

UNIVERSIDADE DE PASSO FUNDO
Programa de Pós-Graduação em
Computação Aplicada

Dissertação de Mestrado

**UMA ABORDAGEM PARA
IDENTIFICAR VIABILIDADE DE UM
LOCAL PARA IMPLANTAÇÃO DE
ULTRAFILTRAÇÃO DE ÁGUA DE
CHUVA COM A UTILIZAÇÃO DE
DEEP LEARNING**

ELIAS MORAIS PEREIRA



UNIVERSIDADE DE PASSO FUNDO
INSTITUTO DE CIÊNCIAS EXATAS E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

**Uma abordagem para Identificar Viabilidade
de um Local para Implantação de
Ultrafiltração de Água de Chuva com a
utilização de Deep Learning**

Elias Morais Pereira

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Computação Aplicada na Universidade de Passo Fundo.

Orientador: Prof. Roberto dos Santos Rabello

Coorientador: Prof. Vandrê Barbosa Brião

Passo Fundo

2021

CIP – Catalogação na Publicação

P436a Pereira, Elias Moraes

Uma abordagem para identificar viabilidade de um local para implantação de ultrafiltração de água de chuva com a utilização de Deep Learning [recurso eletrônico] / Elias Moraes Pereira. – 2021
4.7 MB ; PDF.

Orientador: Prof. Dr. Roberto dos Santos Rabello.

Coorientador: Prof. Dr. Vandr e Barbosa Bri o.

Disserta o (Mestrado em Computa o Aplicada) –
Universidade de Passo Fundo, 2021.

1. Redes neurais (Computa o). 2. Precipita o (Meteorologia). 3.  gua - Filtra o. 4. Aprendizagem profunda. 5. Chuvas. I. Rabello, Roberto dos Santos, orientador. II. Bri o, Vandr e Barbosa, coorientador. III. T tulo.

CDU: 004.8

Cataloga o: Bibliotec ria Jucelei Rodrigues Domingues - CRB 10/1569

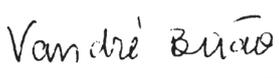
ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO DO ACADÊMICO

ELIAS MORAIS PEREIRA

Aos vinte e seis dias do mês de março do ano de dois mil e vinte e um, às 16 horas, realizou-se, de forma on-line, por meio de videoconferência, a sessão pública de defesa do Trabalho de Conclusão de Curso “Uma abordagem para Identificar Viabilidade de um Local para Implantação de Ultrafiltração de Água de Chuva com a utilização de Deep Learning”, de autoria de Elias Morais Pereira, acadêmico do Curso de Mestrado em Computação Aplicada do Programa de Pós-Graduação em Computação Aplicada – PPGCA. Segundo as informações prestadas pelo Conselho de Pós-Graduação e constantes nos arquivos da Secretaria do PPGCA, o aluno preencheu os requisitos necessários para submeter seu trabalho à avaliação. A banca examinadora foi composta pelos doutores Roberto dos Santos Rabello (Orientador), Dr. Vandré Barbosa Brião (Coorientador), Carlos Amaral Hölbig (UPF) e Márcia de Fátima Brondani Binelo (UNIJUÍ). Concluídos os trabalhos de apresentação e arguição, a banca examinadora considerou o candidato **APROVADO**. Foi concedido o prazo de até quarenta e cinco (45) dias, conforme Regimento do PPGCA, para o acadêmico apresentar ao Conselho de Pós-Graduação o trabalho em sua redação definitiva, a fim de que sejam feitos os encaminhamentos necessários à emissão do Diploma de Mestre em Computação Aplicada. Para constar, foi lavrada a presente ata, que vai assinada pelos membros da banca examinadora e pela Coordenação do PPGCA.



Prof. Dr. Roberto dos Santos Rabello – UPF
Presidente da Banca Examinadora
(Orientador)



Prof. Dr. Vandré Barbosa Brião - UPF
(Coorientador)



Prof. Dr. Carlos Amaral Hölbig – UPF
(Avaliador Interno)

p.p.

Prof. Dra. Márcia de Fátima Brondani Binelo – UNIJUÍ
(Avaliadora Externa)



Prof. Dr. Carlos Amaral Hölbig
Coordenador do PPGCA

UMA ABORDAGEM PARA IDENTIFICAR VIABILIDADE DE UM LOCAL PARA IMPLANTAÇÃO DE ULTRAFILTRAÇÃO DE ÁGUA DE CHUVA COM A UTILIZAÇÃO DE DEEP LEARNING

RESUMO

As redes neurais, que são inspiradas pelo conceito de neurônios biológicos, são normalmente utilizadas em muitas aplicações, incluindo no campo da previsão do tempo. A irradiação solar em um local específico pode ajudar a prever a quantidade de eletricidade que será gerada através de painéis solares e uma previsão precisa pode ajudar a calcular a dimensão do sistema. Na mesma linha, as medidas hidrometeorológicas são uma das tarefas mais desafiantes da natureza, na qual a precipitação tornou-se o fator mais significativo e técnico, onde abordagens de redes neurais fornecem resultados promissores para ajudar no processo de tomada de decisão para a previsão de precipitação. Neste contexto, o presente trabalho estabelece uma abordagem para previsão de precipitação e irradiação solar utilizando modelos de deep learning e na escolha de um local adequado para instalação de um sistema de ultrafiltração de água de chuva utilizando energia fotovoltaica. Esses modelos foram empregados com a finalidade de prever precipitação e irradiação solar para os próximos seis meses, de acordo com o último mês de dados históricos coletado. Modelos lineares e de machine learning foram testados e confrontados com os modelos de deep learning a fim de traçar uma linha de testes e determinar o RMSE de cada modelo. O modelo skillfull escolhido e empregado neste trabalho foi o long short-term memory (LSTM). Esse modelo teve um RMSE de 42,53 para precipitação e 0,45 para irradiação solar. Sendo superior a outros modelos de deep learning, como a MLP e ConvNet. Cálculos de normalização e combinação média, utilizando três pesos (0,7; 1,0 e 1,3) foram utilizados para corroborar com a viabilidade do modelo. Uma aplicação web foi desenvolvida para a apresentação dos resultados. Segundo o experimento realizado, a abordagem apresentou-se adequada, e pode servir como tomada de decisão em cálculos de dimensão de um sistema de energia fotovoltaica, quantidade de água de chuva predita, bem como para o sistema de ultrafiltração de água de chuva.

Palavras-Chave: Precipitação, Irradiação Solar, Aprendizado Profundo, Redes Neurais.

AN APPROACH TO IDENTIFY LOCATION FEASIBILITY FOR RAINWATER ULTRAFILTRATION WITH THE USE OF DEEP LEARNING

ABSTRACT

Neural networks, which are inspired by the concept of biological neurons, are commonly used in many applications, including in the field of weather forecasting. The solar irradiation at a specific location can help predict the amount of electricity that will be generated through solar panels, and an accurate prediction can help calculate the size of the system. In the same line, hydrometeorological measurements are one of the most challenging tasks in nature, in which precipitation has become the most significant and technical factor, where neural network approaches provide promising results to help in the decision-making process for precipitation forecasting. In this context, the present work establishes an approach for predicting precipitation and solar irradiance using deep learning models and in choosing a suitable site for installing a rainwater ultrafiltration system using photovoltaics. These models were applied to predict precipitation and solar irradiation for the next six months, according to the last month of historical data collected. Linear and machine learning models were tested and compared with the deep learning models in order to draw a line of tests and determine the RMSE of each model. The skillfull model chosen and employed in this work was the long short-term memory (LSTM). This model had an RMSE of 42.53 for precipitation and 0.45 for solar irradiation. It was superior to other deep learning models, such as MLP and ConvNet. Normalization and average combining calculations using three weights (0.7, 1.0, and 1.3) were used to corroborate the model's viability. A web application was developed to present the results. According to the experiment performed, the approach proved to be adequate, and can serve as a decision making tool in calculations of the size of a photovoltaic system, predicted amount of rainwater, as well as for a rainwater ultrafiltration system.

Keywords: Precipitation, Solar Irradiation, Deep Learning, Neural Networks.

LISTA DE FIGURAS

1	Porcentagem de Abastecimento de Água por rede nas Macrorregiões [2] .	22
2	Sistema de abastecimento de água com captação superficial e subterrâneo [11].	22
3	Diagrama do equipamento de filtração tangencial [19].	26
4	Exemplo de um sistema de coleta de energia fotovoltaica ligada a rede [23].	27
5	Exemplo de um sistema de coleta de energia fotovoltaica off-grid [24]. . . .	28
6	Matriz energética brasileira - Ano base 2019. Adaptado de [30]	30
7	Diagrama de Venn. Adaptado de [31].	31
8	Os ingredientes que contribuem para a ciência de dados [35].	32
9	Modelo básico de uma Rede Neural Artificial [4].	36
10	Partes constituintes da Deep Learning [40].	37
11	Diagrama multilayer perceptrons [48].	39
12	Diagrama long short-term memory [49].	39
13	Esquema do processo de revisão sistemática [Autoria própria].	45
14	Exemplos dos quatro diferentes modelos de algoritmos utilizando aprendizado de máquina [5]	49
15	Fluxograma dos modelos testados [6].	50
16	Diagrama conceitual do modelo de aprendizado de máquina baseado em DMLP para estimativa de precipitação de radar [54].	52
17	Diagrama conceitual do modelo de aprendizado de máquina baseado em MLP para estimativa de precipitação TRMM PR [8].	53
18	Arquitetura proposta [47].	54
19	Fluxograma da metodologia aplicada [61].	55
20	Fluxo metodológico [Autoria própria]	60
21	Estações meteorológicas automáticas do Estado do Rio Grande do Sul [Autoria própria].	61
22	Dashboard do frontend [Autoria própria].	63
23	Estrutura do frontend [Autoria própria].	63
24	Diagrama de uma RNN LSTM [Autoria própria].	66
25	Testes de hiperparâmetros em relação à irradiação solar [Autoria própria].	67
26	Médias de irradiação global horizontal e normal climatológica de precipitação anual. Adaptada de [29]	70

LISTA DE TABELAS

1	Padrões de qualidade de água da chuva estabelecidas pela NBR 15527/07 [12].	23
2	Top 10 países classificados com potencial solar [21].	27
3	Totais diários de irradiação global horizontal para cada região brasileira 2017 [29].	31
4	Resumo das técnicas de ciência de dados utilizadas para a precipitação e irradiação solar.	46
4	Resumo das técnicas de ciência de dados utilizadas para a precipitação e irradiação solar.	47
5	Teste de performance dos métodos [4]	48
6	Resultado dos cálculos para as cidades escolhidas	70
7	Valores máximos de precipitação e irradiação solar dos locais adequados no período verificado x média combinada.	74
8	Resultados do Grid Search para a precipitação	75
9	Resultados do Grid Search para a irradiação	75
10	Resultados de RMSE para os modelos lineares	76
11	Resultados de RMSE para o modelo de ML Random Forest	76
12	Resultados de RMSE para os modelo de DL	77
13	Resultados da sumarização das estatísticas descritivas com os melhores e piores locais	78

LISTA DE SIGLAS

ABNT – Associação Brasileira de Normas Técnicas
ACM – Association for Computing Machinery
ANA – Agência Nacional de Águas
ANFIS – Adaptive Neuro-Fuzzy Inference System
ANN – Artificial Neural Network
API – Application Programming Interface
ARIMA – Auto Regressive Integrated Moving Average
ARMA – Autoregressive Moving Average
ASIC/TPU – Tensor Processing Units
AVNNET – Averaged Neural Network
CNN – Convolutional Neural Network
CPU – Central Process Unit
CSS – Cascading Style Sheets
CSV – Comma-separated Values
CONVNET – Convolutional Network
DMLP – Deep Multiplayer Perceptron
DNA – Deoxyribonucleic Acid
DP – Deep Learning
DRNN – Deep Recurrent Neural Network
DT – Decision Tree
EEAB – Estação Elevatória de Água Bruta
EEAT – Estação Elevatória de Água Tratada
FNN – Feedforward neural network
GPU – Graphics Processing Unit
HTML – HyperText Markup Language
IA – Inteligencia Artificial
IDE – Integrated Development Environment
IEEE – Institute of Electrical and Electronic Engineers
INMET – Instituto Nacional de Meteorologia
JSON – JavaScript Object Notation
KWH – kilowatts Hora

LSTM – Long Short-Term Memory
MAPE – Mean Absolute Percentage Error
MATLAB – MATrix LABoratory
MBE – Mean Bias Error
MLP – Multilayer Perceptron
ML – Machine Learning
MS – Ministerio da Saúde
MW – Megawatts
NBR – Norma Brasileira
NIR – Near Infrared Radiation
PCA – Análise de Componentes Principais
PPGCA – Programa de Pós-Graduação em Computação Aplicada
PPGCTA – Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos
PR – Precipitation Radar
PVC – Policloreto de Vinilo
RF – Random Forests
RMSE – Root Mean Square Error
RNA – Ribonucleic Acid - Ácido ribonucleico
RNC – Redes Neurais Convolucionais
RNR – Redes Neurais Recorrentes
RS – Revisão Sistemática
SARIMA – Seasonal Auto Regressive Integrated Moving Average
SNIS – Sistema Nacional de Informações de Saneamento
SVD – Decomposição de Valor Singular
SVM – Support Vector Machines
SVM – Support Vector Machines
TRMM – Tropical Rainfall Measuring Mission
UF – Ultra Filtração
UPF – Universidade de Passo Fundo

SUMÁRIO

1	INTRODUÇÃO	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	ABASTECIMENTO DE ÁGUA	21
2.1.1	Sistemas de Abastecimento de Água	22
2.1.2	Água de Chuva	22
2.1.3	Qualidade da Água de Chuva e Seus Usos	23
2.1.4	Água de Chuva para Abastecimento Potável	25
2.2	SISTEMAS DE ENERGIA SOLAR	26
2.2.1	Energia Solar Fotovoltaica	27
2.3	PARÂMETROS METEOROLÓGICOS	28
2.3.1	Precipitação	29
2.3.2	Irradiação Solar	29
2.4	TÉCNICAS APLICADAS PARA CIÊNCIA DE DADOS	31
2.4.1	Aprendizado de Máquina	33
2.4.1.1	Aprendizado supervisionado	33
2.4.1.2	Aprendizado não-supervisionado	34
2.4.1.3	O processo de aprendizagem de máquina	35
2.4.2	Redes Neurais	36
2.4.2.1	Aprendizado Profundo ou Deep Learning	36
2.4.3	Softwares Utilizados	40
2.4.3.1	TensorFlow	40
2.4.3.2	Keras	41
3	TRABALHOS RELACIONADOS	43
3.1	INFORMAÇÕES GERAIS	43
3.2	IDENTIFICAÇÃO DOS ESTUDOS	43
3.3	CRITÉRIOS DE ELEGIBILIDADE E ESCOLHA DOS TRABALHOS	44
3.4	PROCEDIMENTOS DE SELEÇÃO DOS TRABALHOS	44
3.5	EXTRAÇÃO DOS DADOS	45
3.6	PREVISÃO DE IRRADIÂNCIA SOLAR USANDO REDES NEURAS PRO- FUNDAS	48

3.7	POTENCIAL DE QUATRO DIFERENTES ALGORITMOS DE APRENDIZADO DE MÁQUINA NA MODELAGEM DE RADIAÇÃO SOLAR GLOBAL DIÁRIA	49
3.8	PREVISÃO DE IRRADIAÇÃO SOLAR COM APRENDIZADO DE MÁQUINA: MÉTODO DE SELEÇÃO DE MODELOS DE PREVISÃO DEPENDENDO DA VARIABILIDADE CLIMÁTICA	50
3.9	PREVISÃO DE ENERGIA SOLAR MÉDIA DIÁRIA USANDO TÉCNICAS DE REGRESSÃO DE APRENDIZADO DE MÁQUINA	51
3.10	UM MODELO DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DE PRECIPITAÇÃO DE RADAR BASEADO EM OBSERVAÇÕES DE MEDIDOR	51
3.11	UM SISTEMA DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DE PRECIPITAÇÃO A PARTIR DE RADARES ESPACIAIS E TERRESTRES	52
3.12	MODELOS DE APRENDIZAGEM PROFUNDA PARA A PREVISÃO DE PRECIPITAÇÃO	53
3.13	COMPARAÇÃO DE QUATRO ALGORITMOS DE APRENDIZAGEM DE MÁQUINAS PARA A SUA APLICABILIDADE EM RECUPERAÇÕES DE PRECIPITAÇÃO ÓPTICA BASEADOS EM SATÉLITE	54
3.14	DISCUSSÕES	55
4	MATERIAIS E MÉTODOS	59
4.1	ESTUDO E ESCOLHA DAS TÉCNICAS	60
4.2	COLETA E ANÁLISE DE DADOS	61
4.2.1	Limpeza dos dados	61
4.3	DESENVOLVIMENTO DA APLICAÇÃO	62
4.3.1	Frontend	62
4.3.2	Backend	64
4.4	TESTES E VALIDAÇÃO	66
4.4.1	Walk-Forward	67
4.4.2	Grid Search	68
4.4.3	Resample e Cálculo das Variáveis Meteorológicas	69
4.4.4	Testes e Validação dos Hiperparâmetros	69
4.4.5	Eficiência Energética de Painéis Fotovoltaicos X Consumo de Água por Habitante	69
4.4.5.1	Normalização e Combinação Média	70
5	ANÁLISE E DISCUSSÃO DOS RESULTADOS	73
5.1	APRESENTAÇÃO DOS RESULTADOS	75

5.2	CONCLUSÃO E TRABALHOS FUTUROS	79
	REFERÊNCIAS	81

1. INTRODUÇÃO

O consumo de água potável é uma das necessidades básicas do ser humano e vem crescendo em taxas superiores às suportadas pelo planeta. Cerca de 75% da terra está coberta de água, no entanto, apesar do fato de que três quartos da terra serem compostos por água, menos de 3% é água fresca [1]. Além disso, da água doce existente, nem toda ela está disponível para o consumo humano. No Brasil, segundo o Sistema Nacional de Informações de Saneamento (SNIS 2017) [2], números mostram que as dificuldades no sistema de distribuição de água potável são enormes, principalmente nas regiões mais pobres e isoladas do país. Apenas 83,5% dos brasileiros são atendidos com abastecimento de água tratada e quase 35 milhões de brasileiros não detêm desse serviço básico.

Ao encontro com a questão dos problemas de água potável no Brasil, está o uso de energias renováveis - sistema de energia fotovoltaica - o qual tem atraído a atenção mundial nas últimas duas décadas devido às propriedades intrinsecamente abundantes, globalmente sustentáveis, livres e limpas da irradiação solar [3].

Na previsão da precipitação e irradiação solar, nota-se o massivo emprego de redes neurais aliadas ao poder computacional, no qual apontam para uma quebra de paradigma onde a análise dos dados como um todo, não mais se dá com padrões e técnicas manuais, mas sim com o auxílio das tecnologias e inovações que áreas da tecnologia disponibilizam [4] [5].

Modelos de aprendizado profundo permitem analisar pequenos e grandes volumes de dados, independente da complexidade, de forma rápida e precisa. Sua utilização já está difundida em diversos segmentos como detecção de fraudes, análise de crédito, modelos de previsão de falhas, detecção de anomalias e previsão de series temporais. Porém, diante da variedade de alternativas, é essencial que se conheça as características individuais de cada método e qual o melhor cenário para sua utilização [6] [7] [8].

Neste contexto, é apresentado o seguinte problema de pesquisa: É possível o uso de modelos de deep learning para identificar um local adequado para implantação de um sistema de ultrafiltração de água de chuva, através de previsão de precipitação e irradiação solar?. O principal benefício desta solução é a abordagem como um todo.

A partir de técnicas computacionais com base nesse cenário, o presente trabalho apresenta uma metodologia para a previsão de precipitação de irradiação solar, utilizando modelos de deep learning, bem como a visualização dos locais identificados em uma aplicação web. O escopo geográfico escolhido é o estado do Rio Grande do Sul. Fatores como, quantidade de dados históricos, poder computacional e tempo de treinamento do modelo foi determinante para a não escolha de todo o território Brasileiro.

A estrutura do trabalho está organizada com os seguintes capítulos:

- Capítulo 2: Descreve a fundamentação teórica com conceitos sobre água, qualidade e seus usos. Utilização de energia solar, sistemas de energia fotovoltaica. Conceitos sobre parâmetros meteorológicos que explicam a precipitação e irradiação/insolação solar;
- Capítulo 3: Apresenta os trabalhos relacionados de acordo com o protocolo de revisão sistemática. A utilização de técnicas de inteligência artificial aplicadas na previsão de precipitação e irradiação solar resultam em 25 trabalhos selecionados na revisão sistemática que são brevemente apresentados e classificados em dois grupos: precipitação e irradiação solar;
- Capítulo 4: Descreve a metodologia utilizada na previsão de precipitação e irradiação solar, bem como o uso de modelos lineares, machine learning e de deep learning. Posteriormente, são apresentadas as principais características presentes no sistema desenvolvido e os componentes de software empregados na sua concepção;
- Capítulo 5: Detalha e discute os resultados encontrados a partir da concepção dos modelos de deep learning. Nesse capítulo é descrito a forma de confronto entre a precipitação e irradiação para a validação da proposta e também as conclusões finais obtidas.

2. FUNDAMENTAÇÃO TEÓRICA

No contexto das cidades, problemas relacionados à falta de água potável, estão sendo considerada como o grande problema do próximo milênio. As justificativas são muitas, entre elas podemos citar que, do total de água do mundo apenas 2,5% é água doce e só 0,03% do total se encontra em superfícies acessíveis [1]. O consumo de água potável é uma das necessidades básicas do ser humano, crescendo em taxas superiores às suportadas pelo planeta a médio prazo contribui para encontrar as melhores condições para viabilizar a purificação da água da chuva.

Projetos com energia solar têm crescido ao longo dos anos e estão permitindo que a luz solar possa ser utilizada, ajudando assim em uma economia para as pessoas, bem como para o Brasil. Quando nossa capacidade energética é excedida, usinas de carvão são utilizadas para suprir a demanda populacional, pois do contrário, poderíamos entrar em colapso em nossa rede elétrica.

De encontro a estas dificuldades, novas ideias e tecnologias surgem, capazes de viabilizar projetos e inovações para que a água de chuva possa ser disponibilizada ao consumo humano, bem como o uso da energia solar através de sistemas como Energia Solar Fotovoltaica e Fototérmica, gerando economia e sustentabilidade ao nosso país.

2.1 ABASTECIMENTO DE ÁGUA

Apesar da alta disponibilidade hídrica do país, precisamos ressaltar que muitas regiões/cidades ainda possuem índices precários no quesito de atendimento de água.

