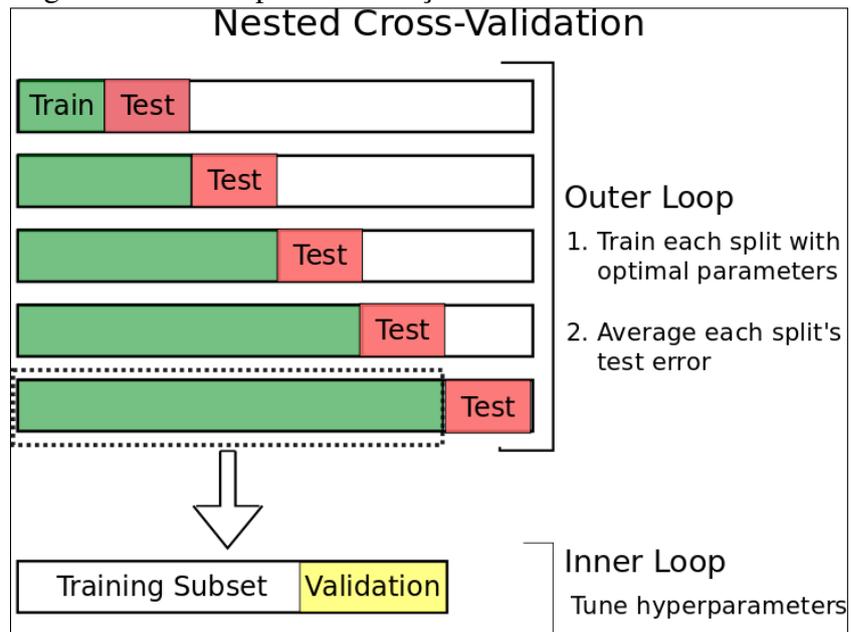
O modelo de AP depende do ajuste de vários hiperparâmetros. Para tanto, é imprescindível que se encontre o conjunto desses valores que permita gerar as melhores predições, conforme as métricas escolhidas. De acordo com Albon (2018), a melhor maneira de fazer isso é por CV, cuja ideia é treinar um modelo em um subconjunto de dados e validar o modelo treinado nos dados restantes. Deve-se repetir esse processo várias vezes e calcular a média do erro de forma a obter uma estimativa do desempenho de generalização do modelo. Pode ser utilizado, por exemplo, k -fold CV, descrito abaixo:

1. Definir os hiperparâmetros a serem ajustados;
2. Realizar a divisão do conjunto de dados em k subconjuntos;
3. Treinar o modelo utilizando $k - 1$ subconjuntos;
4. Testar o modelo com o subconjunto não utilizado na etapa de treinamento;
5. Repetir as etapas 3 e 4 até que todos os subconjuntos tenham sido testados uma vez;
6. Repetir da etapa 1 a 5 para todos os valores possíveis dos hiperparâmetros;
7. Guardar o conjunto de hiperparâmetros que produziu o melhor resultado.

Entretanto, Cawley e Talbot (2010) apontam que, para avaliar Séries Temporais, a CV, ao selecionar séries no mesmo período de tempo para testar e validar pode enviesar os resultados. Desta forma, é necessário um conjunto de teste com período de tempo diferente do que foi utilizado para treinar. Uma forma de resolver esse problema é utilizando a NCV, mostrada na Figura 11. Assim, se a predição ocorre a partir do instante de tempo t_i , utiliza-se $T + P$ amostras de dados imediatamente anteriores a t_i para gerar o modelo. Para prever os P valores a partir de t_i , constrói-se o modelo utilizando T valores anteriores a t_i como histórico. Durante o treinamento, t_i é ajustado progressivamente no tempo considerando as amostras disponíveis.

Baseando-se no modelo preditivo definido, após as etapas de treinamento, o próximo passo é gerar a predição climática. Esse passo consiste em pré-processar os dados, aplicá-los ao

Figura 11 – Exemplo de validação cruzada.



Fonte: Cochrane (2018).

modelo preditivo e receber como resultado a predição, a qual será ajustada para tentar encontrar resultados melhores do que aos que já existem.

5 TRABALHOS RELACIONADOS

Este capítulo discorre sobre alguns trabalhos que abordaram a mesma problemática que o presente trabalho, assim como descreve rapidamente suas soluções propostas.

5.1 *Deep forecast: Deep learning-based spatio-temporal forecasting*

Ghaderi *et al.* (2017) propõe realizar previsões espaço-temporais da velocidade do vento utilizando AP. Ele dá ênfase ao fato de que recursos de energias renováveis são de natureza aleatória e, portanto, sua integração é facilitada com previsões de curto prazo. Ele busca melhorar significativamente as previsões de curto prazo.

Para tornar isso possível, o autor montou uma estrutura de grafo cujos nós são entidades geradoras de dados e suas bordas modelam como esses nós interagem uns com os outros. Com isso tem-se acesso a dados reais a cada hora e, portanto, com base nos dados realiza previsões para a velocidade do vento para as próximas h horas, onde h é quantidade de horas que deseja-se prever.

O autor descreve o modelo da seguinte forma: em diferentes etapas de tempo dos dados, tem-se diferentes tipos de entrada. Para a primeira etapa, temos dados reais para todas as entradas mas para o próximo passo, temos dados reais para todas as entradas com exceção de um. Para esse que ficou de fora ele utiliza não um dado real, mas uma previsão do dado da etapa passada. Esses passos se repetem por todos os h passos.

O algoritmo desenvolvido utiliza a técnica de AP por meio do uso de RNR com arquitetura LSTM. O autor utiliza em seu modelo uma taxa de aprendizado de 0.001, utiliza a ReLu como função de ativação, os dados são normalizados para um intervalo entre 0 e 1. Assim como o presente trabalho, ele utiliza o TensorFlow e o Keras.

Estudos do resultado desse trabalho mostram que a proposta do algoritmo apresentou melhoras significativas na previsões de curto prazo da velocidade do vento em relação a um conjunto de modelos de benchmarks amplamente utilizados.

5.2 *A deep hybrid model for weather forecasting*

Nestes trabalho, Grover *et al.* (2015) propõe um modelo híbrido para realizar previsões do tempo. Ele utiliza o método de AP para treinar seu modelo preditivo. No trabalho ele também mostra como o modelo básico pode ser aprimorado utilizando interpolação espacial de

longo alcance.

Para o experimento ele utiliza dados coletados provindos de observações de balões feitas em 60 estações de predição climática nos EUA. Os balões transmitem informações sobre direção do vento, temperatura, ponto de orvalho e diversas outras variáveis climáticas. Os dados são repassados para uma entidade curadora de dados que tem a responsabilidade de resolver as possíveis inconsistências das variáveis. E de todas as variáveis observadas ele considera formalmente quatro dessas para aplicar ao modelo: velocidade do vento, pressão atmosférica, temperatura e ponto de orvalho.

A abordagem em questão para a construção do modelo quer garantir as seguintes diretrizes:

1. Mineração temporal: o modelo deve ser capaz de identificar e aprender com os padrões climáticos recorrentes.
2. Interpolação espacial: a influência dinâmica das leis atmosféricas nos fenômenos climáticos precisa ser explicada na previsões.
3. Interação entre variáveis: as interdependências locais entre as variáveis climáticas devem ser capturadas pelo modelo.

Descrito dessa forma ele ainda fala sobre seu modelo possuir três principais componentes. O primeiro componente é um conjunto de preditores individuais para as variáveis meteorológicas treinadas utilizando dados históricos. O segundo ele descreve como sendo um componente que trabalha para refinar as inferências produzidas. E por último o terceiro componente se trata de uma RNR que busca soluções que respeitem as estatísticas conjuntas esperadas das variáveis meteorológicas.

Um conjunto de experimentos foram realizados e mostraram que essa metodologia pode fornecer melhores resultados do que os parâmetros de referência da *National Oceanic and Atmospheric Administration* (NOAA). Por fim ele demonstrou o interesse ainda em investigar os aumentos no poder preditivo que ainda podem ser alcançados através da integração de dados adicionais ao modelo proposto.

6 RESULTADOS

Após contextualização, que apresenta os principais conceitos relacionado ao trabalho, este capítulo apresenta os resultados computacionais obtidos e a análise crítica frente ao procedimento adotado. A Seção 6.1 se trata do cenário no qual os experimentos foram realizados, na Seção ?? abordará um análise estocástica do modelo, a Seção 6.3 se trata de uma análise de algumas técnicas de normalização e por fim a Seção 6.4 abordará um análise que visa verificar os acertos do modelo proposto.

