



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA**

**JHEKLOS GOMES DA SILVA**

**Sensoriamento Virtual da Temperatura da Água para uso na Gestão  
Hídrica e na Aquicultura 4.0**

**Dissertação de Mestrado**

**Recife, 2024**



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA**

## **Sensoriamento Virtual de Temperatura da Água para uso na Gestão Hídrica e na Aquicultura 4.0**

Dissertação apresentada ao Programa de Pós-graduação em Informática Aplicada da Universidade Federal Rural de Pernambuco como exigência parcial à obtenção do título de Mestre em Informática Aplicada.

**Área de Concentração: Engenharia de Software Aplicada**  
**Orientador: Prof. Dr. Ricardo André Cavalcante de Souza**  
**Coorientador: Prof. Dr. Obionor de Oliveira Nóbrega**

**Recife, 2024**



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA**

## **Sensoriamento Virtual de Temperatura da Água para uso na Gestão Hídrica e na Aquicultura 4.0**

**JHEKLOS GOMES DA SILVA**

Dissertação julgada adequada como requisito parcial para obtenção do título de Mestre em Informática Aplicada (área de concentração: Engenharia de Software) pelo Programa de Pós-Graduação em Informática Aplicada da Universidade Federal Rural de Pernambuco. A dissertação foi aprovada por unanimidade em sua forma final em sessão pública de defesa em 08/02/2024.

**Orientador:**

---

Prof. Dr. Ricardo André Cavalcante de Souza  
Universidade Federal Rural de Pernambuco

**Banca Examinadora:**

---

Prof. Dr. Marco Antônio de Oliveira Domingues  
Instituto Federal de Pernambuco

---

Prof. Dr. Fernando Antônio Aires Lins  
Universidade Federal Rural de Pernambuco

Dados Internacionais de Catalogação na Publicação Universidade Federal Rural de Pernambuco Sistema Integrado de Bibliotecas  
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

XXXX DA SIVA, JHEKLOS GOMES

Sensoriamento Virtual de Temperatura da Água para Gestão Hídrica e Aquicultura 4.0/ JHEKLOS  
GOMES DA SILVA. - 2023.

145 f. : il.

Orientador: RICARDO ANDRE CAVALCANTE DE SOUZA.

Inclui referências.

Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em  
Informática Aplicada, Recife, 2023. 1.Sensores Virtuais. 2. Temperatura da Água. 3. Aprendizagem de Máquina. 4. Aquicultura  
4.0. I. SOUZA, RICARDO ANDRE CAVALCANTE DE, orient. II. Título

*Dedico este trabalho a todos os que me ajudaram e apoiaram ao longo desta caminhada.*

## **Agradecimentos**

Primeiramente, gostaria de expressar minha gratidão a Deus por me guiar e abrir essa oportunidade para mim. Sua orientação e força foram fundamentais em cada passo deste caminho.

À minha família, por sempre me apoiaram, encorajaram e acreditaram em mim, mesmo nos momentos mais desafiadores.

Aos meus queridos professores, em especial aos meus orientadores, Dr. Ricardo Souza e Dr. Obionor Oliveira, minha gratidão é imensa. Sua dedicação, sabedoria e orientação foram cruciais para o sucesso desta dissertação. Suas valiosas contribuições não apenas enriqueceram meu trabalho, mas também moldaram meu crescimento acadêmico e pessoal.

Além disso, quero expressar minha profunda apreciação à minha equipe de pesquisa. Vocês não apenas foram colegas, mas também orientadores para mim. Nossas discussões, colaborações e esforços compartilhados enriqueceram minha perspectiva e ajudaram a moldar meu trabalho de forma significativa. Agradeço a cada um de vocês por sua dedicação e apoio inestimável.

À Agência Pernambucana de Águas e Climas (APAC), gostaria de estender meu sincero agradecimento por conceder os dados necessários para a realização de uma parte significativa deste trabalho. Sua colaboração foi fundamental para o sucesso desta pesquisa, e estou verdadeiramente grato pela generosidade e suporte fornecidos.

Gostaria de expressar minha gratidão à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro concedido. Sem o suporte da CAPES, alcançar meus objetivos acadêmicos teria sido muito mais desafiador. Sua contribuição foi fundamental para minha formação e para o sucesso deste projeto de pesquisa. Sou imensamente grato por esta oportunidade e pelo investimento em minha trajetória acadêmica.

Àqueles que, de uma forma ou de outra, contribuíram para esta conquista, meu sincero obrigado.

“A ciência é mais que um corpo de conhecimento, é uma forma de pensar, uma forma cética de interrogar o universo, com pleno conhecimento da falibilidade humana.”

(Carl Sagan)

## Resumo

A temperatura da água assume um papel primordial na preservação dos ecossistemas aquáticos, representando um fator crucial para a sua vitalidade e equilíbrio. O aumento da temperatura da água reduz a solubilidade do oxigênio, prejudicando os organismos aquáticos. Além disso, influencia negativamente a taxa metabólica, afetando o crescimento, a reprodução, a ingestão alimentar e, conseqüentemente, a eficiência alimentar e o consumo de ração na aquicultura. Nos cenários aquáticos, como rios e reservatórios voltados ao consumo humano, essa elevação de temperatura pode até ocasionar a proliferação de organismos indesejados, tais como algas nocivas. Embora dispositivos de medição sejam ferramentas essenciais para monitorar a temperatura da água, existem certas restrições que podem dificultar sua implementação, como infraestrutura inadequada, acesso à área remota e falha no sensor. Diante dessas limitações, é importante explorar abordagens alternativas para monitorar a temperatura da água. Sensores virtuais baseados em modelos de aprendizado de máquina, têm sido desenvolvidos como alternativa aos sensores físicos que apresentam defeitos ou falhas em diversas áreas, inclusive no monitoramento da temperatura da água, como também qualidade do solo e do ar. Essa abordagem visa garantir que o monitoramento não seja interrompido. Este estudo examinou quatro modelos distintos - Regressão Linear, Modelo Estocástico, árvores extremamente aleatórias e Rede Neural Perceptron Multicamada - para estimar a temperatura de corpos de água que requerem monitoramento contínuo. A avaliação do desempenho desses modelos foi realizada através de quatro métricas estatísticas: Erro Médio Quadrático (MSE), Raiz do Erro Quadrático Médio (RMSE), Média Absoluta do Erro (MAE) e Coeficiente de Determinação ( $R^2$ ). A contribuição do trabalho incluiu a identificação do modelo mais eficiente, com o Perceptron demonstrando uma acurácia superior. Além disso, foram desenvolvidas soluções acessíveis. Para a aquicultura, foi implementada uma plataforma web de monitoramento em tempo real, enquanto para o monitoramento dos recursos hídricos, foi criado um mapa interativo que possibilita o acompanhamento da temperatura da água em rios e reservatórios.

Palavras-chave: Aquicultura 4.0; Aprendizado de Máquina; Sensor Virtual; Modelos Estatísticos.

## **Abstract**

Water temperature plays a key role in the preservation of aquatic ecosystems, representing a crucial factor for their vitality and balance. Increasing water temperature reduces oxygen solubility, harming aquatic organisms. Furthermore, it negatively influences the metabolic rate, affecting growth, reproduction, food intake and, consequently, feed efficiency and feed consumption in aquaculture. In aquatic scenarios, such as rivers and reservoirs intended for human consumption, this increase in temperature can even cause the proliferation of unwanted organisms, such as harmful algae. Although measuring devices are essential tools for monitoring water temperature, there are certain restrictions that can hamper their implementation, such as inadequate infrastructure, remote area access, and sensor failure. Given these limitations, it is important to explore alternative approaches to monitoring water temperature. Virtual sensors based on machine learning models have been developed as an alternative to physical sensors that present defects or failures in several areas, including monitoring water temperature, as well as soil and air quality. This approach aims to ensure that monitoring is not interrupted. This study examined four distinct models – Linear Regression, Stochastic Model, extremely random trees, and Multilayer Perceptron Neural Network – to estimate the temperature of bodies of water that require continuous monitoring. The performance evaluation of these models was carried out using four statistical metrics: Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination ( $R^2$ ). The contribution of the work included the identification of the most efficient model, with Perceptron demonstrating superior accuracy. Furthermore, affordable solutions were developed. For aquaculture, a real-time monitoring web platform was implemented, while for monitoring water resources, an interactive map was created that makes it possible to monitor water temperatures in rivers and reservoirs.

**Keywords:** Aquaculture 4.0; Machine Learning; Virtual Sensor; Statistical Models.

## LISTA DE FIGURAS

<b>Figura 1. Atividades do trabalho de pesquisa</b>	24
<b>Figura 2. Mapa Mental com as categorias de modelos e seus respectivos autores</b>	33
<b>Figura 3. Mapa Mental dos principais métodos para estimar a temperatura da água</b>	53
<b>Figura 4. Boia com sensor de temperatura da água para coleta de dados</b>	57
<b>Figura 5. Estação meteorológica B2K para coleta de dados</b>	58
<b>Figura 6. Correlação das variáveis Estação Meteorológica das API</b>	60
<b>Figura 7. Correlação das variáveis Estação Meteorológica B2K</b>	61
<b>Figura 8. Multi-layer Perceptron Estrutura</b>	66
<b>Figura 9. Gráficos das métricas estatísticas dos dados da API coletados na web</b>	70
<b>Figura 10. Gráfico do teste de validação da estação API todas variáveis de entrada</b>	71
<b>Figura 11. Dispersão do teste de validação da estação API todas variáveis de entrada</b>	72
<b>Figura 12. Gráfico das métricas estatísticas dos dados da estação b2k coletados</b>	73
<b>Figura 13. Gráfico do teste de validação da estação B2K com todas as variáveis de entrada</b>	75
<b>Figura 14. Dispersão do teste de validação da estação B2K com todas variáveis de entrada</b>	76
<b>Figura 15. Gráficos das métricas estatísticas da correlação moderada da API</b>	78
<b>Figura 16. Gráfico do teste de validação da estação API variáveis com correlação moderada</b>	80
<b>Figura 17. Gráfico das métricas estatísticas da correlação moderada B2k</b>	81
<b>Figura 18. Gráfico do teste de validação da estação B2K variáveis com correlação moderada</b>	82
<b>Figura 19. Gráficos das métricas estatísticas do teste de validação nos rios</b>	84
<b>Figura 20. Estimativas e observações para o conjunto de validação do rio</b>	86
<b>Figura 21. Gráficos das métricas estatísticas do teste de validação nos rios</b>	87
<b>Figura 22. Estimativas e observações para o conjunto de validação de reservatórios</b>	89
<b>Figura 23. Interface com o Usuário do mapa interativo</b>	94
<b>Figura 24. Diagrama de Componentes do Sensor Virtual para Rios e Reservatórios</b>	95
<b>Figura 25. Interface com o usuário do aplicativo Web</b>	96
<b>Figura 26. Diagrama de Componentes do Sensor Virtual para Aquicultura 4.0</b>	96

## **LISTA DE TABELAS**

<b>Tabela 1. Protocolo para o Mapeamento Sistemático da Literatura</b>	<b>28</b>
<b>Tabela 2. String de Busca</b>	<b>29</b>
<b>Tabela 3. Quantidade de EP selecionados na etapa de Extração de Dados</b>	<b>30</b>
<b>Tabela 4. Avaliação da Qualidade</b>	<b>31</b>
<b>Tabela 5. Avaliação qualitativa dos EP</b>	<b>31</b>
<b>Tabela 6. Panorama dos EP analisados</b>	<b>32</b>
<b>Tabela 7. EP analisados pelos critérios qualitativos</b>	<b>34</b>
<b>Tabela 8. Modelos com melhores desempenho de acordo com as escalas de tempo</b>	<b>37</b>
<b>Tabela 9. Estrutura do Modelo de Maturidade de Pearson</b>	<b>59</b>
<b>Tabela 10. Correlação de temperatura do ar e da água</b>	<b>61</b>
<b>Tabela 11. Performance dos modelos com dados API coletados na web</b>	<b>69</b>
<b>Tabela 12. Performance dos modelos com dados da estação meteorológica B2K</b>	<b>73</b>
<b>Tabela 13. Desempenho do modelo com dados da API coletados na Web usando variáveis de correlação moderada</b>	<b>77</b>
<b>Tabela 14. Desempenho do modelo com dados da estação B2K usando variáveis de correlação moderada</b>	<b>81</b>
<b>Tabela 15. Desempenho dos modelos com os dados coletados nos Rios</b>	<b>83</b>
<b>Tabela 16. Desempenho dos modelos com os dados coletados nos Reservatórios</b>	<b>86</b>