De acordo com o Sistema Nacional de Informações sobre Saneamento (SNIS) 2017 [2], no Brasil, 93,0% da população urbana é atendida por rede de abastecimento de água, representando um contingente de cerca 160 milhões de pessoas atendidas por esse serviço. Dentre as macrorregiões do País, a com maior índice de atendimento é a Sul (98,4%), seguida pela Centro-Oeste (98,1%), Sudeste (95,9%) e Nordeste (88,8%). Já o índice no Norte (70,0%) encontra-se bem distante dos observados nas demais macrorregiões, como nos mostra a figura 1, o que indica a necessidade do aprimoramento em políticas de saneamento.

Cabe salientar que os níveis de atendimento com água referem-se ao atendimento feito exclusivamente por redes públicas, logo não inclui soluções individuais como, por exemplo, poços ou nascentes.

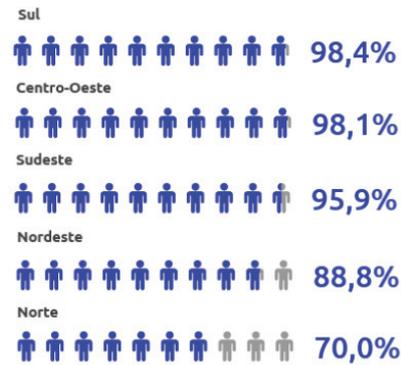


Figura 1: Porcentagem de Abastecimento de Água por rede nas Macrorregiões [2]

Ainda em SNIS 2017 [2], esses números mostram que as dificuldades no sistema de distribuição são enormes, principalmente nas regiões mais pobres e isoladas do país. Na média do país, o índice de atendimento total é de 83,5%. Nesta conta, 16,5%, ou seja, quase 35 milhões de brasileiros não detém desse serviço básico.

2.1.1 Sistemas de Abastecimento de Água

O sistema de abastecimento de água compreende algumas etapas antes que a água esteja própria para o consumo humano. Um sistema de abastecimento de água, resumidamente, engloba a captação, adução, recalque, tratamento, reservatório e rede de distribuição [9] [10] [11]. A figura 2 nos dá uma ideia deste sistema.

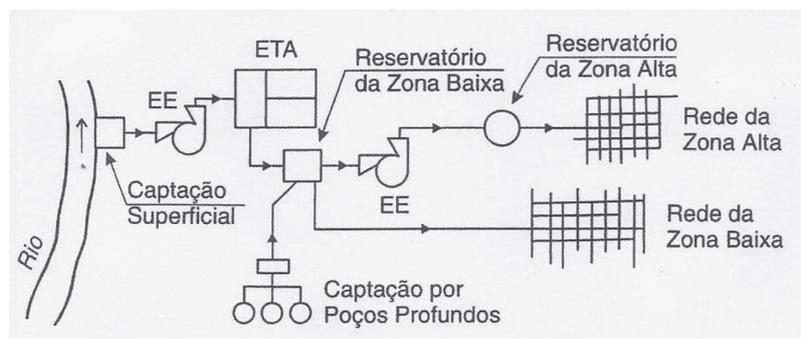


Figura 2: Sistema de abastecimento de água com captação superficial e subterrâneo [11].

2.1.2 Água de Chuva

A água de chuva é um bem valioso, mas que está sendo mal aproveitada. No Brasil, o uso de água de chuva ocorre há séculos, mas a captação imediata no meio urbano, visando ao aproveitamento em usos diversos, ainda é pequena, e no meio rural somente na última década é que a construção de cisternas vem se disseminando para abastecimento difuso de água.

De acordo com o que cita a Associação Brasileira de Normas Técnicas na NBR 15527 [12], define-se água de chuva como a “água resultante de precipitações atmosféricas coletada em coberturas, telhados, onde não haja circulação de pessoas, veículos ou animais”. Concomitantemente, tal água é considerada como não potável devido às suas características e, conseqüentemente, não atendendo a Portaria nº 518 do Ministério da Saúde [13] sobre o controle de qualidade de água para consumo humano. A tabela 1 apresenta as características da qualidade necessárias mínimas para a água de chuva.

Tabela 1: Padrões de qualidade de água da chuva estabelecidas pela NBR 15527/07 [12].

Parâmetro	análise	Valor
Coliformes Totais (NMP/100mL)	Semestral	Ausência em 100mL
Coliformes Termotolerantes (NMP/100mL)	Semestral	Ausência em 100mL
Cloro Residual Livre (mg/L)	Mensal	0,5 a 3,0
Turbidez (UNT)	Mensal	5,0 (usos menos restritivos)
Cor Aparente (uH)	Mensal	15
pH	Mensal	6,0 a 8,0

A água da chuva pode ser captada de diversos modos, sendo um dos mais utilizados a coleta por calha (telhado). Em consonância à este meio de captação encontra-se pátios, rochas e áreas impermeáveis, tomando-se cuidado operacional de que a água passe por um filtro com o objetivo de reter as impurezas sólidas como folhas e pedras, o qual sua manutenção deve ser periódica a fim de não obstruir o fluxo de água, ocasionando problemas [14].

Para Andrade [15], existe amplo conhecimento tecnológico internacional e nacional que deve ser utilizado de forma mais competente para o aproveitamento imediato da água de chuva, sobretudo para maximizar a relação custo/benefício, aumentar a segurança sanitária dos sistemas e assegurar a qualidade da água.

2.1.3 Qualidade da Água de Chuva e Seus Usos

A qualidade da água de chuva depende muito do local onde é coletada. O adequado funcionamento deste sistema de aproveitamento de água depende da quantidade da mesma que poderá ser captada e utilizada para o atendimento das respectivas demandas, os quais dependem das características pluviométricas da região, da área impermeável de captação e do volume do reservatório de armazenamento.

Embora não haja uma legislação federal específica no Brasil para estabelecer os padrões de qualidade de água pluvial captada e utilizada para fins não potáveis, é comum utilizar como base os preceitos da Portaria Nº 2.914/2011 do Ministério da Saúde (MS).

A demanda crescente por água tem feito o reuso planejado ser um tema atual e de grande importância, tal reuso, muitas vezes aplicado em processos industriais, preconiza uma melhor manutenção e utilização desta água a um nível de tratamento recomendado, respeitando e adotando diversos critérios de segurança, custos de capital, de operação e de manutenção associados.

O estudo de Andrade [15], mostra as vantagens do aproveitamento imediato da água de chuva para diversos usos, apresenta e comenta avanços, e avalia e discute a segurança da água e a proteção sanitária do sistema de captação e armazenamento. Destaca a atenção que deve ser dispensada aos primeiros milímetros da cada chuva, pois estes têm grande influência sobre a qualidade da água captada em cisternas; já que esta quantidade é, geralmente, suficiente para lavar a atmosfera e a superfície de captação.

O aproveitamento da água da chuva e a utilização da água gerada pelos aparelhos de ar condicionado como fontes alternativas é um estudo proposto por Eduarda [16], no qual apresenta aspectos qualitativos e quantitativos favoráveis ao desenvolvimento de um projeto de sistema de aproveitamento de água no Setor de Aulas IV da Universidade Federal do Rio Grande do Norte, já que a área de captação para água da chuva na edificação em questão é de 4.582,71 m² e a produção mensal gerada pelos aparelhos de ar condicionado, cujo funcionamento atinge 15h/dia, oscila em torno de 57.703,7 L/mês.

Os estudos abordados nesta seção nos mostram que é possível o uso e reuso de água da chuva para vários fins. A potabilidade da água da chuva depende de alguns parâmetros que são definidos na Portaria 518 do Ministério da Saúde, de 25 de março de 2004, e devem obedecer a critérios mínimos quanto aos padrões:

- (a) microbiológico de potabilidade da água para consumo humano (ausência de *Escherichia coli* e coliformes);
- (b) turbidez para água pós-filtração ou pré-desinfecção, ou seja, medição da resistência da água à passagem da luz por conta de material fino em suspensão na água;
- (c) de potabilidade para substâncias químicas que representam risco à saúde, sendo eles inorgânicos (Fluoreto 1,5 mg/L, Cobre 2,0 mg/L, Arsênio 0,01 mg/L e Chumbo 0,01 mg/L), orgânicos (Acrilamida 0,5 ug/L, Benzeno 5 ug/L, Cloreto de vinila 5 ug/L e Tetracloroeto de carbonato 2 ug/L) ou desinfetantes (Cloro livre: 5 mg/L).

Em geral já existe um consenso sobre os requisitos de qualidade que a água da chuva deve ter para cada tipo de uso. O conhecimento desses limites ajuda a definir para que fins a água da chuva pode ser utilizada, em função da sua qualidade e sem causar prejuízos aos usuários e, quando necessário, o tipo de tratamento que deve ser realizado, para que a água atinja qualidade compatível com o uso que se pretende fazer dela.

2.1.4 Água de Chuva para Abastecimento Potável

Uma questão que tem sido abordada nos sistemas de aproveitamento de águas pluviais é a qualidade da água obtida, principalmente se for para atender ao consumo humano. De um modo geral, a água pluvial, devido ao processo de destilação natural que sofre, pode ser considerada de boa qualidade, mas não apta ao consumo, pois dependendo da região, a chuva pode apresentar poluentes, principalmente em regiões próximas aos grandes centros urbanos ou a áreas bastante industrializadas [17].

Visando métodos alternativos de abastecimento de água, a potabilização de água de chuva se mostra satisfatória e vem crescendo nos últimos anos. Não distante, será o principal meio de obtenção de água e para que possamos utilizá-la para o consumo humano, será necessário métodos de filtração eficazes e que ao mesmo tempo sejam de baixo custo, para que localidades onde a rede normal de distribuição de água não atenda, possam ter um meio seguro de obtenção de água potável [18].

O estudo de Miorando *et al.* [19], utilizou a ultrafiltração (UF) para o tratamento e caracterização da água de chuva armazenada a partir da coleta em telhado. A água foi submetida ao processo de UF em várias condições experimentais, ajustando seus parâmetros físico-químicos e microbiológicos para a potabilidade. Os testes foram realizados em duas membranas de UF com duas pressões diferentes para tratar água de chuva coletada em um prédio da Universidade de Passo Fundo (UPF). Duas etapas foram escolhidas para a caracterização do estudo:

- (a) ultrafiltração tangencial, em duas pressões de operação e dois tipos de membrana (membranas de 4 e 50 kDa). Nessa etapa foram utilizados equipamentos de bancada tangencial com vazões de 200 L.h^{-1} , conforme figura 3. As duas membranas utilizadas nesta etapa passaram por uma limpeza química com uma solução alcalina (pH 10,5) de hidróxido de sódio, e sanitização utilizando-se uma solução de hipoclorito de sódio. O equipamento da bancada é constituído de um reservatório de aço inoxidável de 10L de capacidade e uma bomba pneumática para alimentação da água na carcaça da membrana;
- (b) experimentos de longa duração com a membrana de melhor desempenho na etapa anterior para avaliação da estabilidade da qualidade da água e do fluxo permeado. Nesta etapa foi utilizado um equipamento de bancada que trabalha com membranas submersas e sob vácuo utilizando a membrana que teve o melhor desempenho na etapa anterior. O peso molecular de corte da membrana utilizada nessa etapa possui 50.000 Da (de polietersulfona em configuração de fibra oca e área filtrante de $0,22 \text{ m}^2$). Uma bomba a vácuo auxilio no processo de filtração da parte externa para o interior da membrana.

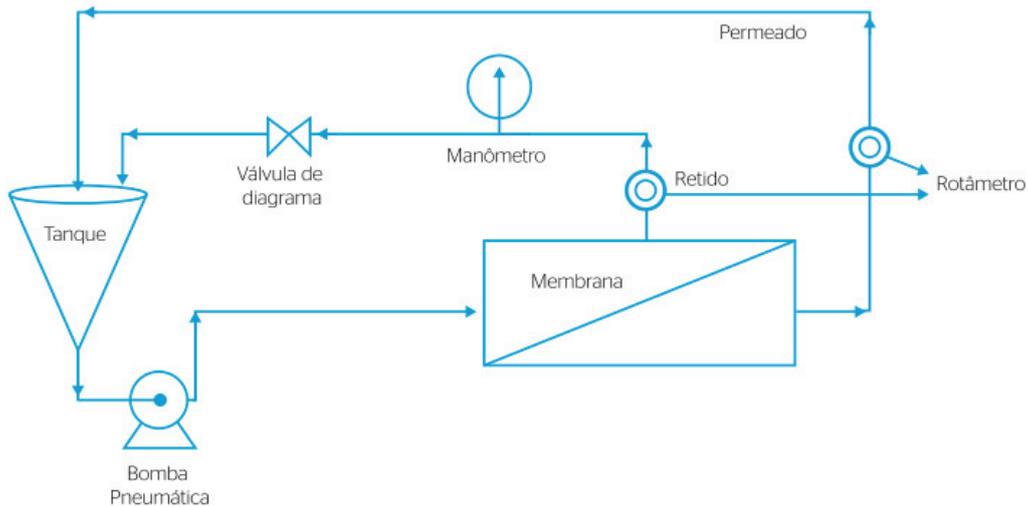


Figura 3: Diagrama do equipamento de filtração tangencial [19].

Os resultados mostram o grande potencial da UF na filtração de água de chuva para o consumo humano. Com o avanço da captação de água de chuva pra muitos fins, munidos de tecnologias e estudos na filtração para o abastecimento e consumo humano, se torna um necessidade para que tenhamos uma sociedade mais sustentável. "A água de chuva representa uma fonte de fácil captação e de razoável qualidade, e que apenas o processo de UF seria suficiente para torná-la potável".

2.2 SISTEMAS DE ENERGIA SOLAR

Entre essas opções bem conhecidas de energia renovável, a conversão de energia da iluminação solar tem atraído a atenção mundial nas últimas duas décadas devido às propriedades intrinsecamente abundantes, globalmente sustentáveis, livres e limpas da irradiação solar [3].

A energia solar é uma das fontes de energia renováveis que mais cresce. No entanto, a flutuação, a intermitência e a aleatoriedade da energia solar são as principais barreiras para integrá-la à rede pública, na qual a previsão precisa da produção de energia solar é crítica [20].

A utilização da energia solar traz vários benefícios a longo prazo, proporcionando desenvolvimento de regiões remotas, onde o custo pelo uso da rede convencional é relativamente alto em relação ao retorno financeiro do investimento.

O estudo de Kommeh *et al.* [21], tabela 2, fez um ranking com maior recursos solares por classe e país com base em dados estatísticos para mostrar o potencial da energia solar. Com o declínio no custo, a energia solar tornou-se a opção suplementar entre os formas renováveis de energia. Isso é essencial, pois mostra quais países em todo o mundo têm recursos solares em grandes quantidades e cujos benefícios podem ser explorados através de financiamento e investimentos.

Tabela 2: Top 10 países classificados com potencial solar [21].

País	MWh/ano	Rank
Russia	30586340907	1
Antarctica	29799042216	2
China	27373606560	3
Australia	25097791333	4
Brazil	24993114081	5
United States	24557081452	6
Canada	21214183621	7
India	9877095200	8
Sudan	8702766347	9
Algeria	8162220322	10

2.2.1 Energia Solar Fotovoltaica

A energia Solar Fotovoltaica é a energia da conversão direta da luz em eletricidade (Efeito Fotovoltaico). O efeito fotovoltaico é o aparecimento de uma diferença de potencial nos extremos de uma estrutura de material semicondutor, produzida pela absorção da luz. A célula fotovoltaica é a unidade fundamental do processo de conversão [22].

A energia solar fotovoltaica ligada à rede pode ser gerada por centrais solares fotovoltaicas ou sistemas fotovoltaicos residenciais ou comerciais. As usinas de energia solar fotovoltaica são grandes instalações de geração, e algumas delas excedem 100 MW. Essas grandes usinas estão conectadas a sistemas de transmissão. No entanto, a maioria dos sistemas fotovoltaicos é residencial (até vários kW), exemplificado na figura 4 e escala comercial (até vários MW) conectados a redes de distribuição [23].

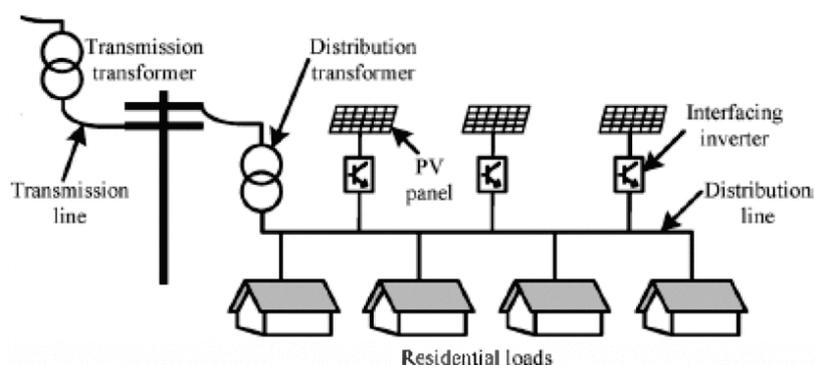


Figura 4: Exemplo de um sistema de coleta de energia fotovoltaica ligada a rede [23].

Sistemas de energia solar fotovoltaica off-grid, figura 5, convertem a radiação solar em corrente elétrica. A tensão na saída desses módulos varia ao longo do dia, já o banco de baterias, possui uma tensão mais estável, com valores nominais normalmente de 12V, 24V, 36V ou 48V. Para resolver esse problema, utiliza-se um equipamento chamado de

controlador de carga. Por fim, o banco de baterias é conectado em um aparelho chamado inversor, que converte a corrente e tensão contínuas oriundas dos módulos fotovoltaicos e baterias, em corrente e tensão alternadas [24].

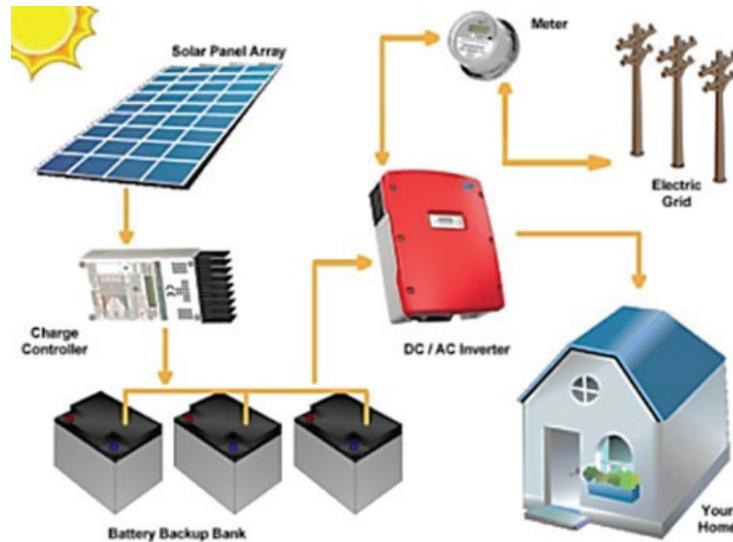


Figura 5: Exemplo de um sistema de coleta de energia fotovoltaica off-grid [24].

Outro ponto importante relacionado a esse sistema de energia renovável, é o atendimento de comunidades isoladas. Isto tem impulsionado a busca e o desenvolvimento de fontes renováveis de energia. Segundo Lavezzo [22], no Brasil, por exemplo, 15% da população não possui acesso à energia elétrica. Coincidentemente, esta parcela da população vive em regiões onde o atendimento por meio da expansão do sistema elétrico convencional é economicamente inviável. Trata-se de núcleos populacionais esparsos e pouco densos, típicos das regiões Centro-Oeste, Nordeste e Norte.

No Brasil a geração de energia elétrica por conversão fotovoltaica teve um impulso notável [25], através de projetos governamentais e também privados, atraindo interesse de fabricantes pelo mercado brasileiro. A quantidade de radiação incidente no Brasil é outro fator muito significativo para o aproveitamento da energia solar.

2.3 PARÂMETROS METEOROLÓGICOS

Para Liu *et al.* [26], parâmetros meteorológicos são as representações numéricas do clima e desempenham um papel importante no desenvolvimento de novas tecnologias. Esses parâmetros influenciam no comportamento termodinâmico da atmosfera, também chamadas de variáveis meteorológicas, tais como temperatura, umidade do ar, pressão, vento, evaporação, irradiação e precipitação. Dentre os parâmetros citados, abaixo sintetizo os dois principais desta fundamentação teórica.

2.3.1 Precipitação

Para Schiavetti e Camargo [27], a precipitação ou popularmente a chuva ocorre a partir de complexos fenômenos de aglutinação e respectivo crescimento das pequenas gotículas em nuvens com a presença significativa de umidade e núcleos de condensação (partículas em suspensão), formando grande quantidade de gotas com tamanho e peso suficientes para que a força gravitacional supere a turbulência natural ou os movimentos ascendentes da atmosfera.

A precipitação é toda água proveniente da atmosfera que atinge a superfície. O estado em que a água se encontra é o fator que vai diferenciar as formas de precipitação [28].

- (a) garoa ou chuvisco: precipitação leve e uniforme, com grande quantidade de gotículas, cujo diâmetro varia entre 0,1 mm e 0,5 mm;
- (b) chuva: precipitação cujas gotículas de água são maiores que 0,5 mm;
- (c) neve: precipitação em forma de cristais de gelo;
- (d) granizo: blocos de gelo com diâmetro que varia de 5 a 50 mm, podendo ser maior em alguns casos raros;
- (e) orvalho: gotas de água que se formam por meio de condensação do vapor-d'água próximo a superfície sob específicas condições de temperatura;
- (f) geada: o mesmo princípio do orvalho, mas o vapor-d'água se condensa diretamente em gelo.

As principais características da precipitação são volume, duração e distribuição espaço-temporal [28]. É possível determinar os locais mais úmidos, mais secos, estações chuvosas ou secas de um continente ou país baseado em registros históricos de monitoramento.

Outro fator importante dentro da precipitação é a possibilidade de sua medição. Esse monitoramento é de grande importância para o estudo e entendimento dos processos da natureza. Essas medições possibilitam que estudos mais amplos em relação ao nosso clima e tempo sejam possíveis.

2.3.2 Irradiação Solar

A irradiação é a intensidade (ou potência) da luz do sol medida em W/m^2 . A potência é um valor instantâneo, e portanto a irradiação é a intensidade da luz do sol em um determinado momento. A irradiação muda ao longo do dia à medida em que o sol se desloca no céu ou é encoberto pelas nuvens.

O Brasil é o país que mais recebe irradiação solar em todo o mundo [29]. Por estar localizado próximo à linha do Equador, o país recebe alta incidência de sol durante todo o dia, com pouca variação ao longo das estações do ano, em função das características de translação do planeta. O país recebe, durante todo o ano, mais de 3.000 horas de brilho do sol, correspondendo a uma incidência solar diária que pode ir de 4,5 a 6,3 kWh/m².