6.1 Cenário

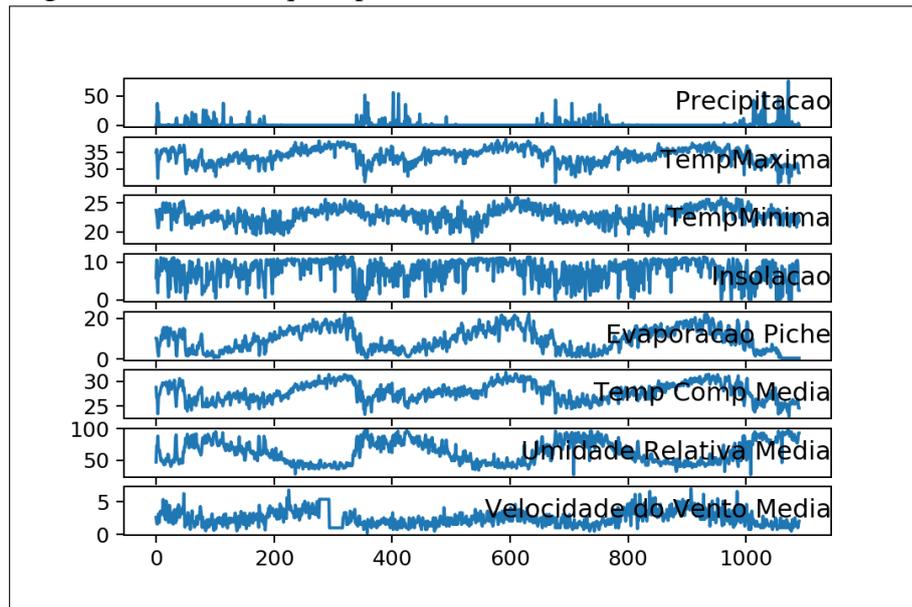
A implementação do modelo foi desenvolvido utilizando a linguagem de programação Python, versão 3.6. Foram utilizadas bibliotecas de pré-processamento e análise de dados e uma biblioteca para construção do modelo de AP, o Keras (CHOLLET *et al.*, 2015), utilizando o TensorFlow (ABADI *et al.*, 2016) como *back end*. Todas as análises e testes foram executadas na ferramenta Jupyter Notebook (KLUYVER *et al.*, 2016).

Para modelos cuja entrada inclui valores previstos, precisamos aumentar a capacidade de aprendizagem do modelo para ajudar a superar o chamado *overfitting*, ou excesso de ajuste, bem como o *underfitting* ou subajuste. Pensando nisso, implementou-se CV, como mencionado na Seção 4.3, utilizando um dia à frente para validar o modelo. Além disso, utilizou-se a estratégia *Early Stopping* ou Parada Precoce, para interromper o processo de aprendizagem quando o erro de validação não melhora nas próximas k ($k = 10$, neste trabalho) épocas.

O modelo de AP do presente trabalho foi definido com as seguintes características: foi utilizada a arquitetura de LSTM, como já citado na Seção 4.2; como função de ativação da camada de saída foi utilizada a *RELU*; para realizar o treinamento da rede neural, utilizou-se como algoritmo de otimização *Adam*; a função de custo utilizada no processo de otimização foi a Erro Absoluto Médio (MAE); a taxa de aprendizagem foi reduzida em 10% a cada 7 épocas sem melhoria de desempenho medido sobre o dado de validação; como parâmetro e como métrica para fins de avaliação durante o treinamento utilizou-se o MSE.

Foram utilizadas bases de 6 cidades de 4 estados do Brasil para o treinamento e validação do modelo. Cada uma das bases foi utilizada separadamente. Ou seja, cada uma das bases foi utilizada para um treinamento, o que gera um modelo de predição diferente. A princípio foi utilizada a base de dados climáticos da cidade de Crateús/CE, mas na Seção 6.4 e

Figura 12 – Gráfico que representa a base de dados da cidade de Crateús.



Fonte: Próprio autor.

no Apêndice B estão expostos resultados para outras cidades. Pode-se observar na Figura 12 a representação gráfica dessa base. O eixo x representa a quantidade de dados que a base possui. Nesse caso possui-se um histórico de pouco mais de mil pontos de dados, que em outras palavras cada ponto de dado representa um dia. Já o eixo y se trata do intervalo em que cada variável está. Por exemplo, a temperatura máxima está num intervalo que começa em pouco menos de 30 graus e vai até aproximadamente 36 graus.

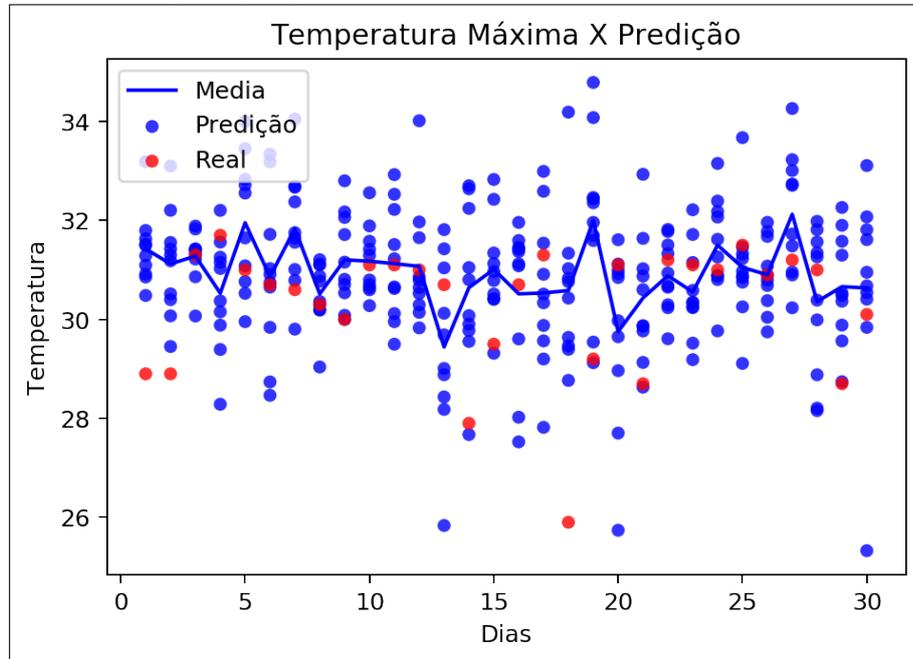
6.2 Análise de Comportamento

Assim como Karlin (2014) cita em seu livro, por mais que se conheça a condição inicial de um processo, há diversas maneiras de evolução deste. Ou seja, por mais que se execute um procedimento milhares de vezes, dificilmente ele irá resultar na mesma solução. Porém, a média das soluções são normalmente bem próximas se o modelo estiver bem definido. Por esse motivo foi importante verificar como o presente modelo se comporta dada a mesma entrada. Dessa forma o modelo foi executado dez vezes para a mesma entrada.

A Figura 13 explica bem como o modelo se comportou para a base de dados da cidade de Crateús. Observe que o eixo x do gráfico diz respeito aos dias aos quais se buscou uma predição e já o eixo y denota as temperaturas preditas e reais para cada dia. Os círculos vermelhos denotam a temperatura real e os azuis denotam as predições. Já que o modelo foi executado dez vezes, cada dia têm-se onze círculos, dez para as predições e um para o valor real.

Já a curva em azul corresponde a média das predições, para que assim se possa observar como o algoritmo está bem definido.

Figura 13 – Gráfico de dispersão entre a temperatura máxima real e sua predição para a cidade de Crateús.



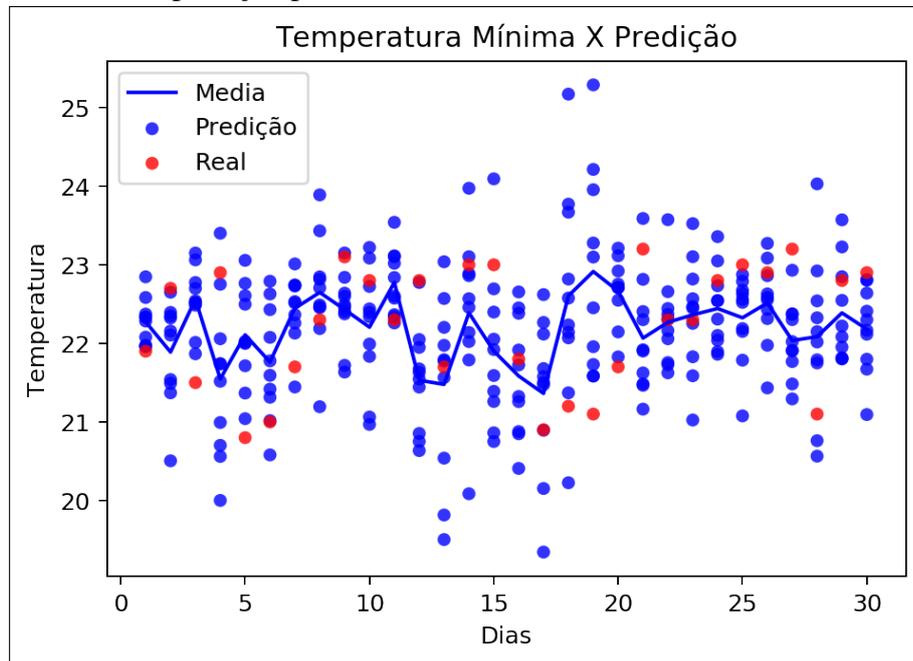
Fonte: Próprio autor.

Observou-se que, na média, o modelo chega próximo do valor real, o que significa que o problema foi bem modelado. Apesar disso ainda tem-se alguns pontos de melhoria. Quando a base apresenta alguns *outliers*, pela natureza aleatória das variáveis climáticas, o modelo se distancia bastante do valor real, o que não é exclusivamente um problema do modelo. *Outliers* são dados que fogem do padrão e não conseguimos explicar o que causou isso.