## LISTA DE ABREVIATURAS E SIGLAS

<b>Abreviação</b>	<b>Significado</b>
<b>EP</b>	Estudos Primários
<b>CQ</b>	Critérios Qualitativos
<b>IA</b>	índice de aceitação
<b>AT</b>	Atende
<b>NA</b>	Não atende
<b>AA</b>	Artigos aceitos
<b>AR</b>	Artigos rejeitados
<b>IoT</b>	Internet das Coisas
<b>B2K</b>	Estação Be Weather
<b>API</b>	Interfaces de programação de aplicativos
<b>MLPNN</b>	Rede Neural Multicamada Perceptron
<b>MSE</b>	Erro quadrático médio
<b>RMSE</b>	Raiz do erro quadrático médio
<b>MAE</b>	Erro médio absoluto
<b>R<sup>2</sup></b>	Coefficiente de determinação
<b>APAC</b>	Agência Pernambucana de Águas e Climas
<b>HTML</b>	Linguagem de Marcação de HiperTexto
<b>CEP</b>	Código de Endereçamento Postal
<b>MSL</b>	Mapeamento Sistemático da Literatura
<b>BPMN</b>	Modelagem de processos de negócio
<b>SIGRH</b>	Sistema Integrado de Gerenciamento dos Recursos Hídricos
<b>UML</b>	Linguagem de Modelagem Unificada

## SUMÁRIO

1	INTRODUÇÃO	16
1.1	Visão Geral	17
1.2	Justificativa e Motivação	19
1.3	Objetivos de Pesquisa	22
1.4	Contribuições Esperadas do Trabalho	23
1.5	Método de Pesquisa	23
1.6	Estrutura do Trabalho	26
2	MAPEAMENTO SISTEMÁTICO DA LITERATURA	27
2.1	Visão Geral	28
2.2	Definição do Escopo do Mapeamento	28
2.3	Busca por Estudos Primários	29
2.4	Extração de Dados	30
2.5	Análise e Síntese	31
2.6	Resultados	32
2.7	Discussões	36
2.8	Considerações Finais	38
3	ARCABOUÇO TEÓRICO	41
3.1	Visão Geral	42
3.2	Aquicultura	42
3.3	Internet das Coisas (IoT)	44
3.4	Aprendizado de máquinas	46
3.5	Sensores Virtuais	48
3.6	Gestão dos Recursos Hídricos	50
3.7	Estações Meteorológicas	51
3.8	Considerações Finais	52
4	MODELOS DE APRENDIZADO DE MÁQUINA PARA O DESENVOLVIMENTO DE SENSORES VIRTUAIS NA ESTIMATIVA DA TEMPERATURA DA ÁGUA	55
4.1	Visão Geral	56
4.2	Materiais e Métodos	56
4.2.1	Coleta de Dados no Tanque de Aquicultura	56
4.2.2	Coleta dos Dados nos Reservatórios e Rios	58
4.2.3	Correlação das Variáveis	59
4.2.4	Normalização dos Dados	61
4.3	Modelos de Aprendizado de Máquinas	62

4.3.1	Regressão Linear	62
4.3.2	Descida do Gradiente Estocástico	63
4.3.3	Árvores Extremamente Aleatórias	64
4.3.1	Rede Neural Multicamada Perceptron	65
4.4	Desempenho e Avaliação do Modelo	66
4.4.1	MSE	66
4.4.2	RMSE	67
4.4.3	MAE	67
4.4.4	R <sup>2</sup>	67
4.5	Método e tecnologias utilizadas para implementação dos algoritmos de aprendizado de máquina	68
4.6	Resultados	69
4.6.1	Resultados com dados da API coletados na Web utilizando todas as variáveis	69
4.6.2	Resultados com dados da estação B2K utilizando todas as variáveis	72
4.6.3	Resultados com dados da API coletados na Web utilizando variáveis com correlações moderadas	77
4.6.4	Resultados com Dados da Estação B2K utilizando Variáveis com Correlações Moderadas	78
4.6.5	Resultados dos testes de validação dos rios	83
4.6.6	Resultado dos testes de validação dos reservatórios	85
4.7	Considerações Finais	89
5	UTILIZAÇÃO DE SENSORES VIRTUAIS PARA APOIAR A AQUICULTURA 4.0 E A GESTÃO DOS RECURSOS HÍDRICOS	91
5.1	Visão Geral	92
5.2	Tecnologias utilizadas para a interface dos Sensores Virtuais	92
5.2.1	Tecnologias da interface Web para rios e reservatórios	92
5.2.2	Tecnologias da interface Web para tanques	93
5.3	Desenvolvimento das Interfaces para suporte tecnológico	93
5.3.1	Aplicativo Web Para os Sensores Virtuais nos Rios e Reservatórios	94
5.3.2	Aplicação Web Para os Sensores Virtuais nos Tanques de Aquicultura	95
5.4	Sensores Virtuais aplicados no Governo e na Iniciativa Privada	97
5.4.1	A startup Smart Rural	97
5.4.2	A Agência Pernambucana de Águas e Clima	98
5.5	Considerações Finais	99
6	CONCLUSÃO	100
6.1	Visão Geral	101

6.2	Dificuldades e Lições Aprendidas	103
6.3	Contribuições do Trabalho	104
6.4	Limitações e Trabalhos Futuros	106
	REFERÊNCIAS	107

## **1 INTRODUÇÃO**

A pesquisa científica é o resultado de um inquérito ou exame minucioso, conduzido com o propósito de resolver um problema, empregando métodos científicos (GERHARDT; SILVEIRA, 2009). Este capítulo apresenta uma visão abrangente do estudo, destacando a justificativa e motivação que o embasam, os objetivos propostos, as contribuições esperadas, a metodologia de pesquisa adotada e a estrutura do trabalho desenvolvido.

## 1.1 Visão Geral

A temperatura da água é um dos fatores mais importante para os ambientes aquáticos e os organismos que vivem nela. Os peixes são exotérmicos e dependem da temperatura dentro de sua faixa de tolerância (MCGRATH et al., 2017). A temperatura, além de influenciar o oxigênio dissolvido na água e a capacidade dos animais obterem oxigênio para sobreviver, afeta também o metabolismo das espécies, conseqüentemente influenciando na reprodução, alimentação e no crescimento.

Segundo Christ e Wenli (2014), a temperatura da água é uma medida da energia cinética da água e é expressa em graus Fahrenheit (F) ou Celsius (C). A temperatura da água varia de acordo com a estação, a profundidade e, em alguns casos, a hora do dia, especialmente em corpos d'água mais rasos e menos estáveis, a hora do dia pode ser um fator significativo na variação da temperatura da água. Nos ambientes mais profundos a hora do dia tem uma influência mínima. Além disso, as condições climáticas, como a quantidade de luz solar, a intensidade dos ventos e a cobertura de nuvens, têm influência significativa na temperatura da água. Outras variáveis meteorológicas que exercem forte influência são a temperatura do ar, especialmente em corpos d'água rasos, onde a transferência de calor entre a água e o ar ocorre mais rapidamente. A precipitação, como a chuva, pode resfriar a superfície da água. A umidade do ar também desempenha um papel, influenciando a taxa de evaporação da água e, conseqüentemente, sua temperatura. Além disso, a intensidade e direção dos ventos são relevantes: ventos fortes podem causar mistura na água, redistribuindo o calor ao longo da coluna de água. A presença de vegetação ao redor de um corpo d'água também desempenha um papel, fornecendo sombra e reduzindo a exposição direta à luz solar. Outrossim, as descargas de água para consumo humano em reservatórios podem afetar a temperatura da água. De acordo com Quartieri (2019), a temperatura é uma indicação da agitação das moléculas de água, onde temperaturas mais altas correspondem a moléculas mais agitadas e temperaturas mais baixas correspondem a moléculas menos agitadas.

A temperatura da água exerce um papel fundamental em múltiplos aspectos dos ecossistemas aquáticos e das atividades humanas relacionadas à água, uma vez que está intimamente ligada à qualidade desse recurso natural. A temperatura da água em córregos e rios é um atributo importante da qualidade da água e controla a saúde dos ecossistemas de água doce (MORRILL et al., 2005). Essa relação tem um impacto significativo nos sistemas de abastecimento público, influenciando diretamente o uso dos recursos hídricos.

O aumento no aquecimento global tem causado mudanças climáticas que afetam diretamente a temperatura da água. Segundo Nobre (2011), por conta do acúmulo de gases de efeito estufa de origem antrópica na atmosfera, o aquecimento global vem causando o aumento das temperaturas do ar e dos oceanos. Esses impactos fazem com que seja necessário tomar medidas para proteger a biodiversidade aquática.

Uma das estratégias que podem ser adotadas é monitorar a temperatura da água continuamente a fim de avaliar se as condições estão dentro dos limites adequados, uma vez que a temperatura desempenha um papel crítico na saúde e no equilíbrio dos ecossistemas aquáticos. A monitorização de ecossistemas aquáticos desempenha um papel crucial na gestão ambiental, uma vez que fornece dados que contribuem para diagnosticar a qualidade dos corpos de água em movimento (lóticos) e parados (lênticos), permitindo identificar as principais causas dos impactos que levam à degradação dos recursos hídricos (BUZELLI; CUNHA, 2013).

Mudança na temperatura da água pode levar a um desequilíbrio no ecossistema e desenvolvimento das espécies, resultando em perda para muitas atividades relacionadas à criação de espécies aquáticas. Em vista disso, é importante monitorar a temperatura da água para garantir a saúde e identificar tendências ameaçadoras para que seja possível tomar decisões para manter os ecossistemas seguros para os seres vivos que dependem dela.

O monitoramento da temperatura da água oferece aos gestores de recursos hídricos informações essenciais para tomar decisões embasadas, detectar eventos extremos, avaliar a saúde do ecossistema aquático e planejar estratégias de adaptação às mudanças climáticas. Essa prática desempenha um papel fundamental na proteção da biodiversidade aquática e na gestão sustentável dos recursos hídricos.

Para as espécies cultivadas nos tanques de aquicultura, a temperatura ideal pode variar de acordo com a espécie e com as condições ambientais. Segundo Lourenço et al. (1999), o acompanhamento da temperatura da água deve ser diário devido a sua grande variação ao longo do dia, com pelo menos três leituras ao dia. Dessa forma os produtores devem manter a temperatura dentro da faixa aceitável para garantir o bem-estar dos animais e conseqüentemente uma melhor qualidade para a produção das espécies.

A saúde de uma espécie destinada à alimentação é fundamental para assegurar a segurança dos consumidores. O monitoramento adequado da temperatura da água permite garantir que as espécies estejam recebendo as condições ideais de crescimento e reprodução, resultando em uma produção

mais eficiente, o que diminui os custos de produção e conseqüentemente os preços finais para o consumidor, possibilitando um maior acesso à proteína animal pela população.

Embora a aquicultura seja uma fonte importante de proteína animal para a alimentação, existem muitos desafios tecnológicos para serem superados e torná-la mais eficiente e rentável. A falta de tecnologia para controlar e monitorar os problemas da área torna difícil garantir condições ideais. Com isso, torna-se necessário investir em pesquisas para melhorar as produções de cultivo em tanques ou ambientes controlados.

Diante dessa perspectiva, o presente trabalho de pesquisa busca fornecer uma solução inteligente para a estimativa da temperatura da água. A era da tecnologia digital e da conectividade global, juntamente com o avanço da Internet das Coisas (IoT), tem ampliado a demanda por desenvolvimentos tecnológicos que auxiliem os produtores e gestores de recursos hídricos. Essa abordagem é fundamental para garantir o êxito da Aquicultura 4.0 e o adequado gerenciamento dos ecossistemas aquáticos.

## **1.2 Justificativa e Motivação**

A temperatura da água é medida através de sensores de temperatura que normalmente são compostos por resistência, termistor ou termopar (LIMA, 2018). Esses dispositivos requerem a aplicação de corrente elétrica para medir a resistência ou a tensão, o que implica na necessidade de uma fonte de energia no local de medição. No entanto, há situações em que pode dificultar a medição direta da temperatura da água por meio desses sensores, como em áreas remotas ou em casos de pequenos produtores que não possuem acesso a esses dispositivos. Soluções de baixo custo para monitorar a temperatura da água através de sensores físicos, torna essa tecnologia mais acessível e amplamente disponível (GARBOSSA et al., 2021; PARRA et al., 2018). Apesar da importância dos sensores físicos, é importante reconhecer que eles podem falhar devido a vários fatores, como desgaste, calibração inadequada ou interferências externas (THIYAGARAJAN et al., 2018). Essas falhas podem resultar na interrupção da coleta de dados e na impossibilidade de monitorar adequadamente a temperatura da água. É nesse contexto que a pesquisa sobre sensoriamento virtual se torna relevante. O objetivo não é substituir os sensores físicos, mas sim oferecer uma alternativa servindo como redundância quando esses sensores se tornam inoperáveis por falha em alguma parte do equipamento eletrônico ou em áreas onde eles ainda não estão instalados. Diante disso, é essencial buscar alternativas inovadoras para estimar a temperatura da

água, especialmente em atividades como a aquicultura, que desempenha um papel significativo na busca pela erradicação da fome.

