UNIVERSIDADE DE CAXIAS DO SUL

**ROAN ROBERTO BROLESE** 

# PREVISÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAICA UTILIZANDO MÉTODO DE APRENDIZADO DE MÁQUINA

CAXIAS DO SUL 2019

## **ROAN ROBERTO BROLESE**

# PREVISÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAICA UTILIZANDO MÉTODO DE APRENDIZADO DE MÁQUINA

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do título de Bacharel em Engenharia Elétrica pela Universidade de Caxias do Sul.

Orientador Prof. Me. André Bernardes Michel

CAXIAS DO SUL 2019

## **ROAN ROBERTO BROLESE**

# PREVISÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAICA UTILIZANDO MÉTODO DE APRENDIZADO DE MÁQUINA

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do título de Bacharel em Engenharia Elétrica pela Universidade de Caxias do Sul.

Aprovado em 04/07/2019

**Banca Examinadora** 

Prof. Me. André Bernardes Michel (orientador) Universidade de Caxias do Sul – UCS

Prof. Ma. Andréa Cantarelli Morales Universidade de Caxias do Sul – UCS

Prof. Me. Tiago Cassol Severo Universidade de Caxias do Sul – UCS

Dedico este trabalho aos meus pais, Roberto e Angela, pelo amor, carinho e apoio, principalmente pela confiança nas decisões e escolhas que fiz, acreditando nas minhas possibilidades.

#### AGRADECIMENTOS

Ao longo desta jornada, algumas pessoas nos marcam pelo caminho, servem como exemplos, nos ajudam diante de qualquer situação.

Agradeço primeiramente a Deus por tudo.

Agradeço aos meus pais Roberto André Brolese e Angela da Silva Faria, pelo amor, incentivo, e pela confiança em mim depositada. Agradeço a minha irmã Luane Faria Brolese, por todo o apoio e cumplicidade. Agradeço aos meus avós Primo Brolese (in memoriam) e Juracy Amélia Conte Brolese pela criação, pelo carinho e pelos valores transmitidos.

Agradeço a Leonir Zanotti, pelo apoio, pela atenção e pela ajuda em momentos variados.

A minha namorada Nathália Grazziotin, pela paciência, pelo carinho, por ser meu porto seguro, a pessoa com quem divido minhas angústias, minhas dúvidas, que entende meus anseios, estando sempre ao meu lado nas minhas decisões, este caminho seria mais complicado de se trilhar sem você por perto.

Agradeço ao meu orientador Me. André Bernardes Michel, pelos conhecimentos transmitidos, pelo auxílio, pela paciência e pela excelência na orientação deste trabalho, bem como nas aulas por ele ministradas.

A empresa Akron Eletro Eletrônica Ltda, pelo suporte, pela flexibilidade e pelo apoio ao longo desta trajetória.

Agradeço, por fim, aos amigos e colegas, pelos conhecimentos, pelo coleguismo e pelo auxílio ao longo do curso.

"Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus, que um dia você chega lá. De alguma maneira você chega lá." (Ayrton Senna)

#### **RESUMO**

Com o aumento da inserção de geração fotovoltaica juntamente com o crescimento de conexões junto a rede, diferentes métodos e algoritmos, capazes de prever esta geração em diversos horizontes vem sendo aplicados pelos pesquisadores. Dentre estes algoritmos destaca-se a massiva utilização do aprendizado de máquina através de redes neurais artificiais, devido a sua capacidade de lidar com elementos não lineares. O presente trabalho intenta em realizar uma previsão de geração fotovoltaica para a cidade de Garibaldi, RS, para um horizonte de seis meses a frente, por meio da modelagem de uma rede neural artificial. Os dados de entrada desta rede são compostos por conjuntos de dados históricos climáticos em complemento a um conjunto de dados históricos de geração solar fotovoltaica que serão usados como alvos. A classificação das variáveis de entrada é realizada através da regressão stepwise, que elenca as variáveis com maior relação com a resposta. A modelagem conta com o auxílio do software numérico Matlab, para a construção do algoritmo e a classificação dos dados. Para o treinamento desta rede neural artificial é empregado o método de aprendizado supervisionado de retropropagação do erro, utilizando a função de treinamento bayesian regularization, visando reduzir o erro da previsão. A rede neural artificial exibiu um bom desempenho, atingindo um MAPE de 12,97 %. Também neste contexto são apresentadas as comparações de desempenho das redes neurais para diferentes horizontes, com resultados alcançados estando de acordo com os encontrados na bibliografia.

**Palavras – chave:** Geração fotovoltaica, Previsão, Rede neural artificial, Aprendizado de máquina.

### ABSTRACT

Due to the rise of photovoltaic generation together with the growth of network connections, different methods and algorithms, which are able to predict this generation in many time horizons, have been used by researchers. The main algorithm adopts the machine learning using artificial neural networks. This method is used because of its capability of working with nonlinear elements. This work has the objective of developing a photovoltaic power generation forecasting in Garibaldi, RS, with a time horizon of six months, creating an artificial neural network. The input data of this network is composed of climatic historical data sets in addition to a set of historical solar photovoltaic generation data, which will be used as targets. The input variable classification is performed through stepwise regression, which lists the highest relation between input variables and the response. The modeling is accomplished by the numerical software Matlab. The software develops the algorithm and ranks the data. To train the artificial neural network the supervised learning method of error backpropagation was used. This method applies the training function bayesian regularization, aiming to reduce the forecast error. The artificial neural network obtained a good performance, reaching a 12.97% MAPE. The comparisons of neural network performance for different time horizons are also presented in this context. The results achieved are in agreement with those found in the bibliography.

**Keywords:** Photovoltaic power generation, Forecasting, Artificial neural network, Machine learning.

# LISTA DE ILUSTRAÇÕES

Figura 1 - Fluxograma de Redes Neurais Artificiais	21
Figura 2 - Fluxograma Máquina de Vetores de Suporte	22
Figura 3 - Fluxograma K-ésimo Vizinho Mais Próximo	23
Figura 4 - Fluxograma Floresta Aleatória	24
Figura 5 - Mapa do Rio Grande do Sul de Irradiação Global Horizontal	26
Figura 6 - Influência de Fatores Climáticos	27
Figura 7 - Erro relacionado aos dois modelos de RNA	32
Figura 8 - Comparação entre valor real e o previsto	
Figura 9 - Comparação entre os Modelos de Previsão	34
Figura 10 - Diagrama da RNA Auto Regressiva	35
Figura 11 - Desempenho da Previsão com RNA Recorrente	35
Figura 12 - Diagrama Estrutural uma Rede Neural Artificial	
Figura 13 - Método de Regressão Stepwise	
Figura 14 - Curva característica função sigmoide	45
Figura 15 - Interface Gráfica do Power Data Access Viewer	49
Figura 16 - Fluxograma do Modelo de RNA Proposto	52
Figura 17 - Esquemático Rede Neural PDAV	56
Figura 18 - Regressão Treinamento PDAV	57
Figura 19 - Saída Treinamento PDAV	58
Figura 20 - Regressão Previsão PDAV	58
Figura 21 - Saída Previsão PDAV	59
Figura 22 - Esquemático Rede Neural PC	60
Figura 23 - Regressão Treinamento PC	61
Figura 24 - Saída Treinamento PC	61
Figura 25 - Regressão Previsão PC	62
Figura 26 - Saída Previsão PC	63
Figura 27 - Esquemático Rede Neural Mista	64
Figura 28 - Regressão Treinamento Mista	65
Figura 29 - Saída Treinamento Mista	65
Figura 30 - Regressão Previsão Mista	66
Figura 31 - Saída Previsão Mista	66
Figura 32 - Previsão Realizada para julho, MAPE 22,96 %	69

Figura 33 - Previsão Realizada	para novembro, MAPE 9,09 %	69
--------------------------------	----------------------------	----

## LISTA DE TABELAS

Tabela 1 - Regressão Stepwise Rede Neural PDAV	55
Tabela 2 - Regressão Stepwise Rede Neural PC	59
Tabela 3 - Regressão Stepwise Rede Neural Mista	63
Tabela 4 - Comparativo dos erros de treinamento entre as três RNAs	67
Tabela 5 - Comparativo dos erros de previsão entre as três RNAs	67
Tabela 6 - Tabela Comparativa de MAPE	68
Tabela 7 - MAPE Calculado nos Diferentes Treinamentos	70
Tabela 8 - Comparação de MAPE para Previsão a Médio Prazo	70
Tabela 9 - Comparação de MAPE para Previsão a Longo Prazo	71
Tabela A.1 - Amostra de Dados Obtidos pelo Power Data Access Viewer	80
Tabela B.1 - Amostra dos Dados Obtidos pelo Portal Clima	81
Tabela C.1 - Amostra de Dados Históricos de Geração	

## LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
ABSOLAR	Associação Brasileira de Energia Solar Fotovoltaica
ARMA	Auto Regressive Moving Average
ARIMA	Auto Regressive Integrated Moving Average
INMEP	Instituto Nacional de Meteorologia
BDMEP	Banco de Dados Meteorológicos para Ensino e Pesquisa
KNN	K-Nearest Neighbor
MAPE	Mean Absolute Percentual Error
MSE	Mean Square Error
NARX	Nonlinear Auto Regressive exogenous
NASA	National Aeronautics and Space Administration
NOCT	Nominal Operating Cell Temperature
ONS	Operador Nacional do Sistema Elétrico
RMSE	Root Mean Square Error
RNA	Rede Neural Artificial
SARIMA	Seasonal Auto Regressive Integrated Moving Average
SIN	Sistema Interligado Nacional
STC	Standard Test Conditions
SVM	Support Vector Machine

1	INTRODUÇÃO14
1.1	OBJETIVOS
1.1.1	Objetivo Geral15
1.1.2	Objetivos Específicos15
1.2	ORGANIZAÇÃO DO TRABALHO16
1.3	LIMITAÇÔES16
2	FUNDAMENTAÇÃO TEÓRICA17
2.1	PREVISÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAÍCA17
2.2	MÉTRICAS DE VERIFICAÇÃO17
2.3	APRENDIZADO DE MÁQUINA18
2.4	MÉTODOS DE PREVISÃO DE ENERGIA FOTOVOLTAICA19
2.4.1	Series Temporais Lineares20
2.4.2	Redes Neurais Artificiais20
2.4.3	Máquina de Vetor de Suporte21
2.4.4	K-ésimo Vizinho Mais Próximo23
2.4.5	Floresta Aleatória24
2.5	PARÂMETROS PARA A PREVISÃO DE GERAÇÃO FOTOVOLTAICA25
2.5.1	Irradiação Solar25
2.5.2	Condições Climáticas26
2.6	SOFTWARES UTILIZADOS PARA APRENDIZADO DE MÁQUINA28
2.6.1	Matlab28
2.6.2	GNU Octave29
2.6.3	Neural Designer
2.6.4	Python
2.6.5	Weka
2.7	REVISÃO DO ESTADO DA ARTE
3	METODOLOGIA

3.1	MÉTODOS E HORIZONTE DE PREVISÃO A SEREM UTILIZADOS	36
3.2	REDES NEURAIS ARTIFICIAIS	37
3.2.1	Estrutura	37
3.2.2	Classificação	38
3.2.3	Pré-processamento de Dados	39
3.2.3.1	Classificação de Variáveis	39
3.2.3.2	Limpeza de Dados	43
3.2.3.3	Normalização	43
3.2.3.4	Função de Ativação	44
3.2.4	Treinamento	45
3.2.4.1	Função de Treinamento	46
3.2.5	Validação Cruzada	48
3.3	ESTUDO DE CASO	49
3.3.1	Aquisição de dados climáticos	49
3.3.2	Dados históricos de geração fotovoltaica	50
3.3.3	Parametrização de Dados	51
3.3.4	Modelagem das RNAs	52
4	RESULTADOS	55
4.1	PREVISÕES REDE NEURAL PDAV	55
4.2	PREVISÕES REDE NEURAL PC	59
4.3	PREVISÕES REDE NEURAL MISTA	53
4.4	ANÁLISE COMPARATIVA DE PRECISÂO ENTRE AS REDES NEURAIS.	67
4.5	PREVISÕES REALIZADAS PARA DIFERENTES TEMPORADAS	58
4.6	COMPARAÇÃO COM A BIBLIOGRAFIA	70
5	CONCLUSÕES	73
REFERÊN	NCIAS	74
APÊNDIC	CE A - DADOS OBTIDOS pelo POWER DATA ACCESS VIEWER	80
APENDIC	LE B - DADOS OBTIDOS PELO PORTAL CLIMA	51

APÊNDICE C - DADOS HISTÓRICOS DE GERAÇÃO	82
--	----

## 1 INTRODUÇÃO

A geração de energia elétrica através da energia solar vem crescendo globalmente. Este crescimento se deve em grande parte pela busca por energias renováveis em substituição ao uso de combustíveis fosseis para a geração elétrica. (ALZAHRANI et al., 2017). Segundo Sansa et al., (2014), o aumento vertiginoso na utilização de fontes de energias renováveis no mundo acompanha o crescimento da economia juntamente com o aumento da demanda.

De acordo com Chaturvedi e Isha (2016), até 2050 a demanda de energia elétrica no mundo será maior que o dobro da demanda atual. Em virtude desta previsão é fundamental a ampliação de geração de energia elétrica através de fontes renováveis. Para Haque, Nehrir e Mandal (2013), com o advento das redes inteligentes, os sistemas de geração provenientes de fontes renováveis de energia são conectados à rede de forma mais eficiente. Em 2012 através da norma regulamentadora número 482 a ANEEL, Agência Nacional de Energia Elétrica, elaborou critérios para a integração de sistemas de micro e minigeração distribuída. (ANEEL, 2012).

Segundo Barros, Suaia e Koloszuk (2019), o Brasil possui mais de 2000 MW de potência instalada em usinas centralizadas de geração solar fotovoltaica. Ainda, segundo os autores, existem mais de 1700 MW em projetos no Brasil, tanto em fase de desenvolvimento quanto em construção, com previsão de início de operação em 2022. De acordo com a ABSOLAR (2019), o Brasil possui cerca de 49.177 sistemas fotovoltaicos conectados à rede, totalizando 500 MW de potência instalada em sistemas de microgeração e minigeração distribuída. Minas Gerais é o estado brasileiro com maior capacidade instalada, com 109,5 MW. O segundo colocado é o estado do Rio Grande Sul com 78,8 MW, o que corresponde a 15,7 % da potência instalada em sistemas de micro e minigeração no país.

Liu e Zhang (2016), enfatizam que com o aumento da inserção de energia fotovoltaica à rede, aumenta também a imprevisibilidade do sistema, causando problemas no planejamento de médio e longo prazo, além de problemas relacionados à operação. Ao passo que, para Singh et al. (2015) a integração de um sistema de geração fotovoltaico a rede, visto que esta geração depende de fatores climáticos, principalmente da radiação solar, pode auxiliar no planejamento, na operação e na manutenção destes sistemas. Ainda, segundo Cheng, Ge e Cao (2012), as oscilações e a variabilidade na geração dos sistemas fotovoltaicos conectados à rede impactam diretamente nos sistemas elétricos na sua regulação e na sua proteção. A realização de uma previsão precisa de geração pode auxiliar no controle de operações e nas manutenções relacionadas a sistemas com fontes de geração fotovoltaicos a ele conectados. Majumder, Behera e Nayak (2017), também relatam a importância das previsões para projetos de novas usinas, analisando dados climáticos de uma determinada região, sendo assim possível planejar a inserção destas novas usinas através de estimativas prévias.

Com a ampliação destas fontes geradoras fotovoltaicas e o aumento de conexões na rede, surge assim uma necessidade de prever a energia que será gerada, buscando garantir a estabilidade no sistema de distribuição. (XIYUN; SONG, 2014). Diferentes métodos vêm sendo utilizados pelos pesquisadores para a realização desta previsão. Usualmente estes métodos utilizam algoritmos para a mineração de dados relevantes a geração fotovoltaica. Fatores climáticos como temperatura ambiente, umidade relativa, radiação solar e nebulosidade são alguns exemplos de condições influentes na geração fotovoltaica. (ABEDIN et al.,2017).

Ainda segundo Abedin et al. (2017), a mineração destes dados pode ser realizada através de métodos de aprendizado de máquina, que procuram identificar padrões entre os dados analisados no intuito de aprender com estes dados, traduzindo este conhecimento adquirido em modelos preditivos. Para Liu e Zhang (2016) há inúmeras previsões realizadas utilizando estes métodos de aprendizado, sendo os modelos mais utilizados os auto-regressivos, como o ARMA, as redes neurais artificiais e as máquinas de vetores de suporte.