6.3 Análise de normalização

O propósito de uma normalização de dados é minimizar os problemas oriundos do uso de unidades e dispersões distintas entre as variáveis (QUACKENBUSH, 2002). A normalização de dados facilita o trabalho da rede neural encontrar padrões. Existem diversas técnicas para normalização de dados. O modelo foi executado com quatro tipos diferentes de técnicas de normalização para que se pudesse realizar uma comparação dos resultados do modelo para cada uma das técnicas. Vale ressaltar que a normalização depende muito da base de dados que está sendo utilizada. Portanto pode ser que uma mesma técnica apresente diferentes resultados para bases diferentes com o mesmo modelo.

Figura 14 – Gráfico de dispersão entre a temperatura mínima real e sua predição para a cidade de Crateús.



Fonte: Próprio autor.

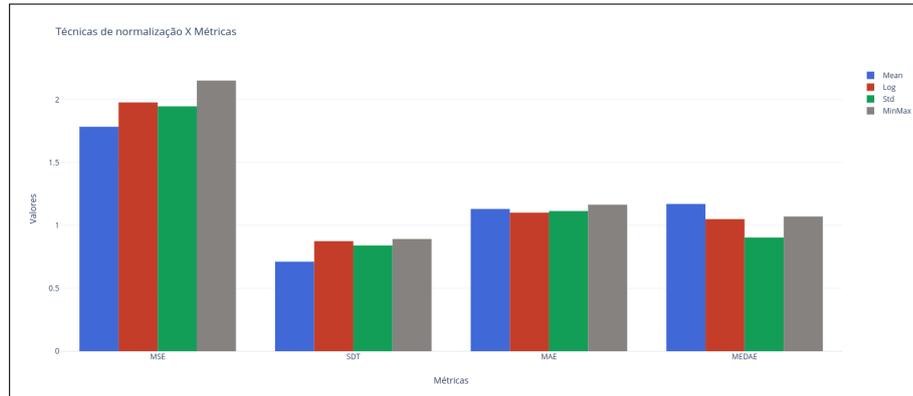
A Figura 15 e a Figura 16 representam graficamente o desempenho do modelo em relação ao tipo de normalização. Observe que o eixo x possui todas as métricas de avaliação utilizada no modelo. O MSE representa o Erro Quadrático médio, o STD representa o desvio padrão, o MAE é o Erro Absoluto Médio e o MEDAE representa o Erro Absoluto Mediano. O eixo y, compõe os valores atingidos para cada tipo de normalização. A técnica de normalização *Mean* normaliza os dados em torno da média de todos dados da base, a *Log* utiliza-se da função logarítmica para tal, a *STD* normaliza com base no desvio padrão dos dados e por ultimo a *MinMax* que recebe dois valores limitantes para efetuar essa normalização. Os valores limitantes utilizados foram os valores padrões 0 e 1.

Para as predições de temperatura máxima, concluiu-se que para a base de dados de Crateús, a técnica que apresentou melhores resultados para o MSE e desvio padrão foi técnica *Mean*, porém perde para a técnica de desvio padrão nas métricas MAE e MEDAE. Já para as predições de temperatura mínima, observou-se que a técnica de desvio padrão é a melhor em quase todas as métricas.

6.4 Análise de precisão

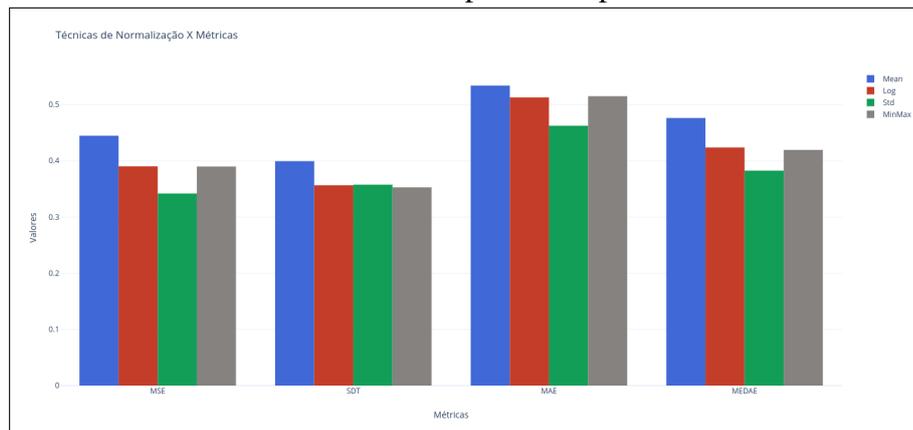
Uma outra forma de validação é a verificação da precisão de um modelo. Nessa Seção, será analisado a quantidade de acertos que o modelo teve para cada um dos dias em

Figura 15 – Representação gráfica do desempenho do modelo quando aplicado a diferentes tipos de normalização de dados para a Cidade de Crateús/CE para a temperatura máxima.



Fonte: Próprio autor.

Figura 16 – Representação gráfica do desempenho do modelo quando aplicado a diferentes tipos de normalização de dados para a Cidade de Crateús/CE para a temperatura mínima.



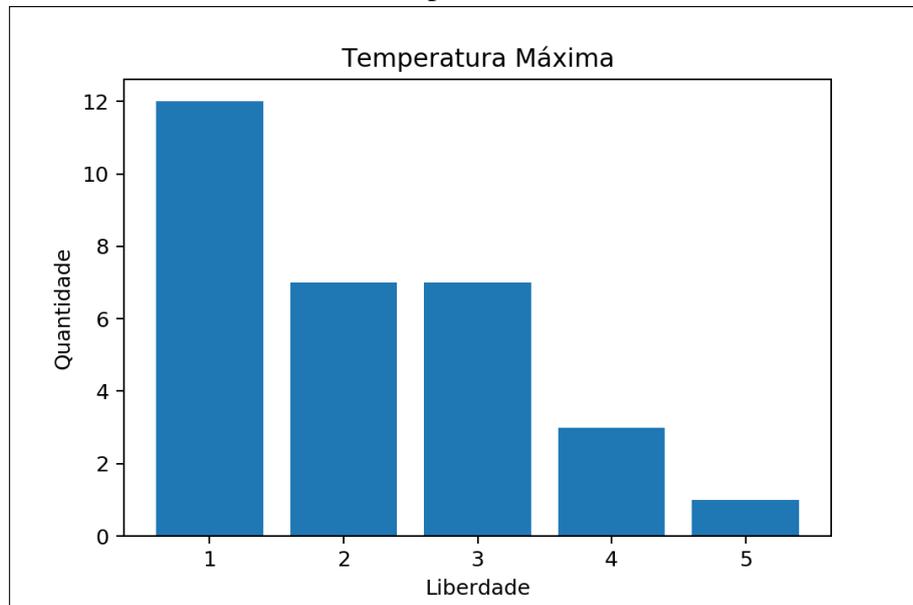
Fonte: Próprio autor.

relação ao grau de liberdade dado. Por exemplo, é interessante saber quantas vezes o modelo errou até 1 ponto para mais ou para menos. É isso que iremos verificar logo a seguir.

Veja que a Figura 17 e a Figura 18 retratam a quantidade de vezes, dado um grau de liberdade, o modelo errou. A primeira figura diz respeito às predições de temperatura máxima, já a segunda figura diz respeito às predições de temperatura mínima. O eixo das abcissas no primeiro gráfico demonstra o grau de liberdade dado ao modelo, já o eixo das ordenadas denota quantidade de vezes que modelo errou.

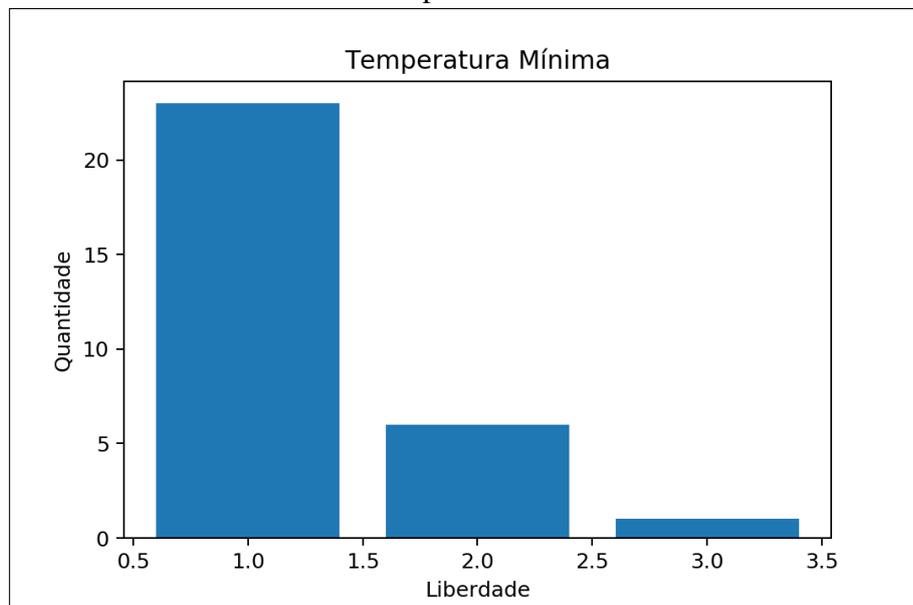
Pelas representações gráficas foi possível verificar que a grande maioria das predições errou até no máximo 3 pontos para mais ou para menos com respeito às predições de temperatura máxima. Foi ainda melhor o desempenho de predições para as temperaturas mínimas. Observe que 80% das predições erraram até no máximo 1 ponto para mais ou para menos.