Um comparativo à ser feito é em relação a Alemanha, que é o país que mais explora a energia fotovoltaica em todo o mundo, recebe aproximadamente 40% menos luz solar em sua região de maior potencial, em comparação com a incidência brasileira. Mesmo assim, a energia solar ainda é pouco aproveitada no Brasil, correspondendo a pouco mais de 1,4% da nossa matriz energética de acordo com a Associação Brasileira de Energia Solar Fotovoltaica - ABSOLAR [30], como mostra a figura 6.

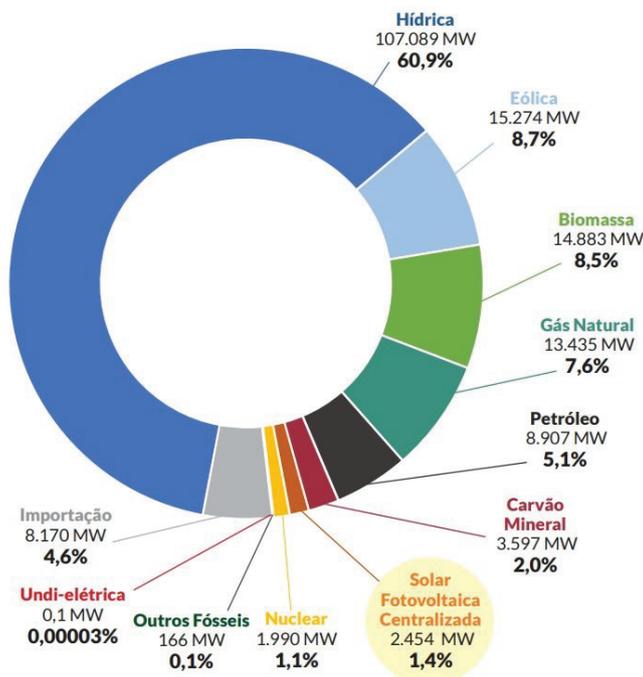


Figura 6: Matriz energética brasileira - Ano base 2019. Adaptado de [30]

O índice de irradiação solar de cada região é medido em Wh/m², unidade de medida padrão da ABNT. Para medi-lo, é usado um instrumento chamado de piranômetro, que mede a radiação total que chega a determinado ponto, sendo ela direta, difusa ou refletida. Outros aparelhos também podem ser usados para medir a radiação incidente, como o pireliômetro e o heliógrafo [29].

O Brasil tem um enorme potencial no cenário de energia solar mundial, como mostra a tabela 3, onde os valores ultrapassam a média de 3 kWh/m². Com iniciativas governamentais, o Brasil vem aumentando a participação da energia solar em seu espectro energético, dominado pelos métodos hidrelétrico e termelétrico. O incentivo por parte do governo e com a participação da população, possui as condições necessárias para fazer

do país um grande produtor de energia solar, podendo assim se tornar uma referência na transformação energética e sustentável do planeta.

Tabela 3: Totais diários de irradiação global horizontal para cada região brasileira 2017 [29].

Região	Irradiação Global Horizontal Média Observada (kWh/m ²)
Norte	4,8
Nordeste	5,5
Centro-Oeste	5,1
Sudeste	5,0
Sul	4,4
Médio	5,1

2.4 TÉCNICAS APLICADAS PARA CIÊNCIA DE DADOS

As técnicas aplicadas para a ciência de dados já auxiliam e a cada momento novos rumos são alcançados nessa inovação tecnológica. Técnicas de IA encontraram o ambiente ideal para concretizar sua potencialidade, tanto para a organização dos dados não estruturados (voz, imagens e textos) como para correlações e sugestões voltadas para a criação de cenários de risco e para a tomada de decisão.

De acordo com o diagrama de Venn [31], figura 7, a ciência de dados está na interseção de alguém com habilidades em hacking, conhecimento de estatística e matemática e conhecimento do domínio em análise.



Figura 7: Diagrama de Venn. Adaptado de [31].

Em alto nível, a ciência de dados é um conjunto de princípios fundamentais que apoiam e orientam a extração de informações e o conhecimento de dados baseado em

princípios [32]. O conceito mais relacionado à ciência de dados é a mineração de dados - a extração real do conhecimento dos dados por meio de tecnologias que incorporam esses princípios.

Segundo Zegura *et al.* [33], ciência de dados é um campo interdisciplinar que extrai *insights* de dados por meio de um processo em várias etapas de coleta, análise e uso de dados. Quando a ciência de dados é aplicada para o bem social, uma variedade de partes interessadas é introduzida no processo com a intenção de informar políticas ou programas para melhorar o bem-estar.

A ciência dos dados engloba todas as fases de coleta de dados, engenharia, análise e arquitetura [34]. Para Van der Aalst [35], a ciência de dados é uma fusão/liga de diferentes (sub) disciplinas parcialmente sobrepostas e com tamanhos variados, como é exemplificado na figura. Além disso, os limites não são claros e parecem mudar com o tempo. "Considere, por exemplo, a diferença entre mineração de dados e aprendizado de máquina ou estatística. Suas raízes são muito diferentes: a mineração de dados emergiu da comunidade de bancos de dados e o aprendizado de máquina surgiu da comunidade de IA, ambos completamente desconectados da comunidade de estatística. Apesar das diferentes raízes, as três (sub) disciplinas estão definitivamente se sobrepondo" [35].

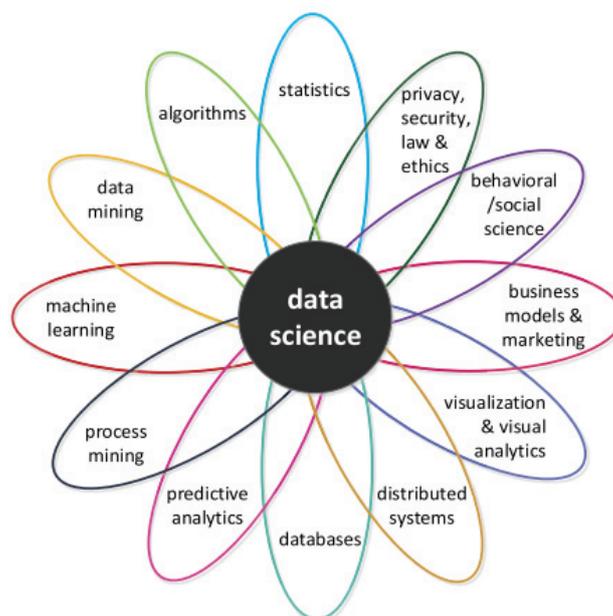


Figura 8: Os ingredientes que contribuem para a ciência de dados [35].

As máquinas são capazes de aprender através do uso de técnicas avançadas oriundas da Ciência de Dados. No futuro espera-se que as aplicações e sistemas hoje existentes se tornem ainda mais rápidos, inteligentes e intuitivos.

2.4.1 Aprendizado de Máquina

Segundo Vanderplas [36], o aprendizado de máquina é frequentemente categorizado como um subcampo de inteligência artificial, surgindo em pesquisas nesse contexto, mas, na aplicação da ciência de dados dos métodos de aprendizado de máquina, é mais útil pensar no aprendizado de máquina como um meio de construir modelos de dados.

Sobretudo, o aprendizado de máquina envolve a construção de modelos matemáticos para ajudar a entender os dados. "Aprendizado" se destaca quando damos a esses modelos parâmetros ajustáveis que podem ser adaptados aos dados observados. Desta maneira, o programa pode ser considerado como "aprendizado" a partir dos dados [36]. Uma vez que esses modelos tenham se adaptado em dados vistos anteriormente, eles podem ser usados para prever e entender aspectos dos dados recém-observados.

É possível classificar os algoritmos de aprendizagem em aprendizado supervisionado, aprendizado não-supervisionado, aprendizado de reforço e aprendizado evolutivo [37].

- (a) aprendizado supervisionado ou também chamado de aprender com exemplos, recebe um conjunto de treinamento com as respostas corretas (metas) e, com base nesse conjunto de treinamento, o algoritmo "generaliza" para responder corretamente a todas as entradas possíveis;
- (b) aprendizado não-supervisionado é também conhecido como estimativa de densidade. Ao contrário do aprendizado supervisionado, não recebe entradas com respostas corretas. O algoritmo busca identificar semelhanças entre as entradas, de modo que as entradas que têm algo em comum sejam categorizadas juntas;
- (c) aprendizado de reforço é um meio termo entre o aprendizado supervisionado e o não supervisionado. O algoritmo é informado quando a resposta está errada, mas não é contada como corrigi-lo. Ele tem que explorar e experimentar diferentes possibilidades até descobrir como acertar a resposta;
- (d) aprendizado evolutivo pode ser definido como um processo de aprendizado, onde os organismos biológicos se adaptam para melhorar suas taxas de sobrevivência e chance de ter filhos em seu ambiente.

Dentre os quatro modelos, os mais utilizados são o Aprendizado supervisionado e o não supervisionado.

2.4.1.1 Aprendizado supervisionado

O aprendizado supervisionado pode ser dividido em classificação e regressão [36].

(a) classificação consiste em tomar vetores de entrada e decidir a qual das classes N eles pertencem, baseado no treinamento de exemplares de cada classe. O ponto mais importante sobre o problema de classificação é que ele é discreto - cada exemplo pertence exatamente a uma classe e o conjunto de classes abrange todo o espaço de saída possível. Algoritmos comumente utilizados na classificação:

- (1) support vector machines são mais popularmente usados para lidar com problemas relacionados à segmentação de imagens, mercado de ações, categorização de textos e ciências biológicas;
- (2) decision trees constroem uma árvore de decisões preditivas feitas com base nos valores reais dos atributos nos dados. As árvores de decisão são usadas para problemas de classificação e regressão;
- (3) naïve bayes são usado principalmente em mineração de texto, análise de sentimentos, categorização de documentos, filtragem de spam, previsão de doenças. O classificador Naïve Bayes baseia-se no Teorema de Probabilidade de Bayes e assume a independência dos atributos, também conhecidos como pressupostos de independência condicional.

(b) Regressão: se o resultado desejado consiste em uma ou mais variáveis contínuas, a tarefa é chamada de regressão. Um exemplo de um problema de regressão seria a previsão do comprimento de um salmão em função de sua idade e peso. Alguns algoritmos mais usados na regressão podem ser:

- (1) regressão linear é utilizada em problemas em que o rótulo é de natureza contínua. Ex: Vendas de uma cadeia de varejo. Consiste no método dos mínimos quadrados ordinários que se ajusta à melhor linha que minimiza a soma dos erros quadrados entre os pontos de dados previstos e reais. Esses algoritmos são usados principalmente para o processo de tomada de decisão;
- (2) regressão logística é utilizada em problemas em que o rótulo é de natureza categórica ou discreta, usado *log odds ratio* para otimizar problemas de negócios, como pontuar clientes e, em seguida, prever aqueles com maior probabilidade de inadimplência no pagamento do empréstimo ou prever taxas de resposta mais altas. determinada estratégia de marketing.

2.4.1.2 Aprendizado não-supervisionado

O aprendizado não-supervisionado é dividido em *clustering* e redução da dimensionalidade [36].

(a) *Clustering*: O objetivo é descobrir grupos de exemplos semelhantes dentro dos dados; Exemplos de alguns algoritmos utilizados:

- (1) *K-Means clustering* agrupa seus pontos de dados em um número (K) de *clusters* mutuamente exclusivos. Muita da complexidade envolve como escolher o número certo para K;
 - (2) *clustering* hierárquico agrupa seus pontos de dados em *clusters* pai e filho. Você pode dividir seus clientes entre idades mais jovens e mais velhas e dividir cada um desses grupos em seus próprios *clusters* individuais;
 - (3) *clustering* probabilístico agrupa seus pontos de dados em *clusters* em uma escala probabilística.
- (b) redução da dimensionalidade significa combinar partes de seus dados de maneiras únicas para transmitir significado.

2.4.1.3 O processo de aprendizagem de máquina

Para Vanderplas [36], o processo de aprendizagem de máquina consiste em examinar brevemente o processo pelo qual os algoritmos de aprendizado de máquina podem ser selecionados, aplicados e avaliados para o problema. Podemos elencar os seguintes processos:

- (a) coleta de Dados e preparação: Se o problema for completamente novo, de modo que os dados apropriados possam ser escolhidos, esse processo deve ser mesclado com a próxima etapa da seleção de recursos, de modo que apenas os dados necessários sejam coletados. Os algoritmos de aprendizado de máquina precisam de quantidades significativas de dados, de preferência sem muito ruído, mas com o aumento do tamanho do conjunto de dados aumentam os custos computacionais e o ponto ideal em que há dados suficientes sem sobrecarga computacional excessiva é geralmente impossível de prever;
- (b) seleção de recurso: Consiste em identificar os recursos mais úteis para o problema em exame e também é necessário que os recursos possam ser coletados sem despesas ou tempo significativos, e que eles sejam robustos ao ruído e à corrupção dos dados que possam surgir no processo de coleta;
- (c) escolha do algoritmo: Consiste em selecionar o algoritmo mais adequado para tratar o problema que está sendo estudado;
- (d) seleção de parâmetros e modelos: Alguns parâmetros precisam ser definidos e ajustados manualmente ou que exigem experimentação para identificar valores apropriados;
- (e) treinamento: Consiste no uso de recursos computacionais para construir um modelo dos dados, a fim de prever os resultados em novos dados;

- (f) avaliação: O sistema precisa ser avaliado quanto a sua acurácia para os dados que foi treinado.

2.4.2 Redes Neurais

É um tipo de técnica computacional baseada em um modelo matemático e assemelha-se ao cérebro humano, criando uma rede neural artificial que através de um algoritmo, permite que o computador aprenda com inserção de novos dados [4]. Também podemos descrever como um método computacional que utiliza um enorme grupo de neurônios artificiais. Esses neurônios são inexatamente equivalentes ao axônio em um cérebro biológico [38]. As redes neurais são usadas em vários campos, tais como aprendizagem de máquina, processamento de imagens, processamento de sinais, series temporais, etc [39].

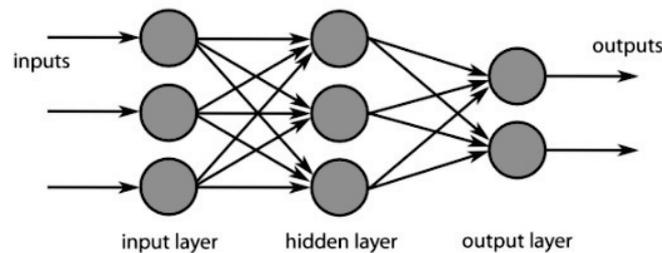


Figura 9: Modelo básico de uma Rede Neural Artificial [4].

2.4.2.1 Aprendizado Profundo ou Deep Learning

É uma subárea de aprendizado de máquina que estuda como solucionar problemas intuitivos. Este tipo de solução permite que computadores aprendam a partir de experiências anteriores e compreendam o mundo em termos de uma hierarquia de conceitos, no qual os conceitos mais complexos são definidos e compreendidos em termos de sua relação com conceitos mais simples e já conhecidos [40]. Envolve um tipo particular de modelo matemático que pode ser pensado como uma composição de blocos simples (composição da função) de um determinado tipo e onde alguns desses blocos podem ser ajustados para prever melhor o resultado final.

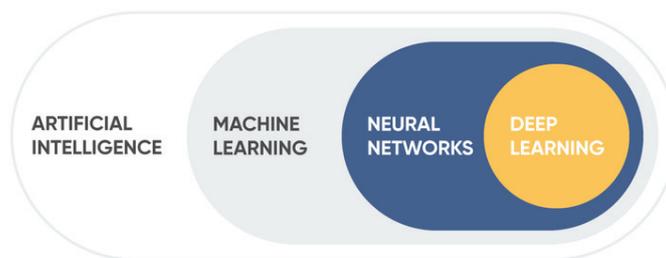


Figura 10: Partes constituintes da Deep Learning [40].

Embora existam muitos algoritmos de inteligência artificial, as redes neurais são capazes de realizar o que hoje em dia é chamado de *Deep Learning* ou aprendizado profundo. Esse novo modelo de aprendizado possui diferentes tipos de arquiteturas.

- (a) redes neurais convolucionais [41] (RNC) contêm cinco tipos de camadas: de entradas, de convolução, de agrupamento, as completamente conectadas e as de saída. Cada camada tem um propósito específico, como de resumo, conexão ou ativação. As redes neurais convolucionais popularizaram a classificação de imagens e a detecção de objetos. Entretanto, RNCs também foram aplicadas em outras áreas como previsão e processamento de linguagem natural;
- (b) redes neurais recorrentes [4] (RNR) usam informações sequenciais, como dados de registro de data e hora de um sensor ou uma frase dita. Essas informações são compostas por uma sequência de termos. Diferentemente das redes neurais tradicionais, as entradas de uma rede neural recorrente não são independentes umas das outras, e os resultados para cada elemento dependem da computação dos elementos precedentes. RNRs são utilizadas na previsão e aplicação de séries temporais, análise de sentimento e outras aplicações de texto;
- (c) redes neurais *feedforward* [4] nas quais cada perceptron em uma camada é conectado a todo perceptron da camada seguinte. A informação é entregue de maneira antecipada de uma camada à seguinte seguindo sempre em frente. Não há *loops* de *feedback*;
- (d) redes neurais *autoencoder* [42] são utilizadas para criar abstrações chamadas *encoders*, criados a partir de um conjunto estipulado de entradas. Apesar de similares às redes neurais mais tradicionais, *autoencoders* procuram modelar as entradas por si só e, portanto, o método é considerado não supervisionado. A premissa dos *autoencoders* é diminuir a sensibilidade ao que é irrelevante e aumentar ao que é. Conforme camadas são adicionadas, outras abstrações são formuladas em camadas mais altas (camadas mais próximas ao ponto onde uma camada decodificadora é introduzida). Essas abstrações podem, então, ser usadas por classificadores lineares ou não lineares.

Os métodos a seguir podem ser aplicados a todas essas arquiteturas e auxiliam na procura do melhor modelo ou do modelo ideal para determinado problema a ser aplicado uma rede neural.

- (a) *back propagation* [43] é simplesmente um método para calcular as derivadas parciais (ou gradiente) de uma função, que tem a forma como uma composição da função (como em Redes Neurais). Quando você resolve um problema de otimização usando um método baseado em gradiente (descida de gradiente é apenas um deles), você quer computar o gradiente da função a cada iteração;
- (b) *dropout* [44] trabalha na eliminação aleatória de neurônios durante o processo de aprendizagem, para evitar a sobreaptação aos dados (*overfitting*). Isto evita que as unidades se adaptem demais;
- (c) *batch normalization* [45], regulariza gradientes de distração para *outliers* e flui em direção ao objetivo comum (normalizando-os) dentro de uma faixa do mini lote. Naturalmente, as redes neurais, incluindo as redes profundas, requerem um ajuste cuidadoso dos parâmetros de inicialização e aprendizagem do peso. A normalização dos lotes ajuda a flexibilizá-los um pouco;
- (d) *max pooling* [45] é um processo de discretização baseado em amostras. O objetivo é reduzir uma representação de entrada (imagem, matriz de saída de camada oculta, etc.), reduzindo sua dimensionalidade e permitindo que sejam feitas suposições sobre as características contidas nas sub-regiões armazenadas;
- (e) *transfer learning* [46] é quando você pega uma CNN treinada em um conjunto de dados, corta a(s) última(s) camada(s), recria a(s) última(s) camada(s) dos modelos em um conjunto de dados diferente. Intuitivamente, você está re-treinando o modelo para reconhecer diferentes características de nível superior. Como resultado, o tempo de treinamento é muito reduzido, portanto, a transferência do aprendizado é uma ferramenta útil quando você não tem dados suficientes ou se o treinamento requer muitos recursos;
- (f) *multilayer perceptrons (MLP)* [5] [47] [48] é um perceptron que se une a outros perceptrons, empilhados em várias camadas, para resolver problemas complexos. O diagrama abaixo mostra um MLP com três camadas. Cada perceptron da primeira camada à esquerda (a camada de entrada), envia saídas para todos os perceptrons da segunda camada (a camada oculta), e todos os perceptrons da segunda camada enviam saídas para a camada final à direita (a camada de saída);

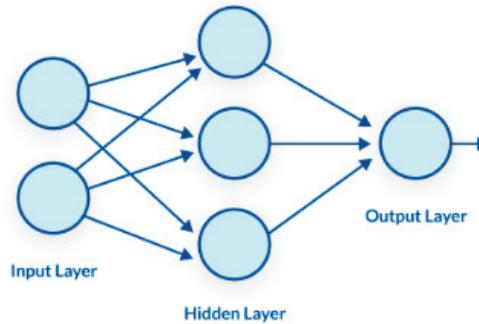


Figura 11: Diagrama multilayer perceptrons [48].

(g) *long short-term memory (LSTM)* [47] [4] [49] são um tipo de rede neural recorrente capaz de aprender a dependência da ordem de aprendizagem em problemas de previsão de seqüência. Este é um comportamento necessário em domínios de problemas complexos como tradução automática, reconhecimento da fala, series temporais e muito mais. As LSTMs são uma área complexa de aprendizado profundo. Pode ser difícil entender o que são LSTMs e como termos como bidirecional e seqüência-a-sequência se relacionam com o campo. Uma rede LSTM possuem 3 aspectos principais que a diferenciam de um neurônio comum em uma rede neural recorrente:

- (1) ela tem o controle de decidir quando deixar a entrada de dados entrar no neurônio;
- (2) ela tem o controle de decidir quando lembrar o que foi computado na etapa de tempo anterior;
- (3) tem controle sobre a decisão de quando deixar a saída passar para o próximo carimbo de tempo.

A decisão é baseada na entrada atual em si, como mostra o diagrama a seguir:

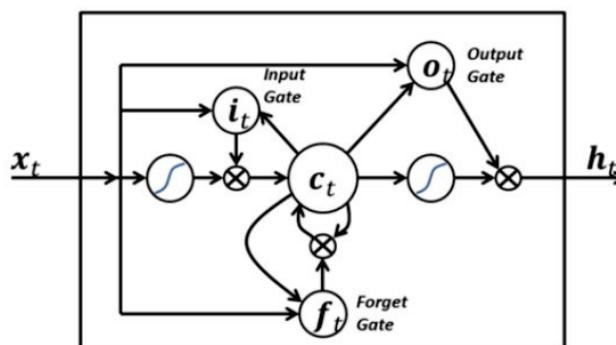


Figura 12: Diagrama long short-term memory [49].

A entrada inicial x_t decide todos os 3 pontos acima, onde o *input gate* i_t toma uma decisão para o ponto 1. O *forget gate* f_t toma uma decisão sobre o ponto 2 e o *output*

gate o_t toma uma decisão sobre o ponto 3. Somente a entrada é capaz de tomar todas estas três decisões. Isto é inspirado pela forma como nossos cérebros trabalham e pode lidar com interruptores de contexto repentinos com base na entrada.