#### 1.1 OBJETIVOS

#### 1.1.1 Objetivo Geral

Este trabalho tem como objetivo realizar uma previsão de geração fotovoltaica para a região de Garibaldi utilizando redes neurais artificiais, baseado em um método de aprendizado de máquina.

### 1.1.2 Objetivos Específicos

Durante a execução deste trabalho existem alguns temas que serão abordados de maneira detalhada, visando a construção de um modelo de previsão de geração. Estes temas envolvem determinados objetivos específicos elencados a seguir.

- a) Analisar métodos de previsão de geração fotovoltaica para a integração em um modelo de redes neurais;
- b) Elencar parâmetros que envolvem a previsão de geração fotovoltaica;
- c) Pesquisar software relacionado ao aprendizado de máquina a ser utilizado, comparando e justificando o mesmo;
- d) Elaborar uma conceituação teórica do método escolhido, caracterizando e justificando a escolha deste método;
- e) Modelar um sistema de previsão de geração fotovoltaica para a região de Garibaldi.

#### 1.2 ORGANIZAÇÃO DO TRABALHO

Este texto é constituído por 5 capítulos, no primeiro capítulo é apresentada a justificativa do trabalho bem como os objetivos geral e específicos. O capítulo seguinte aprofunda-se no desenvolvimento do tema com a revisão bibliográfica a respeito dos principais métodos estatísticos utilizados para a previsão de geração fotovoltaica, incluindo os modelos auto regressivos bem como os principais métodos de aprendizado de máquina aplicáveis a este tema. O estudo presente neste texto também relaciona os parâmetros climáticos geralmente abordados pelos autores e os *softwares* numéricos que permitem a construção dos modelos de aprendizado de máquina. No terceiro capítulo é apresentado o estudo de caso juntamente com a metodologia empregada. No capítulo 4 são apresentados os resultados obtidos, e por fim, as considerações finais são apresentadas no capítulo 5.

## 1.3 LIMITAÇÔES

O presente trabalho abordará somente um método de previsão, foge do escopo deste trabalho a comparação entre resultados utilizando diferentes métodos. Os dados necessários para a construção do modelo de previsão serão obtidos com o auxílio de sistemas de monitoramento externo. Já os dados históricos necessários serão adquiridos através de sistemas de monitoramento meteorológicos, como também através de monitoramento de inversores, não sendo criados ou capturados pelo autor. A análise dos dados limita-se a região de Garibaldi.

## 2 FUNDAMENTAÇÃO TEÓRICA

## 2.1 PREVISÃO DE GERAÇÃO DE ENERGIA FOTOVOLTAÍCA

Com a expansão no consumo de energia elétrica e o incremento de novas fontes de energia renovável na geração distribuída, a geração fotovoltaica vem se destacando mundialmente. Neste cenário, os métodos de previsão de geração ganham maior relevância devido à necessidade do aumento da confiabilidade para os sistemas de geração solar, porém a dependência com as condições climáticas, principalmente irradiação solar, criam empecilhos na precisão desta previsão. (YANG; CAO; XIU, 2014). Esta precisão traz vantagens para o sistema de geração melhorando a distribuição, a eficiência e o monitoramento. (GIGONI et al., 2018).

Segundo Majumder, Behera e Nayak (2017), há diferentes horizontes de previsão que podem ser utilizados, sendo eles classificados justamente conforme o período em que a previsão será realizada. Desta forma, cada um destes horizontes tem a sua relevância em áreas distintas envolvidas na geração de energia fotovoltaica. Os horizontes comumente utilizados para previsão e suas aplicações são descritos abaixo. (MAJUMDER; BEHERA; NAYAK, 2017).

- a) Previsão a Curto Prazo: O período de previsão vai desde poucas horas até um ou dois dias a frente. Muito utilizado no mercado de energia e nas operações no sistema elétrico de potência;
- b) Previsão a Médio Prazo: Com um período de previsão que pode se estender desde poucos dias até semanas a frente. Emprega-se este método para agendar manutenção nos sistemas de geração;
- c) Previsão a Longo Prazo: Abrange um período ainda mais amplo de previsão que pode variar desde poucos meses até anos à frente. Este tipo de previsão ajuda na avaliação de futuros parques solares através de uma série de dados históricos como irradiação solar entre outras condições climáticas.

## 2.2 MÉTRICAS DE VERIFICAÇÃO

Na realização de qualquer previsão é fundamental avaliar a qualidade dos resultados obtidos, para isso, existem técnicas que são executadas para a análise destes resultados que calculam o erro entre os valores previstos e os valores atuais. Entre as técnicas utilizadas para

esta análise se destacam o MAPE, erro percentual absoluto médio, e o RSME, erro da raiz média quadrática, utilizadas tanto em Qiu et al. (2014) quanto Lahouar, Mejri e Slama (2017). O MAPE está representado pela Equação 1 e o RMSE, definido pela Equação 2. Em Alzahrani (2017), outra técnica utilizada é o erro da média quadrática, expresso pela Equação 3.

$$MAPE = 100 \times \frac{1}{n} \sum_{i=n}^{n} \frac{|y_i - y_i'|}{y_i}$$
(1)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=n}^{n} (y_i - y_i)^2}$$
(2)

$$MSE = \frac{1}{n} \times \sum_{i=1}^{n} (y_i - y_i)^2$$
(3)

Onde:

- ·  $y'_i$  é o valor previsto;
- ·  $y_i$  é o valor atual;
- · n é o número de observações.

A métrica de avaliação usualmente utilizada nas previsões tanto de demanda quanto de geração é o MAPE, devido basicamente a simplicidade de interpretação da dimensão do erro quanto este é representado de forma proporcional a um valor real. (SANGRODY et al., 2017). Assim sendo, o erro absoluto percentual médio será a métrica utilizada neste trabalho para a quantificação e validação dos resultados obtidos.

## 2.3 APRENDIZADO DE MÁQUINA

Conforme Liu e Zhang (2016), a definição de aprendizado de máquina é basicamente a modelagem de um algoritmo, através de auxilio computacional, com o objetivo de tornar esse algoritmo capaz de aprender por experiência. Estes métodos vêm sendo utilizados na área de previsão de energia fotovoltaica, encontrando padrões entre a série de dados históricos inseridos afim de prever dados futuros.

Para Lopez (2010), o aprendizado pode ser classificado quanto ao seu tipo, podendo ser supervisionado, não supervisionado e por reforço. As características de cada um são expostas abaixo:

- a) Aprendizado supervisionado: os resultados obtidos no processo são comparados com dados reais, normalmente comparados a dados utilizados como entrada do sistema. A saída do sistema é monitorada durante o processo de treinamento. (LOPEZ, 2010).
- b) Aprendizado não supervisionado: este tipo de aprendizado é usualmente aplicado onde não há necessidade de comparação dos resultados, ou seja, a saída do sistema não tem um resultado certo ou errado. É frequentemente empregado na classificação de dados. (PALUSZEK; THOMAS, 2017).
- c) Aprendizado por reforço: a avaliação dos resultados obtidos é caracterizada pela correção de forma heurística, não existe treino o sistema se auto corrige conforme o seu desempenho. (LOPEZ, 2010).

O treinamento é um processo empregado em qualquer modelo de aprendizado supervisionado onde parte dos dados adquiridos são reservados para o treino e parte para a validação do modelo construído. Geralmente o conjunto de dados destinados ao treinamento é maior que o conjunto de validação, e o modelo é treinado repetidas vezes até que o valor esteja abaixo do erro esperado. (PALUSZEK; THOMAS, 2017).

## 2.4 MÉTODOS DE PREVISÃO DE ENERGIA FOTOVOLTAICA

A previsão de energia fotovoltaica pode ser dividida em três métodos distintos, os físicos, que utilizam dados de satélites e de sensores para realizar previsões, os métodos estatísticos, caracterizados pela utilização de dados históricos e análise de padrões. Este último pode ser classificado em dois grupos, séries temporais e aprendizado de máquina. O terceiro método é híbrido, sendo uma junção de dois ou mais métodos de previsão, com o objetivo de melhorar a precisão eliminando as fraquezas das previsões individuais (CHATURVEDI; ISHA, 2016).

Como o cerne deste trabalho está voltado aos métodos estatísticos de previsão de energia fotovoltaica, a seguir será demonstrado alguns dos diferentes métodos que podem ser utilizados para a previsão desta geração.

#### 2.4.1 Series Temporais Lineares

São métodos de previsão empregados em modelos lineares. Estes métodos manipulam um conjunto de dados históricos caracterizados por uma sazonalidade de aquisição. A Equação 4 representa um modelo genérico de séries temporais. (SWAMI; MENDEL, 1990).

$$\sum_{k=0}^{p} a(k) y(n-k) = \sum_{k=0}^{q} b(k) u(n-k)$$
(4)

Onde:

· u(n) é entrada do sistema;

· os dados  $p \in q$  são conhecidos;

· y(n) é a saída do sistema.

Entre alguns dos métodos utilizados nos cálculos de séries temporais estão, AR do inglês *Auto Regressive*, MA do inglês *Moving Average*, ARMA, *Auto Regressive Moving Average*, sendo este um composto dos dois primeiros. (ALANAZI; ALANAZI; KHODAEI, 2016).

ARIMA, Auto Regressive Integrated Moving Average, e SARIMA, Seasonal Auto Regressive Moving Average, são outros exemplos de séries temporais, sendo o segundo uma variação sazonal do primeiro. Ainda, o ARIMA é um dos métodos de série temporal mais implementados, sendo uma variante mais ampla do ARMA, onde a diferença está nos dados tratados. Por último, o ARMA lida com séries temporais estacionárias e ARIMA manipula séries temporais não estacionárias. (SORKUN; PAOLI; INCEL, 2017).

#### 2.4.2 Redes Neurais Artificiais

A ideia das redes neurais artificiais, ANN, *artificial neural network*, é simular o comportamento dos circuitos neurais do cérebro humano, com a finalidade de aprender com erros, descobrir padrões, com capacidade de se auto organizar. Assim como no cérebro humano, o principal responsável pelo processamento de dados são os seus neurônios. As redes neurais artificiais são comumente utilizadas como método de previsão de geração solar, isto se justifica pela capacidade das redes neurais lidarem com sistemas não lineares e com múltiplas variáveis de entrada. (ABEDIN et al., 2017).

No algoritmo de redes neurais os dados são adicionados a rede, sendo que cada dado fica à reponsabilidade de um neurônio. O conjunto desses neurônios é chamado camada de entrada, e no caso da previsão de geração existirá apenas um neurônio na camada de saída. Existe uma terceira camada chamada camada interna ou oculta, onde os dados são processados. O funcionamento deste algoritmo será detalhado na seção 2.7, pois este será o algoritmo utilizado neste trabalho. Um fluxograma, representado na Figura 1, mostra o funcionamento básico do modelo de redes neurais.





Fonte: Adaptado de (ABEDIN et al., 2017).

#### 2.4.3 Máquina de Vetor de Suporte

Conforme Zhu, Zhou e Fan (2016), umas das principais características das SVM, support vector machine, está em proporcionar uma previsão ótima com um baixo número de amostras, tornando o processamento de dados mais rápido. O fluxograma exibido na Figura 2 tem como objetivo demonstrar a metodologia aplicada no processo de treino e validação realizados através da identificação de padrões no conjunto de amostras e comparados como dados passados.



Figura 2 - Fluxograma Máquina de Vetores de Suporte

Fonte: Adaptado de (ZHU; ZHOU; FAN, 2016).

Ainda segundo os autores, este modelo utiliza um método de aprendizado supervisionado. Desta forma, as amostras que apresentam um determinado padrão são separadas durante o treino, com o intuito de reproduzir os valores desejados de previsão, ou seja, valores que são parecidos com os dados passados.

#### 2.4.4 K-ésimo Vizinho Mais Próximo

Outro método de aprendizado de máquina que pode ser implementado para realizar previsões de geração fotovoltaica é o chamado k-ésimo vizinho mais próximo, *k-nearest neighbor*, ou KNN. A sua implementação, cujo fluxograma do algoritmo está ilustrado na Figura 3. Em comparação com redes neurais artificiais e máquina de vetor de suporte este método é menos complexo, e o processo de treino demanda um tempo menor. O processo de cálculo segrega os dados de acordo com padrões de leitura durante as fases de treino e teste. (LIU; ZHANG, 2016).



Figura 3 - Fluxograma K-ésimo Vizinho Mais Próximo

Fonte: Adaptado de (AHMED et al., 2018).

Após a inicialização dos dados o algoritmo procura por dados com o mesmo padrão de valores, esta procura se repete até que todos os dados estejam agrupados com seus respectivos vizinhos, visto que o objetivo é encontrar uma solução ótima para todo o conjunto de dados colocados no sistema. Este algoritmo pode ser utilizado tanto em situações quando os dados

obtidos não são classificados, ou quando existem dados referentes a diferentes fatores em um mesmo aglomerado. (LIU; ZHANG, 2016).

## 2.4.5 Floresta Aleatória

O nome floresta aleatória é a tradução literal de *random forest*. Recebe este nome por ser composto por um conjunto de pequenos métodos de regressão denominado de árvore de decisões. Assim como as RNA, SVM e KNN este também é um algoritmo de aprendizado de máquina. A principal vantagem deste modelo em relação aos demais citados anteriormente é a sua capacidade de filtrar os dados de entrada que não agregam na previsão, escolhendo aleatoriamente uma amostra de dados do treino e comparando essas amostras com a árvore de decisões, conforme fluxograma da Figura 4, para que ao final de n amostras selecionadas seja possível obter uma árvore de previsões. (LAHOUAR; MEJRI; SLAMA, 2017).





Fonte: Adaptado de (CHEN; TSAU; LIN, 2010).

É possível entender o funcionamento básico deste algoritmo, que ao receber a série de dados temporais para treino, os separa em grupos aleatoriamente. Cada grupo passa por uma árvore de decisões que classifica cada elemento do grupo e os divide conforme suas características e repetição de padrões. Neste momento qualquer dado insignificante para o objetivo final é desclassificado. Neste método as amostras de testes recebem o mesmo tratamento, porém os padrões que o algoritmo irá buscar já foram definidos pelo treino. (CHIANG et al., 2017).

## 2.5 PARÂMETROS PARA A PREVISÃO DE GERAÇÃO FOTOVOLTAICA

Para a construção dos algoritmos de aprendizado de máquina são necessários parâmetros de entrada, ou seja, parâmetros que tenham relação com geração de energia solar para que seja possível encontrar uma relação entre os dados inseridos e os dados obtidos. Um levantamento histórico destes dados se faz necessário para a realização das previsões. Alguns dos principais parâmetros serão aqui expostos.

### 2.5.1 Irradiação Solar

A geração de energia fotovoltaica está fortemente influenciada pelas condições climáticas, sendo que a principal delas é a irradiação solar. Porém os níveis de nebulosidade, temperatura ambiente, umidade relativa do ar e velocidade do ar também afetam diretamente a qualidade de irradiação solar em uma região. Em virtude desta dependência, resultados precisos em uma previsão estão relacionados a qualidade dos dados adquiridos. Cada aspecto meteorológico aqui citado tem um grau de relevância neste momento. (WAN et al., 2015).

A irradiação solar é a quantidade de energia por metro quadrado em um determinado intervalo de tempo. É o fator determinante para a implementação de fontes geradoras de energia solar, sendo a principal característica climática a ser analisada em qualquer previsão. No Brasil existe uma quantidade significativa de irradiação solar incidente, ilustrada no Atlas Brasileiro de Energia Solar, que disponibiliza os mapas de irradiação solar no país. (PEREIRA et al., 2017). Na Figura 5, são apresentados os diferentes níveis de irradiação global horizontal ao longo do território do estado do Rio Grande do Sul. (HAAG et al., 2018).



Figura 5 - Mapa do Rio Grande do Sul de Irradiação Global Horizontal

Fonte: Adaptado de (HAAG et al., 2018).

De acordo com Pereira et al. (2017), a média anual de irradiação global horizontal no Brasil varia de 3,5 kWh/m<sup>2</sup> a 6,25kWh/m<sup>2</sup>, e neste cenário no estado do Rio Grande do Sul, e com base nos dados disponíveis em Haag et al. (2018), a média anual varia entre 4,36 kWh/m<sup>2</sup> e 5,01 kWh/m<sup>2</sup>. Para Oliveira et al. (2013), esta média de irradiação solar no Rio Grande do Sul é maior que a média alemã, um dos países que mais investem em energia solar.