Figura 17 – Gráfico que representa os erros de predições de temperatura máxima do modelo para a Cidade de Crateús/CE.



Fonte: Próprio autor.

Figura 18 – Gráfico que representa os erros de predições de temperatura mínima do modelo para a Cidade de Crateús/CE.



Fonte: Próprio autor.

Além da base de dados da cidade de Crateús, testou-se o modelo para bases de dados de outras cidades em execução única. Pode-se observar pela tabela 3 as métricas resultantes para cada uma das cidades. Assim como na Seção 6.3, utilizou-se todas as métricas lá descritas.

Os resultados das métricas presentes na Tabela 3 demonstram que o modelo mantém o mesmo desempenho quando treinado com bases diferentes.

Tabela 3 – Métricas de resultado do treinamento da cidade de Maceió no estado de Alagoas para as temperaturas mínimas.

Cidade	MSE	STD	MAE	MEDAE
Maceio/AL	1.97	0.87	1.11	1.04
Campos Sales/CE	5.09	1.44	1.73	1.68
Turiacu/MA	4.22	1.26	1.613	1.33
Belem/PA	3.15	1.02	1.457	1.18
Monte Alegre/PA	4.45	1.33	1.63	1.23

Fonte: Próprio autor.

Tabela 4 – Métricas de resultado do treinamento da cidade de Maceió no estado de Alagoas para as temperaturas mínimas.

Cidade	MSE	STD	MAE	MEDAE
Maceio/AL	0.39	0.35	0.51	0.42
Campos Sales/CE	1.52	0.85	0.88	0.62
Turiacu/MA	0.26	0.33	0.39	0.32
Belem/PA	0.80	0.616	0.651	0.38
Monte Alegre/PA	1.47	0.70	0.98	0.96

Fonte: Próprio autor.

7 CONCLUSÕES E TRABALHOS FUTUROS

Existem na literatura diversos modelos relevantes de Aprendizagem Profunda para predição de séries temporais, tais como: Ghaderi *et al.* (2017), que busca prever a velocidade do vento para aumentar a eficiência de geração de energia eólica nos EUA; e o trabalho de Grover *et al.* (2015), que busca realizar previsões do tempo por meio de um modelo híbrido que utiliza AP e obtém resultados que são totalmente relevantes para a área e que chega a oferecer melhores resultados até do que o modelo da NOAA. Após verificação das abordagens existentes, a presente pesquisa teve por objetivo investigar modelos de AP a fim de melhorar a previsão do tempo. Vale ressaltar que o código fonte do presente trabalho pode ser encontrado em github.com.

Observando os dados coletados, o modelo foi desenvolvido e verificado por meio de análises estatísticas que demonstraram resultados pertinentes à área. Com a análise, foi possível verificar que quanto melhor as bases de dados são, melhor é o desempenho do modelo.

Como trabalho futuro, observa-se a necessidade de refinamento do modelo, para que seja capaz de lidar com previsões não só para o próximo dia, mas para os próximos k dias.

REFERÊNCIAS

- ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDI, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M. *et al.* Tensorflow: Large-scale machine learning on heterogeneous distributed systems. **arXiv preprint arXiv:1603.04467**, 2016.
- ALBON, C. **Machine Learning With Python Cookbook**. [S.l.]: O'Reilly Media, Inc., 2018.
- BARROS, P. **Aprendizagem de Máquina: Supervisionada ou Não Supervisionada?** 2016. Disponível em: shorturl.at/xHRT0. Acesso em: 1 abr. 2019.
- BERGER, J. O. **Statistical decision theory and Bayesian analysis**. [S.l.]: Springer, 2013.
- BERNARDO, J. M.; BERGER, J. O. **Bayesian Statistics 6**. [S.l.]: Oxford University Press, 1998. v. 6.
- BOCANEGRA, C. W. Procedimentos para tornar mais efetivo o uso das redes neurais artificiais em planejamento de transportes. **Escola de Engenharia de São Carlos**, 2002.
- BUENO, R. de Losso da S. **Econometria de Séries Temporais**. [S.l.]: Cengage Learning, 2008.
- CASAGRANDE, A.; JUNIOR, P. S.; MENDONÇA, F. Mudanças climáticas e aquecimento global: controvérsias, incertezas e a divulgação científica. **Revista Brasileira de Climatologia**, v. 8, 2011.
- CAWLEY, G. C.; TALBOT, N. L. On over-fitting in model selection and subsequent selection bias in performance evaluation. **Journal of Machine Learning Research**, v. 11, n. Jul, p. 2079–2107, 2010.
- CHOLLET, F. *et al.* Keras: Deep learning library for theano and tensorflow.(2015). **There is no corresponding record for this reference**, 2015.
- COCHRANE, C. **Time Series Nested Cross-Validation**. 2018. Disponível em: <https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9>. Acesso em: 26 jun. 2019.
- COSTA, H. C. **Séries Temporais Multivariadas**. 2018. Disponível em: <https://rpubs.com/hudsonchavs/seriesmultivariadas>. Acesso em: 11 maio 2019.
- COSTA, V. J.; FERREIRA, M.; CORDEIRO, M. T. A. Análise de séries temporais climáticas. **Revista de Ciências Agroveterinárias**, v. 14, n. 2, p. 169–177, 2015.
- EHLERS, R. **Análise de Séries Temporais**. 2012. Disponível em: <http://bit.ly/2Jic2nA>. Acesso em: 11 maio 2019.
- ESCOBAR, G. C. J. Padrões sinóticos associados a ondas de frio na cidade de São Paulo. **Revista Brasileira de Meteorologia**, v. 22, n. 2, p. 241–254, 2007.
- FERREIRA, M. S.; GALO, M. Influência dos dados de entrada na classificação a partir de rede neural artificial. **III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação—III SIMGEO**. Anais, p. 27–30, 2010.

FOLHES, M. T.; FISCH, G. Caracterização climática e estudo de tendências nas séries temporais de temperatura do ar e precipitação em taubaté (sp). **Ambiente & Água-An Interdisciplinary Journal of Applied Science**, Universidade de Taubaté, v. 1, n. 1, p. 61–71, 2006.

GHADERI, A.; SANANDAJI, B. M.; GHADERI, F. Deep forecast: Deep learning-based spatio-temporal forecasting. **arXiv preprint arXiv:1707.08110**, 2017.

GONÇALVES, L. **Características das séries temporais**. 2018. Disponível em: <http://bit.ly/2WzH9hu>. Acesso em: 1 maio 2019.

GONÇALVES, V. H. **Dados da estação automática de Crateús**. 2017.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. 367,368,369 p. <http://www.deeplearningbook.org>. Acesso em: 26 jun. 2019.

GROVER, A.; KAPOOR, A.; HORVITZ, E. A deep hybrid model for weather forecasting. In: ACM. **Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. [S.l.], 2015. p. 379–386.

HAYKIN, S. **Redes Neurais: Princípios e Práticas**. [S.l.]: Cengage Learning, 2001.

IMRAM, A. A. **Intuition of Gradient Descent for Machine Learning**. 2017. Disponível em: <https://medium.com/abdullah-al-imran/intuition-of-gradient-descent-for-machine-learning-49e1b6b89c8b>. Acesso em: 02 jun. 2019.

INMET. **Dados da estação automática de Crateús**. 2019. Disponível em: <http://bit.ly/2JNnlNh>. Acesso em: 22 maio 2019.

KARLIN, S. **A first course in stochastic processes**. [S.l.]: Academic press, 2014.

KLUYVER, T.; RAGAN-KELLEY, B.; PÉREZ, F.; GRANGER, B.; BUSSONNIER, M.; FREDERIC, J.; KELLEY, K.; HAMRICK, J.; GROUT, J.; CORLAY, S.; IVANOV, P.; AVILA, D.; ABDALLA, S.; WILLING, C. Jupyter notebooks – a publishing format for reproducible computational workflows. In: LOIZIDES, F.; SCHMIDT, B. (Ed.). **Positioning and Power in Academic Publishing: Players, Agents and Agendas**. [S.l.], 2016. p. 87 – 90.

MALIK, H.; DAVIS, I. J.; GODFREY, M. W.; NEUSE, D.; MANSKOVSKII, S. Connecting the dots: anomaly and discontinuity detection in large-scale systems. **Journal of Ambient Intelligence and Humanized Computing**, Springer, v. 7, n. 4, p. 509–522, 2016.

MAXWELL, J. C. **Métodos de Previsão de Séries Temporais**. 2018.

MORETTIN, P. A.; TOLOI, C. M. **Análise de Series Temporais**. [S.l.]: Blucher, 2006.

NOAA. **The First Climate Model**. 2017. Disponível em: https://celebrating200years.noaa.gov/breakthroughs/climate_model/welcome.html#model. Acesso em: 1 maio 2019.