O Aprendizado Profundo é fortemente focado na técnica. Não há muitas explicações concretas para cada uma das novas ideias e, a maioria das novas ideias surgiram com resultados experimentais para provar que elas funcionam.

2.4.3 Softwares Utilizados

A grande demanda e necessidade do desenvolvimento de sistemas utilizando técnicas e modelos de aprendizado profundo em um curto espaço de tempo fez com que a comunidade desenvolvesse bibliotecas de software que pudessem ser utilizadas nas mais diversas situações. Seria impraticável o desenvolvimento de sistemas para a solução de problemas complexos como os que existem hoje sem a utilização dos componentes de software disponíveis. Nessa seção abordaremos alguns destes softwares.

2.4.3.1 TensorFlow

Como um dos principais softwares para aprendizado de máquinas, o TensorFlow [50] é um sistema que opera em larga escala e em ambientes heterogêneos. Utilizá-se de grafos de fluxos de dados para representar o cálculo, o estado compartilhado, e as operações que alteram esse estado. Através de um conjunto de clusters, ele mapeia os nós de grafos dentro de uma máquina através de múltiplos dispositivos computacionais, incluindo CPUs de múltiplos núcleos, GPUs de uso geral, e ASICs projetados sob medida conhecidos como Tensor Processing Units (TPUs).

O TensorFlow permite que os vértices representem cálculos que possuem ou atualizam o estado mutável. As bordas transportam tensores (matrizes multidimensionais) entre os nós, e TensorFlow insere transparentemente a comunicação apropriada entre as sub-computações distribuídas. Esta arquitetura dá flexibilidade ao desenvolvedor da aplicação: enquanto no servidor de parâmetros anterior o gerenciamento do estado compartilhado é incorporado ao sistema, TensorFlow permite aos desenvolvedores experimentar novas otimizações e algoritmos de treinamento.

TensorFlow suporta uma variedade de aplicações, com foco no treinamento e inferência em redes neurais profundas.

2.4.3.2 Keras

Keras [51] é uma API de aprendizado profundo escrito em Python, executado em cima da plataforma de aprendizado da máquina TensorFlow. Foi desenvolvido com o foco em permitir uma rápida experimentação. Possibilitando partir da ideia e chegar em um resultado o mais rápido possível.

Sua arquitetura permite que engenheiros e pesquisadores aproveitem ao máximo a escalabilidade e as capacidades de cross-platform do TensorFlow 2.0: você pode executar o Keras em TPU ou em grandes grupos de GPUs, e pode exportar seus modelos Keras para executar no navegador ou em um dispositivo móvel.

Possui quatro princípios orientadores:

- (1) Modularidade: Um modelo pode ser entendido como uma sequência ou um gráfico apenas. Todas as preocupações de um modelo de aprendizado profundo são componentes discretos que podem ser combinados de forma arbitrária.
- (2) Minimalismo: A biblioteca fornece apenas o suficiente para alcançar um resultado, sem adornos e maximizando a legibilidade.
- (3) Extensibilidade: Novos componentes são intencionalmente fáceis de adicionar e usar dentro da estrutura, destinados a pesquisadores para experimentar e explorar novas ideias.
- (4) Python: Sem arquivos de modelo separados com formatos de arquivo personalizados. Tudo é Python nativo.

3. TRABALHOS RELACIONADOS

De acordo com Nakagawa *et al.* [52], a sumarização dos resultados dos estudos primários por meio de estudos secundários é um mecanismo de pesquisa valioso por prover conhecimento de um dado tópico e apoiar na identificação de temas para pesquisa futura. Ainda como um aspecto importante, a RS apresenta uma avaliação considerada justa sobre um tópico de pesquisa, uma vez que utiliza uma forma de revisão rigorosa, confiável e passível de auditoria.

As seções a seguir apresentam o protocolo da revisão sistemática utilizado e os resultados obtidos a partir da utilização deste método para obtenção dos trabalhos selecionados.

3.1 INFORMAÇÕES GERAIS

O objetivo é identificar na literatura trabalhos relacionados ao uso de técnicas que fazem arte da Ciência dos Dados para a obtenção de um modelo de previsão de precipitação e na irradiação solar.

Questões de Pesquisa: Esta revisão sistemática pretende responder as seguintes questões:

- (a) Qual a finalidade da Ciência de Dados em estudos meteorológicos?
- (b) Quais resultados relevantes podem ser alcançados com o uso de técnicas de Ciência de Dados em estudos meteorológicos que abordem as variáveis precipitação e irradiação/insolação solar?

3.2 IDENTIFICAÇÃO DOS ESTUDOS

As bases de dados escolhidas para esta revisão são: ACM (Association for Computing Machinery), ScienceDirect, Scopus, Springer e IEEE (Institute of Electrical and Electronic Engineers).

Para o levantamento de trabalhos, definiu-se a seguinte expressão de busca e suas respectivas palavras-chave: **(precipitation OR rainwater OR rainfall OR "solar irradiation"OR "solar insolation") AND ("recommendation system"OR "machine learning"OR "data science"OR "data analysis"OR "deep learning"OR "artificial intelligence"OR "big data")**

Não houve alteração na expressão de busca nas bases de dados. O intervalo de busca foi de 4 anos, ou seja, de 2016 a 2019. Os estudos analisados e considerados nessa

revisão devem estar escritos na língua inglesa em artigos, desconsiderando capítulos de livros, anais e resumos de eventos e seminários.

3.3 CRITÉRIOS DE ELEGIBILIDADE E ESCOLHA DOS TRABALHOS

Segundo Nakagawa *et al.*, uma das características que diferem a RS de outros métodos é o rigoroso critério de seleção dos estudos, que abrangem tantos critérios de inclusão quanto exclusão.

- (a) O estudo deve fazer uso de alguma técnica da ciência dos dados alinhado com o estudo da precipitação;
- (b) O estudo deve fazer uso de alguma técnica da ciência dos dados alinhado com o estudo da insolação ou irradiação solar;
- (c) O estudo propõe ou relata um processo, método, técnica, relacionada a precipitação e insolação/irradiação solar para apoiar este estudo;

3.4 PROCEDIMENTOS DE SELEÇÃO DOS TRABALHOS

A avaliação dos trabalhos é realizada através das seguintes etapas: identificação, triagem (Fase 1 e 2) e elegibilidade.

- (a) Identificação: a expressão de busca é utilizada em cada base de dados para obter os trabalhos de interesse. Os resultados são consolidados para a remoção de itens duplicados;
- (b) Triagem - Fase 1: é realizada a avaliação de título e do resumo quanto ao enquadramento na área de interesse do presente estudo;
- (c) Triagem - Fase 2: é realizada a avaliação da introdução. Verifica se objetivos, justificativas e contextualização atendem os critérios definidos. Caso não atendam, descarta-se o trabalho;
- (d) Elegibilidade: é avaliado se o estudo atende aos critérios de elegibilidade previstos e se os resultados e a conclusão obtidos foram satisfatórios. Considera-se também a qualidade da descrição da técnica em análise e se há comparação com outros métodos similares.

3.5 EXTRAÇÃO DOS DADOS

Para cada um dos 25 estudos selecionados para essa revisão sistemática, serão destacadas as seguintes informações: autores, objetivos gerais, descrição do método utilizado para análise dos dados de precipitação de água de chuva, descrição do método utilizado para análise dos dados de insolação solar e os principais resultados obtidos.

Foram obtidos 595 artigos a partir da busca nas bases de dados utilizando, sempre que disponível, filtrando por metadados, pois é um recurso que permite filtrar e localizar dinamicamente estes conteúdos. Na triagem foram selecionados 54 artigos para a leitura completa e verificação dos resultados. Dentre estes 54 artigos, 25 foram selecionados para o escopo desta revisão, como mostra a figura 13.

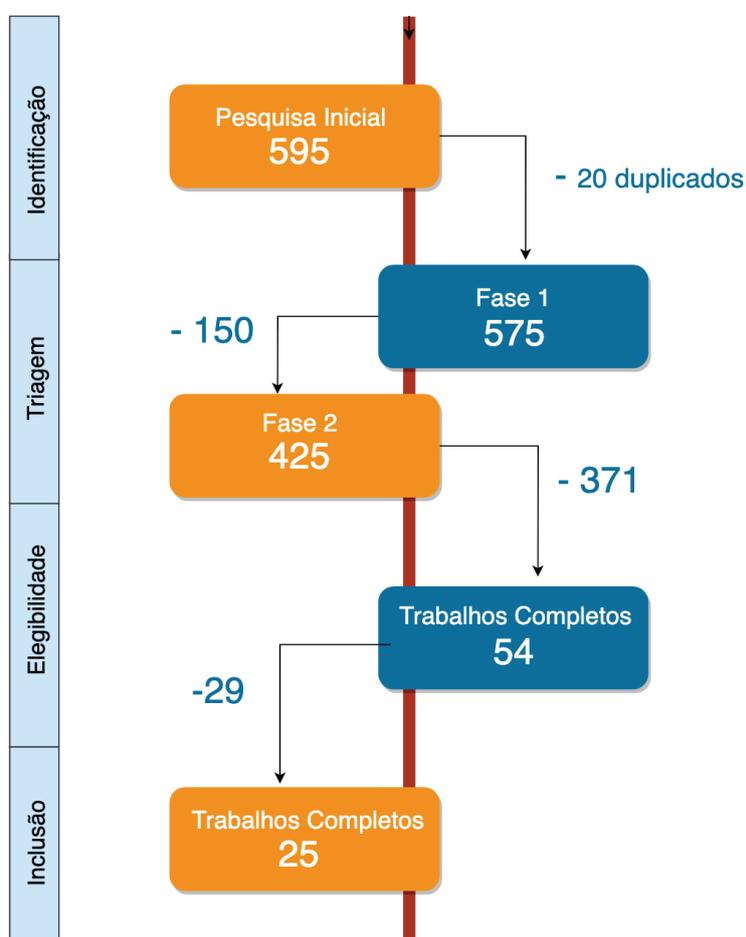


Figura 13: Esquema do processo de revisão sistemática [Autoria própria].

Os 25 estudos analisados por completo, envolveram cerca de 27 técnicas para ciência de dados. Estas técnicas mostram várias possibilidades, bem como o uso na predição de variáveis e construção de melhores modelos meteorológicos.

A tabela 4 mostra um resumo das técnicas de ciência dos dados utilizados para as variáveis meteorológicas de precipitação e irradiação solar incluídas na revisão sistemática.

Título	Ano	Subconjunto de I.A	Abreviação	Métodos Extra, Técnica ou Algoritmo	Grupo
Forecasting Based neural Networks	2018	Deep Learning	DL	Convolutional Deep Belief Networks(CDBNs)	Precipitation
Model For Radar Gauge Observations	2017	Machine Learning	ML	Dynamic Multilayer Perceptron (DMLP)	Precipitation
tem For Rainfall And Ground Radars	2017	Machine Learning	ML	Multilayer Perceptron (MLP)	Precipitation
arning To An Early t-term Heavy Rainfall	2019	Machine Learning	ML	Logistic Regression (LR)	Precipitation
struction And er Prediction	2019	Machine Learning	ML	Convolutional Neural Networks (CNNs) Recurrent Neural Networks (RNNs) Long-short term memory (LSTM) Long-Short Term Memory (LSTM)	Precipitation
els For The ainfall	2018	Deep Learning	DL	Convolutional Neural Network (ConvNet/CNN) All deep learning architectures were trained using the back propagation through time (BPTT)	Precipitation
Rainfall Using pproaches	2019	Machine Learning	ML	Decision Tree (DT), Random Forest (RF) Linear Regression (LR), Neural Network (NN)	Precipitation
Rainfall Using echniques	2018	Machine Learning	ML	Linear Regression (LR)	Precipitation
ed On Weather ainfall Based On ficial Neural Network	2017	Deep Learning	DL	Multilayered Perceptron (MLP) trained using Lavenberg Marquardt (LM)Algorithm	Precipitation
casting Using epton	2018	Deep Learning	DL	Multi-Layer Perceptron (MLP) with Principal Component Analysis (PCA) to reduce the dimension of thirteen physical factors and a greedy algorithm to determine de structure of MLP.	Precipitation
Learning Algorithms ellite-based Optical als	2016	Machine Learning	ML	Random Forests (rf) Neural Networks (nnet) Averaged Neural Networks (avnnet) Support Vector Machines (svm)	Precipitation
rests And Support me Radar-derived sting	2017	Machine Learning	ML	Random Forests (RF) Support Vector Machine (SVM)	Precipitation
del From Satellite Neural Network	2018	Deep Learning	DL	Convolutional Neural Network (ConvNet - CNN)	Precipitation
ediction Model al Neural Networks	2017	Deep Learning	DL	Multi-task Convolutional Neural Network (convnet/cnn)	Precipitation

o das técnicas de ciência de dados utilizadas para a precipitação e irradiação solar.

Título	Ano	Subconjunto de I.A	Abreviação	Métodos Extra, Técnica ou Algoritmo	Grupo
Comparison Algorithm For Precipitation From Satellite Weather Derivatives	2016	Machine Learning	ML	Compare The Performance Of Our Hybrid Gp/ga, Against Mcrp Radial Basis Function And Gp Without Decomposition	Precipitation
Forecasting On The Intensity Of Precipitation Using A Lte/4g Mobile Network	2018	Deep Learning	DL	Probabilistic Neural Network (PNN)	Precipitation
Forecasting Precipitation For Data-driven Modelling	2017	Machine Learning	ML	Mean-based residuals, Min/Max-gased residuals	Solar Irradiation
Forecasting Solar Irradiation Using Deep Learning	2017	Deep Learning	DL	Deep Recurrent Neural Networks (DRNN) with Long-short term memory (LSTM) Feedforward neural networks (FNN)	Solar Irradiation
Forecasting Solar Irradiation In Power Generation: A Study On Solar Irradiation Forecasting	2019	Deep Learning	DL	Long-Short Term Memory (LSTM) Linear Regression (LR)	Solar Irradiation
Forecasting Solar Irradiation Using A Deep Learning Program	2019	Deep Learning	DL	Back Propagation Artificial Neural Network (BP-ANN)	Solar Irradiation
Forecasting Solar Irradiation Using A Machine-learning Approach	2017	Machine Learning	ML	Multi-layer perceptron (MLP) adaptive neuro-fuzzy inference system (ANFIS) Support Vector Machines (SVM) Decision Trees (DT) algorithms	Solar Irradiation
Forecasting Solar Irradiation Using A Machine-learning Approach With Forecasting Models	2018	Machine Learning	ML	Naive Models, Classical Machine Learning Models And Regression Trees-based Models.	Solar Irradiation
Forecasting Solar Power Generation Using Regression Techniques	2017	Machine Learning	ML	Compare Linear Regression (LR) Artificial Neural Netorkw (ANN)	Solar Irradiation
Forecasting Solar Power Generation Using Angstrom-prescott And Ann Techniques	2017	Machine Learning	ML	Compare Angstrom-Prescott (A-P) model and two Machine Learning techniques (ML) Support Vector Machine (SVM) Artificial Neural Network (ANN)	Solar Irradiation
Forecasting Solar Power Generation In Brazil Using A Deep Learning Approach	2017	Deep Learning	DL	Multilayer Perceptron-based Ann Model	Solar Irradiation

A seguir é apresentado os principais trabalhos abordados na revisão sistemática. Ao final, será feita uma síntese geral dos restantes dos artigos deste tema que não foram abordados inicialmente.

3.6 PREVISÃO DE IRRADIÂNCIA SOLAR USANDO REDES NEURAI PROFUNDAS

A irradiância pode ser prevista usando métodos estatísticos como ANN, SVM ou ARMA. No entanto, eles não têm precisão porque não conseguem capturar dependência de longo prazo ou não podem ser usados com big data devido à escalabilidade.

Diante disto, para prever a irradiância solar, Alzahrana *et al.* [4], apresentou um método usando DRNN. Esse método adiciona complexidade ao modelo sem especificar que forma a variação deve tomar e permitir a extração de recursos de alto nível.

Os dados da irradiação solar foram normalizados através de uma função do MATLAB, *Mapminmax*. Após a normalização, os dados foram divididos em três partes: (a) conjunto de dados de treinamento; (b) conjunto de dados de teste e (c) conjunto de dados de validação com 70%, 15% e 15%, respectivamente. A arquitetura utilizada foi DRNN com unidades de LSTM com duas camadas ocultas e 35 neurônios ocultos. Outro modelo foi implementado como um modo de referência. O modelo de referência usa um FNN. A FNN foi treinada usando o método *Backpropagation*. Os neurônios ocultos e o número de camadas foram escolhidos por tentativa e erro para obter o máximo desempenho da FNN.

Esse método forneceu resultados promissores, já que o RMSE é o mais baixo. A comparação foi feita no conjunto de dados de teste. Uma rede neural feedforward foi implementada usando uma camada oculta com 21 neurônios, e para o treino utilizou-se o algoritmo de Levenberg-Marquardt. O treinamento foi interrompido cedo para evitar os problemas de overfitting. A FNN teve o pior desempenho entre os outros métodos. A máquina de vetores de suporte teve melhor desempenho que a FNN. No entanto, o LSTM mostrou que as redes neurais de aprendizagem profunda podem superar todos os outros métodos, como indicam os testes de desempenho mostrados na tabela 5.

Tabela 5: Teste de performance dos métodos [4]

Metodo	RMSE	MBE
FNN	0,16	0,005
SVR	0,11	0,0042
Deep Learning (LSTM)	0,86	0,004

3.7 POTENCIAL DE QUATRO DIFERENTES ALGORITMOS DE APRENDIZADO DE MÁQUINA NA MODELAGEM DE RADIAÇÃO SOLAR GLOBAL DIÁRIA

Estimar a quantidade de radiação solar incidente é o primeiro passo na concepção ou análise do desempenho de qualquer sistema solar. No entanto, as medições da radiação solar são muito caras e raras. Mesmo nos países desenvolvidos, as estações solares são muito poucas comparadas às estações meteorológicas, especialmente aquelas com alta resolução temporal.

Hassan *et al.* [5] adotaram quatro (MLP, ANFIS, SVM e DT) diferentes modelos de algoritmos utilizando aprendizado de máquina. Todos os modelos foram agrupados em quatro categorias: modelos baseados em luz do sol, temperatura, parâmetros meteorológicos e número do dia, como mostra a figura 14.

O objetivo deste estudo foi comparar o desempenho do algoritmo de soft-computing com o de modelos empíricos bem conhecidos, a fim de examinar em que medida esses modelos podem melhorar as previsões de Gd e determinar qual é o algoritmo mais adequado [5].

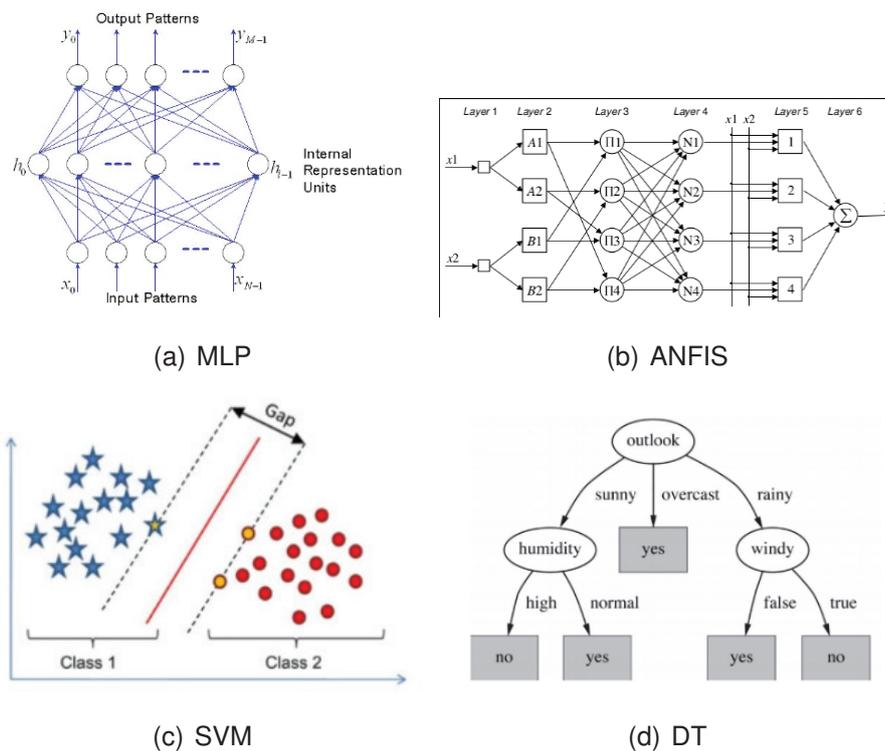


Figura 14: Exemplos dos quatro diferentes modelos de algoritmos utilizando aprendizado de máquina [5]

Os resultados mostraram que os modelos MLP se sobressaem na estimativa da irradiação global com erro quadrático médio inferior ao dos melhores modelos de regressão correspondentes em 4,75 e 31,69%, dependendo da categoria do modelo. Seguido

por modelos ANFIS (se cuidadosamente validados) e modelos SVM. Além disso, o estudo avaliou a capacidade das árvores de decisão em modelar a radiação solar. Apesar de sua simplicidade, os méritos dos modelos baseados em temperatura e número de dias são demonstrados, com coeficientes de determinação superiores a 85%, a serem usados em caso de indisponibilidade de registros de luz solar.

3.8 PREVISÃO DE IRRADIAÇÃO SOLAR COM APRENDIZADO DE MÁQUINA: MÉTODO DE SELEÇÃO DE MODELOS DE PREVISÃO DEPENDENDO DA VARIABILIDADE CLIMÁTICA

Neste estudo, Voyant *et al.* [6] apresentou comparações de onze modelos de aprendizado de máquina - figura 15 - para prever a irradiação solar global e comparar seus desempenhos em três locais diferentes. O principal objetivo foi propor uma metodologia de priorização para destacar o melhor método de previsão entre os testados de acordo com o nível de variabilidade da irradiação.