## 2.5.2 Condições Climáticas

A qualidade de irradiação solar disponibilizada em um determinado local é influenciada por outros fatores climáticos, representados pela Figura 6, como a nebulosidade ou a quantidade de nuvens no céu por exemplo, pois as nuvens funcionam como barreiras da irradiação solar. (PEREIRA et al., 2017). Este parâmetro pode ser considerado o mais complexo de ser obtido com exatidão, uma vez que pesquisas costumam utilizar dados baseados na captura de imagens diárias do local em que a previsão é baseada. Esses dados também podem ser obtidos através de centros meteorológicos. (CHATURVEDI; ISHA, 2016).



Figura 6 - Influência de Fatores Climáticos

Fonte: Adaptado de (PEREIRA et al., 2017).

A umidade relativa do ar, a temperatura e a velocidade do vento servem como fatores de emparelhamento entre os dados de irradiação históricos introduzidos na previsão e os resultados obtidos. A utilização única e exclusivamente de dados históricos de irradição solar reduz a precisão do cálculo o que acarreta em um erro maior. Deste modo, com alta umidade relativa do ar, com alta velocidade do vento e com baixas temperaturas os paineis fotovoltaicos tem a sua melhor eficiência, ou seja, a irradição incidente será transformada em energia elétrica com menos perdas. Já no pior caso, com altas temperaturas, baixa umidade e baixa velocidade do vento, a eficiência dos geradores de energia fotovoltaica será reduzida consideravelmente. (CANTOR, 2018).

Segundo Silva et al. (2018) períodos com temperaturas elevadas tendem a aumentar a irradiação incidente, porém as placas solares perdem eficiência sob altas temperaturas. Entretanto, com uma umidade relativa do ar elevada a irradiância solar direta diminui, pois a

quantidade de água existente na atmosfera aumenta a possibilidade da radição solar no momento em que incide sobre uma goticula da água sofrer reflexão, refração ou difração. Por outro lado, a velocidade do vento é benéfica na relação de troca de calor com o ambiente, diminuindo a temperatura na superficie dos módulos fotovoltaicos, aumentando sua eficiência.

Os módulos fotovoltaicos são normalmente apresentados pela sua potência de pico (Wp) definida pelas condições-padrão de ensaio STC, *Standard Test Conditions*. Estas condições geralmente não representam situações reais. Para isto, é definida a temperatuara nominal de operação das células existentes nos módulos, afim de aproximar as características elétricas das situações reais. Esta temperatura é determinada por meio da exposição em circuito aberto a uma irradiância de 800 W/m<sup>2</sup>, a uma temperatura do ar de 20°C e a uma velocidade do vento de 1 m/s. A temperatura nominal dos módulos é identificada pela sigla NOCT, *Nominal Operating Cell Temperature*. Sendo assim, quanto menor a NOCT, menores serão as perdas provenientes da temperatura e melhor será o desempenho do módulo. (PINHO; GALDINO, 2014).

## 2.6 SOFTWARES UTILIZADOS PARA APRENDIZADO DE MÁQUINA

Os algoritmos de aprendizado de máquina exigem a utilização de programas específicos para seu cálculo. A seguir alguns desses *softwares* serão apresentados com o objetivo de avaliar a empregabilidade de cada um, evidenciando as suas diferenças para demonstrar a melhor opção.

## 2.6.1 Matlab

Desenvolvido essencialmente para a comunidade científica, o Matlab possui ferramentas que possibilitam a análise de dados e criação de algoritmos. Este *software* é capaz de trabalhar com aprendizado supervisionado e não supervisionado, e com um grande volume de dados, sendo assim possível a execução dos modelos de aprendizado de máquina citados anteriormente. (MATHWORKS, 2018).

Atualmente existem inúmeras ferramentas no Matlab que podem ser empregadas em diferentes situações de uso do aprendizado de máquina. Neste trabalho é importante frisar duas ferramentas a *Statistics and Machine Learning Toolbox* e a *Deep Learning Toolbox*. A primeira

possui ferramentas que possibilitam o uso de métodos de classificação, de regressão e de agrupamento, onde os modelos de SVM, KNN e Random Forest podem ser empregados.

Já a *Deep Learning Toolbox* é uma ferramenta desenvolvida para projetar e modelar redes neurais artificiais. Os métodos de classificação e regressão podem ser utilizados nesta ferramenta. Contudo, quando existe uma grande quantidade de dados a serem analisados é possível dividir o processamento do computador entres os núcleos do processador e do processador gráfico, reduzindo assim o tempo de execução. (PALUSZEK; THOMAS, 2017).

As vantagens na utilização do Matlab estão no número de recursos disponíveis, na bibliografia, visto que é possível encontrar inúmeras referências, no tempo de processo que pode ser reduzido e na possibilidade de visualizar os processos. O desenvolvimento deste trabalho utilizará este *software* devido a bibliografia referente a modelagem de algoritmos de aprendizado de máquina para previsões de geração fotovoltaica.

#### 2.6.2 GNU Octave

O GNU Octave é um *software* de cálculo numérico de código aberto, que utiliza uma linguagem parecida com a do Matlab, e assim como tal é possível implementar algoritmos avançados para o cálculo de previsão, como os métodos de aprendizado de máquina. Dessa forma, este programa surge como uma opção gratuita, porém sem a mesma experiência do software da Mathworks. (KOUSIOURIS et al., 2010).

Para este software existe uma ferramenta *open source* desenvolvida para o uso de aprendizado de máquina, chamada nnet. A nnet é uma biblioteca C++ que permite a modelagem de redes neurais artificiais, incluindo a aprendizado supervisionado com retropropagação. Entre as desvantagens envolvidas na utilização do Octave estão os poucos recursos presentes nestes *softwares*, a ferramenta nnet não sofre revisão desde 2016 e não possui memória interna para a alocação dos dados.

#### 2.6.3 Neural Designer

O Neural Designer foi criado para a modelagem de redes neurais artificiais, e possui uma versão gratuita, porém esta versão tem um número limitado de dados para analisar, sendo possível executar apenas 5000 instâncias por vez, o que pode comprometer a previsão dependendo da quantidade de dados disponível. Este *software* não utiliza linhas de comando, possuindo uma interface gráfica para a implementação das redes, sendo apto para averiguar um conjunto de dados significativo em pouco tempo.

Desenvolvido pela Artelnics, o Neural Designer permite a implementação de redes neurais complexas com facilidade e agilidade, pois o *software* segmenta todos os processos de construção da rede, simplificando sua aplicação. Em comparação aos demais programas a sua principal desvantagem está na exclusividade no uso das RNAs e a capacidade limitada de instâncias executadas na sua versão gratuita.

#### 2.6.4 Python

Python é uma linguagem de programação de alto nível orientada a objetos, capaz de realizar múltiplas chamadas de sistemas e bibliotecas, sendo extensível em C e C++. (PHYTON SOFTWARE FOUNDATION, 2018). Em Holmgren e Groenendyk (2016) foi utilizada a biblioteca *open source* PVLIB – Python, desenvolvida para previsão de geração fotovoltaica, com o objetivo de facilitar o acesso a dados de previsões climáticas. A modelagem do modelo de previsão é dividida em duas etapas, aquisição de dados climáticos e a conversão destes dados em previsão fotovoltaica. Este algoritmo é baseado na biblioteca PVLIB do Matlab, e utiliza o modelo de Liu-Jordan para a previsão, um modelo que estima a irradiação tanto difusa quanto direta baseado no índice de claridade. (SCOLAR; MARTINS; ESCOBEDO, 2003).

No entanto a linguagem Python permite a implementação de diversos modelos de previsão através de suas bibliotecas de mineração de dados e aprendizado de máquina. Como exemplo a biblioteca *scikit-learn*, que proporciona funções para a implementação de diferentes algoritmos de aprendizado de máquina, como as citadas anteriormente. A linguagem Python tem como vantagem a flexibilidade para a implementação destes algoritmos, através de seu acervo *open source* de bibliotecas destinadas para esta função.

#### 2.6.5 Weka

Em Meenal et al. (2018) os autores utilizam o modelo Floresta Aleatória de aprendizagem de máquina para a previsão dos recursos de irradiação solar nos 29 estados

indianos. O software escolhido para esta aplicação foi o WEKA. Este programa funciona como explorador de dados utilizando técnicas de agrupamento, classificação ou regressão.

Criado pela Universidade de Waikato na Nova Zelândia, WEKA é um *software open source* de mineração de dados que contém um acervo de algoritmos de aprendizado de máquina. (MEENAL et al., 2018). As vantagens deste programa estão no fato de ser gratuito e de ter sido desenvolvido exclusivamente para trabalhar com mineração de dados e aprendizado de máquina, incluindo todos os métodos apresentados anteriormente, logo as desvantagens que podem ser citadas estão no baixo referencial bibliográfico e interface complexa.

### 2.7 REVISÃO DO ESTADO DA ARTE

Devido ao incremento da inserção de energias renováveis em sistemas de geração distribuída, o desenvolvimento de métodos de previsão de geração de energia fotovoltaica tem se tornado um assunto recorrente entre os pesquisadores. Em Sansa et al. (2014), dois diferentes modelos de redes neurais artificias foram empregados para previsão de geração fotovoltaica em curto, médio e longo prazo em uma localidade ao norte de Barcelona, um estático e um dinâmico. A diferença na construção do modelo dinâmico está na relação entre a saída e a entrada de dados, uma vez que os valores calculados ao longo do processo de treinamento são usados como um parâmetro de entrada. Existe assim uma realimentação entre a entrada e saída da rede neural, com isso o algoritmo reduz o erro de previsão a cada etapa do treinamento.

Na previsão de longo prazo realizada para cinco dias a frente, visto que a estrutura da rede neural artificial estática é modelada com dois neurônios na camada de entrada, o primeiro destinado aos dados de irradiação solar e o segundo para os dados de temperatura ambiente, os resultados mostraram, conforme ilustrado na Figura 7, um erro MSE de 17% enquanto os resultados provenientes da rede neural dinâmica composta por três neurônios de entrada, o terceiro uma realimentação da saída, mostrou um erro de 10%.



Figura 7 - Erro relacionado aos dois modelos de RNA

Fonte: (SANSA et al., 2014).

Na publicação de Alanazi, Alanazi e Khodaei (2016), os autores conduziram uma pesquisa procurando diminuir as incertezas relacionadas as previsões de longo prazo. As dificuldades encontradas se deram ao realizar uma previsão de irradiação solar, dificuldades estas influenciadas pelas mudanças climáticas e a forte relação entre o aumento da precisão de uma previsão quando em dias de céu aberto criam a necessidade da busca de dados complementares como nebulosidade, temperatura do ar e velocidade do vento.

Os dados para a previsão foram obtidos através do NREL, *National Renewable Energy* Laboratory. Antes do processamento, os dados dispensáveis para a pesquisa, como as horas noturnas, são retirados. Os autores criam um modelo de redes neurais utilizando o NARX, *nonlinear auto-regressive exogenous model*, onde são inseridos os parâmetros no modelo de rede neural para treino, o qual é realizado até o erro entre a previsão e os dados atuais ser minimizado. A métrica utilizada para a verificação da previsão é o MAPE que trouxe um erro de 9,97%, conforme mostra a Figura 8. Para o método proposto, as curvas em azul representam a irradiação real para um determinado momento e as curvas em vermelho representam a sua respectiva previsão.



Figura 8 - Comparação entre valor real e o previsto

Fonte: Adaptado de (ALANAZI; ALANAZI; KHODAEI, 2016).

8060

8080

hora

8100

8120

50

o

Já Anderson e Yakimenko (2018), trabalharam na previsão de geração de energia fotovoltaica na ilha de Eigg, na Escócia utilizando redes neurais artificiais com o auxílio da *Neural Network Toolbox* do Matlab. Os dados históricos de clima e irradiação foram obtidos através do serviço de satélite SoDa, *solar radiation data*, contendo dados climáticos como umidade relativa, pressão atmosférica, velocidade do vento e direção, chuvas, neve e profundidade de neve.

Os autores modelaram redes neurais diretas, com 10 neurônios na camada oculta cada uma, e um neurônio na camada de saída. Na primeira rede a previsão foi realizada utilizando como dado de entrada apenas uma componente de irradiação solar, a irradiação global horizontal, na segunda etapa a rede neural foi alimentada com os nove componentes da irradiação solar, irradiação direta inclinada, irradiação difusa inclinada, irradiação refletida, irradiação inclinada global, irradiação horizontal direta, irradiação difusa horizontal, irradiação global horizontal, céu aberto e irradiação extraterrestre. Os resultados obtidos mostraram diferença mínima entre a previsão apenas com irradiação global horizontal e a previsão com as nove componentes. Na terceira rede foram utilizados os dados climáticos para simular situações em que os dados de irradiação não possam ser obtidos. Na Figura 9 é possível observar a comparação entre o modelo com a componente de irradiação global horizontal e o modelo com as nove componentes de irradiação, como também a comparação entre o primeiro modelo e o modelo com os dados climáticos.



Figura 9 - Comparação entre os Modelos de Previsão

Fonte: Adaptado de (ANDERSON; YAKIMENKO, 2018).

As previsões realizadas com as componentes de irradiação solar mostraram um melhor resultado do que as realizadas apenas com componentes climáticas. Para melhorar o desempenho desta última, um novo parâmetro de entrada foi inserido, os dados históricos de irradiação solar. Com a inserção destes dados, a configuração da rede neural foi modificada para uma rede neural auto regressiva. Esta configuração apresenta uma realimentação entre o neurônio de saída e as entradas da rede, a Figura 10 apresenta um esquemático do modelo construído, utilizando como parâmetros de entrada oito conjuntos de dados climáticos incluindo a realimentação interligando a camada de saída e a camada entrada da rede.

Os resultados obtidos são apresentados pela Figura 11, que mostra a relação entre os valores da saída do treinamento comparados aos valores esperados, as saídas representadas conforme a legenda, apresentaram um erro absoluto de 0,92% em relação aos alvos, ou seja, em relação a valores reais de geração. (ANDERSON; YAKIMENKO, 2018).


Figura 10 - Diagrama da RNA Auto Regressiva

Fonte: O autor (2018).



Figura 11 - Desempenho da Previsão com RNA Recorrente

Fonte: (ANDERSON; YAKIMENKO, 2018).

## **3 METODOLOGIA**

# 3.1 MÉTODOS E HORIZONTE DE PREVISÃO A SEREM UTILIZADOS

O horizonte de estudo deste trabalho será de longo prazo, com o objetivo realizar a previsão de geração fotovoltaica na região de Garibaldi, para os próximos 6 meses, com um horizonte de aproximadamente 184 dias a frente. As previsões a longo prazo são caracterizadas por projetar a disponibilidade de geração de energia fotovoltaica em determinada área ao longo de meses até anos. (MAJUMDER; BEHERA; NAYAK, 2017).

O fator primordial para uma implementação confiável está na qualidade dos dados históricos obtidos. A aquisição destes dados será abordada no item 3.2. O período em que determinado parâmetro é monitorado também tem influência na previsão, desta forma, o objetivo deste estudo é adquirir dados com o maior período possível de monitoramento. (CHATURVEDI; ISHA, 2016). Devido a quantidade de estudos relacionados a previsão de energia que utilizam redes neurais artificiais para suas previsões, aliada a capacidade das redes neurais de lidar com diferentes tipos de dados e aspectos não lineares, Abedin et al. (2017), este algoritmo será o utilizado nesta previsão.

Estudos mostram que a precisão de uma previsão com rede neural é incrementada de acordo com a qualidade dos dados. (SINGH et al., 2015). A precisão é aperfeiçoada através do processo de treinamento e pela arquitetura da rede neural construída. (ALANAZI; ALANAZI; KHODAEI, 2016). Desta forma, será necessário realizar diversas etapas de treinamento para encontrar uma arquitetura ótima para esta previsão. Para a construção da rede neural artificial e para a implementação do algoritmo de previsão o *software* utilizado será o Matlab, devido a quantidade de trabalhos na literatura que trabalham com este *software*, pelo suporte disponibilizado pela fabricante para seus recursos e pela confiabilidade e integridade de seu sistema. (ANDERSON; YAKIMENKO, 2018). No Matlab será utilizada a *Deep Learning Toolbox*, uma biblioteca que permite criar, dimensionar e visualizar redes neurais artificiais. Após realizada a precisão, os valores gerados serão testados através do MAPE, com o objetivo de comparar os valores obtidos com valores atuais de geração.

A modelagem de uma rede neural artificial depende dos fatores intrínsecos as características dos dados a serem analisados. A estrutura de uma rede neural artificial e os processos que envolvem a sua modelagem serão apresentados no item 3.2 deste capítulo.