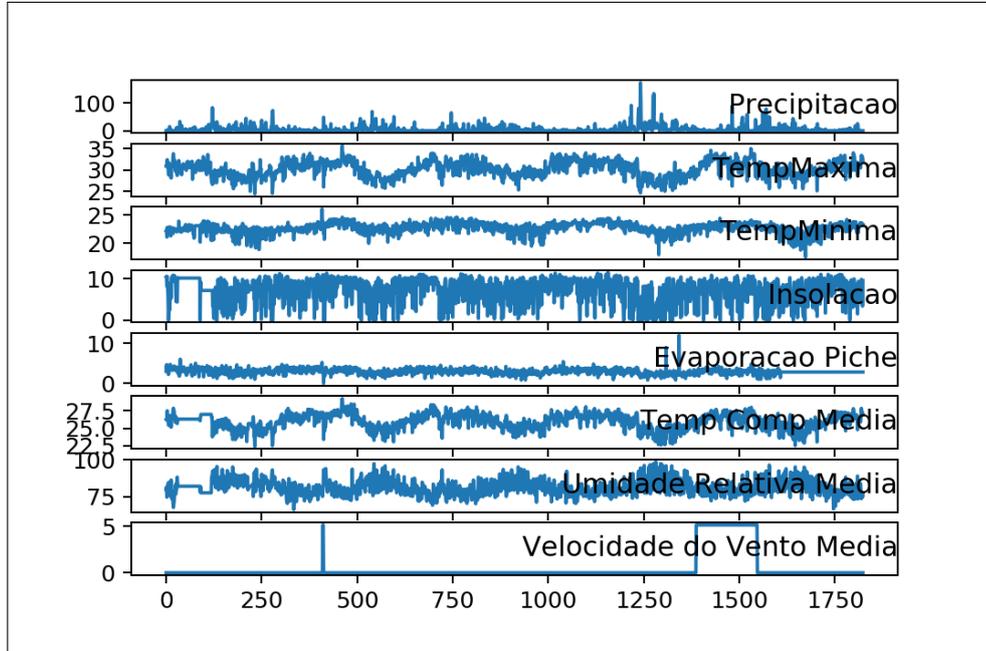
OLIVEIRA, A. C. S. de; SOUZA, A. de; LACERDA, W. S.; GONÇALVES, L. R. Aplicação de redes neurais artificiais na previsão da produção de álcool. **Ciênc. agrotec.**, v. 34, n. 2, p. 279–284, 2010.

PARREIRA, M. **Evidências paleoclimáticas do aquecimento global**. 2011. Disponível em: <http://bit.ly/2JBKrwL>. Acesso em: 13 maio 2019.

- QUACKENBUSH, J. Microarray data normalization and transformation. **Nature genetics**, Nature Publishing Group, v. 32, n. 4s, p. 496, 2002.
- SAMPAIO, G.; DIAS, P. L. da S. Evolução dos modelos climáticos e de previsão de tempo e clima. **Revista USP**, n. 103, p. 41–54, 2014.
- SANTIAGO, R.; CAJAL. **Histologie du système nerveux de l’homme et des vertébrés**. 1911.
- SANTOS, A. S. dos; COSTA, O. A. Sistema de recepção de dados do satélite meteosat-9 na secretaria de meio ambiente e recursos hídricos–sergipe: Implementação e aplicações. 2013.
- SATO, R. C. Gerenciamento de doenças utilizando séries temporais com o modelo arima. **Einstein**, v. 11, p. 128–131, 2013.
- SHEPHERD; KOCH. **Introduction to synaptic circuits**. [S.l.]: Oxford University Press, Oxford, 1990.
- STEWART, M. The limitations of machine learning. **Towards Data Science**, 2019. Disponível em: <https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9>. Acesso em: 02 dez. 2019.
- SUNDERMEYER, M.; SCHLÜTER, R.; NEY, H. Lstm neural networks for language modeling. In: **Thirteenth annual conference of the international speech communication association**. [S.l.: s.n.], 2012.
- VIANNA, L. F. d. N.; MASSIGNAN, Â. M. Distribuição espacial dos metadados temporais e qualidade dos metadados espaciais das estações hidrometeorológicas do estado de santa catarina para climatologia. **Rev. bras. meteorol**, SciELO Brasil, p. 412–425.
- WOOLDRIDGE, J. M. **Introductory econometrics: A modern approach**. Fifth editon. [S.l.]: South-Western College Publishing, a division of Thomson Learning, 2000. 8 p.
- YAN-TAK, A. **What data scientists should know about deep learning**. 2015. 30 p. Disponível em: <https://pt.slideshare.net/ExtractConf/andrew-ng-chief-scientist-at-baidu>. Acesso em: 2 maio 2019.
- ZAMBIASI, S. P. **O Neurônio Artificial**. 2011. Disponível em: <http://bit.ly/2YGQJ2z>. Acesso em: 23 maio 2019.

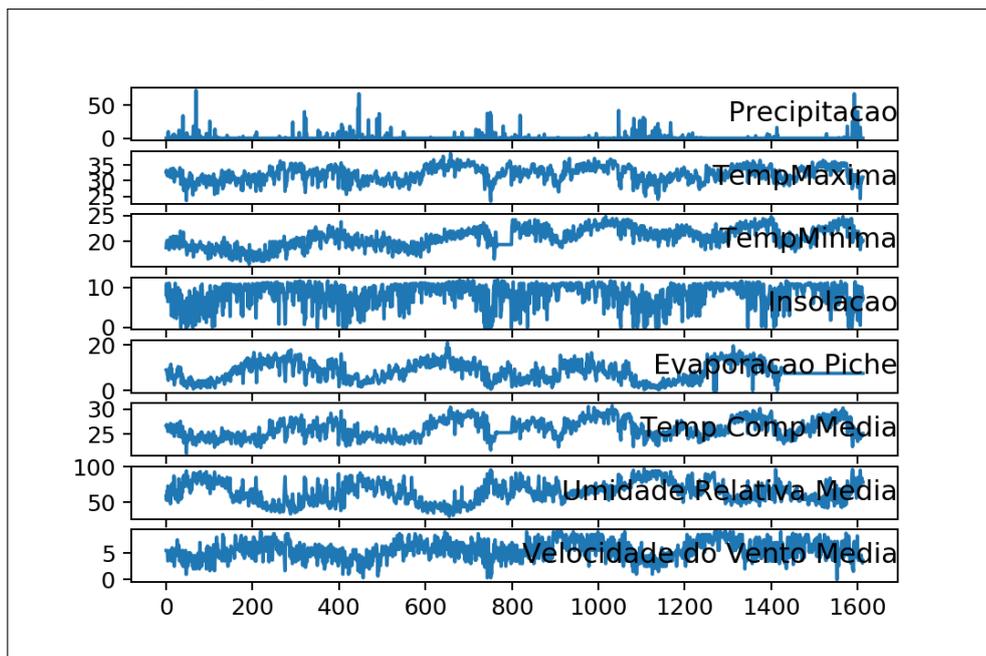
APÊNDICE A – REPRESENTAÇÃO GRÁFICA DA BASES DE DADOS UTILIZADAS

Figura 19 – Gráfico que representa a base de dados da cidade de Maceió/AL.



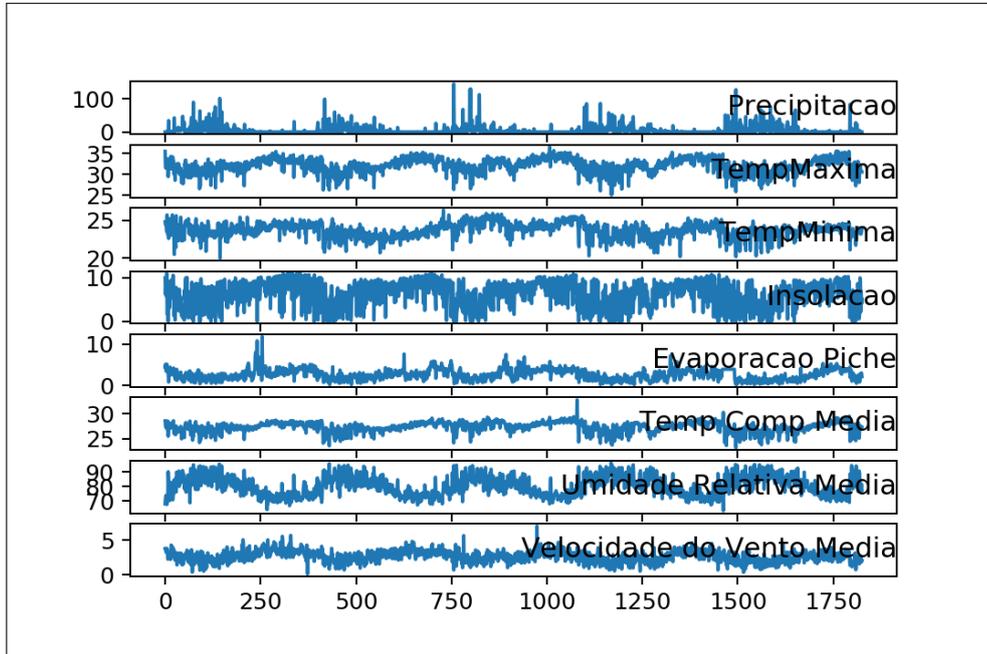
Fonte: Próprio autor.