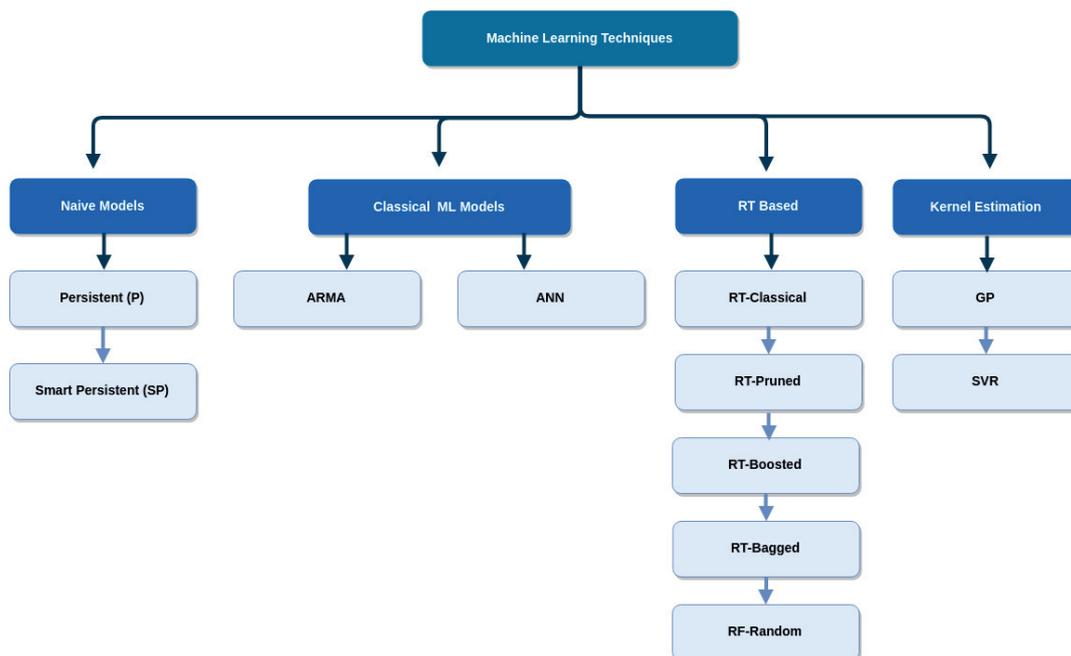


Figura 15: Fluxograma dos modelos testados [6].

Os resultados mostraram que os modelos mais eficientes são selecionados para cada variabilidade e horizonte temporal: (a) para a variabilidade fraca, ARMA e MLP ANN *based* são os mais eficientes, (b) para uma variabilidade média, a ARMA e a RT-Bagged são os melhores preditores. E (c) para uma alta variabilidade, apenas métodos mais complexos podem ser usados eficientemente, tais como RT-Bagged e RT-Random.

3.9 PREVISÃO DE ENERGIA SOLAR MÉDIA DIÁRIA USANDO TÉCNICAS DE REGRESSÃO DE APRENDIZADO DE MÁQUINA

Neste estudo Jawaid e NazirJunejo [7], apresentou uma análise comparativa da previsão através de ANN em contraponto aos algoritmos de regressão padrão. Foram testados quatro técnicas K-NN, Linear Regression, SVM e ANN de aprendizado de máquina comumente usadas com o conjunto de dados selecionados para avaliar desempenhos individuais com a escolha de atributos climáticos. Para a validação dos modelos treinados foi utilizado as técnicas K-fold Cross-Validation (k-fold CV) e Split Validations (k-fold SV).

A técnica de CV foi usada para ordenar os dados em segmentos K. A sequência de treinamento é realizada em segmentos K-1 e o teste é feito no segmento restante. Este ciclo inteiro é repetido K vezes. A técnica de SV foi utilizada para ordenar dados cronológicos.

Os resultados mostraram que o uso da capacidade de técnicas de regressão de aprendizado de máquina, especialmente redes neurais artificiais para prever com precisão a energia solar média diária até um grau muito alto de precisão é satisfatória. Além de usar atributos específicos de clima e época do ano, também foi comprovado a eficácia do uso de ângulos solares (azimute e zênite) para aumentar a precisão de tais previsões.

3.10 UM MODELO DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DE PRECIPITAÇÃO DE RADAR BASEADO EM OBSERVAÇÕES DE MEDIDOR

A estimativa de precipitação baseada em medições de radar foi abordada usando algoritmos paramétricos como a relação Z-R. No entanto, tais relações empíricas podem não ser suficientes para capturar a variabilidade espaço-temporal da precipitação.

Tan, Chandrasekar e Chen [54], apresentam um modelo de aprendizado de máquina baseado em DMLP para estimativa de precipitação usando múltiplas camadas para capturar as abstrações complexas de refletividade de radar em diferentes níveis de altitude.

Em comparação com a abordagem tradicional de aprendizado de máquina baseada em redes neurais, o modelo DMLP possui um número maior de camadas ocultas e uma arquitetura mais complexa para a representação de dados, conforme mostra a figura 16. As informações e conhecimentos estruturados de alto nível podem ser extraídos dos recursos de dados de baixo nível através do processo de aprendizado hierárquico.

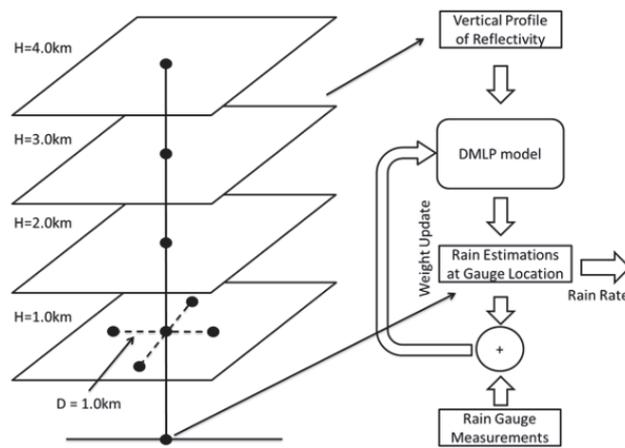


Figura 16: Diagrama conceitual do modelo de aprendizado de máquina baseado em DMLP para estimativa de precipitação de radar [54].

Este modelo baseado em DMLP para estimativa de precipitação usando medidas de refletividade de radar é projetado para capturar as abstrações complexas da refletividade em diferentes níveis de atitude, usando a identificação e extração de múltiplas camadas.

Os resultados do produto de precipitação derivado do modelo DMLP, comparado com um conjunto de dados de pluviômetro independente, mostram um excelente desempenho do novo modelo de precipitação baseado em aprendizado de máquina.

3.11 UM SISTEMA DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DE PRECIPITAÇÃO A PARTIR DE RADARES ESPACIAIS E TERRESTRES

A precipitação medida pelo PR (*Precipitation Radar*) da TRMM (*Tropical Rainfall Measuring Mission*) é importante para o estudo da distribuição de precipitação nas regiões tropicais. A validação do TRMM PR no solo é difícil porque os sistemas de sensoriamento possuem características diferentes em termos de resolução, escala, aspecto de vista e ambientes de detecção.

Um sistema de aprendizado de máquina para treinar radares de solo para estimativa de chuva usando dados de pluviômetro e, subsequentemente, usando a estimativa de precipitação de radar de solo treinada para treinar TRMM PR é mostrado por Chandrasekar, Tan e Chen [8]. Este sistema, figura 17, pode construir uma conexão entre as medições do medidor de solo e observações de radar no solo, e transferir essa conexão para as observações do TRMM PR para a estimativa de precipitação.

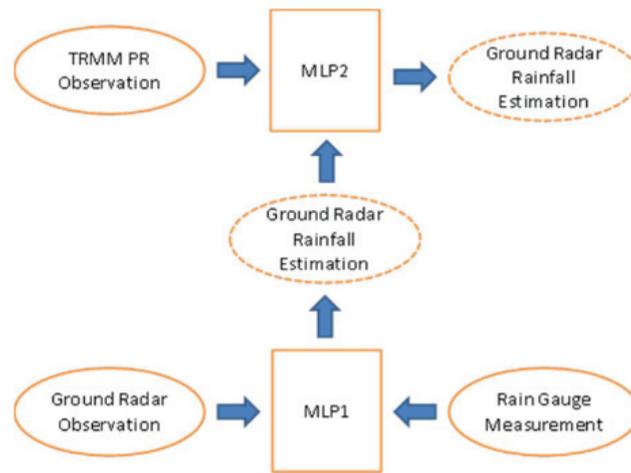


Figura 17: Diagrama conceitual do modelo de aprendizado de máquina baseado em MLP para estimativa de precipitação TRMM PR [8].

O produto de estimativa de precipitação derivado deste novo sistema é comparado com os produtos padrão TRMM, o que mostra a melhoria trazida pelo novo sistema de aprendizado de máquina.

3.12 MODELOS DE APRENDIZAGEM PROFUNDA PARA A PREVISÃO DE PRECIPITAÇÃO

O Deep Learning permitiu que os rótulos de dados de auto-aprendizagem permitissem criar um modelo orientado por dados para um conjunto de dados de séries temporais. Ele permite fazer a detecção de anomalias a partir dos dados da série temporal e também prevê os dados do evento futuro em relação aos eventos ocorridos no passado.

Aswin *et al.* [47], trata da obtenção de modelos de precipitação pluviométrica utilizando Arquiteturas de Aprendizagem Profunda (LSTM e ConvNet) e determina a melhor arquitetura, conforme figura 18, com RMSE de LSTM, afirmando que para qualquer conjunto de dados de séries temporais, modelos Deep Learning serão eficazes e eficientes para os modeladores.

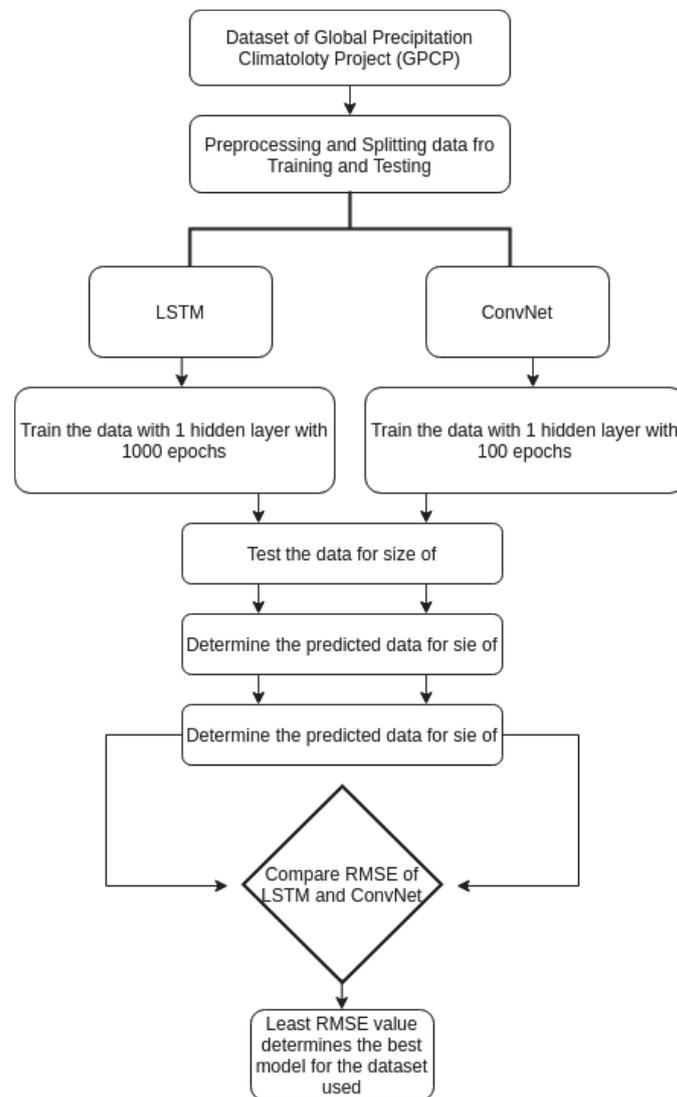


Figura 18: Arquitetura proposta [47].

De acordo com os resultados, aumentando o número de camadas ocultas, os erros de RMSE e MAPE ainda podem ser reduzidos para que os próximos dados previstos sejam confiáveis para fins meteorológicos.

3.13 COMPARAÇÃO DE QUATRO ALGORITMOS DE APRENDIZAGEM DE MÁQUINAS PARA A SUA APLICABILIDADE EM RECUPERAÇÕES DE PRECIPITAÇÃO ÓPTICA BASEADOS EM SATÉLITE

A metodologia, conforme figura 19, aplicada neste estudo, compara quatro algoritmos de aprendizado de máquina (RF, NNET, AVNNET e SVM) para detecção de área de chuva e taxa de precipitação usando dados MSG SEVIRI. Os proxies baseados em satélite para a altura superior da nuvem, a temperatura máxima da nuvem, a fase da nuvem e o caminho da água na nuvem servem como variáveis de previsão [61].

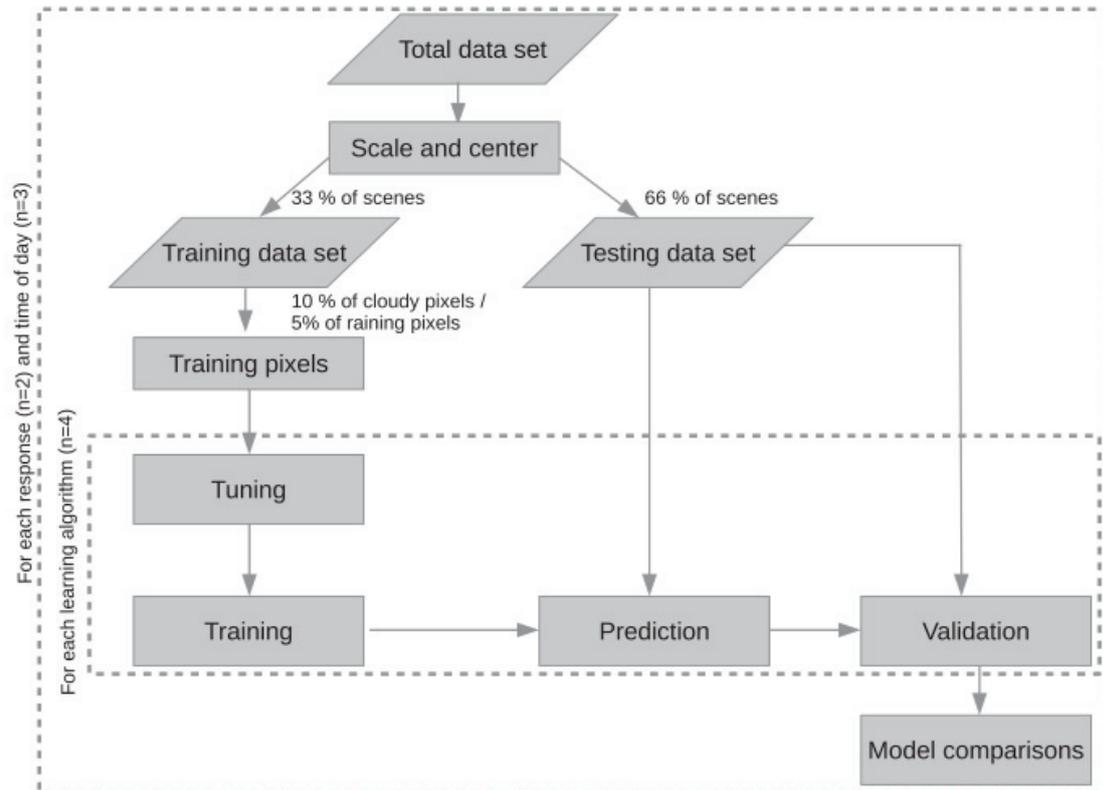


Figura 19: Fluxograma da metodologia aplicada [61].

Os resultados indicam uma superestimação do delineamento da área da chuva, independentemente do algoritmo ML (polarização média = 1,8), mas uma alta probabilidade de detecção variando de 81% (SVM) a 85% (NNET). Em uma base de 24 horas, o desempenho da atribuição da taxa de precipitação produziu valores de R2 entre 0,39 (SVM) e 0,44 (AVNNET). Embora as diferenças no desempenho dos algoritmos fossem bastante pequenas, a NNET e a AVNNET foram identificadas como os algoritmos mais adequados. Em média, eles demonstraram o melhor desempenho na delimitação da área de chuvas, bem como na atribuição de taxa de precipitação. A velocidade computacional da NNET é uma vantagem adicional no trabalho com grandes conjuntos de dados, como em recuperações de precipitação baseadas em sensoriamento remoto.

3.14 DISCUSSÕES

Aprendizado profundo é um subconjunto de Aprendizado de Máquina, mas especialistas utilizam a separação dos temas para uma melhor compreensão, pois suas capacidades são diferentes. Embora os modelos básicos de aprendizado de máquina se tornem progressivamente melhores em qualquer função, eles ainda precisam de alguma orientação. Se um algoritmo de IA retornar uma previsão imprecisa, o programador/engenheiro/cientista de dados precisará intervir e fazer ajustes. Com um modelo de aprendizado

profundo, um algoritmo pode determinar por si próprio se uma previsão é precisa ou não através de sua própria rede neural.

Geralmente, a predição de precipitação com condições diurnas mostrou resultados mais promissores do que as previsões da noite. Isso pode ser devido às informações que estão diretamente relacionadas aos processos de espalhamento solar que podem ser avaliados pelos canais Data of the Visible (VIS) e Near Infrared Radiation (NIR) [61] que não estão disponíveis no crepúsculo ou na noite.

O trabalho de Qiu [41], estudou como utilizar um conjunto de *features* de clima que foram coletadas de múltiplos locais de observação para predição de precipitação. A abordagem proposta é baseada em redes neurais e, aproveita a vantagem de redes neurais convolucionais para extrair automaticamente características profundas a partir de dados de observação de múltiplos locais. Extensivos testes mostraram que a abordagem proposta tem um desempenho significativamente melhor do que os métodos de base, incluindo vários modelos de previsão de chuva baseados em aprendizagem profunda de última geração.

Na predição da irradiação solar um ponto importante que é discutido entre alguns estudos é a questão do céu-claro [67]. Ele tem por objetivo a acurácia e estimativa da irradiância média em um local em um determinado momento na ausência de nuvens. Outro ponto discutido é em relação à métodos comuns de modelagem da irradiância do céu claro, pois são imprecisos ou exigem dados atmosféricos extensos, que tendem a variar de acordo com a localização.

A ciência de dados já está inserida em todos os meios que podemos imaginar. Ela incorpora métodos científicos, processos, algoritmos e sistemas para reunir conhecimento e trabalhar sobre o mesmo. Com modelos mais robustos e trabalhando com conceitos de inteligência artificial, temos as técnicas de aprendizado profundo e de máquina, na qual, se mostraram mais eficazes do que modelos matemáticos básicos. Modelos baseados em previsões de series temporais, como a LSTM, têm como premissa, aprender o contexto necessário, em vez de ter esse contexto pré-especificado e corrigido.

Desta forma, a escolha da técnica de ciência de dados a ser utilizada nesse projeto, se deu através desta revisão sistemática e mostrou o quanto uma revisão é importante para que o pesquisador possa conhecer todo o processo de descoberta, análise e escolha dos trabalhos. O projeto final fez uso do modelo LSTM para a predição de precipitação e irradiação solar, pelo fato de que as células LSTM têm memória que pode armazenar informações de intervalos de tempo anteriores e, dessa forma, aprendem a cada nova execução, pois utilizam dados de séries temporais. Também temos a retro-propagação constante de erros dentro das células de memória, a qual resulta na capacidade da LSTM de superar atrasos de tempo muito longos em caso de problemas.

A revisão sistemática proposta neste trabalho visou a busca de estudos que previssem a precipitação e a irradiação solar de forma combinada e que fizessem uso de técnicas para a ciência dos dados. Estudos seguindo este critério não foram encontrados e, por

este motivo, ambos os estudos foram utilizados separadamente. A avaliação das técnicas de precipitação e irradiação solar será utilizada para o desenvolvimento de uma abordagem onde o objetivo é identificar locais adequados para a implantação de um sistema de ultrafiltração de água de chuva.

4. MATERIAIS E MÉTODOS

Como a pesquisa aborda uma arquitetura que utiliza dados quantitativos de precipitação e irradiação solar, posteriormente, objetiva a análise dessas informações, pode-se classificar o trabalho descrito nesta proposta como uma pesquisa do tipo quantitativa. Já com relação a sua natureza, como neste trabalho a arquitetura proposta objetiva a resolução de um problema prático que é a identificação de um local ou locais adequados para a implantação de um sistema de ultrafiltração de água de chuva com o auxílio da deep learning, pode ser classificada como uma pesquisa do tipo aplicada.

Quanto aos objetivos da presente pesquisa, pode-se afirmar que se encaixam entre o tipo de pesquisa descritiva - já que envolve a coleta de dados meteorológicos - e o tipo de pesquisa explicativa, que visa a caracterização das causas dos fenômenos encontrados através da pesquisa descritiva. Com relação aos seus procedimentos, a presente pesquisa pode ser classificada como experimental, pois passa por uma etapa de validação dos dados, cálculos em relação a energia consumida por painéis fotovoltaicos e aplicação. Finalmente, a classificação final da pesquisa pode ser nominada como pesquisa quantitativa-aplicada-explicativa-experimental.

A metodologia utilizada na previsão de precipitação e irradiação solar faz uso de técnicas de *deep learning*, utilizando modelos de RNN. Um modelo de RNN usa informações sequenciais, como dados de registro de data e hora de um sensor ou variáveis como *features*. Essas informações são compostas por uma sequência de termos e, diferentemente das redes neurais tradicionais, as entradas de uma rede neural recorrente não são independentes umas das outras, e os resultados para cada elemento dependem da computação dos elementos precedentes. No caso do trabalho proposto, a ideia é utilizar uma LSTM para teste, treinamento e previsão.

Uma LSTM é um tipo de rede neural recorrente capaz de aprender a dependência da ordem de aprendizagem em problemas de previsão de sequência. Este é um comportamento necessário em domínios de problemas complexos como tradução automática, reconhecimento da fala, séries temporais e muito mais. Modelos lineares e outros modelos de deep learning poderão ser usados para estabelecer modelos de linha de base.

A figura 20 apresenta o fluxo metodológico no qual descreve os passos a serem seguidos.