## 3.2 REDES NEURAIS ARTIFICIAIS

#### 3.2.1 Estrutura

O principal elemento de uma rede neural artificial é o neurônio. Os neurônios são combinadores lineares que interagem com os dados de entrada, uma vez que a ponderação desses dados é definida pelos seus pesos sinápticos. (ABEDIN et al., 2017). As RNAs são compostas por camadas, sendo que cada uma dessas camadas contém um número de neurônios como mostra a Figura 12. Na primeira camada, conhecida como camada de entrada, são alocados os dados de entrada que serão utilizados para o aprendizado do algoritmo, nesta camada cada neurônio é responsável por um tipo de dado, ou seja, se para uma previsão são utilizados dois tipos de dados, na camada de entrada existirá apenas dois neurônios. Na segunda camada, chamada de camada oculta, acontece o tratamento dos dados, essa camada é composta pelos neurônios ocultos.





Fonte: Adaptado de (CHENG; GE; CAO, 2012).

Para a definição da quantidade ótima de neurônios a serem inseridos na camada oculta de uma rede neural artificial, existem diferentes abordagens. Em Silva (2005) é apresentada uma técnica de determinação da quantidade de neurônios ocultos a partir do teorema de

Kolmogorov, representada pela Equação 5, que relaciona a quantidade de neurônios ocultos com o número de neurônios de entradas e de saída.

$$N_{ocultos} = 2 \times N_{entradas} + 1 \tag{5}$$

Onde:

- · Nocultos é o número de neurônios ocultos;
- ·  $N_{entrada}$  o número de neurônios de entrada.

Já para Haykin (2001), o número ótimo de neurônios ocultos se dá pela relação entre o erro médio quadrático e a quantidade destes neurônios, sendo que o valor ótimo está no modelo que apresentar menor erro com o menor número de neurônios ocultos. Experimentos com um número mínimo de 2 neurônios ocultos são realizados, e a cada novo experimento este número é incrementado, com o objetivo de encontrar a relação ótima entre número de neurônios ocultos e o erro médio quadrático. A saída do sistema se encontra na terceira camada, sendo o número de neurônios proporcional ao número de saídas do sistema. As redes neurais artificiais podem ter múltiplas saídas, e no caso específico para a previsão de geração, existe apenas uma saída, a da própria previsão. (ALZAHRANI et al., 2017).

### 3.2.2 Classificação

Segundo Barreto (2002), é possível classificar as redes neurais artificiais em redes neurais diretas e redes neurais com ciclos. Os modelos mais comuns são as redes neurais diretas, representadas por camadas, como mostrado na Figura 7. Outros modelos são as redes neurais com ciclos, modelo também comumente referenciado como redes neurais recorrentes. Estas últimas também são chamadas de redes neurais com realimentação, sendo versões dinâmicas do modelo direto, originalmente criadas para representar na saída valores próximo aos utilizados na entrada. Para Alzahrani et al. (2017), as redes neurais diretas não são capazes de trabalhar com sequências aleatórias de entradas por não possuírem resposta à valores de saída. Para essas operações são usadas as redes neurais recorrentes ou com realimentação, técnicas que realimentam o sistema com o valor de saída anterior.

### 3.2.3 Pré-processamento de Dados

#### 3.2.3.1 Classificação de Variáveis

Durante a etapa de pré-processamento dos dados, as variáveis preditoras passam por um processo de classificação, este processo determina quais variáveis de entrada tem maior influência para o modelo de previsão. O método de classificação de variáveis *stepwise regression* busca dentre o conjunto total as variáveis com maior importância estatística. Este processo busca avaliar cada uma das possíveis variáveis de entrada de um modelo de previsão, neste caso, as variáveis de entrada de uma rede neural, afim de eliminar variáveis redundantes ou que possam prejudicar a eficiência da previsão. Este processo é ilustrado pela Figura 13. (GHUNEM; ASSALEH; EL-HAG, 2012).





Fonte: Adaptado de (GHUNEM; ASSALEH; EL-HAG, 2012).

No método de classificação *stepwise regression* as variáveis são analisadas uma a uma e comparadas pelo seu *valor-p* ou pelo parâmetro *t-statistic*. Este termo significa o teste de probabilidade de hipóteses nulas. A variável com menor *valor-p* é acrescida ao modelo, em seguida a segunda com menor valor e assim sucessivamente, porém a cada inclusão de variável é realizado um novo teste sobre o *valor-p* do modelo, podendo haver a remoção de uma variável

que não contribua da mesma maneira após a inclusão de uma nova. O processo de inclusão e remoção é definido por limites ( $p < 0.05 \ e \ p > 0.1$ ) repectivamente. (GHUNEM; ASSALEH; EL-HAG, 2012). O parâmetro *t-statistic* é definido através da razão do coeficiente de regressão linear da variável dependente pelo seu erro padrão. A Equação 6 representa a regressão linear.

$$Y = \beta_0 + \beta_1 X_1 \tag{6}$$

Nesta equação:

- Y é a resposta esperada para esta regressão;
- ·  $X_n$  é a variável dependente;
- ·  $\beta_0$  é o coeficiente de regressão de interceptação;
- ·  $\beta_1$  é o coeficiente de regressão de inclinação.

Em Walpole et al. (2012), o coeficiente  $\beta_1$  é representado pela Equação 7.

$$\beta_{1} = \frac{\sum_{i=1}^{n} (X_{i} - \bar{X})(Y_{i} - \bar{Y})}{\sum_{i=1}^{n} (X_{i} - \bar{X})^{2}}$$
(7)

Em que:

- $\overline{X}$  é o valor médio da variável dependente;
- $\overline{Y}$  é o valor médio da variável independente;

Ainda em Walpole et al. (2012), o parâmetro *t-statistic* é definido pela Equação 8.

$$t_{statistic} = \frac{\beta_1}{\sigma_{\bar{X}}} \tag{8}$$

Onde:

·  $\sigma_{\bar{X}}$  é o erro padrão da regressão,

A variável  $\sigma_{\bar{X}}$  é definida pela Equação 9.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \tag{9}$$

Na Equação 9,  $\sigma$  representa o desvio padrão da regressão, aqui definido pela Equação 10.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (X - \bar{X})^2}{n-1}}$$
(10)

A partir do parâmetro *t-statitic*, Rouaud (2013) determina o *valor-p*, através da *t-distribution*. A Equação 11 apresenta o conceito geral para o cálculo da *t-distribution*.

$$t_{distribution} = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \times \sqrt{\nu\pi}} \times \left(1 + \frac{t_{statistic}}{\nu}\right)^{-\frac{\nu+1}{2}}$$
(11)

Nesta equação:

 ν representa o número de degraus de liberdade que é igual ao número de amostras menos 1, aqui representado pela Equação 12.

$$v = n - 1 \tag{12}$$

Substituindo 11 em 12:

$$t_{distribution} = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right) \times \sqrt{(n-1)\pi}} \times \left(1 + \frac{t_{statistic}}{n-1}\right)^{\frac{-n}{2}}$$
(13)

O valor-p é definido pela área da curva de distribuição t a partir do valor de *t-statistic* quando este tende tanto para  $-\infty$  quanto para  $+\infty$ . A equação geral deste valor é representada pela Equação 14.

$$valorp = \int_{-\infty}^{-t_{statistic}} t_{distribution} \cdot dt_{statistic} + \int_{t_{statistic}}^{+\infty} t_{distribution} \cdot dt_{statistic}$$
(14)

Ainda segundo Rouaud (2013), situações em que o valor dos degraus de liberdade é muito grande, o valor de *t-distribution* pode ser definido pela Equação 15 e o *valor-p* pode então ser obtido através da Equação 16.

$$\lim_{k \to +\infty} t_{distribution} = \frac{1}{\sqrt{2\pi}} e^{\frac{t_{statistic}}^2}$$
(15)

$$valorp \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{t_{statistic}}{2}}$$
(16)

Após a definição das variáveis inclusas, o parâmetro *F-statistic* pode ser calculado. Este parâmetro é definido pela Equação 17.

$$F_{statistic} = \frac{SS_{R}\left(\frac{\beta_{i}}{\beta_{0}} + \beta_{1}, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_{n}\right)}{MSE}$$
(17)

Nesta equação:

- ·  $SS_R$  representa a soma dos quadrados da regressão e o termo;
- *MSE* representa o erro médio quadrático.

As equações 18 e 19, representam a soma dos quadrados da regressão.

$$SS_R = \sum_{i=1}^n (Y_i - Y_m)^2$$
(18)

$$Y_{m} = \frac{1}{n} \sum_{i=1}^{n} Y_{i}$$
(19)

Em que:

- ·  $Y_i$  representa a saída da regressão;
- ·  $Y_m$  a média entre as saídas. (ABDELMUTALAB; ASSALEH; EL-TARHUNI, 2016).

#### 3.2.3.2 Limpeza de Dados

Para que o conjunto de dados adquiridos possa ser utilizado como entrada da rede neural artificial é necessário um tratamento para a limpeza destes dados. Quando uma grande quantidade de dados é processada, é comum a existência de erros de leitura por parte de sensores ou no tratamento dos dados. Os métodos empregados para a limpeza de dados com erros de leitura, são a substituição destes dados por uma interpolação linear entre dados anteriores e posteriores. Na Equação 20 está a representação genérica da interpolação linear.

$$y = y_0 + \frac{(y_1 - y_0)(x - x_0)}{x_1 - x_0}$$
(20)

Onde:

- ·  $(y_0, x_0)$ é o ponto que representa o valor anterior;
- $(y_1, x_1)$ é o ponto que representa o valor posterior aos dados com erro de leitura. (ALZAHRANI et al., 2017).

Com relação ao aspecto temporal dos dados que envolvem a geração solar, a remoção das horas noturnas, quando possível, ajuda a melhorar a precisão da previsão. Esta remoção é realizada pela classificação dos dados, ou seja, pelas horas do dia na qual existe geração efetiva de energia. (ALANAZI; ALANAZI; KHODAEI, 2016).

#### 3.2.3.3 Normalização

Antes da entrada dos dados no sistema e logo após a sua limpeza, é necessário efetuar o escalonamento dos mesmos. Para tanto, a normalização dos dados é utilizada para lidar com as diferentes escalas nos dados de entrada. (SINGH et al., 2015). O uso desta técnica reduz o processo de treino e diminui as chances de paradas durante o cálculo. Na normalização, os valores de entrada da rede neural são escalonados sob uma faixa de valores mínimos e máximos. (ALZAHRANI et al., 2017). A normalização mantém a proporcionalidade dos valores no conjunto de dados, além de melhorar a precisão das previsões. (SHI et al., 2011).

Ainda segundo Singh et al. (2015), a normalização de Ponto Decimal, a Z-Score e a normalização de Mínimos e Máximos são algumas das técnicas empregadas neste tipo de

tratamento dos dados. A técnica frequentemente utilizada na normalização de dados relacionados a previsão de geração é a normalização de mínimos e máximos, como em Shi et al. (2011), Singh et al. (2015), Alanazi, Alanazi e Khodaei (2016), Alzahrani et al. (2017). Geralmente os valores são escalonados entre 0 e 1, podendo haver variações nos valores mínimos e máximos. Na Equação 21 está a representação genérica da normalização de mínimos e máximos.

$$P_{nor} = \frac{P_n - P_{\min}}{P_{\max} - P_{\min}}$$
(21)

Onde:

- ·  $P_{\rm n}$  é o valor real, não normalizado;
- ·  $P_{\min}$  é o limite inferior da normalização;
- ·  $P_{\text{max}}$  é o seu limite superior.

Após a normalização, os dados de entrada a serem utilizados na rede neural estarão em uma mesma escala de referência. (ALANAZI; ALANAZI; KHODAEI, 2016).

# 3.2.3.4 Função de Ativação

Os neurônios são somadores ponderados pelos pesos sinápticos. Os resultados desta soma ponderada são aplicados a uma função de ativação. Estas funções são normalmente funções não-lineares e são responsáveis pela resposta do neurônio de saída. (SILVA, 2005). A função de ativação mais utilizada na modelagem de uma rede neural é a função sigmoide, caracterizada pela sua curva em s. A função sigmoide logística assume valores contínuos entre 0 e 1, aqui representada pela Equação 22, enquanto a função tangente hiperbólica é uma função sigmoide equivalente que assume valores contínuos entre -1 e 1, representada pela Equação 23. (HAYKIN, 2001).

$$\varphi = \frac{1}{1 + e^{-av}} \tag{22}$$

$$\varphi = \tanh(v) \tag{23}$$

Onde:

- $\cdot$  a é o parâmetro de inclinação da função sigmoide logística;
- $\cdot v$  é o valor que será ativado a função.

A Figura 8 ilustra a curva característica de uma função sigmoide. (HAYKIN, 2001).



Figura 14 - Curva característica função sigmoide

Fonte: (HAYKIN, 2009).

# 3.2.4 Treinamento

O treinamento utilizado pela maior parte dos autores, como em Yang, Cao e Xiu (2014), Alanazi, Alanazi e Khodaei (2016), Hasan, Munawar e Siregar (2017), Rola (2017), Abedin et al. (2017) é o algoritmo de retropropagação, um modelo de aprendizado supervisionado. Neste modelo, a saída do sistema é comparada a um valor desejado, em seguida é aplicada uma métrica de verificação do erro. Se o erro medido for menor que um determinado limite o treino é encerrado, se o erro for maior, o processo de treino é retomado. (ABEDIN et al., 2017)

Ainda de acordo com Abedin et al. (2017) o processo de treino das redes neurais artificiais busca parametrizar os pesos sinápticos dos neurônios com o objetivo de melhorar a performance. Segundo Rola (2017), o processo de retropropagação acontece quando o erro calculado é propagado da saída do sistema para a sua entrada. Este procedimento utiliza o erro como uma nova entrada do sistema com o objetivo de reduzir o seu valor reajustando os pesos sinápticos dos neurônios. Alzahrani et al. (2017), divide os dados de entrada em três grupos,

sendo que 70% dos dados são utilizados somente no processo de treino, 20% para o teste e o restante para a validação. Este algoritmo é um processo supervisionado responsável por ajustar os pesos sinápticos de cada neurônio comparando o erro gerado a um valor desejado. A Equação 24 representa o modelo utilizado para o ajuste destes pesos. (SILVA, 2005).

$$w_{i,j}(t+1) = w_{i,j} + \eta \delta_j x_i + \alpha (w_{i,j}(t) - w_{i,j}(t-1))$$
(24)

Onde:

- ·  $w_{i,j}$  determina o peso sináptico do neurônio i em relação ao j;
- $\eta$  determina a velocidade dos ajustes;
- ·  $\alpha$  termo de momento utilizado para impedir oscilações com a variação de  $\eta$ ;
- $\delta_j$  em um neurônio da camada de saída, determina a variação entre a saída real e a desejada, e é expresso pela Equação 25.

$$\delta_j = y_j (1 - y_j) (d_j - y_j)$$
 (25)

Quando o neurônio pertencer a camada oculta, então  $\delta_i$  é expresso pela Equação 26.

$$\delta_j = x'_j (1 - x'_j) \sum \delta_k w_{j,k} \tag{26}$$

# 3.2.4.1 Função de Treinamento

A função de treinamento é a responsável pelo aprendizado da rede. Em um aprendizado supervisionado por retropopagação existem algumas funções comumente utilizadas, com a *adaptative learning backpropagation*, a *resilient backpropagation* e a *bayesian regularization backpropagation*, que são escolhidas conforme o seu desempenho para determinada tarefa. No modelo construído, a função de treinamento a ser utilizada será a *bayesian regularization*, que permite atualizar os pesos sinápticos ao longo do treinamento garantindo uma boa generalização da rede neural e prevenindo que a ocorrência de *overfitting*. (TICKNOR, 2013).

Para Saini (2007), o objetivo de uma rede neural é produzir o menor erro possível no treino e no teste, procurando manter o mesmo erro em ambos. O algoritmo desta função de

treinamento foi desenvolvido afim de reduzir a soma do erro quadrático da saída da rede em relação ao seu alvo. A Equação 27, mostra a técnica de regularização entre a soma destes erros entre e a soma dos erros quadráticos dos pesos sinápticos.

$$F = \gamma SSE + (1 - \gamma)SSE_w \tag{27}$$

Onde:

- SSE é a soma do erro quadrático;
- $SSE_w$  é a soma do erro quadrático dos pesos;
- ·  $\gamma$  é o parâmetro de desempenho que está diretamente relacionado como o ritmo do treinamento.