Essa questão assume uma relevância ainda maior quando consideramos o papel crucial da aquicultura no alcance do segundo Objetivo do Desenvolvimento Sustentável - ODS 2 da Agenda 2030 da ONU (2015) é a erradicação da fome e a aquicultura pode desempenhar um papel crucial na segurança alimentar, fornecendo fontes de proteínas acessíveis e nutricionalmente adequadas para a população. A cadeia produtiva de peixes cultivados no Brasil tem aumentado a demanda nos últimos anos. De acordo com a Peixe BR, em 2022 o Brasil produziu 860.355 toneladas de peixes, com receita de cerca de R\$ 9 bilhões, representando um aumento de 48,6% em relação ao ano de 2014 com produção de 578.800 toneladas. Além disso, a piscicultura gera cerca de 3 milhões de empregos diretos e indiretos.

Ao estimar a temperatura da água em locais de produção na aquicultura, é possível monitorar e manter condições ideais que variam de acordo com a espécie de peixe, mas geralmente estão dentro de uma faixa específica para cada espécie, em regra, essas condições incluem a qualidade da água em que o oxigênio é afetado pela temperatura, impedir o stress térmico, peixes estressados devido a variações na temperatura da água podem ser mais suscetíveis a doenças, pois o estresse pode comprometer sua capacidade de lidar com infecções e patógenos, além da eficiência alimentar onde temperaturas ótimas facilitam a digestão e a absorção de nutrientes. Em ambientes com temperaturas elevadas, os peixes podem apresentar redução no apetite, resultando em desperdício de ração e menor eficiência alimentar.

Além disso, o ODS 6 da Agenda 2030 consiste em alcançar o acesso universal e equitativo à água para consumo humano, segura e acessível para todas as pessoas. Uma maneira importante de apoiar esse objetivo é fornecer a estimativa da temperatura da água para o monitoramento em rios e reservatórios que não possuem dispositivos de medição. Isso pode ser uma ferramenta valiosa para o gerenciamento dos recursos hídricos, uma vez que a temperatura da água desempenha um papel fundamental na saúde dos ecossistemas aquáticos e na qualidade da água destinada ao consumo humano afetando diretamente a vida aquática, o ciclo de nutrientes, a oxigenação da água e o controle da proliferação de organismos indesejáveis, como algas nocivas.

A estimativa da temperatura da água para os rios e reservatórios, também pode ser realizada através de imagens de satélites, no entanto, existem alguns desafios significativos como a correção atmosférica que pode distorcer as imagens capturadas pelos sensores ópticos dos satélites, especialmente devido à absorção e dispersão da luz (YANG et al., 2022).

O uso de modelos estatísticos para simular ou prever a temperatura da água em riachos se tornou uma ferramenta importante na área (BENYAHYA et al., 2007). Esses modelos podem fornecer estimativas em locais diversos, viabilizando o monitoramento em áreas onde a utilização de sensores físicos é desafiadora devido a vários fatores locais, como a falta de infraestrutura como o difícil acesso a áreas rurais, falta de eletricidade ou eletricidade limitada, conectividade de rede de internet limitada e as dificuldades de manutenção. Esses modelos, podem ser treinados a partir de grandes conjuntos de dados climáticos, que podem ser obtidos a partir de estações meteorológicas que dispõem de dados acessível na internet, como temperatura do ar, radiação solar e umidade, intensidade e direção do vento, cobertura de nuvem e precipitação. Além de se apoiarem em dados históricos de medições anteriores da temperatura da água. A obtenção de dados de medição da temperatura da água apresenta desafios significativos devido à necessidade de dispositivos físicos de medição no local, que não estão facilmente acessíveis ao público. Além disso, lacunas nos dados podem surgir devido a falhas nos sensores. No entanto, uma vez que as variáveis preditoras (meteorológicas) e as dependentes (temperatura da água) são obtidas, torna-se possível estimar a temperatura da água.

Os modelos estatísticos são uma representação matemática de dados observados que ajudam analistas e cientistas de dados a visualizar as relações e padrões entre conjuntos de dados. Além disso, fornecem uma base sólida para prever e projetar dados para o futuro previsível (JOSHI, 2021). Esses modelos possuem a capacidade de compreender padrões complexos e relações não lineares nos dados, sendo capazes de se adaptar a várias variáveis independentes disponíveis, de acordo com o problema específico e o objetivo da estimativa. Essa abordagem também permite aumentar a acurácia da estimativa de maneira eficiente e rápida, além de oferecer melhoria contínua ao longo do tempo, ou seja, podem ser aprimorados e refinados à medida que mais dados se tornam disponíveis, o que pode ser útil quando se trata de estimar a temperatura da água em tempo real, permitindo uma resposta ágil às mudanças nas condições ambientais.

Segundo Baltar (2000), por definição, uma pesquisa só é necessária quando houver uma dúvida a ser esclarecida. Se não existir dúvida, não é necessário haver pesquisa. O problema constitui-se por uma dúvida que motiva e orienta a pesquisa. Nesse sentido, este trabalho é motivado pela seguinte questão de pesquisa: Qual o modelo estatístico de aprendizagem de máquina mais apropriado para estimar a temperatura da água em diferentes corpos de água, utilizando uma combinação de dados históricos da água e dados meteorológicos? Aqui, a "dúvida" referida é

precisamente essa questão de pesquisa, que busca esclarecimentos sobre a melhor abordagem para a estimativa da temperatura da água.

As variáveis meteorológicas exercem um impacto substancial na temperatura dos corpos d'água, destacando-se como um fator primordial para a escolha da temperatura da água como objeto de estudo nesta pesquisa. As mudanças climáticas, que afetam diretamente essas variáveis meteorológicas, têm repercussões significativas sobre a temperatura da água, influenciando, por sua vez, a biodiversidade dos ecossistemas aquáticos, as áreas de produção da aquicultura e os recursos hídricos destinados ao uso da população.

### **1.3 Objetivos de Pesquisa**

#### **Objetivo Geral:**

- Avaliar modelos estatísticos de aprendizagem de máquina para estimar a temperatura da água em tanques de aquicultura, como também em outras áreas como rios e reservatórios, para ser utilizado como forma alternativa dos sensores físicos de medição, dando suporte ao produtor que não tiver acesso a estes dispositivos como também aqueles que estão em áreas remotas, além de contribuir significativamente para a gestão dos reservatórios e rios, fornecendo informações valiosas para tomadas de decisão embasadas.

#### **Objetivos Específicos:**

- Investigar diferentes modelos estatísticos de aprendizado de máquina com dados históricos de estações meteorológicas e da temperatura da água;
- Observar a correlação da temperatura da água e do ar e de outras variáveis meteorológicas;
- Treinar diferentes tipos de modelos estatísticos e avaliar os que tiveram melhor desempenho nos testes de validações de estimação da temperatura da água;
- Comparar a acurácia dos resultados entre a estação meteorológica mais próxima do tanque e as disponíveis na Internet;
- Avaliar o desempenho dos modelos de aprendizado de máquinas para estimar a temperatura da água dos rios e reservatórios utilizando dados históricos da temperatura da água e temperatura do ar; e

- Desenvolver e aplicar sensor virtual para monitorar e estimar a temperatura do tanque, rios ou reservatórios de acordo com a região e os dados meteorológicos em tempo real.

#### **1.4 Contribuições Esperadas do Trabalho**

As contribuições científicas e técnicas esperadas para este trabalho são:

##### **Contribuições Científicas:**

- Demonstrar a eficácia dos modelos estatísticos na estimativa da temperatura da água em diferentes corpos d'água; e
- Desenvolvimento de uma abordagem inteligente para estimar a temperatura da água de forma contínua e em tempo real por meio de modelos de aprendizado de máquina.

##### **Contribuição Técnica:**

- Desenvolvimento de uma aplicação que torne possível estimar a temperatura da água dos tanques escavados de aquicultura, utilizando dados meteorológico de acordo com a localização; e
- Desenvolvimento de um aplicativo Web de suporte tecnológico por meio de um mapa interativo que permite ao usuário obter a estimativa da temperatura da água dos rios e reservatórios de Pernambuco.

#### **1.5 Método de Pesquisa**

Para realização de uma pesquisa faz-se necessário um método científico para direcionar e garantir a confiabilidade do trabalho realizado. Santos (2012, p. 4), define o método científico como um dispositivo ordenado, um conjunto de procedimentos sistemáticos que o pesquisador emprega para obter o conhecimento adequado do problema que se propõe resolver.

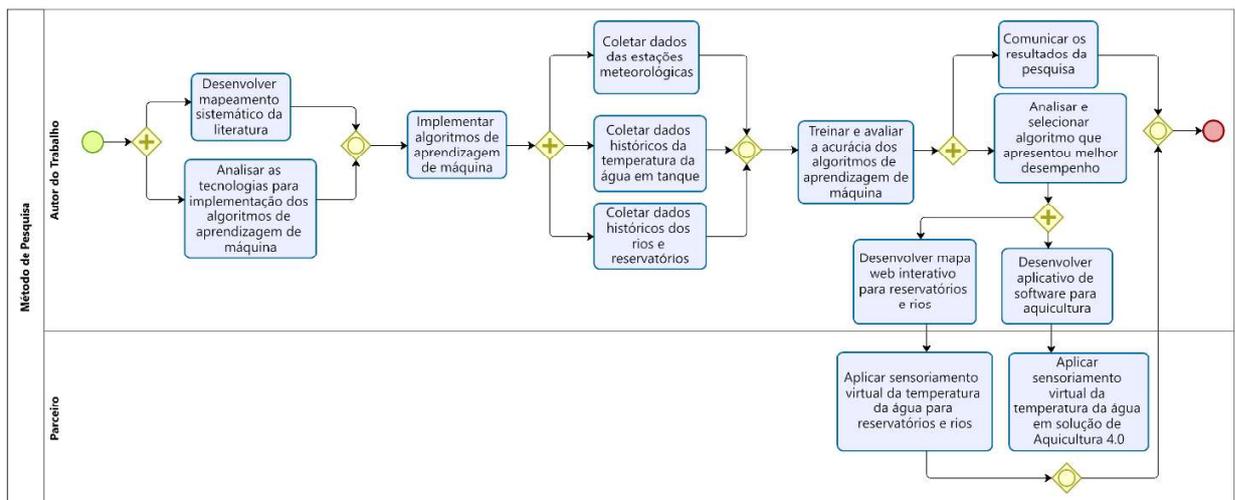
Buscando analisar o problema proposto, este trabalho será fundamentado em uma estratégia experimental, pois envolve a introdução de variáveis independentes para examinar o efeito sobre uma variável dependente. A natureza será quantitativa, pois envolve coletas de dados quantitativos que são analisados e utilizados nos testes de modelos estatísticos, além dos resultados que são apresentados por meios de gráficos ou tabelas, permitindo a avaliação de qual modelo algoritmo teve um melhor desempenho. Além disso, essa pesquisa será de caráter exploratório pois se inicia

a partir de um mapeamento sistemático da literatura a fim de investigar um problema de pesquisa que ainda não foi amplamente estudado.

Em relação aos procedimentos técnicos, esta pesquisa pode ser considerada como uma pesquisa aplicada, experimentada no contexto da produção de aquicultura, além da gestão hídrica de rios e reservatórios.

Com base na delimitação dos objetivos que guiaram o presente estudo, foi viabilizada a construção de um fluxo de atividades, representado pela Figura 1. Esse fluxo delinea as etapas requeridas para a concretização da abordagem proposta, com a finalidade de elaborar tecnologia de sensoriamento virtual da temperatura da água acessível tanto para os produtores quanto para os usuários, a fim de fomentar o avanço da aquicultura 4.0 e fornecer suporte à gestão dos recursos hídricos.

**Figura 1. Atividades do trabalho de pesquisa**



Fonte: Autor (2023).

1. **Desenvolver Mapeamento Sistemático da Literatura** – para identificação e análise dos trabalhos correlatos, além de fornecer uma visão abrangente das principais temáticas, abordagens e lacunas existentes no campo de estudo em questão;
2. **Analisar as tecnologias para implementação dos algoritmos de aprendizagem de máquina** – para identificação das tecnologias e métodos frequentemente utilizados por outros pesquisadores para implementação de algoritmos de aprendizagem de máquina;

3. **Implementar algoritmos de aprendizagem de máquina** – desenvolvimento de algoritmos utilizando diferentes modelos estatísticos para fins de predição ou estimação de temperatura da água;
4. **Coletar dados das estações meteorológicas** – necessário para obter as variáveis preditoras, serão realizadas a coleta na estação mais próxima do tanque e as API disponíveis na internet para a comparação entre as diferentes estações meteorológicas;
5. **Coletar dados históricos em tanque** – Os dados relacionados às variáveis dependentes serão adquiridos por meio dos sensores IoT disponíveis em um tanque específico para fins de aquicultura, situado nas instalações da universidade;
6. **Coletar dados históricos dos rios e reservatórios** – utilizar fontes confiáveis de informações, como bancos de dados especializados para adquirir dados históricos da temperatura da água e meteorológicos em rios e reservatórios;
7. **Treinar e avaliar a acurácia dos algoritmos de aprendizagem de máquina** – fazer comparações e avaliações de desempenho dos algoritmos utilizados, bem como a comparação dos resultados obtidos nos testes de validação com os valores verdadeiros. O objetivo é identificar quais algoritmos, estações meteorológicas e variáveis apresentaram um melhor desempenho na estimação da temperatura da água;
8. **Analisar e selecionar algoritmo que apresentou melhor desempenho** - Através da avaliação de métricas estatísticas, identificar o algoritmo de aprendizagem de máquina que demonstrou a maior acurácia durante os testes de validação. Este algoritmo será escolhido para a implementação nos sensores virtuais, visando otimizar o desenvolvimento e a precisão de estimação;
9. **Desenvolver mapa web interativo para reservatórios e rios** – a partir das informações da temperatura da água estimada em reservatórios e rios, desenvolver e apresentar mapa interativo para apoiar a gestão dos recursos hídricos, permitindo uma visualização clara e acessível das condições térmicas em diferentes corpos d'água;
10. **Desenvolver aplicativo de software para aquicultura** – para suporte tecnológico à soluções destinadas ao monitoramento em tanques de aquicultura, utilizando o melhor algoritmo de aprendizado de máquina treinado para estimar a temperatura da água nos tanques;

11. **Aplicar sensoriamento virtual da temperatura da água em reservatórios e rios** – demonstrar o mapa web desenvolvido à equipe de organização estadual responsável pela gestão de recursos hídricos de reservatórios e rios;
12. **Aplicar sensoriamento virtual da temperatura da água em solução de Aquicultura 4.0** – demonstrar o uso do aplicativo de software desenvolvido para startup com foco na Aquicultura 4.0; e
13. **Comunicar os resultados da pesquisa** – a partir da escrita de artigos científicos e da dissertação de mestrado.