Figura 20 – Gráfico que representa a base de dados da cidade de Campos Sales/CE.



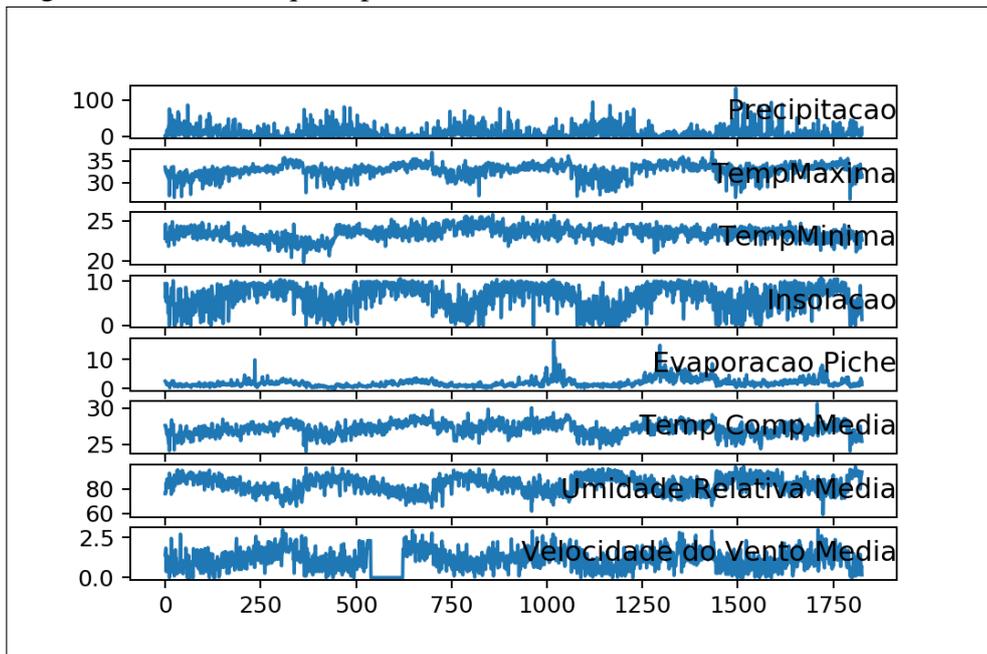
Fonte: Próprio autor.

Figura 21 – Gráfico que representa a base de dados da cidade de Turiçu/MA.



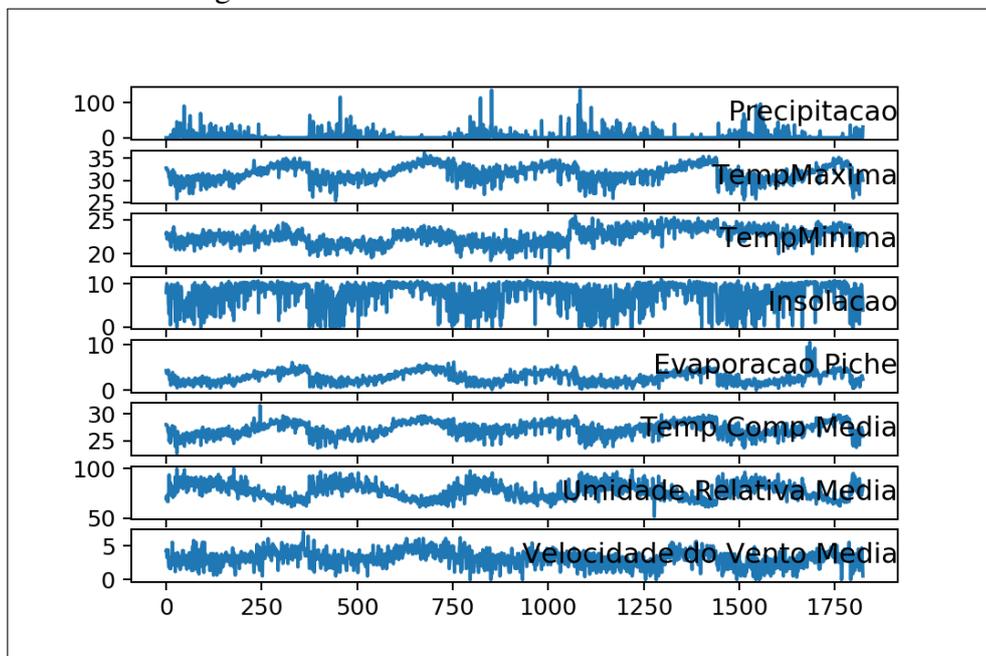
Fonte: Próprio autor.

Figura 22 – Gráfico que representa a base de dados da cidade de Belém/PA.



Fonte: Próprio autor.

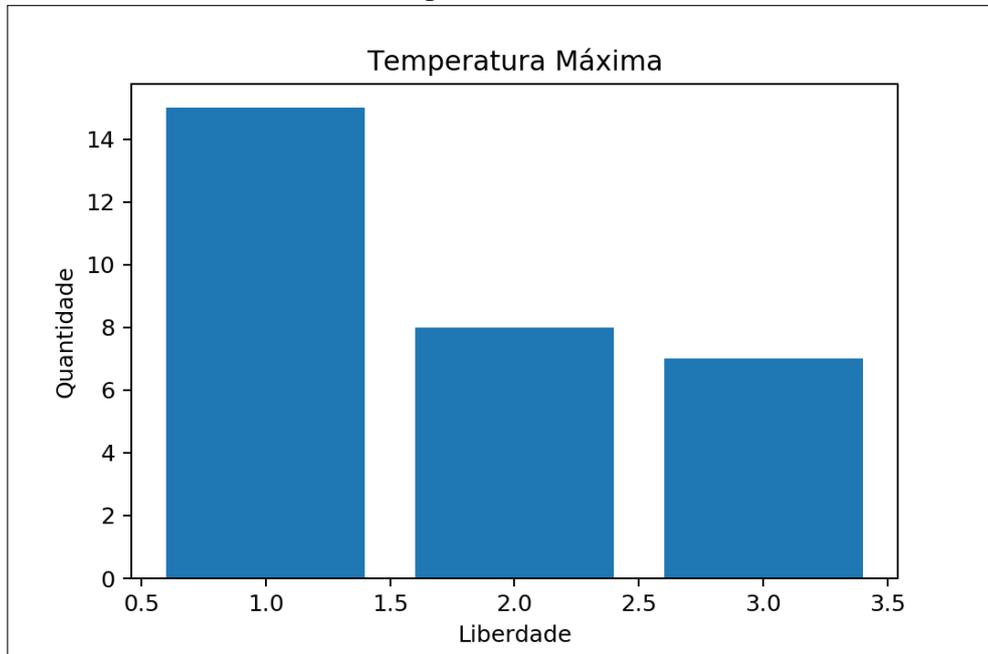
Figura 23 – Gráfico que representa a base de dados da cidade de Monte Alegre/PA.



Fonte: Próprio autor.

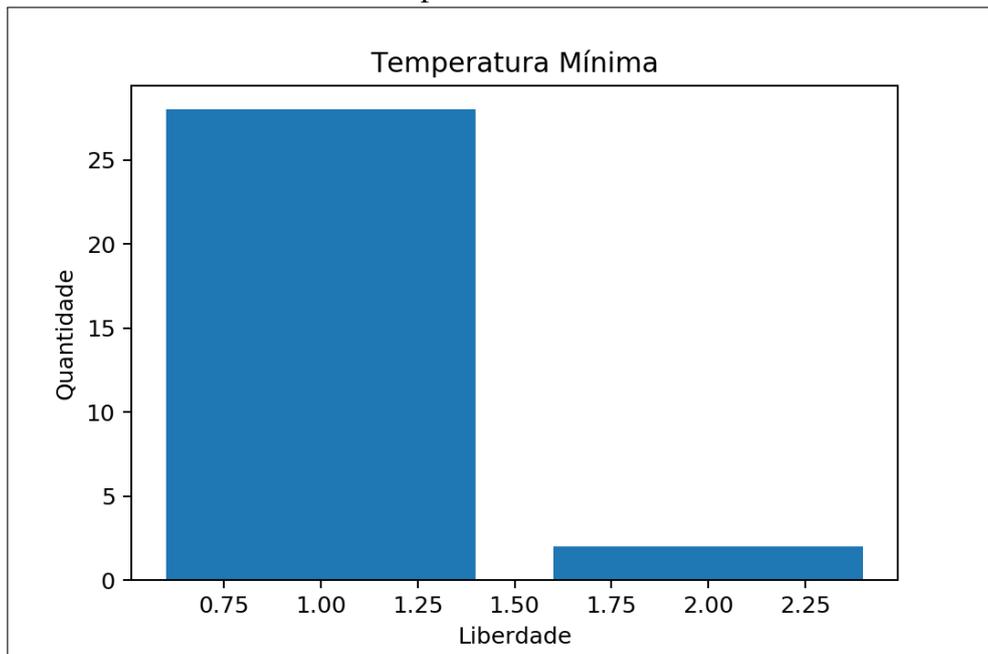
APÊNDICE B – RESULTADOS DE PRECISÃO DAS PREDIÇÕES

Figura 24 – Gráfico que representa os acertos de predições de temperatura máxima do modelo para a Cidade de Maceió/AL.