Como os pesos sinápticos são varáveis que se alteram ao longo do processo de treinamento, esta função de treinamento realiza uma distribuição aleatória destes pesos. Quando o processo de treino se inicia a próxima distribuição obedece a Equação 28.

$$p(w \mid D, \gamma) = \frac{p(D \mid w, \gamma) p(w \mid \gamma)}{p(D \mid \gamma)}$$
(28)

Onde:

- *D* é o conjunto composta pelas entradas e o alvo;
- ·  $p(w|\gamma)$  é a primeira distribuição dos pesos;
- ·  $p(D|w, \gamma)$  é a distribuição desejada;
- ·  $p(D|\gamma)$  é o fator de normalização.

Neste treinamento quanto melhor é o peso encontrado, melhor desempenho da função. O parâmetro de desempenho γ é ajustado pela regra Bayes, descrita pela Equação 29.

$$p(\gamma \mid D) = \frac{p(D \mid \gamma)p(\gamma)}{p(D)}$$
(29)

Onde:

·  $p(\gamma)$  é a densidade prévia assumida pelo parâmetro de regularização.

Desse modo a primeira distribuição de pesos é ajustada de acordo com a atualização da função de verossimilhança descrita pela Equação 30.

$$p(D \mid \gamma) = \left(\frac{\pi}{\gamma}\right)^{-\frac{N}{2}} \left[\frac{\pi}{1-\gamma}\right]^{-\frac{L}{2}} Z_F(\gamma)$$
(30)

Em que:

· L é o número total de parâmetros na rede neural;

·  $Z_F$  é aproximação da função quadrática para a distribuição dos pesos w.

Na Equação 31 é apresentada a função aproximada para  $Z_F$ .

$$Z_F = (2\pi)^{\frac{L}{2}} \det^{-\frac{1}{2}} H e^{(-F(w))}$$
(31)

Onde:

• *H* é uma matriz Hessian derivativa da Equação 27, e representada pela Equação 32.

$$H = \nabla^2 (\gamma SSE + (1 - \gamma)SSE_w)$$
(32)

# 3.2.5 Validação Cruzada

A técnica divide o conjunto de dados em grupos de treinamento, teste e validação como mostram os trabalhos Alzahrani et al. (2017), Abedin et al. (2017). Na validação cruzada o treinamento de uma rede neural é realizado com um conjunto de dados diferente do conjunto utilizado no teste, este por sua vez, também é diferente do conjunto utilizado para a validação. O objetivo deste método está em encontrar uma generalização ótima sem correr o risco de *overfitting*, quando a rede neural é excessivamente treinada, perdendo generalização. (HAYKIN, 2001).

# 3.3 ESTUDO DE CASO

#### 3.3.1 Aquisição de dados climáticos

A NASA, *National Aeronautics and Space Administration*, mantém um projeto de apoio a energias renováveis e afins, o *Power Data Access Viewer*, um sistema de visualização de dados solares e meteorológicos. A sua interface gráfica pode ser visualizada pela Figura 15. Este acervo foi construído através de dados fornecidos por estações meteorológicas e por imagens de satélites, contendo informações acumuladas sobre todas as regiões do mundo, com medições realizadas desde 1981. Este sistema é disponibilizado gratuitamente afim de facilitar e proporcionar o incremento de novas fontes de geração renováveis. (STACKHOUSE et al., 2018). A partir desde sistema serão adquiridos dados meteorológicos diários como o índice incidente de irradiação solar a uma superfície horizontal, o índice de irradiação a céu claro, a velocidade do vento à 10 metros, a temperatura à 2 metros, a umidade relativa a 2 metros, entre outros, ao todo serão coletados 24 dados diferentes. No Apêndice A é apresentada uma amostra destes dados. Estes dados passarão pelo processo de classificação *stepwise* e os dados selecionados serão utilizados como entrada da rede neural.

POWER S	ingle Point (	Data Access	- ×	0
1. Choose a Us	ser Commun	ity	· P + 50	0
SSE-Renewabl	le Energy	<ul> <li>[Energia Renovável]</li> </ul>	t a	
<b>2. Choose a Te</b> ● Daily ◎ Int	emporal Ave terannual 🔘	rage Climatology [Médias Diárias]	albinot	
3. Enter Lat/L	on or Add a	Point to Map	Trav	
9	-29.2575	[Latitude]	essa tuat	
Clear	-51.5142	[Longitude]	cairú	
4. Select Time	Extent		gitali	
Start Date	04/20/2016	[Data Início] - MM/DD/AAAA	a do B	P
End Date	03/06/2019	[ Data Fim ] - MM/DD/AAAA	Tast	
5 Select Outp	ut Eilo Eorm	ate Colori All		

#### Figura 15 - Interface Gráfica do Power Data Access Viewer

Fonte: (NASA, 2018)

A prefeitura municipal de Garibaldi, disponibiliza em Garibaldi (2019), um acervo de dados meteorológicos em períodos horários. O site citado disponibiliza dados de duas estações meteorológicas localizadas em Garibaldi, a Estação Aeroclube e a Estação São Gotardo. Entre os dados fornecidos estão componentes de pressão atmosférica, relativa e absoluta, de umidade, relativa e absoluta, de temperatura, médias, mínimas e máximas, velocidade do vento e precipitação, uma amostra destes dados é apresentada no Apêndice B.

O site *Power Data Access Viewer*, NASA (2018), fornece os dados históricos em formato (.csv), *comma-separeted value*, um modelo de classificação de dados por planilha eletrônica e as séries temporais provenientes doo site Portal do Clima são tratadas em formato (.xlsx). Para a realização da previsão serão utilizados os dados climáticos que melhor apresentarem uma padronização relacionada aos dados de geração fotovoltaica. Esta padronização será verificada através do método classificatório *stepwise regression*.

# 3.3.2 Dados históricos de geração fotovoltaica

Assim como na previsão realizada por Anderson e Yakimenko (2018), na construção de um modelo de previsão fotovoltaica, os dados históricos de geração auxiliam no aprendizado da rede neural. Por ser um aprendizado supervisionado é importante ter conhecimento dos dados de geração para ensinar a rede neural e ajudar na validação dos resultados obtidos, nos quais são aplicadas as métricas de validação que compararam o erro entre os valores previstos e os valores reais de geração.

Para este trabalho, os dados de geração utilizados serão os dados mensurados na região de Garibaldi fornecidos por uma empresa da região que atua no ramo de geração de energia solar, com um período temporal de dois anos e meio de medição, sendo possível até o momento da previsão o acúmulo de três anos de dados históricos. Estes dados de geração serão utilizados como alvos durante o a etapa de treinamento da rede neural. Após realizado o treinamento, os dados de geração serão empregados para a validação da previsão, em que os dados reais serão comparados com os dados previstos. É possível observar uma amostra destes dados no Apêndice C.

#### 3.3.3 Parametrização de Dados

A função de ativação de que será utilizada é a função sigmoide logística, função que é geralmente usada em modelos de redes neurais com múltiplas camadas. (SINGH, 2015). Neste trabalho será realizada a modelagem de uma rede neural com três camadas, a camada de entrada, uma camada oculta e a camada de saída. Outro fator que justifica a escolha deste tipo de função é pela possibilidade de ser treinada pelo algoritmo de retropropagação, por ser uma função diferenciável. (SINGH, 2015).

Para o treinamento da rede neural se faz necessário o uso de algum método de parametrização dos dados, método este responsável por ajustar os pesos sinápticos de cada dado. Neste trabalho será utilizado o algoritmo de aprendizado com retropropagação do erro introduzido anteriormente. Antes do processo de treinamento e ajustes dos pesos sinápticos ser iniciado, o conjunto de dados é dividido.

Seguindo os métodos de validação cruzada, o conjunto utilizado no treinamento será diferente dos utilizados na validação. Usualmente, cerca de 10% do conjunto total de dados é separado para a validação do modelo. Portanto, os dados restantes serão divididos em dois outros conjuntos, o primeiro para o treinamento com 70% do total e o segundo para teste, com os 20% restantes. (HAYKIN, 2001). O processo de treinamento será interrompido quando o erro MAPE atingir o menor valor possível. (SILVA, 2005).

A Figura 16 exibe o fluxograma do modelo de RNA proposto, sendo que após a aquisição dos dados históricos é realizado o pré-processamento dos dados. Aplicando o conceito da validação cruzada, a coleção de dados é dividida. Na sequência é modelada a RNA, configurando suas características como: a quantidade de neurônios ocultos, o método de treinamento e a função de ativação. Com o modelo pronto inicia-se o processo de treinamento da rede, na qual o algoritmo percorrerá os dados e buscará por padrões.

Uma vez que o menor erro é calculado através do MAPE o treinamento é encerrado, se o processo de treinamento não superar o menor valor de erro pré-estabelecido a RNA precisará ser remodelada. Caso contrário, o restante da coleção de dados que não foi utilizado para o treinamento será inserido no algoritmo afim de realizar a previsão. A previsão será avaliada a através da verificação do MAPE entre os valores previstos e valores reais na região de Garibaldi, analisando os níveis de precisão entre a previsão e os dados reais.



Figura 16 - Fluxograma do Modelo de RNA Proposto.

Fonte: O autor (2018).

# 3.3.4 Modelagem das RNAs

Para a previsão serão configuradas três redes neurais, uma utilizando os dados do *Power Data Access Viewer*, que aqui será denominada Rede Neural PDAV, outra utilizando os dados do portal do clima, aqui denominada Rede Neural PC e a terceira com uma mescla dos dados com maior relevância dentre as duas fontes, denominada Rede Neural Mista. Os dados utilizados para o treinamento e validação das redes neurais pertencem ao intervalo de 1052 dias, de 20 de abril de 2016 até 06 de março de 2019.

Estes dados serão separados em conjuntos de vetores, em que cada vetor é um fator climático diferente, estes vetores são agrupados em uma matriz, na qual cada coluna corresponde a um dia, portanto para um conjunto de n variáveis de entrada com m dias amostrados, existe uma matriz de entrada de n x m, e um vetor alvo de 1 x m. A validação será realizada em intervalos diários para um período de 184 dias a frente, 6 meses, os dados utilizados para treinamento vão de 20 de abril de 2016 até 5 de setembro de 2018, enquanto os dados para a validação da rede vão de 6 setembro de 2018 a 6 de março de 2019.

A primeira rede utilizará as variáveis disponíveis através do *Power Data Access Viewer*. Nesta primeira rede neural, das 1052 amostras iniciais 182 foram removidas durante o préprocessamento dos dados, por apresentarem erros de leitura. Para a construção da Rede Neural PDAV foram inseridas na classificação *stepwise* 24 variáveis como previsores em formato matricial 686x24, e um vetor de 686x1 como resposta. Esta resposta corresponde a geração correspondente para cada amostra. As variáveis inseridas foram:

- · Precipitação;
- · Faixa de temperatura a 2 metros;
- · Temperatura a 2 metros;
- · Temperatura máxima a 2 metros;
- · Temperatura mínima a 2 metros;
- · Pressão superficial;
- · Temperatura de solo;
- · Umidade relativa;
- Umidade Específica a 2 metros;
- Ponto de gelo a 2 metros;
- · Velocidade mínima do vento a 50 metros;
- · Velocidade máxima do vento a 50 metros;
- · Faixa de velocidade do vento a 50 metros;
- · Velocidade do vento a 50 metros;
- · Direção do vento a 50 metros;
- · Velocidade mínima do vento a 10 metros;
- · Velocidade máxima do vento a 10 metros;
- · Faixa de velocidade do vento a 10 metros;
- · Velocidade do vento a 10 metros;
- · Direção do vento a 10 metros;
- · Fluxo térmico radiativo infravermelho;
- · Índice de irradiação sobre céu claro;
- · Irradiação de topo da atmosfera;
- · Irradiação incidente sobre uma superfície horizontal.

Através dos dados obtidos no Portal do Clima, nesta segunda rede neural, das 1052 amostras iniciais 226 foram removidas durante o pré-processamento dos dados, por apresentarem erros na sua leitura. Sendo assim, na construção da Rede Neural PC as 9 variáveis previsoras foram inseridas na classificação *stepwise*, também em formato matricial, agora de

642x9, e um vetor de 642x1 como resposta. Assim como na primeira, esta resposta corresponde a geração correspondente para cada amostra. As variáveis inseridas foram:

- · Temperatura média;
- · Temperatura máxima;
- · Temperatura mínima;
- · Sensação térmica;
- · Umidade relativa;
- · Umidade absoluta;
- · Pressão atmosférica absoluta;
- · Velocidade do vento;
- · Precipitação.

Na terceira rede (MISTA), as variáveis das duas fontes foram combinadas, sendo utilizadas as 9 variáveis climáticas da Rede Neural PC, porém, incluídas a estas as componentes de irradiação solar presentes na Rede Neural PDAV. Estas foram introduzidas na classificação *stepwise com 642* amostras de 13 previsores.

#### **RESULTADOS** 4

Neste capítulo serão abordados os resultados obtidos conforme a metodologia exposta no capítulo anterior. Serão expostos os resultados da regressão stepwise para cada rede neural, bem como as previsões realizadas por cada rede com seus respectivos MAPE. Também serão expostas previsões realizadas para meses de inverno e verão, afim de verificar a variação do MAPE. Ainda neste capítulo, será comparada a acuracidade de cada previsão através das métricas de avaliação apresentadas no capítulo 2, com os erros obtidos no processo de treinamento e na validação. Por fim, os resultados atingidos neste texto são comparados com os trabalhos apresentados na bibliografia.

#### 4.1 PREVISÕES REDE NEURAL PDAV

Antes da inserção dos dados na rede neural é realizado o pré-processamento dos dados, como apresentado no item 3.2.3. Para a regressão stepwise são inseridas as 24 variáveis candidatas. A classificação foi realizada através do software Matlab, 7 variáveis serão utilizadas como entrada. O resultado da regressão é apresentado pela Tabela 1.

				(continua)
Variável	Coeficiente de Regressão	Erro Padrão	t-estatístico	valor-p
X1	-0,043151659	0,013060473	-3,303988917	0,001003324
X2	-0,024594196	0,020407662	-1,205145162	0,228568349
X3	301,1588916	87,87789208	3,427015424	0,000647118
X4	0,146128007	0,2714612	0,538301635	0,590545754
X5	0,206756935	0,040442926	5,112313916	4,14E-07
X6	0,01148652	0,084916143	0,135268974	0,892439422
X7	-0,091478386	0,115942242	-0,788999632	0,430388436
X8	0,101891093	0,08490421	1,200071155	0,230531482
X9	0,101140231	0,084923601	1,190955524	0,234088435
X10	0,036480983	0,08867074	0,411420761	0,680894243
X11	0,041662573	0,089252935	0,466792193	0,640798743
X12	-0,002184275	0,154433592	-0,014143783	0,988719437
X13	-0,043175879	0,085160408	-0,50699474	0,612323602
X14	-0,057598577	0,187546702	-0,307115914	0,758849588
X15	-0,004430567	0,077205182	-0,057386914	0,95425392
X16	-0,027673951	0,127970237	-0,216253026	0,828855665

Tabela 1 - Regressão Stepwise Rede Neural PDAV

				(conclusão)
Variável	Coeficiente de Regressão	Erro Padrão	t-estatístico	valor-p
X17	-0,00031049	0,000787451	-0,394297429	0,693485633
X18	-0,000266761	0,000793073	-0,336364184	0,736700486
X19	-0,051653636	0,092675304	-0,557361382	0,577464772
X20	-0,068714699	0,183457022	-0,374554751	0,708108923
X21	2,269365417	0,184304379	12,31313888	1,38E-31
X22	-1,429370766	0,32433892	-4,407028201	1,22E-05
X23	0,590298222	0,10539614	5,600757525	3,10E-08
X24	4,063589399	1,945106616	2,089134532	0,037067887

Fonte: O autor (2019).

As linhas com valores em azul representam as variáveis que apresentaram menor *valorp*. O método classificatório selecionou as variáveis que melhor modelam a saída da rede neural, sendo assim, as variáveis selecionadas foram:

- x1. Precipitação;
- x2. Umidade específica;
- x3. Faixa de temperatura a 2 metros;
- x4. Irradiação incidente sobre uma superfície horizontal;
- x5. Fluxo térmico radiativo infravermelho;
- x6. Irradiação de topo da atmosfera;
- x7. Índice de irradiação sobre céu claro.

Esta rede foi então modelada com 15 neurônios ocultos, segundo teorema de Kolmogorov, com função de ativação sigmoide tangente hiperbólica e função de treinamento *bayesian regularization*. Um esquemático desta rede é apresentado pela Figura 17.

Figura 17 - Esquemático Rede Neural PDAV



Fonte: O autor (2019).