## **1.6 Estrutura do Trabalho**

Além deste capítulo introdutório, o trabalho está estruturado em mais cinco capítulos e referências bibliográficas. O Capítulo 2 apresenta o resultado de um mapeamento sistemático da literatura sobre os trabalhos relacionados aos modelos estatísticos utilizados para estimar a temperatura da água em diversos corpos de água. O Capítulo 3 apresenta as abordagens e teorias que fundamentaram o desenvolvimento do trabalho. O Capítulo 4 apresenta as contribuições científicas na utilização dos modelos de aprendizado de máquina, os métodos e os resultados da validação na estimativa da temperatura da água em tanques de aquicultura e nos rios e reservatórios. O Capítulo 5 apresenta as interfaces tecnológicas dos sensores virtuais desenvolvidas para os usuários terem a estimativa da temperatura da água com base nos dados meteorológicos em tempo real, além das aplicações e contribuições tecnológicas para ser utilizadas pelos parceiros. O Capítulo 6 discorre sobre as considerações finais, limitações descobertas e trabalhos futuros. Finalmente, são apresentadas as referências de trabalhos citados ao longo do texto.

## **2 MAPEAMENTO SISTEMÁTICO DA LITERATURA**

O objetivo de uma revisão da literatura é encontrar o maior número possível de estudos primários relacionados à pergunta de pesquisa possível usando uma imparcial estratégia de busca (KITCHENHAM, 2007).

O Mapeamento Sistemático da Literatura (MSL) neste trabalho, é apoiado pelas diretrizes e etapas prescritas por Petersen et al. (2008). O principal objetivo de um estudo do tipo MSL é fornecer uma visão geral de um assunto e identificar a quantidade e o tipo de pesquisa e os resultados disponíveis nela. (PETERSEN et al., 2008).

## 2.1 Visão Geral

Neste capítulo, serão apresentados os procedimentos e os resultados do mapeamento sistemático da literatura realizado neste trabalho, cujo objetivo principal consistiu em obter uma resposta à pergunta de pesquisa formulada. Além disso, o MSL serviu para identificar os principais estudos relacionados à pesquisa em questão, bem como identificar lacunas e oportunidades para futuras pesquisas, tal como evidenciado por outros pesquisadores.

O MSL foi dirigido pelas seguintes etapas (PETERSEN et al., 2008): (1) Definição das Questões de Pesquisa (Escopo da Pesquisa), para orientar a busca e a seleção dos estudos relevantes identificando o tema e formulando perguntas específicas; (2) Realizar busca de estudos primários, para identificar as bases de dados relevantes e desenvolver uma estratégia de busca combinando termos e palavras chaves; (3) Extração de Dados, para analisar e sintetizar as informações relevantes encontradas nos estudos selecionados; (4) Análise e síntese dos principais estudos, para inclusão e exclusão de artigos relevantes; (5) Resultados dos estudos, para uma análise qualitativa dos principais resultados dos estudos incluídos, principais descobertas, tendências e padrões que surgem da análise comparativa; (6) Discussão, para responder a questão de pesquisa, discutir implicação e indicar lacunas. Nas próximas subseções, serão apresentados os passos detalhados e executados no contexto deste trabalho.

O mapeamento sistemático da literatura descrito neste capítulo foi avaliado por especialistas e aceito para publicação (SILVA, SOUZA e NÓBREGA, 2023a) na revista de publicação científica multidisciplinary Research, Society and Development.

## 2.2 Definição do Escopo do Mapeamento

Nesta etapa, foi elaborado um protocolo (Tabela 1) contendo as questões de pesquisa e as estratégias que foram feitas nas buscas com o objetivo de iniciar o processo. O objetivo é identificar quais os estudos primários mais relevantes que argumentam sobre como medir a temperatura da água por meio de dados estatísticos.

**Tabela 1. Protocolo para o Mapeamento Sistemático da Literatura**

<b>Questão de Pesquisa:</b> 1. Quais as abordagens utilizadas para estimar a temperatura da água através de modelos estatísticos? 2. Quais modelos estatísticos são mais utilizados e com melhores desempenhos?	
<b>Estratégia de Busca</b>	
<b>Bases de Dados Científicas</b>	<b>Equipe Executora</b>
IEEE Xplore Digital Library; Springer Link; ScienceDirect; Scopus; Web of Science; Taylor & FrancisOnline; researchgate; Emprapa Publicações.	<b>1.</b> Estudante de pós-graduação; <b>2.</b> Professor orientador; <b>3.</b> Professor coorientador.

<b>Anos de Publicação</b>	2003 a 2022
<b>Keywords</b>	<b>Palavras-chave</b>
"water temperature", "statistical models temperature", "IOT", "aquaculture water temperature", "water temperature estimation", "water temperature predictions", "water temperature Neural Networks", "Water Temperature Forecasting"	"temperatura da água", "temperatura dos modelos estatísticos", "IOT", "temperatura da água da aquicultura", "estimação da temperatura da água", "previsões da temperatura da água", "temperatura da água Redes Neurais", "Previsão da Temperatura da Água"
<b>Método de Execução da Busca</b>	
1) Pesquisar nas bases de dados através da <i>string</i> de busca (busca automática); 2) Analisar os títulos e resumos dos Estudos Primários (EP) retornados nas buscas e selecionar aqueles que atendem os critérios de inclusão e questão de pesquisa; 3) Analisar os estudos que citam o estudo relacionado ( <i>backward snowballing</i> ); 4) Analisar o corpo do texto dos estudos selecionados e aplicar os critérios de qualidade para identificar os aceitos e rejeitados no contexto do mapeamento sistemático da literatura.	

Fonte: (SILVA, SOUZA e NÓBREGA, 2023a).

Uma das etapas mais importantes no escopo apresentado na Tabela 1 foi a formulação da questão de pesquisa, que proporcionou o ponto de partida para a estratégia de busca em diversas bases e o desenvolvimento de palavras-chave. Assim, foi possível planejar o método de busca a ser executado.

### 2.3 Busca por Estudos Primários

Nesta etapa, as buscas dos estudos foram executadas nas bases de pesquisas a fim de encontrar estudos relacionados com as questões de pesquisas. Os resultados trouxeram os estudos primários mais relevantes e que responderam às Questões de pesquisa. O desenvolvimento da String de busca (Tabela 2) consistiu na definição de sentenças utilizando as palavras-chave nos idiomas português e inglês alinhadas à questão de pesquisa, concatenadas por meio dos conectores lógicos OR e AND, retornando 281 EP na execução da String.

**Tabela 2. String de Busca**

"water temperature" <b>AND</b> "statistical models" <b>OR</b> IOT <b>AND</b> aquaculture <b>AND</b> estimation <b>AND</b> predictions <b>AND</b> Neural Networks <b>OR</b> Water Temperature Forecasting	"temperatura da água" <b>E</b> "modelos estatísticos" <b>OU</b> IOT <b>E</b> aquicultura <b>E</b> estimação <b>E</b> previsões <b>E</b> Redes Neurais <b>OU</b> Previsão da Temperatura da Água
--	---

Fonte: (SILVA, SOUZA e NÓBREGA, 2023a).

Observa-se que a String de busca foi elaborada tanto em português quanto em inglês com o objetivo de identificar estudos realizados por pesquisadores não somente no Brasil, mas em diversos outros países.

## 2.4 Extração de Dados

Uma vez executado a busca dos estudos, iniciou-se o estágio da extração de dados para selecionar os Estudos Primários (EP) que foram retornados de acordo com as informações obtidas nele. Além disso, foram utilizados os critérios de exclusão: EP não apresenta argumentos relacionados a medir a temperatura da água através de cálculos estatísticos; EP não apresenta argumentos relacionados a medir a temperatura utilizando redes neurais; EP não relacionado ao tema da pesquisa; EP repetido; e EP não é um artigo completo. Além disso, foram eliminados os estudos que não tinham no seu título “temperatura da água”, o que levou a 86% dos estudos serem eliminados pois não se tratava a respeito de estimação da temperatura da água.

Além do título foi feita a leitura do resumo dos EP e também aplicados os critérios de exclusão. O quantitativo dos estudos selecionados neste estágio é apresentado a seguir (Tabela 3) por base de pesquisa e ano de publicação.

**Tabela 3. Quantidade de EP selecionados na etapa de Extração de Dados**

Base de busca	Quantidade	Ano de Publicação	Quantidade
web of science	18	2003	1
Scopus	6	2007	2
IEEE	3	2009	1
Sciencedirect	5	2010	2
Springerlink	2	2012	3
Taylor&FrancisOnline	1	2013	5
EmprapaPublicações	1	2014	3
researchgate	1	2015	2
<b>Total</b>	<b>37</b>	2016	5
		2017	2
		2018	2
		2019	2
		2020	2
		2021	3
		2022	2
		<b>Total</b>	<b>37</b>

Fonte: (SILVA, SOUZA e NÓBREGA, 2023a).

Conforme apresentado na Tabela 3, foram selecionados estudos dos últimos 20 anos, sendo a maioria proveniente de bases de dados que possibilitam a realização de buscas avançadas, com o intuito de selecionar os estudos mais aderentes ao protocolo proposto.

## 2.5 Análise e Síntese

Nesta etapa, foram determinados e aplicados critérios qualitativos (CQ) (Tabela 4) que nortearam o estudo qualitativo dos EP.

**Tabela 4. Avaliação da Qualidade**

<b>Crítérios de Qualidade</b>
<ul style="list-style-type: none"><li>• CQ1. EP utiliza alguma abordagem para estimar a temperatura da água semanais ou diárias?</li><li>• CQ2. EP apresenta de forma apropriada os cálculos e métodos para determinar a temperatura?</li><li>• CQ3. EP demonstra conteúdo que pode ajudar no problema proposto nessa pesquisa?</li><li>• CQ4. EP Utiliza modelo com dados de entrada com poucas variáveis (temperatura ar)?</li></ul>

Fonte: (SILVA, SOUZA e NÓBREGA, 2023a).

Importante ressaltar que cada critério estabelecido teve como finalidade a seleção de estudos que apresentassem a possibilidade de estimar a temperatura da água de forma objetiva, utilizando dados de entrada com variáveis possíveis de serem obtidas, como dados meteorológicos, por exemplo, como estabelecido no (CQ4). Apresentar a pesquisa com os passos adequados, como a coleta de dados, avaliação e treinamento do modelo, tem um impacto fundamental na qualidade do (CQ2). Além disso, garantir que o conteúdo esteja alinhado com os problemas propostos nesta pesquisa é essencial para ajudar a resolvê-los (CQ3).

Baseado nos critérios definidos, foi possível estabelecer aceitações para aqueles que apresentassem relações com a questão de pesquisa através da leitura de cada um deles. Para cada CQ atendido foi atribuído o valor 1 e atribuído 0 em caso de não atendimento. O Índice de Aceitação (IA) corresponde à somatória do valor atribuído a cada CQ. A estratégia para a seleção foi estabelecer as condições de atendimento (AT) e não atende (NA). Os artigos aceitos (AA) foram aqueles que tiveram o IA maior ou igual a 3, e os artigos rejeitados (AR) foram os que tiveram o IA menor que 3. O Tabela 5 apresenta os artigos avaliados com os resultados de cada CQ.