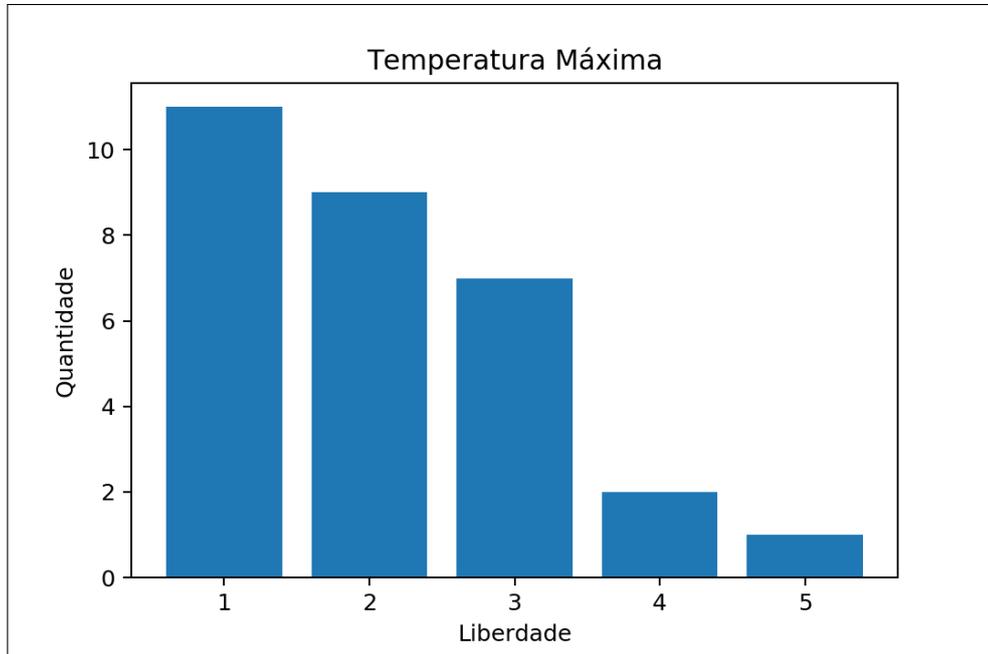
Fonte: Próprio autor.

Figura 25 – Gráfico que representa os acertos de predições de temperatura mínima do modelo para a Cidade de Maceió/AL.



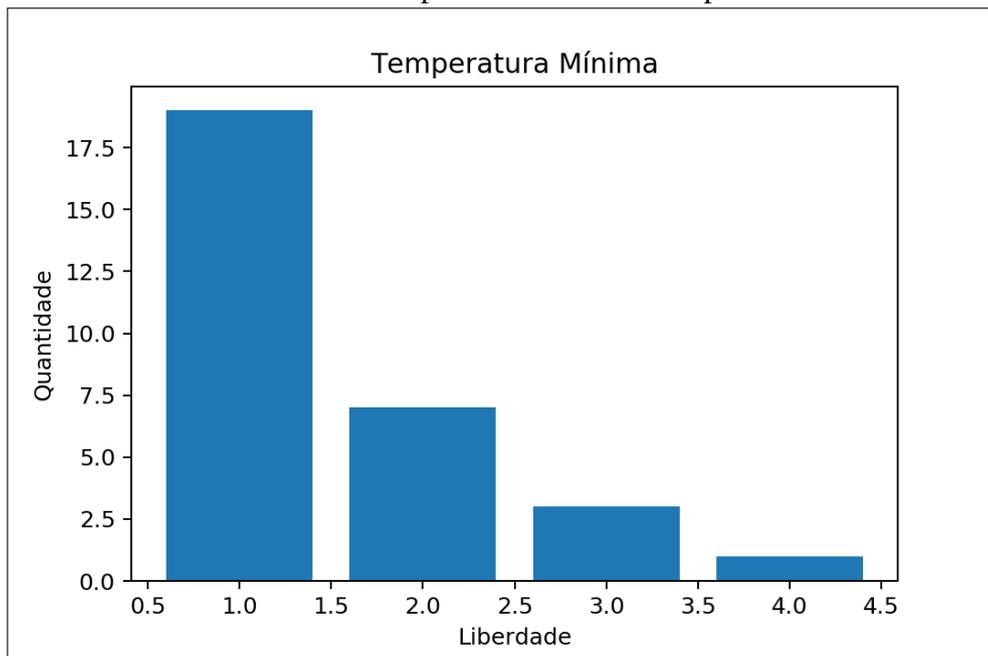
Fonte: Próprio autor.

Figura 26 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de Campos Sales/CE.



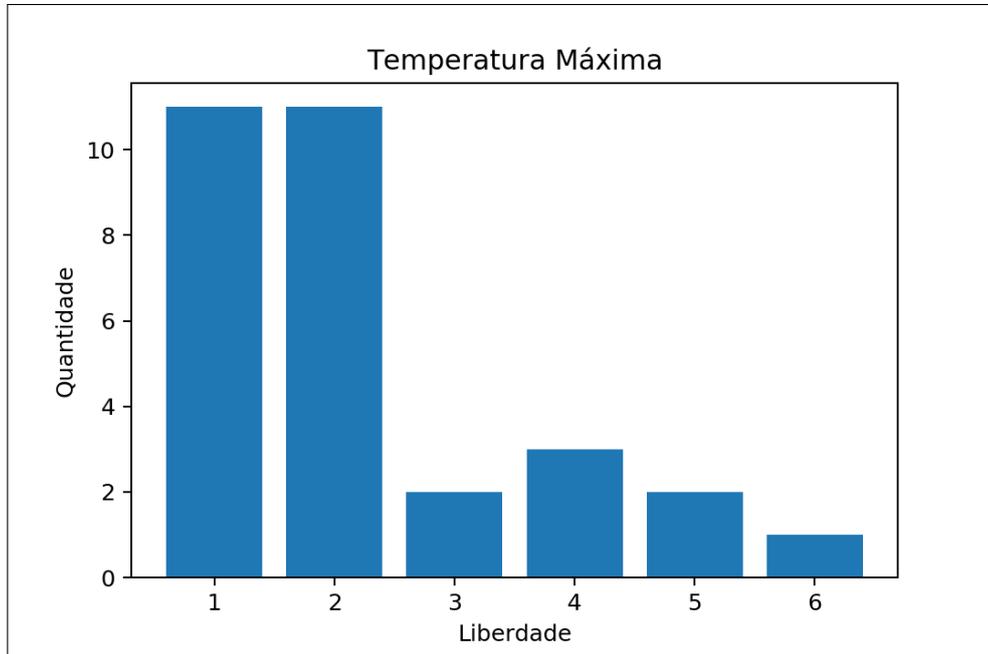
Fonte: Próprio autor.

Figura 27 – Gráfico que representa a precisão de predições de temperatura mínima do modelo para a Cidade de Campos Sales/CE.



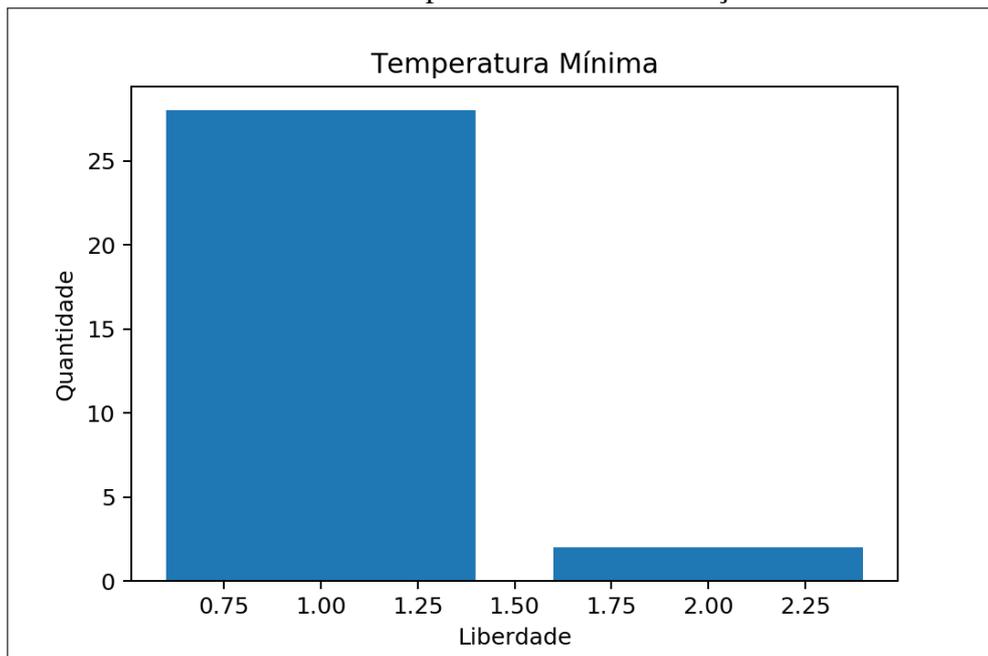
Fonte: Próprio autor.

Figura 28 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de Turiiaçu/MA.