Existem outras funções de treinamento diferentes da *bayesian regularization*, que podem levar a um menor erro de treinamento. Em testes realizados com estas funções, que não possuem regularização, ocorreu excesso de treinamento e a rede generalizou em todos os casos testados. Com isso, as redes testadas apresentavam um baixo erro de treinamento, porém um alto erro na validação.

A inserção dos dados de treinamento foi realizada em formato matricial com uma matriz de 686x7, sendo 686 amostras de cada componentes. Estes dados foram separados conforme o método de validação cruzada. O treinamento foi realizado até o erro MAPE atingir o menor valor possível. A regressão deste treinamento apresentou uma acuracidade de 0,9637 entre os dados utilizados como alvo, e os dados de saída do treinamento, como mostra a Figura 18.



Figura 18 - Regressão Treinamento PDAV

Fonte: O autor (2019).

A saída do treinamento, ilustrada pela Figura 19, mostra o gráfico comparativo entre os dados reais de geração utilizados como alvos da rede neural, estes representados pela curva azul, com os dados de saída do treinamento, estes representados pela curva em vermelho. O MAPE atingido durante o treinamento foi de 14,86%. Como o método de treinamento utilizado previne que ocorra um *overfitting* no treinamento, este foi o menor erro calculado.

Figura 19 - Saída Treinamento PDAV



Fonte: O autor (2019).

Após o treinamento da rede neural, os dados de validação são inseridos como entrada da rede neural. Estes dados são inseridos em formato matricial 184x7. Desta vez, não existe alvo, com a rede neural treinada a saída da rede é por aprendizagem, e por meio dos dados inseridos na entrada a rede prevê os dados de saída. A Figura 20, mostra a regressão entre a saída e os dados de comparação, com uma acuracidade de 0,9513.





Fonte: O autor (2019).

Com um MAPE calculado em 12,97%, a Figura 21 ilustra a previsão realizada para 184 dias a frente. Os dados em azul representam os dados reais de geração, utilizados aqui somente como comparação e os dados em vermelho representam a saída.



Figura 21 - Saída Previsão PDAV

# 4.2 PREVISÕES REDE NEURAL PC

Dentre as 9 iniciais, após a classificação *stepwise*, foram selecionadas 7 variáveis de entrada. Os resultados obtidos nesta regressão são apresentados pela Tabela 2.

Variável	Coeficiente de Regressão	Erro Padrão	t-estatístico	valor-p
X1	4,827289132	7,231407084	0,66754493	0,504667408
X2	2,84379413	0,790744629	3,596349599	0,000347827
X3	-4,160569364	0,762871451	-5,453827587	7,07E-08
X4	-0,311684274	1,089901726	-0,285974658	0,774991092
X5	-0,677712121	0,070834273	-9,567573605	2,39E-20
X6	2,082154207	0,365966071	5,68947334	1,95E-08
X7	-0,013041153	0,005972613	-2,183491964	0,029365486
X8	1,72083659	0,250863311	6,859658291	1,64E-11
X9	0,009270599	0,001293391	7,167669439	2,13E-12

Tabela 2 - Regressão Stepwise Rede Neural PC

Fonte: O autor (2019).

Após a classificação, duas variáveis foram removidas, a temperatura média e a sensação térmica, mostradas em vermelho na Tabela 2. Sendo assim, as componentes utilizadas com entradas nesta rede neural serão:

- x1. Temperatura máxima;
- x2. Temperatura mínima;
- x3. Umidade relativa;
- x4. Umidade absoluta;
- x5. Pressão atmosférica absoluta;
- x6. Velocidade do vento;
- x7. Precipitação.

Ainda segundo o teorema de Kolmogorov, esta RNA será inicializada com 15 neurônios ocultos, com função de ativação sigmoide tangente hiperbólica e função de treinamento *bayesian regularization*, aqui representada pela Figura 22.



Figura 22 - Esquemático Rede Neural PC

Nesta rede neural, o treinamento foi realizado com 642 amostras das diferentes variáveis, assim como na primeira rede, esta matriz 642x7 é dividida para o processo de treinamento em dados de treinamento, teste e validação. A Figura 23 mostra a regressão entre os valores da saída da rede neural com seus respectivos alvos. Esta regressão calculou uma acuracidade de 0,8940.

Fonte: O autor (2019).





Fonte: O autor (2019).

A Figura 24 mostra a saída do treino em comparação com os alvos, os dados em azul são os dados reais de geração, enquanto os dados em vermelho são os valores da saída da rede. O MAPE obtido durante o treinamento foi de 29,82 %, o que não mostrou uma boa aprendizagem.

Figura 24 - Saída Treinamento PC



Fonte: O autor (2019).

A função de treinamento utilizada é a mesma da primeira rede neural, a *bayesin retropropagation*, que previne a ocorrência de um *overfitting* no treinamento. Sendo assim, este foi o menor erro calculado. Utilizando outras funções sem regularização é possível chegar a um menor erro de treinamento, porém com o acesso de treinamento a rede generaliza ocasionando um erro maior na previsão. Na sequência é realizada a inserção da matriz 184x7 para calcular a previsão dos 184 dias a frente. A Figura 25 ilustra a regressão entre os dados de saída da rede neural, representado pelos círculos pretos e a faixa tracejada, que representa a saída quando esta é igual aos dados de geração comparados. Esta regressão apresentou um valor de 0,6842 de acuracidade.



Figura 25 - Regressão Previsão PC

A Figura 26 ilustra a saída da rede neural, os dados em azul representam os dados reais e os dados em vermelho representam a saída da rede neural, como é possível observar, não ocorreu uma boa aprendizagem com os dados inseridos no treinamento e o MAPE calculado para esta previsão foi de 30,17 %.

Figura 26 - Saída Previsão PC



Fonte: O autor (2019).

# 4.3 PREVISÕES REDE NEURAL MISTA

Nesta rede neural foram incluídas às 9 variáveis iniciais da Rede Neural PC, as 4 componentes de irradiação disponíveis na Rede Neural PDAV, fluxo térmico radiativo infravermelho, índice de irradiação sobre céu claro, irradiação de topo da atmosfera e irradiação incidente sobre uma superfície horizontal. A classificação eliminou 6 variáveis, e os resultados são mostrados na Tabela 3.

Tabela 3 - Regressão Stepwise Rede Neural Mista

				(continua)
Variável	Coeficiente de Regressão	Erro Padrão	t-estatístico	valor-p
X1	-0,114535407	0,122326241	-0,936311019	0,349470712
X2	-0,058725577	0,123740206	-0,474587681	0,635244829
X3	-0,146733641	0,115817938	-1,266933636	0,205645695
X4	-0,15565293	0,097509639	-1,596282502	0,110925609
X5	-0,091540338	0,010329774	-8,861794853	7,93E-18
X6	0,227540772	0,061816239	3,680922272	0,000252208
X7	-0,008584455	0,002257894	-3,801973747	0,000157432
X8	0,14993952	0,093169841	1,609313906	0,108047077
X9	0,001602236	0,000558812	2,867219947	0,004278704
X10	2,673757144	0,091062359	29,36182602	3,02E-120

				(••••••••••••••••••••••••••••••••••••••
Variável	Coeficiente de Regressão	Erro Padrão	t-estatístico	valor-p
X11	-1,452655915	0,260884936	-5,568186246	3,81E-08
X12	0,003277474	0,002346042	1,397023161	0,162896913
X13	0,360034544	0,077035833	4,673598387	3,62E-06

Fonte: O autor (2019).

Portanto, como esta rede neural teve 7 variáveis de entrada.

- x1. Umidade relativa;
- x2. Umidade absoluta;
- x3. Pressão atmosférica absoluta;
- x4. Precipitação;
- x5. Irradiação incidente sobre uma superfície horizontal;
- x6. Fluxo térmico radiativo infravermelho;
- x7. Irradiação de topo da atmosfera.

A RNA foi modelada com 15 neurônios ocultos, com função de ativação sigmoide tangente hiperbólica e função de treinamento *bayesian regularization*, como é ilustrada pela Figura 27.

Figura 27 - Esquemático Rede Neural Mista



Fonte: O autor (2019).

Os dados utilizados como entrada desta rede neural são balizados pelos dados presentes na segunda rede neural, com isso, a matriz utilizada como entrada desta rede será de 642x7, contendo 7 variáveis para 642 amostras. A Figura 28 ilustra a regressão entre as saídas do treinamento e seus respectivos alvos. Pode ser observado que com a inclusão dos dados relacionados a irradiação solar a acuracidade desta regressão subiu de 0,894 para 0,9684.

(conclusão)



Figura 28 - Regressão Treinamento Mista



Neste treinamento a rede neural Mista apresentou um erro de 15,37 %, uma redução no erro obtido no treinamento da rede Portal Clima. A Figura 29 mostra a saída do treinamento comparada com os dados reais utilizados para a aprendizagem.





Fonte: O autor (2019).

Com a rede neural treinada são inseridos os dados para a previsão. Na Figura 30 é apresentado o resultado da regressão da rede neural Mista entre a saída da previsão e os dados reais utilizados para a comparação. A acuracidade desta regressão com a inserção dos dados de irradiação melhorou de 0,6842 para 0,9374. A Figura 31 ilustra a previsão realizada.



Figura 30 - Regressão Previsão Mista

Fonte: O autor (2019).

Figura 31 - Saída Previsão Mista



Fonte: O autor (2019).

Em que, na Figura 31, os dados em vermelho representam a previsão e os dados em azul representam os valores reais utilizados para comparação. Nesta previsão um MAPE de 13,87 % foi calculado, resultado similar ao da primeira rede aqui apresentada. O erro foi significativamente reduzido em comparação ao da rede neural Portal Clima. Esta redução devese da inclusão dos dados relacionados a irradiação solar.

# 4.4 ANÁLISE COMPARATIVA DE PRECISÂO ENTRE AS REDES NEURAIS

A Tabela 4 apresenta um comparativo dos erros apresentados durante a fase de treinamento. As métricas utilizadas foram o MAPE, o RMSE, o MAE e o MSE. A última coluna mostra o valor da regressão entre a saída e seus respectivos alvos.

TREINAMENTO	MAPE [%]	RMSE [kWh]	MAE [kWh]	MSE [kWh <sup>2</sup> ]	R
RNA PDAV	14,8694	2,1378	1,5766	4,5700	0,9637
RNA PC	29,8201	3,6118	2,7668	13,0452	0,8940
RNA MISTA	15,3763	2,0062	1,4994	4,0250	0,9684

Tabela 4 - Comparativo dos erros de treinamento entre as três RNAs

Fonte: O autor (2019).

A Tabela 5 mostra um comparativo dos erros apresentados na previsão para 184 dias para as três redes neurais utilizadas.

Tabela 5 - Comparativo dos erros de previsão entre as três RNAs

PREVISÃO	MAPE [%]	RMSE [kWh]	MAE [kWh]	MSE [kWh <sup>2</sup> ]	R
RNA PDAV	12,9794	2,4297	1,8155	5,9036	0,9513
RNA PC	30,1797	6,5000	5,1610	42,2498	0,6842
RNA MISTA	13,8708	2,8537	2,1267	8,1436	0,9374

Fonte: O autor (2019).

É possível observar uma boa aprendizagem tanto para a primeira rede quanto para a terceira, apresentando baixos valores de MAPE, entre 14,86 e 15,37 % no treinamento e entre

12,97 e 13,87 % na previsão. No entanto, a segunda rede neural não apresentou boa aprendizagem, com alto valor de MAPE, praticamente o dobro das outras duas, essa diferença pode ser melhor observada através dos valores do MSE, que vão de 5,9 e 8,14 para 42,24 kWh<sup>2</sup> na segunda rede.

# 4.5 PREVISÕES REALIZADAS PARA DIFERENTES TEMPORADAS

Com o objetivo de verificar as variações entre o erro nas previsões para diferentes épocas do ano, através da verificação do MAPE, foram realizadas previsões para temporadas de inverno e verão. Sendo assim, no primeiro caso observou-se o comportamento para previsões realizadas para os meses de inverno, com dados de validação para 92 dias de inverno a partir de 21 de junho de 2018. No segundo, para a temporada de verão, foram utilizadas 92 amostras de dias quentes, partindo-se de 17 de novembro. Posteriormente, foi abordada a previsão para um único mês, sendo utilizados como validação os meses de julho, novembro e dezembro de 2018. A Tabela 6 mostra o comparativo entre os erros nas previsões realizadas utilizando as três redes neurais.

	PREVI	PREVISÃO MAPE [%]				
	PDAV	PC	MISTA	Intervalo		
INVERNO	21,6665	60,2676	19,8252	92 dias		
VERÃO	12,4300	28,5767	14,5704	92 dias		
JULHO	22,9639	61,0088	28,2650	31 dias		
NOVEMBRO	9,0993	26,2045	9,1215	30 dias		
DEZEMBRO	10,9700	23,2913	11,5193	31 dias		

Tabela 6 - Tabela Comparativa de MAPE.

Fonte: O autor (2019).

Com estas previsões é perceptível o aumento no MAPE em situações de baixa geração de energia. Como o erro é calculado através da diferença absoluta entre o valor real e o previsto dividido pelo valor real, quando o valor real é pequeno o erro tende a aumentar. Na Figura 32 é ilustrado a previsão realizada para o mês de julho, com MAPE de 22,96 % e a Figura 33 mostra a previsão realizada para o mês de novembro com MAPE de 9,09 %.



Figura 32 - Previsão Realizada para julho, MAPE 22,96 %.

Fonte: O autor (2019).



Figura 33 - Previsão Realizada para novembro, MAPE 9,09 %

Fonte: O autor (2019).

Como o horizonte de dias para a validação da previsão era menor, para cada horizonte foi realizado um novo treinamento. Este treinamento permitiu aumentar o número de dados amostrais utilizados. Porém os erros durante o processo de treinamento permaneceram semelhantes, entre 14,37 e 15 %, como é mostrado pela Tabela 7.

TREINAMENTO MAPE [%]					
PDAV	PC	MISTA			
14,8947	31,0230	14,6365			
14,3787	30,2135	14,6040			
14,6470	30,7490	14,4951			
15,0096	31,6445	14,3349			
14,7533	33,5548	14,1208			
14,8694	29,8201	15,3763			
	AMENTO PDAV 14,8947 14,3787 14,6470 15,0096 14,7533 14,8694	AMENTO MAPE [9PDAVPC14,894731,023014,378730,213514,647030,749015,009631,644514,753333,554814,869429,8201			

Tabela 7 - MAPE Calculado nos Diferentes Treinamentos

Fonte: O autor (2019).

# 4.6 COMPARAÇÃO COM A BIBLIOGRAFIA

As Tabelas 8 e 9 mostram as diferentes previsões encontradas na bibliografia comparadas com as melhores previsões das redes elaboradas neste trabalho. Estas previsões foram todas realizadas através de redes neurais artificiais, sendo o MAPE a métrica de validação empregada. Em Lahouar, Mejri e Slama (2017) são realizadas previsões com dados obtidos de uma região da Austrália, em Cheng, GE e Cao (2012) as previsões são realizadas para uma região na China e as previsões de Haque, Nehrir e Mandal (2013) são para a cidade de Ashland no estado do Óregon, Estados Unidos.

Tabela 8 - Comparação de MAPE para Previsão a Médio Prazo

Autor	Uorizonto	Intervalo	
Autor	Horizonite	Intervalo	MAPE [%]
Lahouar, Mejri e Slama (2017)	Médio Prazo	1 a 7 de outubro	10,26
Lahouar, Mejri e Slama (2017)	Médio Prazo	1 a 7 de janeiro	19,16
Lahouar, Mejri e Slama (2017)	Médio Prazo	1 a 7 de abril	28,98
Lahouar, Mejri e Slama (2017)	Médio Prazo	1 a 7 de julho	17,45
O autor (2019)	Médio Prazo	1ª semana julho	22,23
O autor (2019)	Médio Prazo	1 <sup>a</sup> semana novembro	13,20
O autor (2019)	Médio Prazo	1 <sup>a</sup> semana dezembro	9,78

Fonte: O autor (2019).
Autor	Horizonte	Intervalo	MAPE [%]
Cheng, GE e Cao (2012)	Longo Prazo	1º quarto do ano, 90 dias	20,24
Cheng, GE e Cao (2012)	Longo Prazo	2º quarto do ano, 91 dias	19,31
Cheng, GE e Cao (2012)	Longo Prazo	3º quarto do ano, 92 dias	15,52
Cheng, GE e Cao (2012)	Longo Prazo	4º quarto do ano, 92 dias	21,98
Haque, Nehrir e Mandal (2013)	Longo Prazo	Inverno	15,77
Haque, Nehrir e Mandal (2013)	Longo Prazo	Primavera	16,08
Haque, Nehrir e Mandal (2013)	Longo Prazo	Verão	15,41
Haque, Nehrir e Mandal (2013)	Longo Prazo	Outono	19,17
O autor	Longo Prazo	6 meses	12,97
O autor	Longo Prazo	Inverno - 92 dias	19,82
O autor	Longo Prazo	Verão - 92 dias	12,43
O autor	Longo Prazo	1 mês - julho	22,96
O autor	Longo Prazo	1 mês - novembro	9,09
O autor	Longo Prazo	1 mês - dezembro	10,97

Tabela 9 - Comparação de MAPE para Previsão a Longo Prazo

Fonte: O autor (2019).