**Tabela 5. Avaliação qualitativa dos EP**

ID	Referência	CQ1	CQ2	CQ3	CQ4	IA	Resultado
01	(Benyahya et al., 2013)	AT	AT	AT	AT	4	AA
02	(Piccolroaz et al., 2016)	NA	AT	NA	NA	2	AR
03	(Cara et al., 2013)	NA	NA	NA	AT	1	AR
04	(Colombo et al., 2019)	AT	AT	AT	AT	4	AA
05	(Yearsley, 2012)	NA	AT	NA	NA	1	AR
06	(Ali et al., 2015)	AT	AT	AT	NA	3	AA
07	(Caldwell et al., 2014)	AT	NA	NA	NA	1	AR
08	(Ferchichi et al., 2022)	AT	AT	NA	NA	2	AR
09	(Mcgrath et al., 2016)	AT	NA	NA	AT	2	AR
10	(Larnier et al., 2010)	AT	AT	NA	AT	3	AA
11	(Caldwell et al., 2013)	AT	AT	AT	NA	3	AA
12	(Ahmadi-Nedushan et al., 2007)	AT	AT	AT	AT	4	AA

13	(Laanaya et al., 2017)	AT	AT	AT	AT	4	AA
14	(Letcher et al., 2016)	AT	AT	NA	NA	2	AR
15	(Merran e Patterson, 2014)	AT	NA	NA	AT	2	AR
16	(Sahoo et al., 2009)	AT	AT	NA	AT	3	AA
17	(Zhu et al., 2020)	AT	AT	AT	AT	4	AA
18	(Heddam et al., 2020)	AT	AT	AT	AT	4	AA
19	(Zhu et al., 2018)	AT	AT	NA	AT	3	AA
20	(Toffolon e Piccolroaz, 2015)	AT	AT	NA	NA	2	AR
21	(Vassilis et al., 2002)	NA	AT	NA	AT	2	AR
22	(Piccolroaz et al., 2014)	NA	AT	AT	AT	3	AA
23	(Tasnim et al., 2021)	AT	AT	NA	NA	2	AR
24	(Harvey et al., 2011)	AT	AT	AT	AT	4	AA
25	(Saeed et al., 2016)	AT	AT	AT	NA	3	AA
26	(Sun e Daoliang, 2012)	NA	AT	AT	NA	2	AR
27	(Liu e Chen, 2012)	AT	AT	AT	NA	3	AA
28	(Wenxian et al., 2010)	NA	AT	AT	NA	2	AR
29	(Mohr et al., 2021)	NA	NA	NA	AT	1	AR
30	(Zhu et al., 2019)	AT	AT	AT	AT	4	AA
31	(Jiang et al., 2022)	AT	AT	NA	NA	2	AR
32	(St-Hilaire et al., 2012)	AT	NA	NA	NA	1	AR
33	(Zhu et al., 2018)	AT	AT	NA	AT	3	AA
34	(Rabi et al., 2015)	AT	AT	AT	AT	4	AA
35	(Toffolon et al., 2015)	AT	AT	NA	AT	3	AA
36	(Jeong et al., 2016)	AT	AT	AT	AT	4	AA
37	(Pike et al., 2013)	AT	AT	NA	NA	2	AR

Fonte: (SILVA, SOUZA e NÓBREGA, 2023a).

No âmbito deste estudo, foi relevante designar um número de identificação (ID) para cada um dos EP selecionados, o que viabilizou a criação de um repositório desses estudos para consultas futuras, identificando-os de maneira mais eficaz.

## 2.6 Resultados

Nessa seção são apresentados o estudo aprofundado dos 37 artigos selecionados para o mapeamento sistemático. A Tabela 6 apresenta o panorama dos EP selecionados e o resultado dos aceitos e rejeitados conforme foram analisados.

**Tabela 6. Panorama dos EP analisados**

<b>Etapa</b>	<b>EP analisados</b>	<b>EP aceitos</b>	<b>EP rejeitados</b>
Extração de Dados	281	37 (14%)	244 (86%)
Análise Qualitativa	37	20 (55%)	17 (45%)

Fonte: (SILVA, SOUZA e NÓBREGA, 2023a).

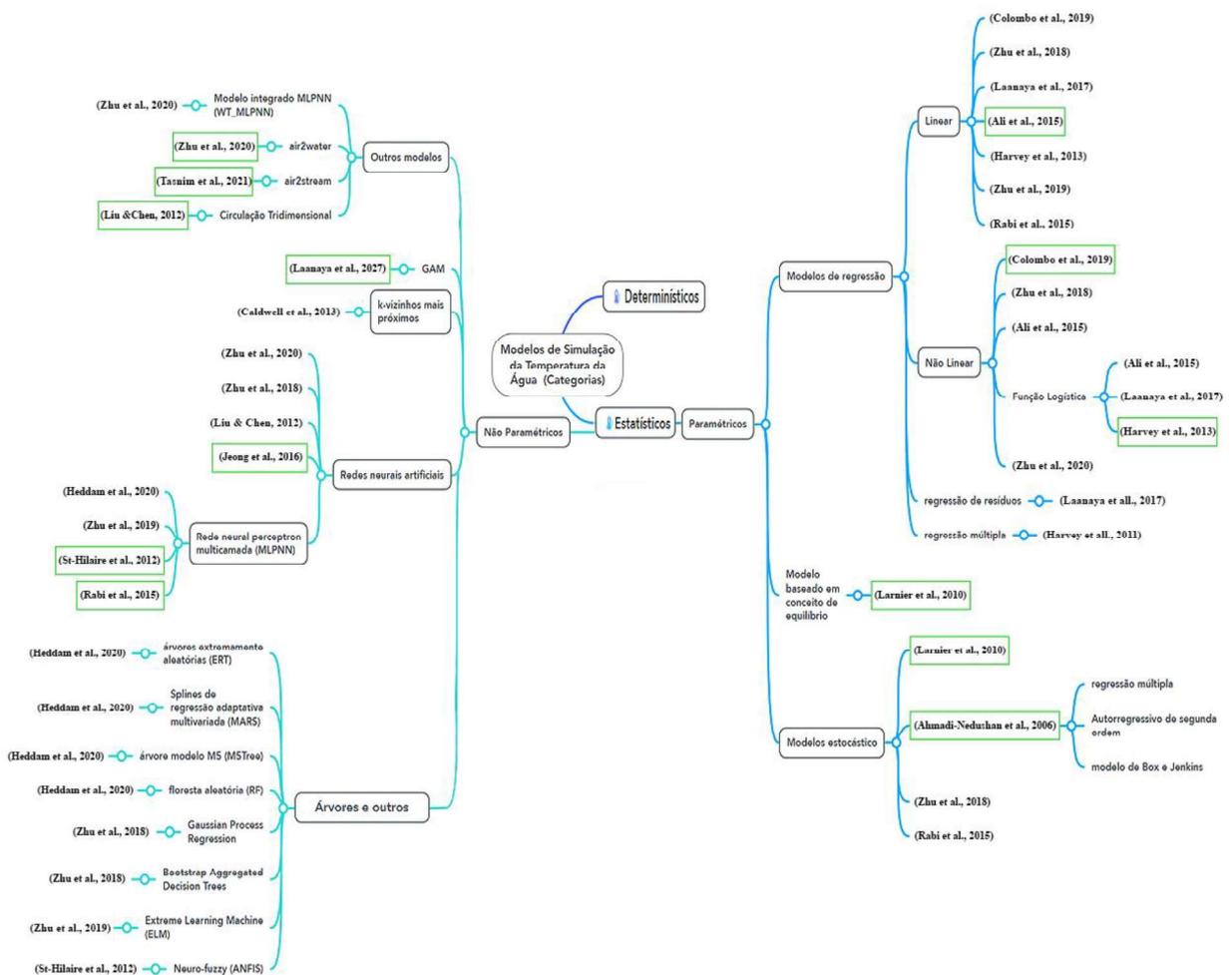
Observa-se que, na análise qualitativa, 45% dos estudos potencialmente relevantes foram rejeitados devido à falta de conformidade com a maioria dos critérios de qualidade estabelecidos.

Para tomar essa decisão, foi realizada uma leitura completa de cada estudo que foram aceitos na extração de dados.

Com o aprofundamento dos estudos analisados foi possível elaborar um mapa mental (Figura 2), com o intuito de organizar e estruturar os principais assuntos descritos em cada EP considerado aceito no contexto do MSL. Os mapas mentais são ferramentas que possibilitam representações da organização do pensamento, de forma livre ou mais elaborada, através de relações entre conceitos, hierarquicamente organizados (CORDOVIL, 2018).

No mapa mental desse estudo, foi feito uma revisão esquemática das categorias dos modelos de temperatura da água e seus respectivos autores que utilizaram as abordagens encontradas nas análises dos EP selecionados. Em cada nó do modelo, estão destacados em verde os autores que indicaram que esse modelo teve o melhor desempenho em suas pesquisas.

**Figura 2. Mapa Mental com as categorias de modelos e seus respectivos autores**



Fonte: (SILVA, SOUZA e NÓBREGA, 2023a).

No mapa mental os modelos foram agrupados em: Estatísticos – utilizados nos EP recomendados e usados nas literaturas por terem exigências mínima; e Determinísticos – apenas foram citados nos estudos não sendo possível utilizá-los porque são complexos e bastante exigentes em termo de dados, os quais, segundo os autores, geralmente não estão disponíveis.

Os modelos Estatísticos foram subdivididos em duas categorias: Paramétricos – utilizam modelo estatístico parametrizado; e Não paramétricos – a estrutura do modelo não é especificada a priori, mas sim determinada a partir de dados (BENYAHYA et al., 2013).

Os modelos Paramétricos também são divididos em categorias: Regressão – usados para modelar a temperatura da água em função, geralmente, da temperatura do ar ou de outras variáveis independentes, o qual foi subdividido em Regressão Linear e Regressão não Linear; Estocásticos – frequentemente baseados em funções estatísticas lineares em que geralmente a única informação necessária é a temperatura do ar (RABI et al., 2015).

Os modelos Não paramétricos foram divididos em: Redes Neurais Artificiais – uma estrutura matemática capaz de descrever relações não lineares complexas entre entrada e saída. (BENYAHYA et al., 2013); K-Vizinhos mais próximos, um método que consiste em encontrar, para um determinado ponto no tempo, um pequeno número de vizinhos mais próximos desse valor, e a previsão é estimada com base nesses vizinhos. Foram encontrados também outros tipos de modelos, como os de árvores.

Os EP considerados aceitos estão descritos na Tabela 7 com suas principais contribuições através de uma síntese que relata quais modelos foram utilizados e quais obtiveram melhores resultados.

**Tabela 7. EP analisados pelos critérios qualitativos**

ID	Síntese da Contribuição
1	Forneceu uma visão geral dos modelos estatísticos de temperatura da água existentes com foco nos modelos estáticos mostrando a sua relatividade simplificada e exigência de dados relativa mínima, além de fazer uma comparação entre as vantagens e desvantagens de cada modelo dependendo do problema e da disponibilidade de dados. Destacou que os modelos estocásticos tem um bom desempenho quando a temperatura do ar é o único dado disponível quando é realizado em intervalo de tempo diário.
4	Explorou formas para estimar a temperatura da água, tendo como correlação a temperatura do ar e água. O estudo utilizou regressões linear e não linear, demonstrou um modelo determinístico simplificado além de utilizar um software SWAT que tem uma equação para estimar a temperatura da água. Por fim, mostra que nesse estudo para o rio Passaúna, a utilização da média móvel de 4 dias teve um melhor ajuste das regressões e que a linear teve o melhor ajuste, independente do intervalo de dias que foi utilizado no cálculo da média.
6	Fez um estudo entre a relação da temperatura do ar-água em uma pequena lagoa aquática com o objetivo de avaliar o impacto das mudanças climáticas em diferentes atributos da lagoa. O estudo fez uma simulação da temperatura da água superficial diariamente, semanalmente, mensalmente. Analisou dados de temperatura da água e do ar de 5 anos, e avaliou dois modelos estatísticos paramétricos. Modelo de regressão linear simples (SLR) e regressão logística não linear de quatro parâmetros (NLR). O modelo de regressão simples