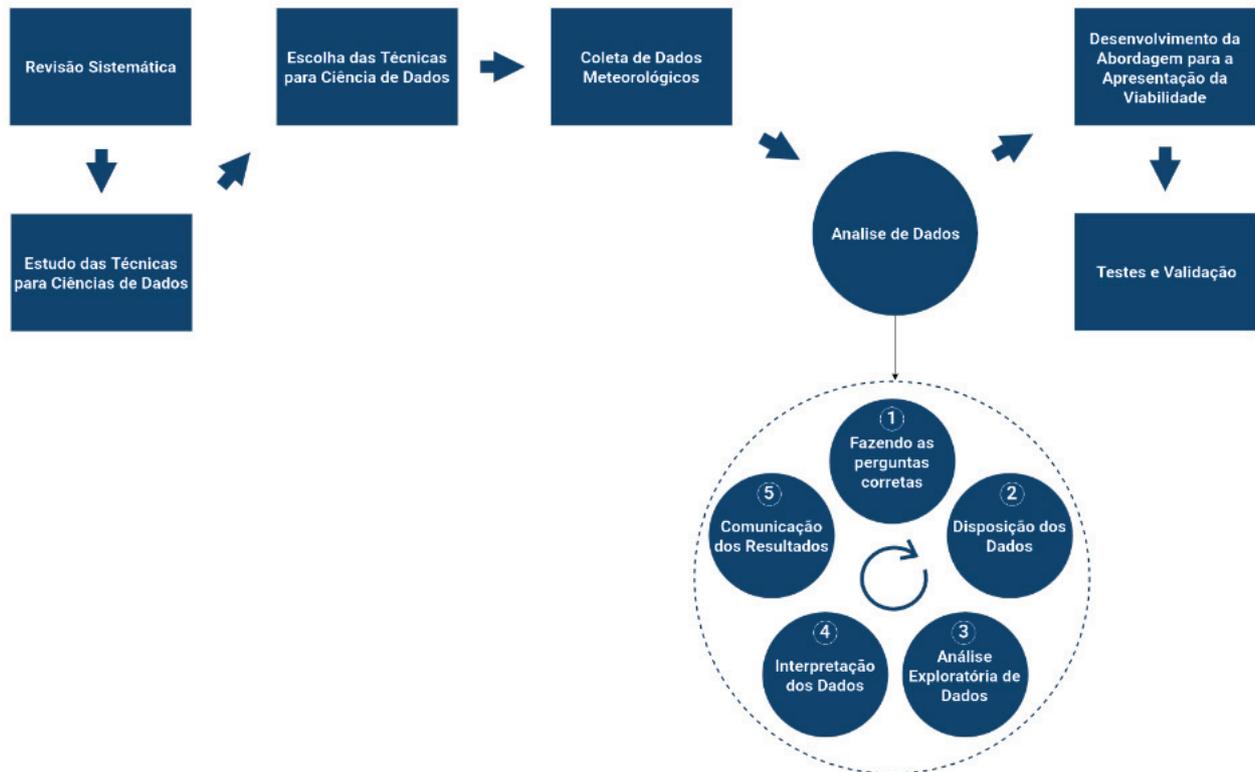


Figura 20: Fluxo metodológico [Autoria própria]

O ambiente de desenvolvimento, bem como de toda a análise de dados utilizou uma CPU Intel Core i5-9300H com 8 GB DDR4 2666MHz de memória RAM, uma GPU NVIDIA GeForce GTX 1650 com 4GB GDDR5 de memória RAM. O sistema operacional utilizado foi o Ubuntu 18.04 LTS com os softwares CUDA 10.1, CuDNN 7.1, PyCharm, python 3.8, keras e tensorflow.

4.1 ESTUDO E ESCOLHA DAS TÉCNICAS

O estudo deu início com uma revisão sistemática, onde, cumprido todas as etapas, foram escolhido 25 trabalhos com critérios pré-definidos na RS. Também teve como base a busca por estudos que fizessem uso de técnicas para ciência de dados como base para esse projeto.

As técnicas e ou modelos mais utilizadas nestes estudos foram divididas em *machine learning* e *deep learning* e, a escolha da RNN LSTM teve como pontos principais: ser um modelo com uma arquitetura desenhada para séries temporais, resultados satisfatórios de RMSE comparados com outros modelos de rede neural e machine learning, vasto conteúdo e conhecimentos básicos do modelo comparados com outros, o que pode ser observado em [4] e [66].

4.2 COLETA E ANÁLISE DE DADOS

A coleta de dados consiste em seguir uma série de etapas capaz de adquirir os dados em APIs, websites, banco de dados e posteriormente serem analisados utilizando dos processos que fazem parte das boas práticas da análise de dados. A análise de dados também conhecida como análise estatística e mineração de dados, busca o entendimento do processo de análise, pela capacidade de diferenciar entre essas diferentes atividades [70] [71].

Foram coletados os dados históricos do Instituto Nacional de Meteorologia - INMET [72] entre os anos de 2000 a janeiro de 2021, oriundos de 44 estações meteorológicas automáticas distribuídas no estado do Rio Grande do Sul. Algumas estações foram instaladas após o ano 2000 e, portanto, não possuem dados anteriores a data da sua instalação.

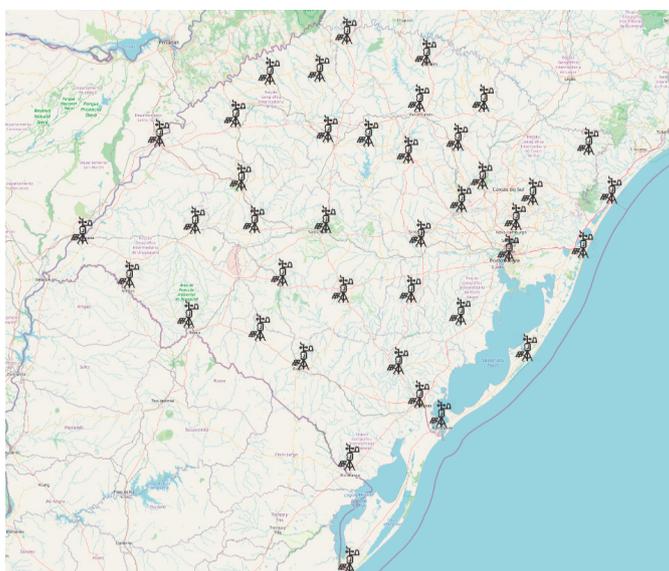


Figura 21: Estações meteorológicas automáticas do Estado do Rio Grande do Sul [Autoria própria].

Por estarem em um formato bruto ou primário, os dados necessitam de tratamento. Os dados coletados do INMET estão divididos por ano e em arquivos zip. Cada arquivo de acordo com o ano, possui os dados das estações em formato CSV.

O conhecimento do conjunto de dados é necessário antes de realizar qualquer análise, pois, primeiro, isto fornece uma breve compreensão e, segundo, impedirá interpretações errôneas ao decorrer da análise.

4.2.1 Limpeza dos dados

A limpeza dos dados significa o processo de identificar a parte incorreta, incompleta, inexata, irrelevante ou ausente dos dados e depois modificá-los, substituí-los ou

excluí-los de acordo com a necessidade. A limpeza de dados é considerada um elemento fundamental da ciência básica dos dados.

Para esse processo foi utilizado alguns scripts na linguagem python, o que torna o processo mais rápido e seguro. Toda a análise e limpeza dos dados é de extrema importância para o processo como um todo. Um pequeno erro no conjunto de dados pode lhe causar muitos problemas e todos os esforços podem ser desperdiçados.

4.3 DESENVOLVIMENTO DA APLICAÇÃO

Através de dados de estações meteorológicas, após toda a análise e tratamento, estes dados serão testados e treinados pelos modelos de *deep learning* e confrontados com modelos básicos de *machine learning*. Após encontrar o modelo hábil, será feita a previsão de irradiação solar e precipitação para 6 meses do ano, contando a partir do último mês de dados fornecidos pelo INMET.

A aplicação foi dividida em *frontend* e *backend*. O *frontend* nos traz as estações meteorológicas e a interação dos dados com cada estação. Ao clicar em cada ponto, dados a respeito da estação e também das previsões são informados na tela. O *backend* é o modelo de *deep learning* em si, no qual após a previsão dos dados meteorológicos, disponibilizará ao *frontend* um *json* com os dados a serem mostrados. Abaixo descrevo com mais detalhes o papel de cada um.

4.3.1 Frontend

Para o desenvolvimento do *frontend* foi utilizado o *framework* Vue.js e as bibliotecas vuetify, bootstrap, vuelayers, openlayers e leaflet. Este *framework* é projetado desde o início para ser adotado de forma crescente. A biblioteca principal é focada apenas na camada de visualização, e é fácil de integrar com outras bibliotecas ou projetos existentes [73].

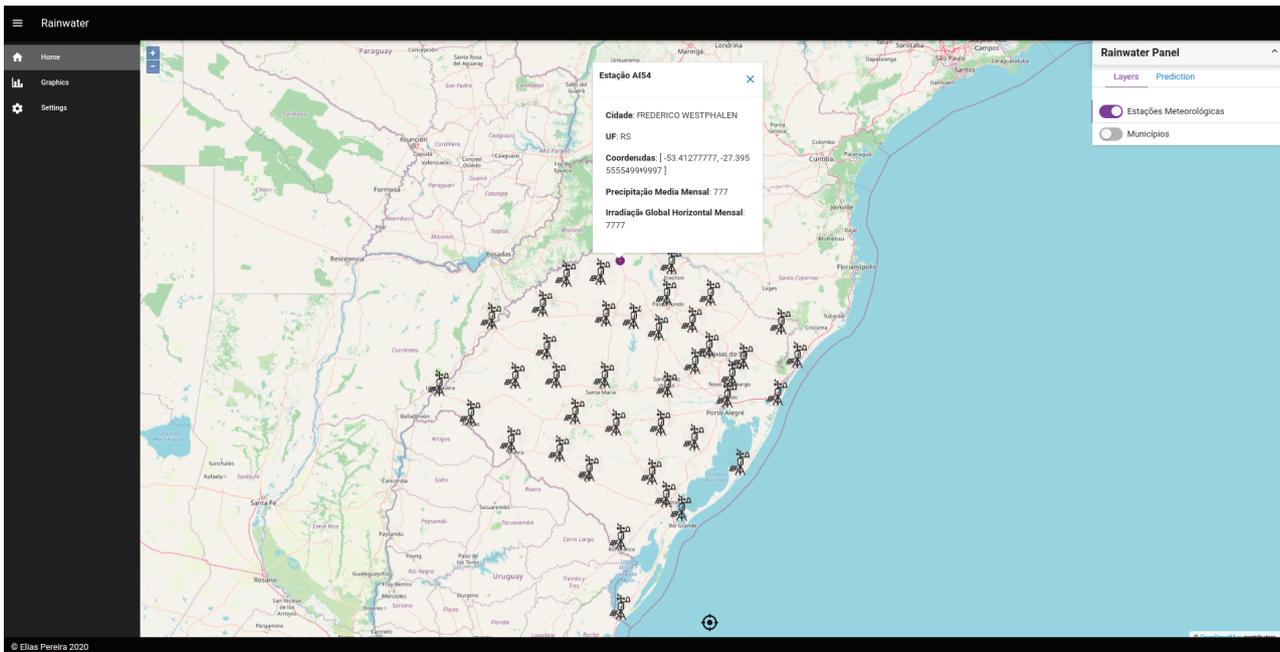


Figura 22: Dashboard do frontend [Autoria própria].

A estrutura do *frontend* é organizada como mostra a imagem 23.

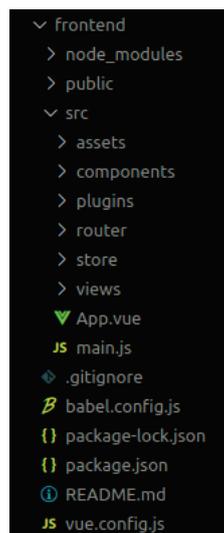


Figura 23: Estrutura do frontend [Autoria própria].

Segue definição dos principais componentes da estrutura.

- (a) *frontend*: pasta raiz para todo conteúdo referente ao *frontend*;
- (b) *node_modules*: pasta de instalação das ferramentas/bibliotecas de construção do projeto.
- (c) *public*: pasta que irá conter o projeto após a realização do build.
- (d) *src*: Engloba diretórios de arquivos estáticos (.css, .js, .png, etc), componentes (outros arquivos .vue), plugins, funções de rotas de acesso da aplicação. Componentes e

views atuam de forma parecida. A principal diferença é que alguns componentes do Vue atuam como visualizações para o roteamento. O roteamento no Vue consiste em definir as rotas a fim de alternar a visualização atual usada no componente `<router-view>`. Essas rotas estão normalmente localizadas em `src/router/routes.js`.

- (e) *App.vue*: é a raiz do seu aplicativo definida no formato de arquivo do componente Vue. Geralmente é algo que define o modelo da sua página: HTML, CSS e JavaScript.
- (f) *main.js*: é o arquivo JavaScript que inicializará esse componente raiz em um elemento da página. Ele também é responsável por configurar plugins e componentes de terceiros que você pode querer usar no aplicativo.
- (g) *package.json*: define quais bibliotecas serão instaladas em `node_modules` quando o comando `npm install` ou `npm i` for executado;
- (h) *vue.config.js*: é um arquivo de configuração opcional que será carregado automaticamente por `@vue/cli-service` se estiver presente na raiz do projeto (próximo a `package.json`). O arquivo deve exportar um objeto contendo opções.

Juntamente com o vue foi utilizado algumas bibliotecas javascript para o desenvolvimento do mapa. Com o auxílio da biblioteca `vuelayers` [74] foi desenvolvido os componentes que utilizam como base, os mapas fornecidos pelo `OpenStreetMap` [75].

4.3.2 Backend

Para o desenvolvimento do backend foi utilizado o framework `PyCharm` e a linguagem de programação `python` na versão 3.8. A instalação do `python` pode ser executada via pacotes de instalação de acordo com a distribuição linux que estiver usando. No caso, este projeto está sendo desenvolvido na distribuição `Ubuntu 18.04`.

A instalação dos pacotes pode ser feita de forma manual, um a um ou pode-se utilizar um arquivo chamado *requirements.txt*, onde se encontram todos os pacotes e suas versões. Esse arquivo é aceito e interpretado pela IDE, possivelmente por outras IDEs e, instalado automaticamente cada pacote configurado.

Exemplo de um arquivo *requirements.txt*

```
1 matplotlib~=3.2.1
2 scikit-learn~=0.22.2.post1
3 numpy~=1.18.4
4 pandas~=1.0.3
5 keras~=2.3.1
6 tensorflow~=2.2.0
```

Alguns modelos padrão de ML foram avaliados juntamente com o modelo de DP, a fim de que se tenha uma base de acurácia e uma melhor escolha de um modelo final. Nesse processo vários testes se fazem necessários para que um modelo seja considerado apto

para a previsão ou que seja considerado "skillful". Um modelo é considerado skillful, quando ele possui um desempenho melhor do que o desempenho de uma simples estratégia de previsão.

Basicamente os modelos possuem a mesma arquitetura. O que diferencia é a forma do treinamento e utilização das diferentes RNNs no processo de `build_model()`.

- (a) `split_dataset`: O dataset é dividido em meses para o teste e treinamento. Por exemplo, para 3 meses de previsão, dividimos o dataset de 3 em 3 meses. Nesse projeto, dividimos de 6 em 6 meses para a previsão de 6 meses no futuro;
- (b) `evaluate_forecasts`: A avaliação da previsão é um dos passos, senão, o passo mais importante de uma modelagem de previsão de séries temporais. Para isso, o modelo utiliza cálculos de MSE e RMSE para avaliar e apresentar os melhores resultados de acordo com os hiperparâmetros. Otimização de hiperparâmetros (às vezes chamada de busca, varredura ou ajuste de hiperparâmetros) é uma técnica para afinar um modelo para melhorar sua precisão final. Os hiperparâmetros comuns incluem o número de camadas ocultas, taxa de aprendizagem, função de ativação e número de épocas.
- (c) `to_supervised`: Converte o histórico de dados em inputs e outputs. Faz o train e reshape, passando ao longo de todo o histórico um passo de cada vez com a função de definir o fim da sequência de entrada e garantir que se tenha dados suficientes para o processo de construção/build modelo;
- (d) `build_model`: O módulo mais importante da RNN, o qual define a técnica e hiperparâmetros que será utilizado para chegar em um modelo ideal ou skillful. O problema de previsão é uma auto-regressão. Isso significa que provavelmente é melhor modelado onde os próximos seis meses possuem alguma função de observações em etapas de tempo anteriores. O número de unidades na camada oculta não está relacionado com o número de etapas de tempo nas sequências de entrada. A camada LSTM é seguida por uma camada totalmente conectada com nós que interpretará as características aprendidas pela camada LSTM. Finalmente, uma camada de saída preverá diretamente um vetor com seis elementos, um para cada mês na sequência de saída. A perda de RMSE funciona como uma boa combinação para a métrica de erro escolhida. Foi utilizada a eficiente implementação ADAM para as epochs/épocas e batch/lote. O tamanho pequeno do batch e a natureza estocástica do algoritmo significa que o mesmo modelo aprenderá um mapeamento ligeiramente diferente de entradas para saídas cada vez que for treinado. Isto significa que os resultados podem variar quando o modelo é avaliado. Pode-se executar o modelo várias vezes e calcular uma média de desempenho. O modelo de construção prepara o treinamento de dados, define o modelo e encaixa no modelo os dados de treinamento, devolvendo o modelo de ajuste pronto para fazer previsões;

- (e) `forecast`: Faz a recuperação dos últimos seis meses de dados de água da chuva e irradiação solar (índice de características 0). Parametriza os dados de treinamento para que o número de meses anteriores usados como entrada pelo modelo possa ser modificado no futuro. Em seguida, reformula a entrada na estrutura tridimensional esperada $[1, n_input, n]$. A função `forecast()` implementa isso e leva como argumentos o ajuste do modelo no conjunto de dados de treinamento, o histórico dos dados observados e o número de etapas de tempo de entrada esperadas pelo modelo.
- (f) `evaluate_model`: Na avaliação do modelo, é utilizado a validação walk-forward. Isto significa que temos as observações disponíveis para o mês anterior, a fim de que seja feito a previsão para o próximo mês. Estas são coletadas em uma série de meses padrão chamadas de *history*. Para prever o próximo mês padrão, precisa-se recuperar os últimos meses de observações. Com os dados de treinamento, deve-se primeiro nivelar os dados históricos para modelar os dados em séries temporais paralelas.

O diagrama na figura 24 exemplifica a estrutura da RNN LSTM utilizada neste trabalho.

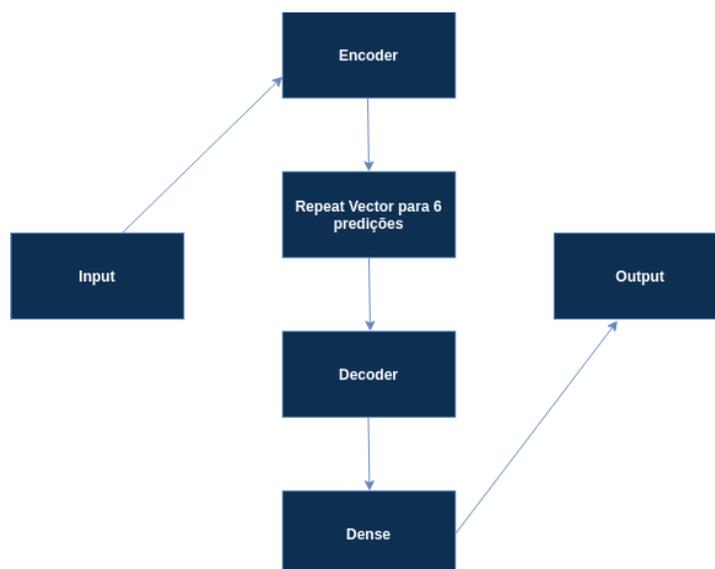


Figura 24: Diagrama de uma RNN LSTM [Autoria própria].

4.4 TESTES E VALIDAÇÃO

Durante o desenvolvimento do modelo de deep learning o qual faz o papel do backend da aplicação, várias etapas, testes e validações foram aplicadas. O *lifecycle* desse tipo de arquitetura necessita que esses passos sejam executados a cada nova modelagem ou mudanças em hiperparâmetros. Normalmente uma previsão em várias etapas é avaliada de duas maneiras, uma é o RMSE geral para todos os períodos de previsão e outra é o

RMSE para cada etapa de tempo de previsão. Para esse último método, é comum que se tenha mais erros em média quanto mais longe no futuro estivermos fazendo previsões.

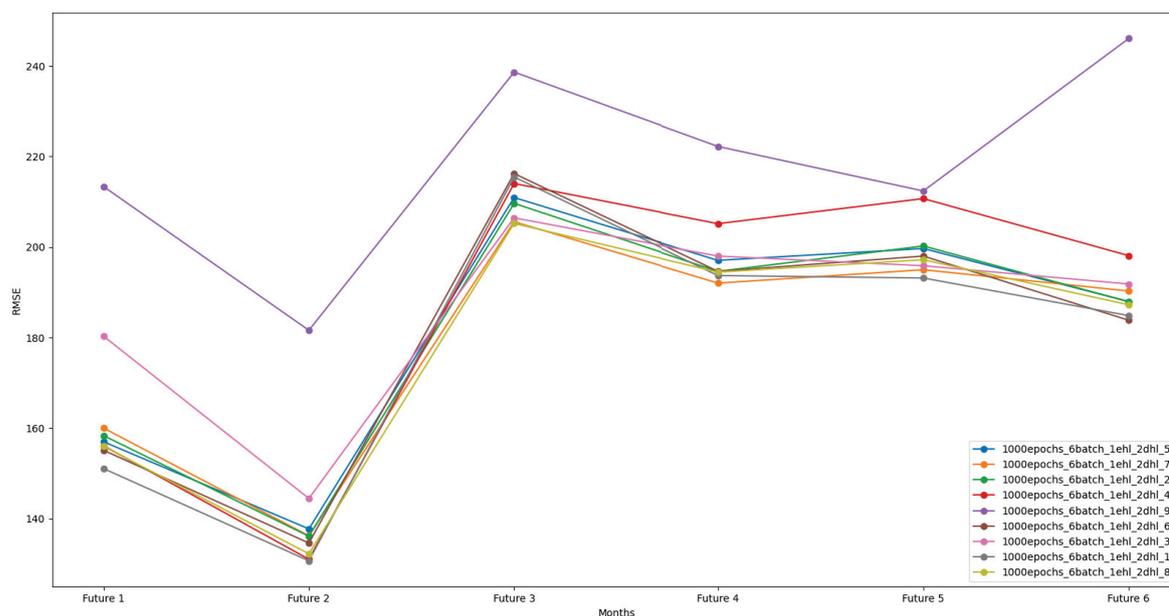


Figura 25: Testes de hiperparâmetros em relação à irradiação solar [Autoria própria].

A figura 25 possui testes iniciais e com valores de hiperparâmetros pequenos, no qual é ajustado ao longo do tempo. Devido a natureza estocástica do algoritmo, os resultados específicos podem variar, mesmo se utilizado os mesmos valores. Isto significa que, dada a mesma configuração de modelo e o mesmo conjunto de dados de formação, um conjunto de pesos internos diferentes resultará cada vez que o modelo formado e que terá um desempenho diferente. Isto é um benefício, permitindo que o modelo seja adaptável e encontre configurações de alto desempenho para problemas complexos. É também um problema ao avaliar o desempenho de um modelo e ao escolher um modelo final a utilizar para fazer previsões.