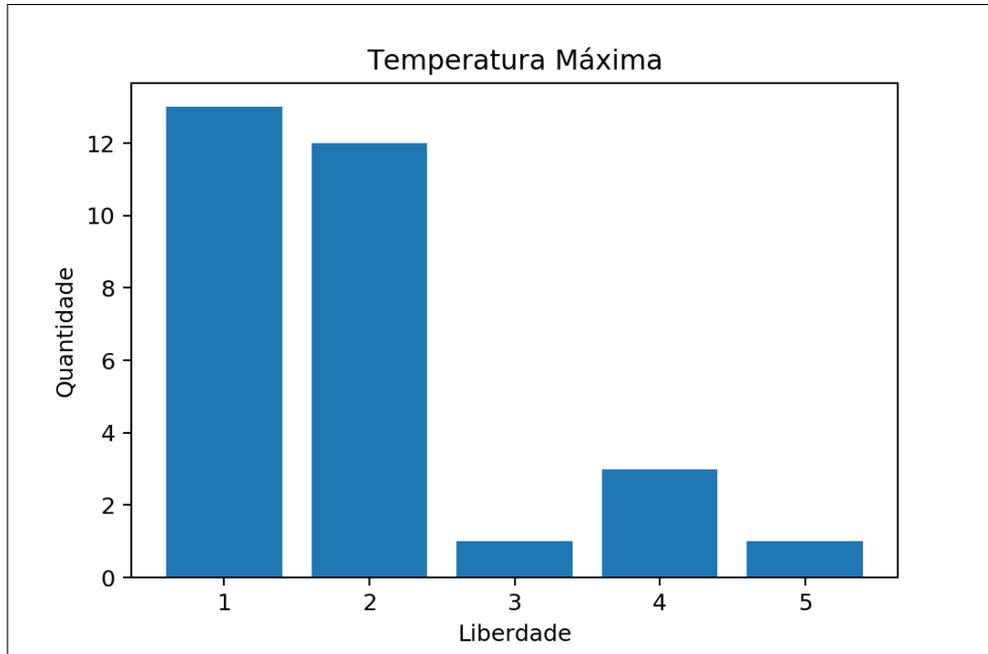
Fonte: Próprio autor.

Figura 29 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de Turiiaçu/MA.



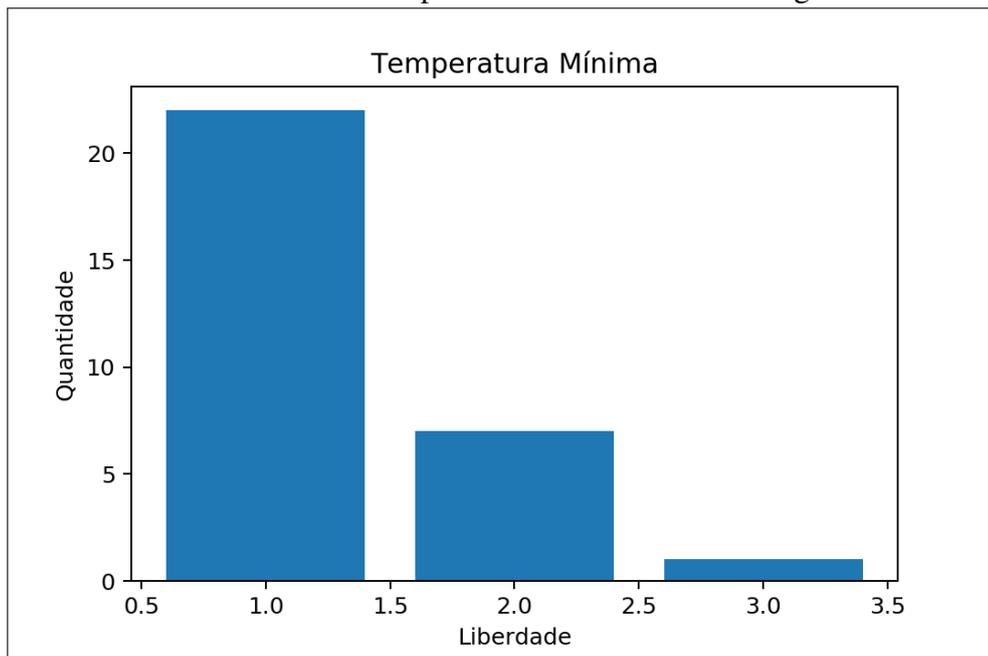
Fonte: Próprio autor.

Figura 30 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de BÉlem/PA.



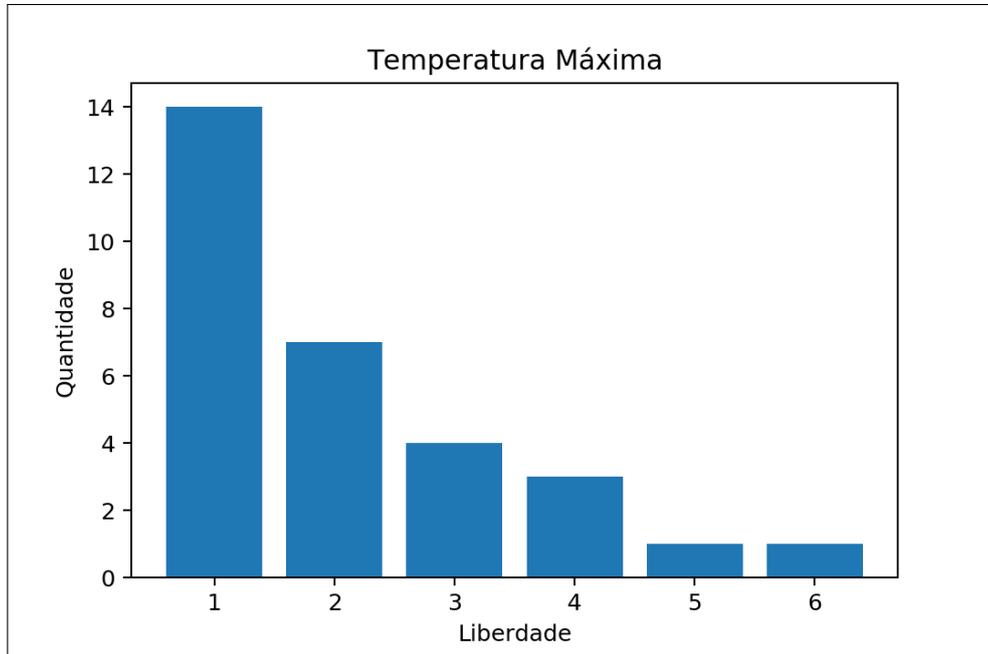
Fonte: Próprio autor.

Figura 31 – Gráfico que representa a precisão de predições de temperatura mínima do modelo para a Cidade de Maceió/Alagoas.



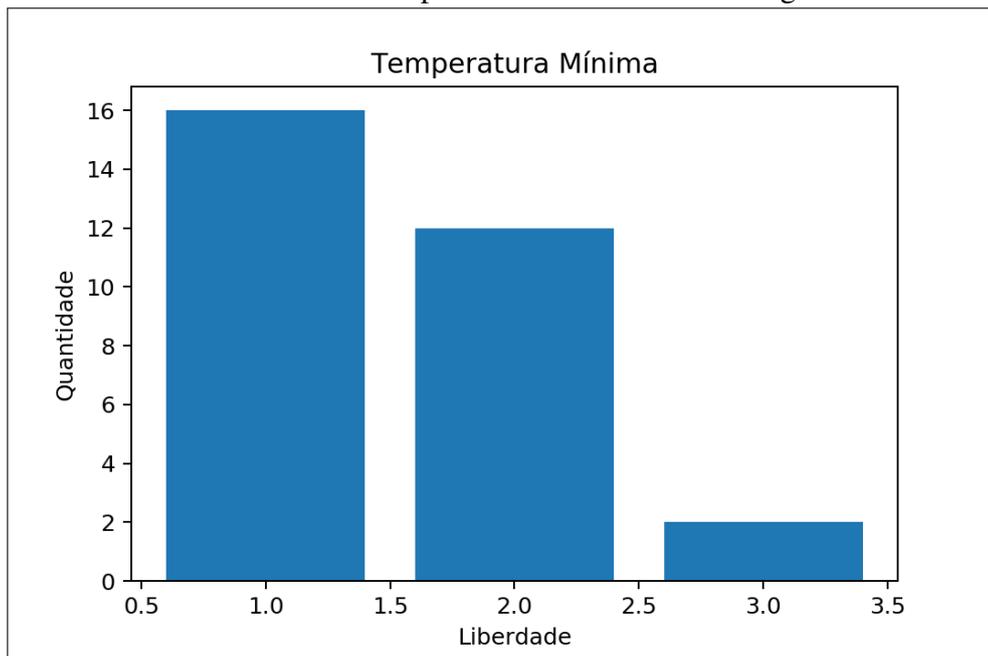
Fonte: Próprio autor.

Figura 32 – Gráfico que representa a precisão de predições de temperatura máxima do modelo para a Cidade de Monte Alegre/PA.



Fonte: Próprio autor.

Figura 33 – Gráfico que representa a precisão de predições de temperatura mínima do modelo para a Cidade de Monte Alegre/PA.



Fonte: Próprio autor.