	apresentou melhor desempenho em escalas de tempo semanais e mensais, para escalas de tempo diárias o erro foi considerado relativamente alto.
10	Utilizou e avaliou a eficiência de dois modelos estatísticos para prever a temperatura da água: modelo baseado em conceito de equilíbrio e os modelos estocásticos autorregressivos com o objetivo de avaliar a influência climáticas na temperatura da água do rio verificando se a temperatura do ar é o principal fator que influenciou o regime térmico do rio nas últimas duas décadas. Ambos os modelos mostraram um bom desempenho ao modelar as temperaturas médias diárias da água para rios grandes, com erro médio quadrático para temperaturas acima de 20°.
11	Desenvolveu uma estrutura de modelagem estatística para prever a temperatura diária e horárias. Utiliza dados 1916-2006. Acoplou um modelo linear generalizado na escala de tempo diária com um algoritmo de reamostragem k-vizinho mais próximo que permite a desagregação de valores diários para estimativas horárias de temperatura do fluxo.
12	Comparou a eficiência de vários modelos estocásticos para prever a temperatura média diária da água usando apenas a temperatura do ar como variável exógena, utilizando dados do rio Moisie, além de explorar a possibilidade de incluir mais uma variável a vazão como variáveis independentes adicionais em modelos estocásticos. Os modelos utilizando apenas a temperatura do ar foram: um modelo de regressão múltipla, um modelo autorregressivo de segunda ordem e um modelo Box e Jenkins. O modelo auto regressivo de segunda ordem teve melhor resultado. Para combinar com a vazão do rio e a temperatura do ar, uma metodologia estatística usando regressão é proposta e mostrou uma pequena melhora em relação aos três primeiros que não incluíram índices de vazão.
13	Avaliou um modelo aditivo generalizado não paramétrico de temperatura de água (GAM) com base na temperatura média diária do ar e na vazão média do rio e fez uma comparação com três modelos: modelos logísticos, de regressão de resíduo e de regressão linear. O GAM teve um desempenho melhor entre os outros. O modelo de regressão de resíduos teve melhor resultado do que o logístico, e o modelo linear foi mais adequado do que o de regressão de resíduo e logístico.
16	Desenvolveu um modelo de redes neurais artificiais, um modelo de regressão linear múltipla e algoritmos dinâmicos não lineares e os examinou para prever a temperatura da água do córrego usando variáveis meteorológicas disponíveis, fez a comparação e sugeriu o melhor modelo otimizado. A eficiência de desempenho de IGA-4BPNN foi o mais alto entre todos os algoritmos.
17	Desenvolveu dois modelos (MLPNN) de redes neurais e (WT_MLPNN). Eles foram comparados com dois outros modelos amplamente utilizados. Air2water e um modelo de regressão não linear. Os resultados mostram que o air2water tem o melhor desempenho, seguido pelo (WT_MLPNN), (MLPNN) e o modelo de regressão não linear que teve o pior desempenho.
18	Compara uma nova família de algoritmos de aprendizado de máquina com o modelo air2stream. Os modelos construídos foram: as árvores extremamente aleatórias (ERT), as <i>splines</i> de regressão adaptativa multivariada (MARS), a árvore modelo M5 (M5Tree), a floresta aleatória (RF) e a rede neural <i>perceptron</i> multicamada (MLPNN). Utilizando a temperatura do ar como variável de entrada. Os resultados mostraram que o modelo air2stream superou todos os outros modelos de aprendizado de máquina.
19	Estimou a temperatura diária do rio através do desenvolvimento de três modelos padrão: regressão linear, regressão não linear e modelos estocásticos. E três modelos de aprendizagem de máquina: <i>Artificial Neural Network</i> (ANN), <i>Gaussian Process Regression</i> (GPR) e <i>Bootstrap Aggregated Decision Trees</i> (BA-DT). Os resultados mostraram que o modelo estocástico superou os modelos linear padrão e os modelos de aprendizado de máquina, superaram o modelo estocástico
22	Desenvolveu um modelo simples baseado em física Air2Water, para estimar a temperatura da água a partir da temperatura do ar. Os resultados mostram que o modelo representa uma boa alternativa para os modelos baseados em processos completos.
24	Desenvolveu modelos de regressão para prever a temperatura média, máxima e mínima da água e a concentração de oxigênio dissolvido nas escalas de tempo mensal, semanal e diária. Regressão linear, dois modelos logísticos não lineares usando a temperatura do ar e regressão múltipla usando temperatura do ar e o nível de estágio. O primeiro modelo logístico foi melhor que o modelo linear ao lidar com temperaturas máximas e mínimas da água. O modelo de regressão múltipla não supera o modelo logístico.
25	Comparou o desempenho de algumas técnicas de computação suave, para prever temperaturas horárias da água do lago. Usando um conjunto de dados de alguns meses do ano 2008. A técnicas foram: programação de expressão gênica (GEP), sistema de inferência <i>neurofuzzy</i> adaptativo (ANFIS) e redes neurais artificiais (RNAs). Os resultados

	demonstraram que o GEP forneceu tendências razoavelmente melhores em diferentes profundidades.
27	Usou um modelo de rede neural artificial (RNA), e um modelo de circulação tridimensional para prever a temperatura da água do lago. Os resultados simulados foram comparados e mostraram que o modelo de circulação tridimensional, fornece uma melhor precisão da temperatura da água.
30	Desenvolveu um modelo baseado em <i>Extreme Learning Machine</i> (ELM) para prever a temperatura diária da água de rios. Utilizou como entrada A temperatura do ar (Ta), a vazão (Q) e o dia do ano (DOY). E comparou com modelos de rede neural multicamada <i>perceptron</i> (MLPNN) e regressão linear múltipla simples (MLR). Três cenários foram testados: 1. Usando apenas a temperatura do ar; 2. Usando a temperatura ar e vazão e 3. Usando as duas primeiras e a temperatura do dia. Os resultados mostraram que a inclusão de três produziu a melhor precisão. Além disso, os modelos ELM e MLPNN superaram o modelo MLR, foi concluído que no geral o modelo ELM pode ser usado para previsão de temperatura da água do rio.
33	Foi desenvolvido um modelo de rede neural <i>perceptron</i> multicamadas (MLPNN) e três modelos sistemas de inferência neuro-fuzzy adaptativos (ANFIS). Algoritmo de agrupamento <i>fuzzy c-mean</i> (ANFIS_FC), ANFIS com método de partição de grade (ANFIS_GP) e ANFIS com método de agrupamento subtrativo (ANFIS_SC), foram implementados para simular a temperatura diária da água do rio. 1. usando apenas a temperatura do ar (Ta), 2. (Ta) e vazão do rio (Q) e 3. (Ta, Q) e os componentes do calendário gregoriano (CGC) como preditores. Os resultados da modelagem mostraram que a inclusão de três entradas como preditores rendeu a melhor precisão de modelagem entre todos os modelos desenvolvidos. Além disso, o modelo MLPNN foi geralmente o que apresentou os melhores desempenhos.
34	Analisou modelos de regressão linear, modelagem estocásticas e modelagem de redes neurais <i>feedforward</i> de <i>perceptron</i> multicamadas (MPL), para modelar a temperatura diária da água do rio usando apenas a temperatura do ar como parâmetro analisando a relação da temperatura média do ar e diária da água coletadas em alguns pontos do rio. Os resultados mostraram que o modelo estocástico é melhor que o modelo de regressão linear e os modelos MPL superam os modelos estocásticos.
35	Formulou versões do modelo air2stream para prever a temperatura média diária da água do rio em função da temperatura do ar e da vazão.
36	Apresentou um modelo de previsão baseado em rede Neural Artificial (RNA) para a temperatura da água do rio, projetado para facilitar a previsão em tempo real. Foi utilizado apenas dados de previsão do tempo (temperatura do ar e precipitação) que pode facilitar a previsão automatizada e em tempo real da temperatura da água.

Fonte: (SILVA, SOUZA e NÓBREGA, 2023a).

As principais contribuições descritas na Tabela 7 consistem na apresentação dos modelos utilizados pelos autores, nas variáveis meteorológicas empregadas e nos modelos que apresentaram melhor desempenho na validação em relação à estimativa da temperatura da água dos rios.

## 2.7 Discussões

Nesta seção são apresentadas as análises dos resultados encontrados para responder à questão de pesquisa descrita diante dos EP que foram selecionados, com a proposta de mostrar abordagem para estimar a temperatura da água através de métodos estatísticos. Diante disso, foi possível identificar diversas categorias de modelos estatísticos capazes de prever a temperatura da água em escalas diárias, semanais, mensais e anuais, utilizando mínimo dados de entrada (temperatura do ar) e em alguns casos outras variáveis independentes.

A maioria dos estudos não utilizam apenas um modelo de temperatura da água, eles fazem uma comparação de resultados e mostram quais tiveram um melhor desempenho para um

determinado rio e uma escala de tempo. Os dados coletados para variáveis de entradas foram de estações próximas dos rios, no entanto, muitas vezes a falta de dados em alguns períodos de tempo foi um desafio para alguns pesquisadores por deixarem algumas lacunas prejudicando uma melhor eficiência na estimação, vale ressaltar também que a temperatura do ar foi destacada como preditor mais importante para a previsão da temperatura da água.

Diante dos resultados do mapeamento sistemático, com resposta à questão de pesquisa formulada em relação aos modelos estáticos paramétricos, temos que: 16 EP utilizaram modelo de regressão, 7 EP utilizaram regressão linear, 7 EP utilizaram regressão não linear, dentre estes últimos 4 baseados no desenvolvimento de uma função logística, 1 de regressão de resíduos e 2 de regressão múltipla. Ainda como modelos paramétricos foi identificado 1 EP utilizando modelo baseado no conceito de equilíbrio e 4 EP que utilizaram modelos estocásticos. Em relação aos modelos não paramétricos, temos que: 1 EP utilizou k-vizinhos mais próximos, 8 EP utilizaram redes neurais artificiais, 4 EP utilizaram modelos híbridos e 8 EP optaram também por aprendizagem de máquina como árvores.

O objetivo em comum dos estudos foi prever a temperatura futura da água a fim de analisar quais os riscos da temperatura alta dos rios pelo aquecimento global e desmatamento, para preservação das espécies e gerenciamento de pesca. Em vista disso, a previsão foi feita baseadas em escalas anual, mensal e diária, e a Tabela 8 é o resultado dos modelos que tiveram melhor desempenho para cada escala de tempo.

**Tabela 8. Modelos com melhores desempenho de acordo com as escalas de tempo**

<b>Anual</b>	<b>Mensal</b>	<b>Semanal</b>	<b>Diária</b>
Regressão linear	Regressão linear	Regressão linear	Regressão Não linear
			Estocástico
			Redes Neurais
			Árvores

Fonte: (SILVA, SOUZA e NÓBREGA, 2023a).

Importante ressaltar que, para esse estudo, as escalas de tempo se referem aos dados meteorológicos extraídos em que os dados diários podem também ser obtidos em horários específicos para execução do treinamento dos modelos estatísticos.

Apesar de relativamente eficientes, os modelos de regressão em escala de tempo menor que semanal são mais difíceis de aplicar devido às autocorrelações na estrutura das séries temporais de temperatura da água. Nesses casos, modelos estocásticos e não paramétricos como Redes Neurais Artificiais (RNA) apresentaram melhores resultados (BENYAHYA et al., 2007; CHENARD;

CAISSIE, 2008). Contudo, o modelo de regressão pode ser usado para simular a temperatura diária da água com dois termos defasados de temperatura da água, temperatura do ar e vazão como variáveis exógenas. A maioria dos estudos de regressão reconhece que existe um termo de erro (BENYAHYA et al., 2013). Outrossim, os modelos de regressão linear podem não prever temperaturas de fluxo em altas temperaturas. Em temperaturas muito altas ou muito baixas, a relação entre a temperatura do ar e da água pode não seguir um padrão linear, mas sim se assemelhar a uma curva em forma de "S". Ou seja, a variação na temperatura do fluxo da água pode não ser bem descrita por uma linha reta ao longo de uma ampla gama de temperaturas. (MOHSENI et al., 1998).

Apesar da previsão da temperatura da água ser essencial para a análise dos riscos para as espécies, o monitoramento em tempo real da temperatura é também de extrema importância. No entanto, apenas um EP analisado tratou de uma situação que pode facilitar a previsão automatizada e em tempo real da temperatura da água. A falta de EPs relacionados à estimação em tempo real é compreendida pelo fato de serem utilizados sensores de monitoramento.

Portanto, faz-se necessário fornecer um meio de estimar a temperatura da água em ambientes de criação aquáticos, que resulte em uma tecnologia acessível onde não houver um equipamento físico de medição. Essa alternativa pode ser aplicada em diversas áreas, como para auxiliar a produção de pequenos produtores de tanques de aquicultura, podendo ser esta lacuna ser vista como uma oportunidade de pesquisa.

Apesar de alguns relatos entre os estudos visarem a gestão da pesca e da aquicultura, observou-se que modelos de previsão de temperatura da água não são desenvolvidos quando para aplicação em tanques de criação. A previsão da temperatura da água nos tanques é um fator crucial que pode ajudar o produtor na tomada de decisão.

## **2.8 Considerações Finais**

Neste capítulo foi apresentado e discutido o planejamento e os resultados de um MSL com o objetivo de encontrar EP que utilizam modelos estatísticos para estimar a temperatura da água a partir de poucas variáveis disponíveis e que esses pudessem demonstrar onde e como estes modelos são utilizados. Dentre os modelos existentes analisados, os de regressão e os de redes neurais foram o mais utilizados e demonstraram bons resultados em diversos estudos.

Fica evidente, a partir dos EP analisados, que o objetivo desses modelos é prever a temperatura futura da água dos rios para tomadas de decisões a fim de preservar as espécies. Alguns estudos

também relatam a aplicação de modelos em costas marinhas, mas não foi encontrado qualquer relato sobre a aplicação dos modelos em outros cenários em que o monitoramento da temperatura da água é importante.

Nos tanques e viveiros de aquicultura, o período de criação é curto e depende de um acompanhamento contínuo em tempo real da temperatura da água. Portanto, constitui-se uma oportunidade de pesquisa desenvolver uma abordagem de estimação da temperatura da água para o pequeno produtor que não possui um instrumento físico de medição.