4.4.1 Walk-Forward

A configuração de um modelo pode ser avaliada através da validação walk-forward, na qual reporta o erro médio ao longo de cada avaliação. A validação walk-forward é um dos meios de avaliar os modelos usados para previsão no momento da sua execução e de contornar problemas de overfitting e underfitting. O objetivo é dividir o conjunto de dados em dois grupos (train e test). Uma diferença quadrática média simples entre os valores observados e previstos fornece uma medida para a precisão da previsão.

4.4.2 Grid Search

Essa técnica é utilizada para testar todas as combinações possíveis dos hiperparâmetros, exaustivamente. Basicamente, irá fornecer alguns valores de input e testar todas as combinações plotando em um plano cartesiano (por isso o nome de grid). Em seguida, selecionará os hiperparâmetros que obtiveram o menor erro.

Neste trabalho os arquivos CSV de cada estação meteorológica foram reformata-dos para dados mensais, ficando muita vezes, com pouco mais de 150 entradas para cada arquivo. Isso se deve ao fato de que a predição será feita para 6 meses, nos fornecendo um panorama a médio e longo prazo da quantidade de precipitação e irradiação solar que estará disponível para o sistema de ultrafiltração.

Para a execução do Grid Search foi utilizado o dataset `PASSO_FUNDO_A839_clean.csv` com a hiperparâmetrização de ambas as variáveis. Os resultados obtidos através do algoritmo de Grid Search nos permite modelar de forma mais ágil o modelo de deep learning.

`n_input`: O número de entradas anteriores a utilizar como entrada para o modelo (por exemplo, 6 ou 12 meses).

`n_nodes`: O número de nós a utilizar na camada oculta (ex. 50).

`n_epochs`: O número de épocas de treino (ex. 1000).

`n_batch`: O número de amostras a incluir em cada mini-lote (ex. 32).

`n_diff`: A ordem da diferença (ex. 0 ou 12).

`relu activation`: A função de ativação linear retificada, ou função de ativação ReLU, é talvez a função mais comum utilizada para camadas ocultas. A função ReLU é calculada com a formula $\max(0,0,x)$

Isto significa que se o valor de entrada (x) for negativo, então um valor 0,0 é retornado, caso contrário, o valor é retornado.

`sigmoid activation`: A função de ativação sigmóide também é chamada de função logística. É a mesma função utilizada no algoritmo de classificação da regressão logística.

A função de ativação sigmóide é calculada com a formula $1.0/(1.0 + e^{-x})$, onde e é uma constante matemática, que é a base do logaritmo natural.

`tanh activation`: A função de ativação tangente hiperbólica também é chamada simplesmente de função Tanh (também "tanh" e "TanH"). Ela é muito semelhante à função de ativação sigmóide e tem até a mesma forma em S.

A função de ativação Tanh é calculada com a formula $(e^x - e^{-x})/(e^x + e^{-x})$, onde e é uma constante matemática que é a base do logaritmo natural.

4.4.3 Resample e Cálculo das Variáveis Meteorológicas

Os dados que o INMET disponibiliza para a precipitação estão em (*mm*). O cálculo para essa variável deve ser a soma horária para um *resample* diário e da mesma forma para um *resample* mensal.

Para a irradiação solar o cálculo é diferente, pois os dados das estações automáticas do INMET estão em KJ/m^2 acumulados por hora, portanto é preciso somar o total diário e dividir por 3,6 para obter o resultado em Wh/m^2 ($1Wh = 3600J$).

4.4.4 Testes e Validação dos Hiperparâmetros

A escolha dos hiperparâmetros é baseada nas diretrizes do *Grid Search* e parametrização manual. Na parametrização manual usando o conhecimento prévio sobre o problema, faz-se "previsões" dos parâmetros e observação do resultado. Com base nesse resultado, ajusta-se os parâmetros. Esse processo é feito até encontrar parâmetros que demonstrem que o modelo é *skillful*. Cada processo é uma execução completa do aplicativo de treinamento com valores para a seleção de hiperparâmetros,

Hiperparâmetros utilizados para as variáveis:

```
1 n_input = 6
2 n_epochs = 130
3 n_neurons = 250
4 n_batch = train_size
```

4.4.5 Eficiência Energética de Painéis Fotovoltaicos X Consumo de Água por Habitante

Este trabalho é fruto de uma colaboração interdisciplinar entre o PPGCTA na utilização de um sistema de ultrafiltração de água de chuva com utilização de painéis fotovoltaicos e do PPGCA com uma aplicação na qual através do uso de *deep learning* irá identificar e apresentar a viabilidade de um local para a implantação deste sistema.

A eficiência energética dos painéis fotovoltaicos no contexto deste trabalho para as predições de irradiação solar e precipitação de água de chuva no período de 6 meses se entrelaçam no momento que o montante de irradiação predita seja suficiente para que o sistema seja autossuficiente e possa abastecer, por exemplo, uma família de 5 pessoas.

Estima-se que no Brasil, de acordo com o levantamento de Contas Econômicas Ambientais da Água (CEAA) [76], o consumo de água por habitante fique em média 116 L/dia. Com essa estimativa, através do cálculo abaixo, pode-se estimar o volume de água que uma família de 5 pessoas utiliza em um mês.

V = Volume; P = Pessoa; L = Litros; d = dia; M = Mês

$$V = 116 \text{ L/P/d} \times 5 \text{ P} = 580 \text{ L/d} \times 30 \text{ d/M} = 17.4 \text{ m}^3/\text{M} \quad (1)$$

Estima-se a energia consumida para tratar 1 m^3 de água através do consumo mensal de energia: 27 kWh/mês. Para produzir 1 m^3 de água é necessário 1.36 kWh de energia total. Os valores acima são dados experimentais produzidos por Baú *et al.* [77].

Com a irradiação solar para o mês de novembro é estimada a energia gerada e a área de painéis necessária para a auto-suficiência energética.

Três cidades (Passo Fundo, Uruguiana e Rio Grande) foram escolhidas pela distribuição das estações meteorológicas no contexto das médias de irradiação global horizontal e normal climatológica de precipitação anual analisando os dados de Labren [29]. Cada uma está em uma região com diferentes médias de irradiação e diferentes padrões de precipitação em seu território.

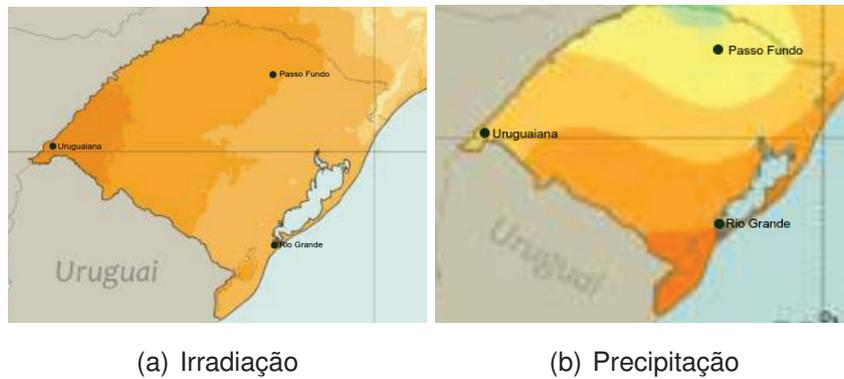


Figura 26: Médias de irradiação global horizontal e normal climatológica de precipitação anual. Adaptada de [29]

Tabela 6: Resultado dos cálculos para as cidades escolhidas

	PRECIPITAÇÃO PREDITA	IRRADIAÇÃO SOLAR PREDITA	QUANT. DE PAINEIS SOLARES
PASSO FUNDO	136 mm	5,8 kWh/m ²	5
URUGUAIANA	73 mm	6,2 kWh/m ²	5
RIO GRANDE	96 mm	4,9 kWh/m ²	4

4.4.5.1 Normalização e Combinação Média

Os cálculos básicos acima nos mostra de uma maneira mais pratica de como a previsão destas variáveis pode auxiliar na tomada de decisão para a instalação do sistema de ultrafiltração. O método de normalização média mostrou-se mais eficaz para verificar o local mais adequado na a instalação do sistema de ultrafiltração, no qual, utiliza a soma ponderada nos dados horários e diários para posteriormente normalizá-los, pois estes dados estão na ordem de milhares e há diferentes unidades (kJ/m² para W/m²) para a comparação em tamanho. A normalização utiliza o valor atual e divide pelo valor máximo para essa

faixa em particular. Foi utilizado três pesos, sendo 0,7, 1,0 e 1,3, variando apenas a importância de cada um na contribuição do resultado. O peso 0,7 significa que a precipitação recebe 30% a menos de peso e há mais contribuição da irradiação. O peso 1.0 é quando tem uma contribuição igual para ambas as variáveis. Com peso de 1,3 a precipitação tem mais valor na contribuição dos resultados.

A formula aplicada é a que segue, onde w é o peso.

$$w \times \frac{prec}{prec_{max}} + \frac{rad}{rad_{max}} \quad (2)$$

5. ANÁLISE E DISCUSSÃO DOS RESULTADOS

Um dos objetivos deste trabalho é mostrar que a DL pode ser utilizada para a tomada de decisão e escolha de locais propícios para a instalação de um sistema de UF de água de chuva. Dentro desse contexto, é importante avaliar através de cálculos de pontuação de erros se o modelo é eficaz ou skillful de acordo com o dataset utilizado. Em problemas de regressão não é utilizada a acurácia para a avaliação, pois é uma medida de problemas de classificação.

De acordo com o que foi apresentado no Capítulo 4, para a previsão de precipitação e irradiação solar foram utilizados dados de 44 estações meteorológicas espalhadas pelo Estado do Rio Grande do Sul. Esses dados foram disponibilizados pelo INMET para download diretamente do seu site. Os dados foram tratados e preparados para a rede neural de acordo com algumas características oriundas de modelos que utilizam dados em séries temporais. De acordo com o modelo final escolhido, os dados foram preparados para garantir uma predição de 6 meses, no qual o dataset foi dividido em janelas de 6 meses para ambos treino e teste. Dado a quantidade pequena de entradas para cada dataset, a quantidade de dados para o teste é baseado nas janelas de 6 meses. O restante dos dados é utilizado para o treino. O modelo LSTM teve um RMSE médio para a precipitação de 42,53 e 0,45 para irradiação solar. Testes e comparações foram realizados com modelos lineares e modelos básicos de ML. Os modelos lineares utilizados foram: Linear Regression, Lasso Lars, Passive Agressive Regression, RANSAC Regressor e SGD Regressor. Para o ML foi utilizado um modelo de Random Forest.

A aplicação web foi desenvolvida para apresentar as previsões, bem como dados das estações meteorológicas, tais como: nome, ID, latitude e longitude.

Cálculos de eficiência energética de painéis fotovoltaicos confrontados com consumo de água por habitante nos mostram que a a DL é capaz de nos apontar um local ou locais em que o sistema de ultrafiltração pode ser instalado. Este sistema depende de alguns números mínimos para o seu funcionamento. Por exemplo os cálculos de Passo Fundo nos mostram que, para uma família de 5 pessoas ser auto-suficiente em energia elétrica, necessita de uma instalação de 5 painéis fotovoltaicos em uma área de $10 m^2$. Da mesma forma, seria auto-suficiente em água potável com uma área de captação de $128 m^2$, sabendo que o tamanho médio de casas no Brasil gira em torno de 70 a $150 m^2$ de construção, para metragens de captação abaixo desse numero, seria necessário outras formas de coleta a água da chuva.

A sumarização e combinação média dos resultados nos mostra os melhores locais para a implantação do sistema de ultrafiltração de água de chuva. Tais valores são a normalização e combinação média dos pesos para cada estação meteorológica. Esta relação

entre as variáveis nos mostra os principais locais onde é possível a instalação do sistema de ultrafiltração. Por exemplo, os 5 melhores locais L9, L27, L18, L11 e L3 correspondem respectivamente as cidades: São Luiz Gonzaga, Uruguaina, Tupanciretã, Cruz Alta e Santo Augusto.

Outro ponto importante após a normalização e combinação média é verificar os números reais dos locais adequados para ver se possui precipitação e irradiação suficientes para que o sistema de ultrafiltração possa filtrar essa água de chuva. Pode ocorrer de algum local menos adequado possuir números maiores de precipitação e não obter uma irradiação suficiente para gerar energia para a ultrafiltração. Da mesma forma pode haver valores maiores para a irradiação e não ter muita precipitação ficando sem energia para a ultrafiltração.

Os valores reais de precipitação e irradiação solar para os locais adequados estão na tabela 7 abaixo.

Tabela 7: Valores máximos de precipitação e irradiação solar dos locais adequados no período verificado x média combinada.

	Local Adequado	Precipitacao (mm)	Irradiacao Solar (kJ/m2)	média Combinada		
				w_0_7	w_1_0	w_1_3
L9	São Luis Gonzaga	144,2	35066,5	1,7	2,0	2,3
L27	Uruguiana	215,0	35377,7	1,6261497527014426	1,8947207326687945	2,1632917126361466
L18	Tupanciretã	86,0	34523,1	1,571426232558008	1,836228223092038	2,1010302136260677
L11	Cruz Alta	120,8	78276,8	1,563942139596683	1,840313674810536	2,116685210024389
L3	Santo Augusto	134,4	34021,0	1,554678606821065	1,8386097425581605	2,122540878295256

Pode-se verificar que a cidade de São Luis Gonzada possui ambas precipitação e irradiação menores que Uruguiana e está como o melhor local. O valor real corresponde ao valor máximo de cada variável para todo o período de dados de cada estação meteorológica e, São Luis Gonzaga é o melhor local, pois é resultado das médias combinadas aplicadas para cada estação, a qual possui uma distribuição mais uniforme de precipitação e irradiação solar para o período de dados.

Da mesma forma a análise nos mostra que podemos intercalar entre a captação da água em dias que a precipitação é maior e utilizar o sistema de ultrafiltração em dias que a irradiação está maior, tornando o sistema autônomo. Outro ponto interessante de analisarmos é em relação a região Sul, a qual foram aplicados esses cálculos, bem como a previsão com o modelo de DL. Essa região é historicamente mais propensa a chuvas e, com uma boa média de irradiação. Desta forma a variável em destaque seria a irradiação, pois teríamos que ter mais produção de energia através dos painéis fotovoltaicos para que o sistema pudesse ultrafiltrar essa água de chuva.

5.1 APRESENTAÇÃO DOS RESULTADOS

As tabelas 8 e 9 indicam os melhores hiperparâmetros encontrados para as variáveis precipitação e irradiação solar.

Tabela 8: Resultados do Grid Search para a precipitação

LSTM		SIGMOID		
n_input	n_nodes	n_epochs	n_batch	rmse
6	250	130	train_size	42,53
LSTM		TANH		
n_input	n_nodes	n_epochs	n_batch	rmse
6	250	130	train_size	53,35
LSTM		RELU		
n_input	n_nodes	n_epochs	n_batch	rmse
6	250	130	train_size	78,77

Tabela 9: Resultados do Grid Search para a irradiação

LSTM		SIGMOID		
n_input	n_nodes	n_epochs	n_batch	rmse
6	250	130	train_size	0,59
LSTM		TANH		
n_input	n_nodes	n_epochs	n_batch	rmse
6	250	130	train_size	0,48
LSTM		RELU		
n_input	n_nodes	n_epochs	n_batch	rmse
6	250	130	train_size	0,45

Para o batch size foi utilizado o algoritmo de aprendizagem BGD - Batch Gradient Descent, o qual utiliza do total de entradas de treinamento para execução. É normalmente utilizado quando os datasets ou samples possuem poucas entradas.

Os modelos lineares (tabela 10) e de ML (tabela 11) apresentam resultados inferiores ao encontrado pelos modelos de DL (tabela 12). O modelo LSTM, o qual foi o modelo final utilizado neste trabalho, teve o melhor resultado. Para todos os modelos foi utilizado o dataset da cidade de Passo Fundo, o qual serviu de matriz para a modelagem e hiperparametrização do modelo final.

Tabela 10: Resultados de RMSE para os modelos lineares

Precipitação		RMSE MENSAL						RMSE Médio ¹
Modelo	Fevereiro	Março	Abril	Mai	Junho	Julho		
Linear Regression	64,0	85,5	106,7	73,5	102,0	95,0	[89,095]	
Lasso Lars	65,1	85,5	105,9	74,5	103,9	90,6	[88,803]	
Passive Agressive Regression	56,7	88,0	106,9	78,5	103,8	89,8	[88,856]	
RANSAC Regressor	69,1	90,7	138,0	92,3	161,6	122,0	[116,5]	
SGD Regressor	60,3	84,6	105,5	73,3	99,0	91,3	[87,0]	

Irradiação		RMSE MENSAL						RMSE Médio
Modelo	Fevereiro	Março	Abril	Mai	Junho	Julho		
Linear Regression	1,0	1,2	1,6	2,7	2,9	2,7	[2,163]	
Lasso Lars	1,1	0,5	0,6	1,6	1,8	1,6	[1,304]	
Passive Agressive Regression	1,4	1,5	1,4	3,1	2,5	3,3	[2,345]	
RANSAC Regressor	0,9	1,2	1,9	3,0	3,2	3,0	[2,406]	
SGD Regressor	0,8	1,0	1,1	2,3	2,5	2,4	[1,826]	

Dentre os modelos lineares, analisando ambas as variáveis, o Lasso Lars teve um desempenho superior que os demais. O RMSE para a precipitação ficou com uma média de 88,3 e a irradiação com uma média de 1,3. Analisando as variáveis em separado, notamos que para a precipitação, o melhor resultado foi para o modelo SGD Regressor. O ideal é que os valores de RMSE sejam o mais próximo de zero, o qual pode indicar problemas de *overfitting* e *underfitting* que podem ser contornados com a validação de dados walk-forward. Esse tipo de modelagem apresenta alguns parâmetros iniciais e apontamentos para o modelo skilfull.

O modelo de machine Learning teve um desempenho superior aos modelos lineares. Esse tipo de abordagem é essencial na modelagem e hiperparametrização de um modelo skilfull.

Tabela 11: Resultados de RMSE para o modelo de ML Random Forest

Precipitação		RMSE Mensal						RMSE Médio
Modelo	Fevereiro	Marco	Abril	Mai	Junho	Julho		
Random Forest	87.4	39.7	29.5	6.2	13.3	112.0	[61.7]	

Irradiação		RMSE Mensal						RMSE Médio
Modelo	Fevereiro	Marco	Abril	Mai	Junho	Julho		
Random Forest	1.2	0.4	1.1	0.9	0.0	0.6	[0.8]	

O modelo de DL utilizado nesse trabalho teve seu RMSE de 42,53 para precipitação e 0,45 para irradiação solar. Sendo superior a outros modelos de DL, como a MLP e ConvNet, como mostra a tabela 12.

¹Os valores de RMSE Médio correspondem ao total de repetições configuradas para os modelos. Neste caso, 30 repetições. Válidos para todas as tabelas que possuem RMSE Médio.

Tabela 12: Resultados de RMSE para os modelo de DL

Precipitação		RMSE Mensal					RMSE Médio
Modelo	Fevereiro	Marco	Abril	Maio	Junho	Julho	
LSTM	25,9	41,1	3,4	5,9	3,4	84,1	[42,53]
Irradiação		RMSE Mensal					RMSE Médio
Modelo	Fevereiro	Marco	Abril	Maio	Junho	Julho	
LSTM	0,3	0,3	0,8	0,0	0,1	0,5	[0,45]
Precipitação		RMSE Mensal					RMSE Médio
Modelo	Fevereiro	Marco	Abril	Maio	Junho	Julho	
MLP	150,3	76,1	89,2	1,5	2,7	26,3	[56,25]
Irradiação		RMSE Mensal					RMSE Médio
Modelo	Fevereiro	Marco	Abril	Maio	Junho	Julho	
MLP	0,4	0,7	1,1	0,3	0,3	0,8	[0,61]
Precipitação		RMSE Mensal					RMSE Médio
Modelo	Fevereiro	Marco	Abril	Maio	Junho	Julho	
CONVNET	36,3	31,7	31,3	35,8	35,4	82,6	[44,74]
Irradiação		RMSE Mensal					RMSE Médio
Modelo	Fevereiro	Marco	Abril	Maio	Junho	Julho	
CONVNET	0,7	0,1	0,7	0,1	0,6	0,6	[0,52]

Os resultados da ponderação ou combinação média mostram algumas poucas diferenças na escolha dos melhores locais para cada peso atribuído. Analisando a tabela 13, os locais L9 e L27 continuam sendo os melhores locais, independente de qual peso seja utilizado. Cabendo ao L18, L11 e L3 a troca de posição de acordo com o peso atribuído. Vale salientar que a variação percentual entre as médias é baixíssima, ficando em torno de 0.3844%.

Tabela 13: Resultados da sumarização das estatísticas descritivas com os melhores e piores locais

MELHORES LOCAIS COM BASE NO PESO W = 0,7				PIORES LOCAIS COM BASE NO PESO W = 0,7			
locations	w_0_7	w_1_0	w_1_3	locations	w_0_7	w_1_0	w_1_3
L9	1,7	2,0	2,3	L34	1,1994308749912879	1,3585717298327924	1,517712584674297
L27	1,6261497527014426	1,8947207326687945	2,1632917126361466	L23	1,2149169373408477	1,4261367826298952	1,6373566279189427
L18	1,571426232558008	1,836228223092038	2,1010302136260677	L29	1,2656922361221887	1,4434271937718635	1,6211621514215384
L11	1,563942139596683	1,840313674810536	2,116685210024389	L41	1,2678020229944182	1,4414005257526945	1,6149990285109708
L3	1,554678606821065	1,8386097425581605	2,122540878295256	L42	1,2855085808084223	1,459100924742526	1,6326932686766291

MELHORES LOCAIS COM BASE NO PESO W = 1,0				PIORES LOCAIS COM BASE NO PESO W = 1,0			
locations	w_0_7	w_1_0	w_1_3	locations	w_0_7	w_1_0	w_1_3
L9	1,7	2,0	2,3	L34	1,1994308749912879	1,3585717298327924	1,517712584674297
L27	1,6261497527014426	1,8947207326687945	2,1632917126361466	L23	1,2149169373408477	1,4261367826298952	1,6373566279189427
L11	1,563942139596683	1,840313674810536	2,116685210024389	L41	1,2678020229944182	1,4414005257526945	1,6149990285109708
L3	1,554678606821065	1,8386097425581605	2,122540878295256	L29	1,2656922361221887	1,4434271937718635	1,6211621514215384
L18	1,571426232558008	1,836228223092038	2,1010302136260677	L14	1,287783922600224	1,4532662978194015	1,6187486730385792

MELHORES LOCAIS COM BASE NO PESO W = 1,3				PIORES LOCAIS COM BASE NO PESO W = 1,3			
locations	w_0_7	w_1_0	w_1_3	locations	w_0_7	w_1_0	w_1_3
L9	1,7	2,0	2,3	L34	1,1994308749912879	1,3585717298327924	1,517712584674297
L27	1,6261497527014426	1,8947207326687945	2,1632917126361466	L41	1,2678020229944182	1,4414005257526945	1,6149990285109708
L3	1,554678606821065	1,8386097425581605	2,122540878295256	L14	1,287783922600224	1,4532662978194015	1,6187486730385792
L11	1,563942139596683	1,840313674810536	2,116685210024389	L29	1,2656922361221887	1,4434271937718635	1,6211621514215384
L18	1,571426232558008	1,836228223092038	2,1010302136260677	L42	1,2855085808084223	1,459100924742526	1,6326932686766291

L9	SAO LUIZ GONZAGA	L34	CAMAQUA
L27	URUGUAIANA	L23	CAMPO BOM
L18	TUPANCIRETA	L29	PORTO ALEGRE
L11	CRUZ ALTA	L41	CANGUCU
L3	SANTO AUGUSTO	L42	RIO GRANDE
		L14	SAO JOSE DOS AUSENTES

Em face dos resultados apresentados, obteve-se um modelo de DL skillful, o qual resultou em um bom RMSE, considerando os outros modelos testados. Corroborando com a previsão através do modelo escolhido, a ponderação ou combinação média teve um papel fundamental na escolha de um local para a instalação de um sistema de UF. Além destes pontos positivos, podemos destacar também que o modelo proposto pode ser utilizado para a previsão de qualquer outro local em território Brasileiro, dependendo apenas de alguns ajustes nas rotinas de análise e limpeza dos dados.