Nas previsões realizadas para médio prazo por Lahouar, Mejri e Slama (2017) assim como os resultados aqui apresentados, mostram a variação no MAPE calculado para as diferentes épocas do ano as quais são realizadas as previsões. Nas previsões realizadas para longo prazo esta variação também ocorre. Os autores Cheng, GE e Cao (2012) e Haque, Nehrir e Mandal (2013) realizaram previsões para diferentes épocas do ano, o primeiro realizou suas previsões para os quatro trimestres do ano analisado, o segundo segregou suas previsões pelas estações do ano. Ambas as previsões, como também as previsões apresentadas neste texto, o MAPE calculado varia de acordo com a época.

Os modelos simulados para previsões entre diferentes horizontes, assim como o modelo de previsão para 6 meses atingiram resultados satisfatórios, similares aos encontrados na bibliografia. É preciso ressaltar o desempenho da métrica de validação MAPE aplicada em situações de baixos valores reais, pois quando há uma divisão por um dado real de baixo valor o erro tende a aumentar. Além disso, as redes neurais mostraram-se realmente capazes de lidar com as intermitências impostas pela geração fotovoltaica. A não linearidade na geração de energia foi superada pela rede neural, em que esta trouxe resultados e variações semelhantes aos dados reais comparados.

O emprego de aprendizado de máquina, neste caso as redes neurais artificiais, qualificaram características climáticas próprias para a quantidade de energia fotovoltaica gerada. Em especial as componentes derivadas da irradiação solar, como a irradiação sobre uma superfície horizontal. A modelagem das redes neurais aqui apresentadas, buscou por meio do processo de classificação de variáveis *stepwise* avaliar dentre as candidatas, aquelas que mais correspondem com a resposta esperada. Esta seleção reforçou as variáveis presentes com maior frequência na bibliografia, dentre as quais podem ser citadas a temperatura máxima do ambiente, a umidade relativa do ar e a irradiação sobre uma superfície horizontal. O processo de aprendizagem atingiu um MAPE entre 14 e 15% no treinamento. Este procedimento foi realizado buscando preservar a capacidade de aprendizagem da rede neural, aplicando uma função de treinamento que prevenisse o excesso de treinamento. O objetivo de realizar uma previsão de 6 meses de geração fotovoltaica foi alcançado com resultados satisfatórios, com um MAPE de 12,97%, erro menor do que as previsões a longo prazo presentes na bibliografia que utilizam redes neurais artificiais.

#### 5 CONCLUSÕES

As redes neurais artificias foram escolhidas como método de aprendizado de máquina para estudo, com ponto central nas suas aplicações relacionadas à previsão de geração fotovoltaica. O emprego da regressão *stepwise* ajudou na classificação das variáveis que melhor representavam o sistema, reduzindo o número de variáveis de entrada da rede neural e consequentemente reduzindo o tempo de treinamento em conjunto com o aumento da acuracidade do modelo.

As variáveis componentes da irradiação solar mostraram grande influência para a precisão da previsão. Quando estas foram incluídas em conjunto com os dados climáticos do Portal Clima, a acuracidade da previsão melhorou significativamente. Quando os valores reais utilizados para comparação são baixos, o MAPE calculado tende a ser maior. As previsões realizadas para os meses de inverno e para o mês de julho apresentaram MAPE elevado de 19,82 % e 22,96 % respectivamente. No entanto, previsões realizadas para os meses de verão, para o mês de novembro e para o mês de dezembro apresentaram MAPE de 12,43 %, 9,09 % e 10,97 %. Com estes resultados, foi possível observar um aumento no MAPE nos dias de baixa geração.

O melhor resultado do objeto deste estudo atingiu um MAPE de 12,97 %, para o horizonte de seis meses, estando estes resultados de acordo com os encontrados na bibliografia. A não linearidade nos dados de geração fotovoltaica analisados e dias com baixa geração culminaram no acréscimo do MAPE nas previsões realizadas. Em trabalhos futuros é possível empregar esta mesma rede neural validada, para estimar a geração de outros locais. Há também a possibilidade de empregar outras técnicas de aprendizado de máquina citadas neste texto, como por exemplo máquina de vetores de suporte. Também podem ser analisadas previsões de curto prazo, para um ou dois dias a frente, utilizando conjuntos de dados segmentados em minutos.

### REFERÊNCIAS

ABDELMUTALAB, Ameen; ASSALEH, Khaled; EL-TARHUNI, Mohamed. Automatic Modulation Classification Using Hierarchical Polynomial Classifier and Stepwise Regression. 2016. Disponível em: <a href="https://ieeexplore.ieee.org/document/7565127">https://ieeexplore.ieee.org/document/7565127</a>. Acesso em: 22 abr. 2019.

ABEDIN, Zainal et al. A model for prediction of monthly solar radiation of different meterological locations of Bangladesh using artificial neural network data mining tool. 2017. Disponível em: <a href="https://ieeexplore.ieee.org/document/7912993/">https://ieeexplore.ieee.org/document/7912993/</a>. Acesso em: 25 ago. 2018.

#### ABSOLAR (Brasil). ENERGIA SOLAR FOTOVOLTAICA ATINGE MARCA HISTÓRICA DE 500 MW EM MICROGERAÇÃO E MINIGERAÇÃO DISTRIBUÍDA

**NO BRASIL.** 2019. Disponível em: <a href="http://www.absolar.org.br/noticia/noticias-externas/energia-solar-fotovoltaica-atinge-marca-historica-de-500-mw-em-microgeracao-e-minigeracao-distribuid.html">http://www.absolar.org.br/noticia/noticias-externas/energia-solar-fotovoltaica-atinge-marca-historica-de-500-mw-em-microgeracao-e-minigeracao-distribuid.html</a>>. Acesso em: 05 jul. 2019.

AHMED, Mohammed M. et al. An Optimized K-Nearest Neighbor Algorithm for Extending Wireless Sensor Network Lifetime. 2018. Disponível em: <http://www.researchgate.net/publication/322704994 An Optimizaed K-Nearest-\_Neigbor\_Algoritm\_for\_Extending\_Wireless\_Sensor\_Network\_Lifetime>. Acesso em: 04 set. 2018.

ALANAZI, Mohana; ALANAZI, Abdulaziz; KHODAEI, Amin. Long-Term Solar Generation Forecasting. 2016. Disponível em: <a href="https://ieeexplore.ieee.org/document/7519883/">https://ieeexplore.ieee.org/document/7519883/</a>>. Acesso em: 21 ago. 2018.

ALZAHRANI, Ahmad et al. **Solar Irradiance Forecasting Using Deep Recurrent Neural Networks**. 2017. Disponível em: <a href="https://ieeexplore.ieee.org/document/8191206/">https://ieeexplore.ieee.org/document/8191206/</a>>. Acesso em: 23 ago. 2018.

ANDERSON, William W.; YAKIMENKO, Oleg A. Using Neural Networks to Model and Forecast Solar PV Power Generation at Isle of Eigg. 2018. Disponível em: <a href="https://ieeexplore.ieee.org/document/8372522/">https://ieeexplore.ieee.org/document/8372522/</a>. Acesso em: 21 ago. 2018.

ANEEL (Brasil). **BIG - Banco de Informações de Geração**. 2018. Disponível em: <a href="http://www2.aneel.gov.br/area.cfm?idArea=8&idPerfil=3>">http://www2.aneel.gov.br/area.cfm?idArea=8&idPerfil=3></a>. Acesso em: 05 nov. 2018.

ANEEL (Brasil). N° **482**: RESOLUÇÃO NORMATIVA Nº 482. 2012. 12 p. Disponível em: <a href="http://www2.aneel.gov.br/cedoc/ren2012482.pdf">http://www2.aneel.gov.br/cedoc/ren2012482.pdf</a>>. Acesso em: 02 nov. 2018.

BARRETO, Jorge M. Introdução às Redes Neurais Artificiais. 2002. Disponível em: <a href="http://www.inf.ufsc.br/~j.barreto/tutoriais/Survey.pdf">http://www.inf.ufsc.br/~j.barreto/tutoriais/Survey.pdf</a>>. Acesso em: 21 ago. 2018.

#### BARROS, Ricardo; SAUAIA, Rodrigo; KOLOSZUK, Ronaldo. **PERSPECTIVAS PARA A GERAÇÃO CENTRALIZADA SOLAR FOTOVOLTAICA NO BRASIL.** 2019.

Disponível em: <http://www.absolar.org.br/noticia/artigos-da-absolar/artigo-perspectivas-para-a-geracao-centralizada-solar-fotovoltaica-no-brasil.html>. Acesso em: 05 jul. 2019.

CANTOR, Guillermo Andrés Rodríguez. **Influência Dos Fatores Climáticos No Desempenho De Módulos Fotovoltaicos Em Regiões De Clima Tropical**. 2018. 176 f. Dissertação (Mestrado) - Curso de Pós-graduação em Energias Renováveis, Centro de Energias Alternativas e Renováveis, Universidade Federal da Paraíba, João Pessoa, 2017. Disponível em:

<a href="http://www.cear.ufpb.br/arquivos/ppger/documentos/versao\_final\_Guillermo.pdf">http://www.cear.ufpb.br/arquivos/ppger/documentos/versao\_final\_Guillermo.pdf</a>>. Acesso em: 13 set. 2018.

CHATURVEDI, D. K.; ISHA. **Solar Power Forecasting: A Review**. 2016. Disponível em: <a href="https://pdfs.semanticscholar.org/a679/0993e18db1b5d488b161194fa07a3d5c139f.pdf">https://pdfs.semanticscholar.org/a679/0993e18db1b5d488b161194fa07a3d5c139f.pdf</a>>. Acesso em: 20 ago. 2018.

CHEN, Austin H; TSAU, Yin-wu; LIN, Ching-heng. **Novel methods to identify biologically relevant genes for leukemia and prostate cancer from gene expression profiles**. 2010. Disponível em: <a href="https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-274">https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-274</a>>. Acesso em: 04 set. 2018.

CHENG, Hang; GE, Peng-jiang; CAO, Wu-shun. Forecasting Research of Long-term Solar Irradiance and Output Power for Photovoltaic generation system. 2012. Disponível em: <a href="https://ieeexplore.ieee.org/document/6301338/">https://ieeexplore.ieee.org/document/6301338/</a>>. Acesso em: 03 set. 2018.

CHIANG, Po-han et al. Forecasting of Solar Photovoltaic System Power Generation using Wavelet Decomposition and Biascompensated Random Forest. 2017. Disponível em: <a href="https://ieeexplore.ieee.org/document/7923968/">https://ieeexplore.ieee.org/document/7923968/</a>>. Acesso em: 17 set. 2018.

CHUPONG, Charnon; PLANGKLANG, Boonyang. Forecasting power output of PV grid connected system in Thailand without using solar radiation measurement. 2011. Disponível em:

<a href="https://www.researchgate.net/publication/251712587\_Forecasting\_power\_output\_of\_PV\_grid\_connected\_system\_in\_Thailand\_without\_using\_solar\_radiation\_measurement">https://www.researchgate.net/publication/251712587\_Forecasting\_power\_output\_of\_PV\_grid\_connected\_system\_in\_Thailand\_without\_using\_solar\_radiation\_measurement</a>>. Accesso em: 20 maio 2019.

DEMUTH, Howard; BEALE, Mark. **Neural Network Toolbox User's Guide:** For Use with MATLAB. 4. ed. Natick: The Mathworks, 2000. 846 p. Disponível em: <ftp://est.ufmg.br/pub/lourdes/ebook-artificial\_neural\_network\_(matlab\_toolbox).pdf>. Acesso em: 15 mar. 2019.

GARIBALDI. PREFEITURA GARIBALDI. **Portal Clima.** 2019. Disponível em: <a href="http://clima.garibaldi.rs.gov.br/historico.aspx?EST\_ID=1>">http://clima.garibaldi.rs.gov.br/historico.aspx?EST\_ID=1></a>. Acesso em: 07 mar. 2019.

GIGONI, Lorenzo et al. **Day-Ahead Hourly Forecasting of Power Generation from Photovoltaic Plants**. 2018. Disponível em: <a href="https://ieeexplore.ieee.org/document/8066360/">https://ieeexplore.ieee.org/document/8066360/</a>. Acesso em: 20 ago. 2018.

GHUNEM, Refat A.; ASSALEH, Khaled; EL-HAG, Ayman H.. **Artificial Neural Networks** with Stepwise Regression for Predicting Transformer Oil Furan Content. 2012. Disponível em: <a href="https://ieeexplore.ieee.org/document/6180233">https://ieeexplore.ieee.org/document/6180233</a>. Acesso em: 15 abr. 2019.

HAAG, Rafael et al (Org.). **ATLAS SOLAR DO RIO GRANDE DO SUL.** 2018. Disponível em: <a href="https://atlassolarrs.com/atlas-solar-ebook">https://atlassolarrs.com/atlas-solar-ebook</a>>. Acesso em: 01 dez. 2018.

HAQUE, Ashraf U.; NEHRIR, M. Hashem; MANDAL, Paras. **Solar PV Power Generation Forecast Using a Hybrid Intelligent Approach**. 2013. Disponível em: <a href="https://ieeexplore.ieee.org/document/6672634">https://ieeexplore.ieee.org/document/6672634</a>>. Acesso em: 26 out. 2018.

HASAN, Hafidh; MUNAWAR, Muhammad Ridha; SIREGAR, Ramdhan Halid. Neural Network-Based Solar Irradiance Forecast for Peak Load Management of Grid-Connected Microgrid with Photovoltaic Distributed Generation. 2017. Disponível em: <https://ieeexplore.ieee.org/document/8253235/>. Acesso em: 27 ago. 2018.

HAYKIN, Simon. **Neural Networks and Learning Machines**. 3. ed. Hamilton: Pearson Prentice Hall, 2009. 906 p. 938 f. Disponível em: <http://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf>. Acesso em: 16 out. 2018.

HAYKIN, Simon. **Redes Neurais - Princípios e práticas**. 2. ed. Porto Alegre: Bookman, 2001. 908 p. Disponível em: <a href="http://www.ncdd.com.br/livros/redes\_neurais\_simon\_haykin.pdf">http://www.ncdd.com.br/livros/redes\_neurais\_simon\_haykin.pdf</a>>. Acesso em: 06 out. 2018.

HOLMGREN, William F.; GROENENDYK, Derek G. **An Open Source Solar Power Forecasting Tool Using PVLIB-Python**. 2016. Disponível em: <a href="https://ieeexplore.ieee.org/document/7749755/">https://ieeexplore.ieee.org/document/7749755/</a>>. Acesso em: 08 set. 2018.

INMET (Brasil). **BDMEP - Banco de Dados Meteorológicos para Ensino e Pesquisa.** 2018. Disponível em: <a href="http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep">http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep</a>. Acesso em: 09 out. 2018.

KOLOSZUK, Ronaldo; SAUAIA, Rodrigo. **Renováveis no Brasil: Maturidades Diferentes para Cada Fonte Exigem Cuidados Especiais.** 2018. Disponível em: <a href="http://www.absolar.org.br/noticia/artigos-da-absolar/renovaveis-no-brasil-maturidades-diferentes-para-cada-fonte-exigem-cuidados-especiais.html">http://www.absolar.org.br/noticia/artigos-da-absolar/renovaveis-no-brasil-maturidadesdiferentes-para-cada-fonte-exigem-cuidados-especiais.html</a>>. Acesso em: 04 nov. 2018.

KOUSIOURIS, George et al. A Service-Oriented Framework for GNU Octave-Based Performance Prediction. 2010. Disponível em: <a href="https://ieeexplore.ieee.org/document/5557222/">https://ieeexplore.ieee.org/document/5557222/</a>>. Acesso em: 29 ago. 2018.