Ao apresentar esses modelos em um contexto de uma revisão de modelos estatísticos para prever a temperatura da água, os autores enfatizaram a natureza estatística do processo de modelagem e previsão. No entanto, atualmente a maioria dos modelos demonstrados são conhecidos popularmente como modelos de aprendizado de máquina, como por exemplo, a aplicação de métodos estatísticos tradicionais como regressão linear, juntamente com abordagens mais avançadas de aprendizado de máquina, como redes neurais perceptron, para fornecer uma visão abrangente da modelagem estatística na previsão da temperatura da água.

A classificação de modelos de aprendizado de máquina como modelos estatísticos muitas vezes destaca as semelhanças na fundamentação teórica e na abordagem de otimização de parâmetros, mesmo que os métodos de aprendizado de máquina representem uma extensão mais flexível e adaptável da estatística tradicional. Diante disso, nos próximos capítulos esses modelos serão descritos como modelos de aprendizagem de máquina.

Como limitação, é importante ressaltar que uma pesquisa desta natureza possui riscos e ameaças à sua validade dos resultados. Para este estudo, apesar de serem utilizados critérios de inclusão e exclusão no processo de seleção dos EPs devido o número de artigos encontrados nas bases de dados, as palavras-chave e string de buscas, mesmo sendo refinadas, pode não ter contemplado todos os estudos existentes e pesquisas importantes podem não ter sido incluídas. Além disso, alguns títulos e resumos não apresentaram de forma adequada o conteúdo que respondesse à questão de pesquisa, impossibilitando encontrar todos os estudos relevantes.

Desta forma, este capítulo empreendeu uma investigação abrangente dos estudos relacionados ao tema, explorando as diversas abordagens de estimação da temperatura da água. O objetivo foi fundamentar as pesquisas realizadas em diversos modelos, visando desenvolver uma abordagem mais eficaz para a estimativa da temperatura da água. Essa iniciativa busca proporcionar uma alternativa que não apenas aprimore a tomada de decisões, mas que também tenha impacto direto na gestão da água e na produção de pescado, promovendo práticas mais sustentáveis. Este esforço

se insere em um contexto mais amplo, abrangendo a transição para a Aquicultura 4.0 e atendendo às demandas de monitoramento em diversas áreas.

### **3 ARCABOUÇO TEÓRICO**

Este capítulo destaca os principais trabalhos relacionados encontrados na literatura atual que convergem com os objetivos propostos nesta pesquisa. São abordados temas como: Aquicultura 4.0 seus benefícios e desafios tecnológicos, além da importância da temperatura da água e do seu monitoramento; Internet das Coisas (IoT), conceituação, importância e coleta de dados; Aprendizado de máquina, os fundamentos e seus diferentes tipos de aprendizado e as técnicas de validação; Sensores Virtuais, definição, vantagens, aplicações e os desafios associados; e Gestão de recursos hídricos, fundamentos, importância e os efeitos da temperatura da água.

### **3.1 Visão Geral**

Os modelos de aprendizado de máquinas são amplamente utilizados em diversas aplicações. Esses modelos se concentram no desenvolvimento de algoritmos que são capazes de aprender a partir de dados e fazer previsões ou tomar decisões com base nesse aprendizado. As aplicações de tecnologias avançadas, como Internet das Coisas (IoT do inglês *Internet of Things*), tem mudado o panorama de diversas áreas, principalmente pelo fato da IoT desempenhar um papel essencial na obtenção de dados em tempo real. Esses dados, uma vez coletados, permitem que os modelos de aprendizado de máquinas possam processar e analisá-los, podendo identificar padrões, tendências e correlações entre os diferentes parâmetros medidos.

A análise de dados complexos desempenha um papel fundamental em áreas como a aquicultura, uma vez que a aplicação dessas tecnologias permite resolver diversos tipos de problemas relacionados e ajuda na tomada de decisões embasadas, como na estimação da temperatura da água por meio de sensores virtuais, contribuindo para a Aquicultura 4.0. Além disso, a análise de dados pode ser aplicada em outros corpos de água como rios e reservatórios, e diversos pesquisadores já fizeram isso como, por exemplo (AHMADI-NEDUSHAN et al., 2007), que compararam a eficiência de vários modelos para prever a temperatura média diária da água utilizando apenas a temperatura do ar como variável exógena, utilizando os dados do rio Moiseie.

Diante disso, como forma de contribuir para gestão dos recursos hídricos, os modelos de aprendizagem de máquina podem fornecer informações valiosas para a tomada de decisões ao analisar os dados históricos da temperatura da água e do ar ou outros dados meteorológicos quando estão disponíveis, bem como desenvolver sensores virtuais como alternativa aos sensores físicos que pode não ser possível serem instalados em áreas remotas.

### **3.2 Aquicultura**

Segundo Rodrigues et al. (2012), a aquicultura é o processo de produção em cativeiro de organismos com habitat predominantemente aquático, tais como peixes, camarões, rãs, algas, entre outras espécies. Esses ambientes controlados podem ser tanques, lagoas ou jaulas submersas em corpos d'água, utilizando diferentes sistemas dependendo das características das espécies a serem cultivadas.

O Brasil, com seu clima tropical e abundância de recursos hídricos, apresenta condições propícias para o desenvolvimento da aquicultura. O país possui aproximadamente 4,2 milhões de hectares de áreas represadas em grandes reservatórios, oferecendo um potencial significativo para

a criação de peixes por meio de tanques-rede. No entanto, é importante destacar que apenas uma pequena parcela dessas áreas tem sido aproveitada para essa finalidade (KUBITZA, 2015).

De acordo com a Embrapa (2022), o Brasil ainda não ocupa uma posição de destaque no cenário mundial da produção aquícola, e os aumentos registrados nos últimos anos não têm sido significativos. A produção de pescado, por exemplo, passou de 578 mil toneladas em 2014 para pouco mais de 802 mil toneladas em 2020. No entanto, é importante ressaltar que, de acordo com o site PeixeBR, a produção de peixe saltou apenas para 860.355 toneladas em 2022.

Esses números evidenciam um progresso lento do país no aumento da produção aquícola, o que é preocupante, considerando a importância desse setor para o fornecimento de proteína animal à população. A demanda por essas proteínas vem crescendo de forma expressiva tanto no Brasil como no mundo, e a aquicultura tem potencial para suprir parte desse aumento, além de reduzir a pressão sobre as proteínas vegetais na alimentação (RODRIGUES et al., 2012). No entanto, é necessário um esforço maior para impulsionar o desenvolvimento da aquicultura no país e alcançar níveis mais significativos de produção.

O crescimento da aquicultura no Brasil é limitado por uma série de fatores. Segundo Kubitzka (2015), produtores de peixes e camarões apontam dificuldades na obtenção de licenças ambientais, ausência de políticas específicas para o desenvolvimento do setor, falta de acesso ao crédito, altos custos de produção e escassez de tecnologia disponível como principais obstáculos para o crescimento da produção aquícola no país. Para impulsionar o crescimento da aquicultura no Brasil é fundamental enfrentar esses desafios, principalmente no que diz respeito à tecnologia, que pode ser um dos principais meios para superar esses fatores limitantes.

A pesquisa aplicada ao setor e a implementação de novas tecnologias exerce um papel fundamental na melhoria da produtividade na aquicultura (RODRIGUES et al., 2012). Dentre as diversas áreas de estudo, destaca-se a importância da qualidade da água na aquicultura, sendo a temperatura um dos fatores mais significativos para o sucesso da atividade. A temperatura da água continua sendo um dos parâmetros mais relevantes, uma vez que influencia diretamente a capacidade produtiva dos ecossistemas aquáticos e afeta o crescimento e desenvolvimento dos organismos aquáticos (SUN; DAOLIANG, 2012).

A temperatura da água dos rios tem impacto nos habitats dos peixes e pode afetar a mortalidade de organismos aquáticos (RABI et al., 2015). O aumento da mortalidade devido ao aumento das temperaturas pode reduzir a abundância de peixes disponíveis para atender aos objetivos de colheita ou de desova (MERRAN; PATTERSON, 2014). O monitoramento da

temperatura da água é um fator considerável para prevenir problemas na aquicultura, como mortes dos peixes e crescimento (LUO et al., 2013). O aumento da temperatura também pode levar a problemas graves, como o aumento na incidência de doenças parasitárias. Isso ocorre devido ao prolongamento da temporada de transmissão, o que aumenta o risco de infecções. (KARVONEN et al., 2010). Garantir uma boa qualidade sustentável da água é fundamental para o sucesso da atividade (NGUEKU, 2014). Realizando um monitoramento contínuo o produtor pode tomar decisões para garantir que a temperatura esteja em boas condições como o uso de sombra que reduz significativamente a temperatura da água durante os meses de verão (GOTHREAUX; GREEN, 2012). Embora o monitoramento manual de dados possa ser utilizado (RAND et al., 2022), é importante ressaltar que algumas características físicas da água, como a temperatura, podem sofrer alterações rápidas. Nesses casos, a coleta de amostras realizada de forma espaçada, como normalmente ocorre diariamente ou semanalmente, pode não ser suficiente para perceber variações que podem levar a prejuízos na criação, como redução no ganho de peso e aumento do risco de doenças, estresse e mortalidade (CARA et al., 2013).

Um sistema de monitoramento de temperatura da água baseado em dispositivos móveis em tempo real permite diminuir a dependência de mão de obra no monitoramento local (BOKINGKITO; LLANTOS, 2017). No entanto, o difícil acesso a tecnologias de ponta representa desafios que colocam a piscicultura nacional em uma situação de desvantagem em relação ao avanço da aquicultura globalmente (RESENDE, 2009). Embora existam alternativa de dispositivos de baixo custo, em algumas áreas a atual aceitação destes dispositivos tem sido lenta (OBIERO et al, 2019).

Como alternativa, têm sido empregados sistemas de monitoramento de baixo custo da qualidade da água, utilizando tecnologias emergentes, como a Internet das Coisas (IoT) e aprendizado de máquina, os quais têm demonstrado um desempenho satisfatório (KODITALA; PANDEY, 2018). Esses avanços tecnológicos oferecem oportunidades promissoras para a aquicultura, permitindo o monitoramento contínuo e preciso da qualidade da água a um custo acessível.

### **3.3 Internet das Coisas (IoT)**

De acordo com Silverio-Fernández et al. (2018), a Internet das Coisas (IoT) é uma rede interconectada de objetos que vão desde simples sensores até smartphones e tablets. O termo IoT geralmente se refere a cenários em que a rede, a capacidade de conectividade e a computação se

estende a objetos, sensores e itens do dia a dia que normalmente não são considerados computadores, permitindo que esses dispositivos gerem, troquem e consumam dados com mínima intervenção humana. (ROSE et al, 2015).

A IoT é uma das tecnologias mais relevantes e abrangentes da economia digital no século XXI (CHAUDHURI, 2018). Com o avanço da conectividade e da capacidade de coleta de dados, a IoT tem o potencial de transformar diversos setores, desde a indústria até a saúde, transporte, agricultura, entre outros. Na indústria de produção de pescados, a aquicultura inteligente ou aquicultura digital envolve o uso de tecnologias avançadas e abordagens baseadas em dados para melhorar a eficiência, sustentabilidade e produtividade, apesar de alguns desafios tecnológicos comuns como limpeza e manutenção de sensores que pode aumentar o custo geral da solução baseada em IoT, além dos desafios específicos como a infraestrutura das fazendas sem acesso à eletricidade ou acesso à Internet (RASTEGARI et al., 2023).

A IoT oferece oportunidades para aprofundar o conhecimento sobre novos negócios que utilizam tecnologias como Big Data para coletar dados relevantes e embasar a tomada de decisões de forma mais informada e precisa (OLIVEIRA; FREITAS., 2016). A capacidade de coletar e analisar dados em tempo real por meio da IoT impulsiona a inovação e o desenvolvimento de soluções inteligentes em diversos setores. As aplicações de IoT têm crescido em números exorbitantes, gerando uma grande quantidade de dados necessários para o processamento inteligente de dados (ADI et al., 2020).

Segundo Kamal (2022), a coleta de dados nos dispositivos IoT ocorre através da captura dos dados na fonte, utilizando sensores incorporados nos próprios dispositivos ou sensores externos instalados em dispositivos que são conectados e monitorados. Esses dados são transmitidos para um sistema onde são armazenados e organizados. Em seguida, esses dados são consumidos e analisados por um sistema, permitindo a geração de insights acionáveis para uma boa tomada de decisão.

Os dados coletados na IoT têm o potencial de serem utilizados no desenvolvimento de novas tecnologias inovadoras, como o aprendizado de máquina. Enormes quantidades de dados brutos são gerados todos os dias em várias áreas. Essa disponibilidade de dados em grandes tamanhos e a quantidade crescente chamam a atenção para o estudo do conceito de aprendizado de máquina. (DOGAN; BIRANT, 2021). Esses dados fornecem informações valiosas que podem ser exploradas por algoritmos de aprendizado de máquina para treinar modelos e melhorar a capacidade de acurácia.