O ganho em termos de eficiência na escolha de um local adequado, bem como a certeza de dados mais confiáveis para o dimensionamento correto do sistema fotovoltaico, também são contribuições deste trabalho.

5.2 CONCLUSÃO E TRABALHOS FUTUROS

Na previsão da precipitação e irradiação solar, nota-se o massivo emprego de redes neurais e suas técnicas - ML, DL - aliadas ao poder computacional, apontam para uma quebra de paradigma onde a análise dos dados não mais se dá com padrões e técnicas manuais, mas sim com o auxílio das tecnologias e inovações que áreas da tecnologia disponibilizam. Não distante, os modelos de DL vivenciaram um grande desenvolvimento devido avanços em sua concepção e no hardware necessário para sua computação.

Neste contexto, nos últimos anos a energia solar ganhou uma importância significativa como fonte de energia renovável, limpa e alternativa. Não distante, o uso da energia solar vem crescendo a cada dia e necessitando cada vez mais de soluções que deixe seu uso mais fácil e ao mesmo tempo entregue um dimensionamento através de dados precisos. Em contrapartida, a instalação de um sistema de ultrafiltração de água de chuva utilizando essa energia solar, depende muito dos padrões climáticos.

O trabalho desenvolvido proporciona uma aplicação utilizando DL para auxílio na identificação de um local adequado para implantação de um sistema de ultrafiltração de água de chuva. Esta metodologia para previsão de precipitação e irradiação solar utilizou diferentes modelos lineares, ML e DL com a finalidade de determinar uma linha base de aferição para a escolha do modelo final de DL. Embora os modelos de DL tenham mostrado muito pouca diferença em seu desempenho, o LSTM provou ser o modelo skillfull, ou seja, o modelo mais adequado. Em média, ele obteve o melhor desempenho de RMSE médio para precipitação com um valor de 42,53 e 0,45 para irradiação solar.

Uma das maiores contribuições deste trabalho é o fato de ter utilizado as duas variáveis em conjunto. Na revisão da literatura, foram encontrados trabalhos em que os autores trabalhavam individualmente com as variáveis. O que possibilitou essa abordagem, foram ambos cálculos de ponderação ou combinação média, confrontados com a previsão mensal de precipitação e irradiação utilizando o modelo final. A vantagem dessa abordagem em relação a apenas analisar as médias históricas de precipitação e irradiação solar se justifica pelo fato de que um modelo LSTM são capazes de "lembrar" e usar informação sobre inputs anteriores, assim essencialmente é um modelo de séries temporais não lineares, onde a não linearidade é aprendida com os dados.

O presente trabalho também possibilitou a produção de um artigo científico submetido a revista de referencia na área de Inteligencia Artificial Aplicada. O artigo encontra-se em avaliação e descreve uma revisão sistemática no contexto de previsão de precipitação e irradiação solar utilizando DL. Embora tenha-se obtido resultados satisfatórios do modelo final em relação a modelos lineares e de ML, acreditamos que buscar novas formas de aquisição de dados históricos nos demais centros de estudos climáticos, utilização de outros modelos de ML e DL além dos utilizados neste trabalho e utilizar a mesma metodologia deste trabalho em regiões diferentes do Brasil podem auxiliar em trabalhos futuros desen-

volvidos a partir do trabalho proposto. Este trabalho apesar de ter atingido seus objetivos, como qualquer pesquisa, ainda pode evoluir em muitos aspectos, como por exemplo o uso destes mesmos modelos em problemas de classificação.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] WHAT Percentage of the Earth's Water Is Drinkable? - WorldAtlas.com. 2019. <https://www.worldatlas.com/articles/what-percentage-of-the-earth-s-water-is-drinkable.html>. Acesso em: 20/06/2019.
- [2] SNIS - Sistema Nacional de Informações Sobre Saneamento. 2019. <http://www.snis.gov.br/component/content/article?id=175>. Acesso em: 17/06/2019.
- [3] WU, X. et al. Photothermal materials: A key platform enabling highly efficient water evaporation driven by solar energy. *Materials Today Energy*, v. 12, p. 277–296, 2019. ISSN 2468-6069. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2468606918303514>>.
- [4] ALZHRANI, A. et al. Solar Irradiance Forecasting Using Deep Neural Networks. *Procedia Comput. Sci.*, Elsevier B.V., v. 114, p. 304–313, 2017. ISSN 18770509. Disponível em: <<https://doi.org/10.1016/j.procs.2017.09.045>>.
- [5] HASSAN, M. A. et al. Potential of four different machine-learning algorithms in modeling daily global solar radiation. *Renew. Energy*, Elsevier Ltd, v. 111, p. 52–62, 2017. ISSN 18790682. Disponível em: <<http://dx.doi.org/10.1016/j.renene.2017.03.083>>.
- [6] VOYANT, C. et al. Prediction intervals for global solar irradiation forecasting using regression trees methods. *Renew. Energy*, Elsevier Ltd, v. 126, p. 332–340, oct 2018. ISSN 18790682. Disponível em: <<https://doi.org/10.1016/j.renene.2018.03.055>>.
- [7] JAWAID, F.; NAZIRJUNEJO, K. Predicting daily mean solar power using machine learning regression techniques. *2016 6th Int. Conf. Innov. Comput. Technol. INTECH 2016*, IEEE, p. 355–360, 2017.
- [8] CHANDRASEKAR, V.; TAN, H.; CHEN, H. A machine learning system for rainfall estimation from spaceborne and ground radars. *2017 32nd Gen. Assem. Sci. Symp. Int. Union Radio Sci. URSI GASS 2017*, v. 2017-Janua, n. August, p. 1–2, 2017.
- [9] GANGHIS, M. P. E. D. Apostila Tratamento de Água. p. 18, 2014. Disponível em: <http://www.cesan.com.br/wp-content/uploads/2013/08/APOSTILA_DE_TRATAMENTO_DE_AGUA-.pdf>.
- [10] MARTINS, T. J. C. Sistemas de Abastecimento de Água para Consumo Humano – Desenvolvimento e Aplicação de Ferramenta Informática para a sua Gestão Integrada. p. 1–100, 2014. Disponível em: <[https://bibliotecadigital.ipb.pt/bitstream/10198/9311/1/Sistemas de Abastecimento de A?gua para Consumo Humano_versa?o=final.pdf](https://bibliotecadigital.ipb.pt/bitstream/10198/9311/1/Sistemas%20de%20Abastecimento%20de%20A?gua%20para%20Consumo%20Humano_versa?o=final.pdf)>.

- [11] TSUTIYA, M. T. *Redução do custo de energia elétrica em sistemas de abastecimento de água*. [S.l.]: ABES, 2001.
- [12] ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 15527: Água de chuva - Aproveitamento de coberturas em áreas urbanas para fins não potáveis - Requisitos. *Rio de Janeiro*, p. 8.
- [13] MINISTÉRIO DA SAÚDE. Portaria n.º 518, de 25 de março de 2017. Procedimentos e Responsabilidades Relativos ao Controle e Vigilância da Água para Consumo Humano e seu Padrão de Potabilidade. DF, Brasília.
- [14] SANTANA, J. S. et al. CARACTERÍSTICAS E VIABILIDADES NO REAPROVEITAMENTO DA ÁGUA DA CHUVA EM AMBIENTES CONSTRUÍDOS – ESTUDO DE CASO EM PEDREIRA. p. 235–253, 2018.
- [15] ANDRADE, C. Aproveitamento Imediato Da Água De Chuva. *Rev. Eletrônica Gestão e Tecnol. Ambient.*, p. 73–86, 2013.
- [16] ALMEIDA, M. E. P. D. ANÁLISE DA VIABILIDADE PARA IMPLANTAÇÃO DE SISTEMA DE APROVEITAMENTO DE ÁGUA DA CHUVA E AR-CONDICIONADO: ESTUDO DE CASO NO SETOR DE AULAS IV DA UFRN. 2018.
- [17] WISBECK, E. et al. Desinfecção de água de chuva por radiação ultravioleta. *Eng. Sanit. e Ambient.*, v. 16, n. 4, p. 337–342, 2017.
- [18] Weiller Dallagasperina, R.; PAVAN, W.; Mauricio Fernandes, J. Universidade De Passo Fundo Instituto De Ciências Exatas E Geociências Programa De Pós-Graduação Em Computação Aplicada Uma Plataforma De Comunicação Universal Utilizando Protocolo Zigbee. 2016.
- [19] MIORANDO, T.; BRIÃO, V. B.; GIRARDELLI, L. Potabilização de água da chuva por ultrafiltração. *Eng. Sanit. e Ambient.*, v. 22, n. 3, p. 481–490, 2017. ISSN 1809-4457. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-41522017000300481&lng=pt&tlng=pt>.
- [20] Liu, Y.; Lu, Z.; Yang, F. The investigation of solar pv models. In: *2018 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. [S.l.: s.n.], 2018. p. 1–5. ISSN 2472-8152.
- [21] KOMMEH, M.; AGRAWAL, R.; OLURIN, A. Ranking solar energy potential by class and country. p. 22–25, 2016.
- [22] LAVEZZO, C. A. L. *FONTES DE ENERGIA*. [S.l.]: UNIFLA, 2016.

- [23] Xiaodong Liang. Emerging power quality challenges due to integration of renewable energy sources. In: *2016 IEEE Industry Applications Society Annual Meeting*. [S.l.: s.n.], 2016. p. 1–9.
- [24] MACHADO, C. T.; MIRANDA, F. S. Energia solar fotovoltaica: Uma breve revisão. *Rev. Virtual Quim.*, v. 7, n. 1, p. 126–143, 2015. ISSN 19846835.
- [25] EPE. Anuário Estatístico de Energia Elétrica 2018 no ano base de 2017. p. 249, 2018. Disponível em: <<http://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-160/topico-168/Anuario2018vf.pdf>>.
- [26] LIU, D.; WANG, W.; LIU, J. Sensitivity Analysis of Meteorological Parameters on Building Energy Consumption. *Energy Procedia*, Elsevier B.V., v. 132, p. 634–639, 2017. ISSN 18766102. Disponível em: <<https://doi.org/10.1016/j.egypro.2017.09.700>>.
- [27] SCHIAVETTI, A.; CAMARGO, A. F. M. *CONCEITOS DE BACIAS HIDROGRÁFICAS - Teorias e aplicações*. [S.l.]: editus, 2002. ISBN 8574550531.
- [28] TUCCI, C. E. et al. *Hidrologia: ciência e aplicação*. São Paulo: Editora da, 2012.
- [29] ATLAS Brasileiro de Energia Solar. 2019. http://ftp.cptec.inpe.br/labren/publ/livros/Atlas_Brasileiro_Energia_Solar_2a_Edicao.pdf. Acesso em: 20/06/2019.
- [30] ABSOLAR. *ABSOLAR - Associação Brasileira de Energia Solar Fotovoltaica*. 2021. <https://www.absolar.org.br/noticia/renovaveis-aumentam-participacao-na-matriz-energetica-em-2019/>. Acesso em 16/04/2021.
- [31] CONWAY, D. *The Data Science Venn Diagram — Drew Conway*. 2019. <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>. Acesso em: 21/06/2019.
- [32] PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. *Big Data*, v. 1, n. 1, p. 51–59, 2013. PMID: 27447038. Disponível em: <<https://doi.org/10.1089/big.2013.1508>>.
- [33] ZEGURA, E.; DISALVO, C.; MENG, A. Care and the Practice of Data Science for Social Good. p. 1–9, 2018.
- [34] GETOOR, L. et al. Computing research and the emerging field of data science. *CRA Bulletin*, 2016.
- [35] Van der Aalst, W. Process mining: Data science in action. *Process Min. Data Sci. Action*, n. April 2014, p. 1–467, 2016.

- [36] VANDERPLAS, J. *Python data science handbook: essential tools for working with data*. [S.l.]: "O'Reilly Media, Inc.", 2016.
- [37] MARSLAND, S. *Machine learning: an algorithmic perspective*. [S.l.]: Chapman and Hall/CRC, 2011.
- [38] RICH, S. K.; TERMAN, J. R. Axon formation, extension, and navigation: only a neuroscience phenomenon? *Current Opinion in Neurobiology*, v. 53, p. 174 – 182, 2018. ISSN 0959-4388. Developmental Neuroscience. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0959438818301363>>.
- [39] BOONYUEN, K.; KAEWPRAPHA, P.; SRIVIHOK, P. Daily rainfall forecast model from satellite image using Convolution neural network. *Proceeding 2018 3rd Int. Conf. Inf. Technol. InCIT 2018*, Mahasarakham University, Faculty of Informatics, p. 1–7, 2018.
- [40] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.
- [41] QIU, M. et al. A short-term rainfall prediction model using multi-task convolutional neural networks. *Proc. - IEEE Int. Conf. Data Mining, ICDM*, v. 2017-Novem, p. 395–404, 2017. ISSN 15504786.
- [42] REYNOLDS, J.; AHMAD, M. W.; REZGUI, Y. Holistic modelling techniques for the operational optimisation of multi-vector energy systems. *Energy and Buildings*, Elsevier, v. 169, p. 397–416, jun 2018. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378778817340240> <https://linkinghub.elsevier.com/retrieve/pii/S0378778817340240>>.
- [43] ARORA, A.; GAUR, P. Comparison of ANN and ANFIS based MPPT Controller for grid connected PV systems. *12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015*, IEEE, p. 1–6, 2016.
- [44] CAO, J.; FAN, Z. Deep learning-based online small signal stability assessment of power systems with renewable generation. *Proceedings - 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovations, SmartWorld/UIC/ATC/ScalCom/CBDCo*, IEEE, p. 216–221, 2018.
- [45] WANG, L. et al. SuperNeurons: Dynamic GPU Memory Management for Training Deep Neural Networks. jan 2018. Disponível em: <<http://arxiv.org/abs/1801.04380> <http://dx.doi.org/10.1145/3178487.3178491>>.
- [46] ZHAO, Y. et al. Bridging the gap between deep learning and sparse matrix format selection. In: *Proceedings of the 23rd ACM SIGPLAN Symposium on*

Principles and Practice of Parallel Programming - PPOPP '18. New York, New York, USA: ACM Press, 2018. p. 94–108. ISBN 9781450349826. Disponível em: <<http://dl.acm.org/citation.cfm?doid=3178487.3178495>>.

- [47] ASWIN, S.; GEETHA, P.; VINAYAKUMAR, R. Deep Learning Models for the Prediction of Rainfall. *Proc. 2018 IEEE Int. Conf. Commun. Signal Process. ICCSP 2018*, IEEE, p. 657–661, 2018.
- [48] SORKUN, M. C.; PAOLI, C.; INCEL, Ö. D. Time Series Forecasting on Solar Irradiation using Deep Learning. *2017 10th International Conference on Electrical and Electronics Engineering (ELECO)*, p. 151–155, 2017. Disponível em: <<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8266215>>.
- [49] MUHAMMAD, A. et al. Deep Learning Application in Power System with a Case Study on Solar Irradiation Forecasting. *1st Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2019*, p. 275–279, 2019.
- [50] ABADI, M. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <<https://www.tensorflow.org/>>.
- [51] CHOLLET, F. et al. *Keras*. 2020. Disponível em: <<https://keras.io/>>.
- [52] NAKAGAWA, E. Y. et al. *Revisão sistemática da literatura em engenharia de software: Teoria e Prática*. [S.l.]: Elsevier Brasil, 2017.
- [53] LI, X.; DU, Z.; SONG, G. A Method of Rainfall Runoff Forecasting Based on Deep Convolution Neural Networks. *Proc. - 2018 6th Int. Conf. Adv. Cloud Big Data, CBD 2018*, IEEE, p. 304–310, 2018.
- [54] TAN, H.; CHANDRASEKAR, V.; CHEN, H. A machine learning model for radar rainfall estimation based on gauge observations. *2017 United States Natl. Comm. URSI Natl. Radio Sci. Meet. Usn. NRSM 2017*, USNC-URSI, p. 1–2, 2017.
- [55] MOON, S. H. et al. Application of machine learning to an early warning system for very short-term heavy rainfall. *J. Hydrol.*, Elsevier, v. 568, n. July 2018, p. 1042–1054, 2019. ISSN 00221694. Disponível em: <<https://doi.org/10.1016/j.jhydrol.2018.11.060>>.
- [56] ZHAN, C. et al. Daily Rainfall Data Construction and Application to Weather Prediction. *2019 IEEE Int. Symp. Circuits Syst.*, IEEE, v. 1, p. 1–5, 2019.
- [57] GOYAL, H.; SHARMA, C.; JOSHI, N. Estimation of Monthly Rainfall using Machine Learning Approaches. *Int. Conf. Innov. Control. Commun. Inf. Syst. ICICCI 2017*, IEEE, p. 1–6, 2019.

- [58] THIRUMALAI, C. et al. Heuristic prediction of rainfall using machine learning techniques. *Proc. - Int. Conf. Trends Electron. Informatics, ICEI 2017*, v. 2018-Janua, p. 1114–1117, 2018.
- [59] HASHIM, F. et al. Prediction of rainfall based on weather parameter using artificial neural network. *J. Fundam. Appl. Sci.*, v. 9, n. 3S, p. 493, 2018.
- [60] ZHANG, P. et al. Short-term Rainfall Forecasting Using Multi-layer Perceptron. *IEEE Trans. Big Data*, PP, n. XX, p. 1–1, 2018. ISSN 2332-7790. Disponível em: <<https://ieeexplore.ieee.org/document/8468083/>>.
- [61] MEYER, H. et al. Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. *Atmos. Res.*, Elsevier B.V., v. 169, p. 424–433, 2016. ISSN 01698095. Disponível em: <<http://dx.doi.org/10.1016/j.atmosres.2015.09.021>>.
- [62] YU, P. S. et al. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *J. Hydrol.*, Elsevier B.V., v. 552, p. 92–104, 2017. ISSN 00221694. Disponível em: <<http://dx.doi.org/10.1016/j.jhydrol.2017.06.020>>.
- [63] CRAMER, S.; KAMPOURIDIS, M.; FREITAS, A. A Genetic Decomposition Algorithm for Predicting Rainfall within Financial Weather Derivatives. *GECCO '16*, p. 885–892, 2016.
- [64] BERITELLI, F. et al. Rainfall Estimation Based on the Intensity of the Received Signal in a LTE/4G Mobile Terminal by Using a Probabilistic Neural Network. *IEEE Access*, IEEE, v. 6, p. 30865–30873, 2018. ISSN 21693536.
- [65] PALANI, K. et al. Blue skies: A methodology for data-driven clear sky modelling. *IJCAI Int. Jt. Conf. Artif. Intell.*, p. 3777–3783, 2017. ISSN 10450823.
- [66] LAOPAIBOON, T. et al. Hour-Ahead Solar Forecasting Program Using Back Propagation Artificial Neural Network. *Proc. Conf. Ind. Commer. Use Energy, ICUE*, Asian Institute of Technology, v. 2018-October, n. October, 2019. ISSN 2166059X.
- [67] FOUILLOY, A. et al. Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability. *Energy*, Elsevier Ltd, v. 165, p. 620–629, 2018. ISSN 03605442. Disponível em: <<https://doi.org/10.1016/j.energy.2018.09.116>>.
- [68] SILVA, M. B. P. da et al. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil. *J. Atmos. Solar-Terrestrial Phys.*, Elsevier Ltd, v. 160, p. 11–23, jul 2017. ISSN 13646826.
- [69] Muhammad Ehsan, R.; SIMON, S. P.; VENKATESWARAN, P. R. Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron. *Neural Comput. Appl.*,

Springer London, v. 28, n. 12, p. 3981–3992, dec 2017. ISSN 09410643. Disponível em: <<http://link.springer.com/10.1007/s00521-016-2310-z>>.

- [70] PICKELL, D. *5 Steps of the Data Analysis Process*. 2019. <https://learn.g2.com/data-analysis-process>. Acesso em: 09/10/2019.
- [71] GOHRANI, K. *The 5 Steps of the Data Analysis Process - Medium*. 2019. https://medium.com/@kunal_gohrani/the-5-steps-of-the-data-analysis-process-2512ba6ac31e. Acesso em: 09/10/2019.
- [72] INSTITUTO Nacional de Meteorologia - INMET. 2020. <https://portal.inmet.gov.br/dadoshistoricos>. Acesso em: 06/10/2020.
- [73] INTRODUCTION — Vue.js. 2020. <https://vuejs.org/v2/guide/>. Acesso em 08/10/2020).
- [74] VERSHININ, V. et al. *VueLayers :: Homepage*. 2020. <https://vuelayers.github.io/#/>. Acesso em 23/10/2020.
- [75] OPENSTREETMAP.ORG. *OpenStreetMap*. 2020. <https://www.openstreetmap.org/#map=5/-15.130/-53.189>. Acesso em 23/10/2020.
- [76] CONTAS Econômicas Ambientais da Água (CEAA). 2019. https://www.ana.gov.br/todos-os-documentos-do-portal/documentos-spr/contas_economicas.pdf. Acesso em: 17/06/2019.
- [77] BAU, S. R. C. *Análise técnica e econômica da ultrafiltração de água da chuva para fins de potabilidade utilizando energia fotovoltaica*. Dissertação (Dissertação de Mestrado) — Universidade de Passo Fundo, 2019.



UPF

UNIVERSIDADE
DE PASSO FUNDO

UPF Campus I - BR 285, São José
Passo Fundo - RS - CEP: 99052-900
(54) 3316 7000 - www.upf.br