LAHOUAR, Ali; MEJRI, Amal; SLAMA, Jaleleddine Ben Hadj. **Importance Based Selection Method for Day-ahead Photovoltaic Power Forecast Using Random Forests**. 2017. Disponível em: <a href="https://ieeexplore.ieee.org/document/8066171/">https://ieeexplore.ieee.org/document/8066171/</a>. Acesso em: 26 ago. 2018.

LIU, Zhao; ZHANG, Ziang. Solar Forecasting by K-Nearest Neighbors Method with Weather Classification and Physical Model. 2016. Disponível em: <a href="https://ieeexplore.ieee.org/document/7747859/">https://ieeexplore.ieee.org/document/7747859/</a>>. Acesso em: 25 ago. 2018.

LOPEZ, Alvaro Gustavo Talavera. **Controle Preditivo com Aprendizado po Reforço para Produção de Óleo em Poços Inteligentes**. 2010. 113 f. Dissertação (Mestrado) - Curso de Pós-graduação em Engenharia Elétrica, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2010. Cap. 2. Disponível em: <a href="http://www2.dbd.puc-rio.br/pergamum/tesesabertas/0812725\_10\_cap\_02.pdf">http://www2.dbd.puc-rio.br/pergamum/tesesabertas/0812725\_10\_cap\_02.pdf</a>>. Acesso em: 17 set. 2018.

MAJUMDER, Irani; BEHERA, Manoja Kumar; NAYAK, Niranjan. **Solar Power Forecasting Using a Hybrid EMD-ELM Method**. 2017. Disponível em: <a href="https://ieeexplore.ieee.org/document/8074179/">https://ieeexplore.ieee.org/document/8074179/</a>>. Acesso em: 03 set. 2018.

MATHWORKS (Estados Unidos). **Statistics and Machine Learning Toolbox**. 2018. Disponível em: <www.mathworks.com/help/stats/index.html>. Acesso em: 29 ago. 2018.

MEENAL, R. et al. **Solar Radiation Resource Assessment using WEKA**. 2018. Disponível em: <a href="https://ieeexplore.ieee.org/document/8398960/">https://ieeexplore.ieee.org/document/8398960/</a>>. Acesso em: 08 set. 2018.

NASA (Estados Unidos). **POWER Data Access Viewer**. 2018. Disponível em: <a href="https://power.larc.nasa.gov/data-access-viewer/">https://power.larc.nasa.gov/data-access-viewer/</a>. Acesso em: 25 set. 2018.

OLIVEIRA, Jivago Schumacher de et al. **Disponibilidade De Irradiância Solar Para Geração De Energia Elétrica Na Alemanha E No Estado Do Rio Grande Do Sul**. 2013. Disponível em:

<a href="https://www.periodicos.unifra.br/index.php/disciplinarumNT/article/view/1330/1262">https://www.periodicos.unifra.br/index.php/disciplinarumNT/article/view/1330/1262</a>. Acesso em: 17 set. 2018.

PALUSZEK, Michael; THOMAS, Stephanie. **Matlab Machine Learning**. New Jersey: Apress, 2017. 326 p. Disponível em:

<a href="https://universalflowuniversity.com/Books/Computer%20Programming/Machine%20Learning%20and%20Deep%20Learning/MATLAB%20Machine%20Learning.pdf">https://universalflowuniversity.com/Books/Computer%20Programming/Machine%20Learning</a>, Acesso em: 14 set. 2018.

PEREIRA, Enio Bueno et al. **Atlas Brasileiro de Energia Solar**. 2. ed. São José dos Campos: E-Book, 2017. 88 p. Disponível em: <a href="http://labren.ccst.inpe.br/atlas\_2017.html">http://labren.ccst.inpe.br/atlas\_2017.html</a>. Acesso em: 11 set. 2018.

PHYTON SOFTWARE FOUNDATION (Estados Unidos). **General Python FAQ**. 2018. Disponível em: <a href="https://docs.python.org/3/faq/general.html#what-is-python">https://docs.python.org/3/faq/general.html#what-is-python</a>. Acesso em: 11 out. 2018.

PINHO, João Tavares; GALDINO, Marco Antonio (Org.). **Manual de Engenharia para Sistemas Fotovoltaicos.** Rio de Janeiro: Cepel - Cresesb, 2014. 530 p. Disponível em: <http://www.cresesb.cepel.br/publicacoes/download/Manual\_de\_Engenharia\_FV\_2014.pdf>. Acesso em: 05 jul. 2019.

QIU, Xueheng et al. Ensemble Deep Learning for Regression and Time Series Forecasting. 2014. Disponível em: <a href="https://ieeexplore.ieee.org/document/7015739/">https://ieeexplore.ieee.org/document/7015739/</a>. Acesso em: 07 set. 2018.

ROLA, Marcelo Coleto. **Previsão Da Geração De Energia Elétrica No Médio Prazo Para O Estado Do Rio Grande Do Sul Empregando Redes Neurais Artificiais**. 2017. 94 f. Dissertação (Mestrado) - Curso de Pós-graduação em Engenharia Mecânica, Escola de

Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Disponível em: <a href="https://www.lume.ufrgs.br/handle/10183/157828">https://www.lume.ufrgs.br/handle/10183/157828</a>>. Acesso em: 20 set. 2018.

ROUAUD, Mathieu. **Probability, Statistics and Estimation:** Propagation of Uncertainties in Experimental Measurement. Boudiguen: Creative Commons, 2013. 191 p. Disponível em: <a href="http://www.incertitudes.fr/book.pdf">http://www.incertitudes.fr/book.pdf</a>>. Acesso em: 25 maio 2019.

SAINI, Lalit Mohan. Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks. 2007. Disponível em: <a href="https://www.sciencedirect.com/science/article/abs/pii/S0378779607002258">https://www.sciencedirect.com/science/article/abs/pii/S0378779607002258</a>. Acesso em: 01 maio 2019.

SANGRODY, Hossein et al. **On the Performance of Forecasting Models in the Presence of Input Uncertainty**. 2017. Disponível em: <a href="https://ieeexplore.ieee.org/document/8107379/">https://ieeexplore.ieee.org/document/8107379/</a>>. Acesso em: 07 set. 2017.

SANSA, Ines et al. **PV power forecasting using different artificial neural networks strategies**. 2014. Disponível em: <a href="https://ieeexplore.ieee.org/document/6835397/">https://ieeexplore.ieee.org/document/6835397/</a>. Acesso em: 11 set. 2018.

SCOLAR, José; MARTINS, Dinival; ESCOBEDO, João Francisco. Estimativa da irradiação total sobre uma superfície inclinada a partir da irradiação global na horizontal. 2003. Disponível em:

<a href="http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S0102-261X2003000300004">http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S0102-261X2003000300004</a>>. Acesso em: 04 dez. 2018.

SHI, Jie et al. Forecasting Power Output of Photovoltaic System Based on Weather Classification and Support Vector Machine. 2011. Disponível em: <a href="https://ieeexplore.ieee.org/document/6074294/">https://ieeexplore.ieee.org/document/6074294/</a>>. Acesso em: 19 set. 2018.

SILVA, Jéssica Alice Alves da et al. **TRATAMENTO E ANÁLISE DE DADOS SOLARIMÉTRICOS DA ESTAÇÃO METEOROLÓGICA DA EMC/UFG.** 2018. Disponível em: <a href="http://anaiscbens.emnuvens.com.br/cbens/article/view/725/725">http://anaiscbens.emnuvens.com.br/cbens/article/view/725/725</a>. Acesso em: 06 dez. 2018.

SILVA, Renato Maia. **Redes Neurais Artificiais aplicadas à Detecção de Intrusão em Redes TCP/IP**. 2005. 144 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Centro Técnico Científica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005. Cap. 4. Disponível em: <a href="https://www.maxwell.vrac.puc-rio.br/7335/7335\_5.PDF">https://www.maxwell.vrac.puc-rio.br/7335/7335\_5.PDF</a>. Acesso em: 06 out. 2018.

SINGH, Vikas Pratap et al. Forecasting of 5MW Solar Photovoltaic Power Plant Generation Using Generalized Neural Network. 2015. Disponível em: <a href="https://ieeexplore.ieee.org/document/7489107/">https://ieeexplore.ieee.org/document/7489107/</a>>. Acesso em: 19 set. 2018.

SONODA, Shimpei et al. **Application of Stepwise Multiple Regression to Design Optimization of Electric Machine.** 2007. Disponível em: <a href="https://ieeexplore.ieee.org/document/1632920">https://ieeexplore.ieee.org/document/1632920</a>>. Acesso em: 20 abr. 2019. SORKUN, Murat Cihan; PAOLI, Christophe; INCEL, Özlem Durmaz. **Time Series Forecasting on Solar Irradiation using Deep Learning**. 2017. Disponível em: <a href="https://ieeexplore.ieee.org/document/8266215/">https://ieeexplore.ieee.org/document/8266215/</a>>. Acesso em: 22 ago. 2018.

STACKHOUSE, Paul W. et al. **POWER Release 8 (with GIS Applications) Methodology** (**Data Parameters, Sources, & Validation**). 2018. Disponível em: <a href="https://power.larc.nasa.gov/documents/POWER\_Data\_v8\_methodology.pdf">https://power.larc.nasa.gov/documents/POWER\_Data\_v8\_methodology.pdf</a>). Acesso em: 25 set. 2018.

SWAMI, Ananthram; MENDEL, Jerry M. **ARMA Parameter Estimation Using Only Ouput Cumulants**. 1990. Disponível em: <a href="https://ieeexplore.ieee.org/document/57554/">https://ieeexplore.ieee.org/document/57554/</a>. Acesso em: 03 set. 2018.

TICKNOR, Jonathan L. A Bayesian regularized artificial neural network for stock market forecasting. 2013. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417413002509>. Acesso em: 25 abr. 2019.

WALPOLE, Ronald E. et al. **Probability & Statistics for Engineers & Scientists.** 9. ed. Boston: Pearson Prentice Hall, 2012. 812 p. Disponível em: <https://fac.ksu.edu.sa/sites/default/files/probability\_and\_statistics\_for\_engineers\_and\_scienti sst.pdf>. Acesso em: 24 maio 2019.

WAN, Can et al. **Photovoltaic and Solar Power Forecasting for Smart Grid Energy Management**. 2015. Disponível em: <a href="https://ieeexplore.ieee.org/document/7377167/">https://ieeexplore.ieee.org/document/7377167/</a>. Acesso em: 26 ago. 2018.

XIYUN, Yang; SONG, Chen. A combination method in photovoltaic power forecasting based on the correlation coefficient. 2014. Disponível em: <a href="https://ieeexplore.ieee.org/document/6933665/">https://ieeexplore.ieee.org/document/6933665/</a>>. Acesso em: 04 nov. 2018.

YANG, Zhengqiu; CAO, Yapei; XIU, Jiapeng. **Power Generation Forecasting Model for Photovoltaic Array Based on Generic Algorithm and Bp Neural Network**. 2014. Disponível em: <a href="https://ieeexplore.ieee.org/document/7175764/">https://ieeexplore.ieee.org/document/7175764/</a>. Acesso em: 19 ago. 2018.

ZHU, Ziming; ZHOU, Doudou; FAN, Zhong. Short term forecast of wind power generation based on SVM with pattern matching. 2016. Disponível em: <a href="https://ieeexplore.ieee.org/document/7514064/">https://ieeexplore.ieee.org/document/7514064/</a>>. Acesso em: 25 ago. 2018.

# APÊNDICE A - DADOS OBTIDOS PELO POWER DATA ACCESS VIEWER

Ao todo amostras de 1052 dias foram adquiridas em NASA (2018). A Tabela A.1 trás uma amostra destes dados, as 24 variáveis candidatas.

Data	22/04/16	25/04/16	26/04/16	27/04/16	28/04/16
Precipitação [mm]	15,51	75,22	20,73	1,41	0
Umidade Relativa [%]	91,98	94,74	95,28	78,49	72,41
Umidade Específica a 2 metros [kg]	0,012755	0,014693	0,008693	0,005758	0,005046
Pressão Superfície a 2 metros [kPa]	97,67	96,48	97,23	97,93	98,26
Faixa de Temperatura a 2 metros [C]	7,73	6,05	8,26	6,22	9,53
Temperatura de Solo [C]	18,94	20,31	12,37	9,4	8,32
Pontos de Geada de Orvalho a 2 metros [C]	17,35	19,35	11,17	5,43	3,64
Temperatura Máxima a 2 metros [C]	23,15	22,93	16,6	12,91	13,77
Temperatura Mínima a 2 metros [C]	15,43	16,88	8,34	6,7	4,24
Temperatura a 2 metros [C]	18,72	20,31	12,16	9,03	8,32
Faixa de Velocidade do Vento a 50 metros [m/s]	1,24	2,91	2,48	2,43	3,52
Faixa de Velocidade do Vento a 10 metros [m/s]	1,18	1,8	1,8	1,58	1,32
Velocidade do Vento Mínima a 50 metros [m/s]	3,05	3,5	3,08	2,6	0,81
Velocidade do Vento Mínima a 10 metros [m/s]	1,34	1,76	1,57	1,36	0,53
Velocidade do Vento Máxima a 10 metros [m/s]	4,29	6,42	5,56	5,03	4,33
Velocidade do Vento Máxima a 50 metros [m/s]	2,51	3,55	3,38	2,95	1,85
Direção do Vento a 50 metros [graus]	128,94	285,54	208,79	269,62	261,01
Direção do Vento a 10 metros [graus]	130,62	268,96	207,91	269,53	256,45
Velocidade do Vento a 50 metros [m/s]	3,64	4,27	4,01	3,88	2,51
Velocidade do Vento a 10 metros [m/s]	1,87	2,3	2,23	1,99	1,4
Irradiação Incidente Superfície Horizontal [kWh/m²/dia]	2,1	1,72	0,48	3,76	4,79
Fluxo Radiativo Infravermelho [kWh/m²/dia]	9,67	9,77	8,66	7,51	6,44
Irradiação de Topo de Atmosfera [kWh/m²/dia]	7,05	6,89	6,83	6,78	6,73
Indice de Irradiação a Céu Claro	0,3	0,25	0,07	0,55	0,71

Tabela A.10 - Amostra de Dados Obtidos pelo Power Data Access Viewer

Fonte: O autor (2019).

## **APÊNDICE B - DADOS OBTIDOS PELO PORTAL CLIMA**

Em Garibaldi (2019), amostras para 1052 dias foram obtidas. A Tabela B.1 trás uma amostra das variáveis presentes.

Data	22/04/2016	25/04/2016	26/04/2016	27/04/2016	28/04/2016	29/04/2016
Temperatura Média [C]	17,2413	18,5583	10,0433	7,1883	6,9717	7,5708
Temperatura Máxima [C]	17,4329	18,9888	10,3692	7,5963	7,4450	8,1138
Temperatura Mínima [C]	17,0679	18,2675	9,6550	6,9188	6,5204	7,0767
Sensação térmica [C]	16,6708	18,6383	8,2533	4,5104	5,1083	7,0554
Umidade Relativa [%]	100,0000	99,6150	100,0000	90,1508	80,2875	81,8013
Umidade Absoluta [g/m³]	16,7338	18,0529	10,9296	8,0754	7,0425	7,5267
Pressão Atmosferica Relativa [Pa]	936,8508	926,1392	931,7533	938,0442	941,2867	940,6992
Velocidade Vento [m/s]	4,8883	3,7175	3,6842	4,1438	2,7400	1,6063
Precipitação [mm]	83,3650	166,7458	196,1233	207,3817	207,4000	207,4000

Tabela B.11 - Amostra dos Dados Obtidos pelo Portal Clima

Fonte: O autor (2019).

# APÊNDICE C - DADOS HISTÓRICOS DE GERAÇÃO

A Tabela C.1 é uma amostra dos dados de geração em que o processo de treinamento se baseou.

Data	Geração Fotovoltaica (kWh)
22/04/16	4,6
25/04/16	6,2
26/04/16	2,3
27/04/16	12,3
28/04/16	18,5
29/04/16	14,8
30/04/16	5
10/05/16	6,8
11/05/16	6,5
12/05/16	3,9
13/05/16	6,5
14/05/16	14,4
15/05/16	1,2
16/05/16	5
17/05/16	14,8
25/05/16	12
26/05/16	8,7
27/05/16	5,3
28/05/16	11,8
29/05/16	3,4
30/05/16	7,1
31/05/16	10,4
01/06/16	10,4
02/06/16	15,3
03/06/16	6,2
04/06/16	13,3
05/06/16	13,5
06/06/16	8,1
07/06/16	14,4

Tabela C.12 - Amostra de Dados Históricos de Geração

Fonte: O autor (2019).