Atualmente, o uso de modelos de aprendizado profundo para análise de dados em aplicativos IoT tem experimentado um crescimento significativo. Isso se deve ao fato de que os modelos de aprendizado profundo são especialmente adequados para lidar com a análise de dados altamente complexos gerados por dispositivos IoT (SALEEM; CHISHTI, 2019), considerando que a capacidade de extrair padrões e informações de alto nível a partir desses dados complexos é um dos principais motivos para crescente adoção desta abordagem.

### **3.4 Aprendizado de máquinas**

Segundo Naga e Murphy (2015), um algoritmo de aprendizado de máquina é um processo computacional que usa dados de entrada para realizar uma tarefa desejada sem ser literalmente programada para produzir um resultado específico. Esses algoritmos são, de certo modo, "*soft coded*", ou seja, não são rígidos para produzir um resultado específico. Em vez disso, alteram ou adaptam automaticamente sua arquitetura por meio da repetição, para que se tornem cada vez melhores na realização da tarefa desejada. Essas interações repetidas são conhecidas como treinamento com o objetivo de melhorar seu desempenho aprimorando suas capacidades.

Os modelos de aprendizado de máquina também são conhecidos como modelos estatísticos por utilizarem métodos estatísticos para inferir relações entre variáveis e aprender com os dados fornecidos. No contexto da temperatura da água, (BENYAHYA et al., 2013) forneceram uma visão geral desses modelos quando aplicados à temperatura da água, destacando a sua relativa simplicidade e exigência na entrada de dados. A contribuição desse estudo supracitado forneceu informações valiosas para pesquisas futuras, uma vez que destaca as vantagens e desvantagens de cada modelo dependendo do contexto específico.

No contexto do aprendizado de máquina, existem diferentes tipos de aprendizado, sendo os mais conhecidos o aprendizado supervisionado e o aprendizado não supervisionado. Cada um possui características e aplicações específicas, e a escolha entre eles depende do tipo de problema em questão, bem como da disponibilidade de dados e recursos disponíveis. Para Sathya e Abraham (2013), o aprendizado supervisionado é baseado no treinamento de uma amostra de dados da fonte com classificação correta já atribuída. Ou seja, o modelo é treinado usando um conjunto de dados em que as saídas desejadas já são conhecidas, desse modo, o objetivo do modelo é mapear corretamente os dados de entradas para saídas correspondentes.

O aprendizado não supervisionado, ao contrário do aprendizado supervisionado, utiliza um conjunto de dados não rotulados. O objetivo é encontrar padrões, estruturas ou agrupamentos nos

dados sem a necessidade de saídas pré-determinadas. Segundo Dayan (1999), o aprendizado não supervisionado estuda como os sistemas podem aprender a representar determinados padrões de entrada de uma maneira que reflita a estrutura estatística da coleção geral de padrões de entrada.

Neste trabalho, foi adotado o aprendizado supervisionado devido à disponibilidade de dados rotulados, ou seja, com saídas desejadas conhecidas. Além disso, a quantidade de amostras utilizadas foi relativamente baixa, devido a falha nos sensores e na transmissão de dados, o que reforçou a motivação para a escolha do aprendizado supervisionado como abordagem principal. De acordo com Mahesh (2020), “se você tem uma quantidade menor de dados e dados claramente rotulados para treinamento, opte pelo Aprendizado Supervisionado”.

Os algoritmos de aprendizagem de máquina também são divididos em grupos, dentre os principais estão os algoritmos de classificação e de regressão. De acordo com Terra (2023), o objetivo do algoritmo de regressão é encontrar a função de mapeamento para que possamos mapear a variável de entrada  $X$  para a variável de saída contínua, ou seja, um valor número real com base nas características dos dados de entrada encontrando correlação entre essas variáveis.

Por outro lado, os algoritmos de classificação têm o objetivo de atribuir uma categoria ou classe a um determinado conjunto de dados. Os algoritmos de classificação são usados para estimar valores discretos ou categorias como, por exemplo, 0 ou 1, sim ou não, verdadeiro ou falso, com base em um conjunto específico de variáveis independentes. Eles utilizam técnicas estatísticas e matemáticas para mapear as características dos dados e fazer previsões sobre a probabilidade de ocorrência de um evento, ajustando os dados a uma função logit (TERRA, 2023). Essa função logit transforma a probabilidade em uma escala logística, permitindo a classificação dos dados em categorias distintas.

Para garantir a eficácia e a capacidade de generalização dos modelos de aprendizado de máquina, é essencial avaliá-los por meio de técnicas como a divisão de dados e a utilização de métricas de avaliação. Na divisão de dados, o conjunto de dados é separado em um conjunto de treinamento e um conjunto de testes. O modelo é treinado com o conjunto de treinamento e, em seguida, avaliado com o conjunto de testes para medir seu desempenho. Essa técnica é amplamente utilizada por pesquisadores, como mencionado em estudos como (FERCHICHI et al., 2022) e (Zhu et al., 2018). Ao utilizar essa abordagem, é possível verificar como o modelo se comporta com dados que não foram utilizados durante o treinamento, permitindo avaliar sua capacidade de generalização e o quanto ele realmente aprendeu.

Além da divisão de dados, também é comum o uso de métricas de avaliação para quantificar o desempenho do modelo. Pesquisas feitas utilizando aprendizagem de máquinas para estimar, simular ou prever a temperatura da água, utilizaram comumente o RMSE (*Root Mean Squared Error*) (LAANAYA et al., 2017; AHMADI-NEDUSHAN et al., 2007; CALDWELL et al., 2013) e destacaram bem o uso como forma de avaliar os modelos. O RMSE é a raiz quadrada da média dos erros quadrado e fornece uma medida do erro médio do modelo em relação aos valores reais. quanto menor o valor do RMSE, melhor o desempenho do modelo.

Outra métrica utilizada no estudo de (SAHOO et al. 2009) foi o MSE (*Mean Squared Error*), uma métrica que mede a média dos erros ao quadrado entre as previsões do modelo e os valores reais. Quanto menor o valor do MSE, melhor o desempenho do modelo. (ZHU et al. 2020) utilizou também o MAE (*Mean Absolute Error*), uma métrica que mede a média dos valores absolutos dos erros entre as previsões do modelo e os valores reais. O MAE é menos sensível a valores extremos do que o MSE e o RMSE, tornando-se uma métrica adequada em algumas situações. Uma opção é remover os valores extremos do conjunto de dados antes de treinar o modelo, além disso, utilizar modelos de aprendizagem de máquinas que são menos sensíveis a valores extremos (outliers). Outra métrica amplamente conhecida é o  $R^2$  (*Coefficient of Determination*), métrica que avalia o quão bem o modelo se ajusta aos dados em relação à variação total dos dados. Segundo Yearsley (2012), “o número e o tipo de métricas usadas para avaliar o desempenho do modelo varia de estudo para estudo, mas todos relatam alguma medida do segundo momento da diferença entre o simulado (previsto) e o observado”.

### **3.5 Sensores Virtuais**

Conforme destacado por Raveendranathan et al. (2011), sensores virtuais são abstrações de componentes de sistemas, englobando a amostragem de sensores e tarefas de processamento para fornecer dados em resposta a solicitações externas. Além disso, esses sensores também denominados sensores soft têm sido amplamente empregados para fins de estimação, como observado em estudos como o de Fujiwara et al. (2009). Essa caracterização dos sensores virtuais como abstrações flexíveis e eficientes ressalta sua utilidade em diversas aplicações e destaca sua crescente importância no campo da coleta e análise de dados.

Uma das vantagens notáveis do uso de sensores virtuais é sua frequente adoção como alternativa aos sensores físicos, beneficiando-se da redução dos custos de hardware, enquanto emprega técnicas de aprendizado de máquina como um paradigma virtual de IoT,

(DRAKOULELIS et al., 2021). Além disso, em ambientes que já possuem sensores físicos, os sensores virtuais podem ser empregados para substituir sensores defeituosos, recorrendo a dados históricos e treinando-se automaticamente. Essa abordagem garante que, quando sensores físicos deixam de funcionar, os sensores virtuais entram em ação, assegurando a continuidade do monitoramento sem interrupções (ILYAS et al., 2020). Esse uso estratégico dos sensores virtuais reforça sua eficácia e versatilidade na otimização do monitoramento em diversos contextos.

Os sensores virtuais são frequentemente desenvolvidos através da aplicação de técnicas de aprendizado de máquina, envolvendo a coleta de conjuntos de dados e o treinamento de um preditor cujas entradas consistem nas variáveis capturadas (MASTI et al., 2021). O aprendizado de máquina oferece abordagens para modelar padrões em dados complexos e em big data provenientes de dispositivos de sensoriamento IoT, viabilizando assim a implementação de sensores virtuais (POLANSKI; ANGELOPOULOS, 2022). A integração eficaz entre aprendizado de máquina e desenvolvimento de sensores virtuais destaca sua utilidade na interpretação de dados e na geração de insights valiosos em ambientes de IoT.

Diversas aplicações têm sido implementadas em diversas áreas utilizando sensores virtuais. Por exemplo, na automação, esses sensores têm sido empregados para realizar detecção virtual de alto desempenho e livre de manutenção (KANO; FUJIWARA, 2013).

Além disso, em cenários como turbinas eólicas modernas, várias limitações surgem quando se trata de instalar sensores físicos. Embora essas turbinas tenham muitos sensores instalados que fornecem fluxos constantes de dados, há algumas medidas cruciais em que a instalação de sensores físicos se torna impraticável ou a tecnologia dos sensores ainda não está suficientemente avançada. Por exemplo, a detecção da forma e localização de déficits de vento induzidos pela esteira, causados pelas turbinas que ficam contra o vento, é uma característica com aplicação relevante no controle de parques eólicos. Contudo, é difícil detectar fisicamente essa condição devido à necessidade de varrer o fluxo de ar na frente da turbina em vários pontos, modelos de séries temporais baseados em aprendizado de máquina têm sido aplicados para criar sensores virtuais (DIMITROV; GÖÇMEN, 2022). Essas aplicações ilustram a versatilidade dos sensores virtuais e como eles têm contribuído para a solução de desafios específicos em diferentes setores.

Os sensores virtuais oferecem uma gama de vantagens, notadamente a redução de custos pela eliminação da necessidade de dispositivos físicos. Além disso, eles consomem menos recursos e exigem uma manutenção reduzida, uma vez que não dependem de componentes físicos (ZAIDAN et al., 2023). Essas características tornam os sensores virtuais atrativos para ambientes com recursos

limitados, contribuindo para uma abordagem mais eficaz e econômica na coleta e monitoramento de dados.

A aplicação de sensores virtuais também enfrenta desafios, como a disponibilidade e o acesso a dados, a necessidade de padronização e a implantação da plataforma (MARTIN et al., 2021). No contexto deste trabalho, um dos desafios identificados está relacionado à disponibilidade de dados, que impacta diretamente na acurácia e eficiência dos sensores, podendo comprometer sua precisão.

### **3.6 Gestão dos Recursos Hídricos**

De acordo com Costa et al. (2012), os recursos hídricos são caracterizados como corpos de água superficial ou subterrânea que estão disponíveis para uso humano, como rios e lagos. Os autores destacam a importância desses recursos em termos econômicos, como fonte de energia e para a sobrevivência de todas as formas de vida no planeta.

A água é um recurso natural de extrema importância, e a garantia de sua qualidade por meio de monitoramento constante é essencial para assegurar seu uso seguro (MENON et al., 2012). Isso implica na necessidade de implementar sistemas de monitoramento eficientes para avaliar regularmente a qualidade da água, a fim de proteger a saúde humana e preservar os ecossistemas aquáticos.

Proteger a biodiversidade aquática é de grande importância. O aumento das temperaturas globais, ocasionado pelas mudanças climáticas, têm um efeito significativo nos ecossistemas aquáticos causando impacto na temperatura da água. Existem vários fatores que influenciam a temperatura da água, as condições atmosféricas estão entre os fatores mais importantes e são as principais responsáveis pelos processos de troca de calor que ocorrem na superfície da água. (CAISSIE, 2006).

O aumento da temperatura da água devido às mudanças climáticas pode acarretar riscos significativos, incluindo o aumento do risco de doenças transmitidas pela água. As alterações nas condições térmicas podem proporcionar um ambiente propício para o crescimento e a proliferação de patógenos na água, aumentando assim a possibilidade de contaminação e o potencial de transmissão de doenças (FUNARI et al., 2012). Portanto, é essencial monitorar de perto essas mudanças e adotar medidas adequadas para mitigar os riscos associados ao aumento da temperatura da água.

Segundo Rodrigues et al. (2010), é crucial ressaltar a importância do desenvolvimento de métodos que contribuam para a conservação da biodiversidade aquática. Além disso, compreender os padrões globais que influenciam a qualidade dos sistemas lóticos é fundamental para promover a preservação desses ambientes. De acordo com, Chowdury et al. (2019), o atual sistema de monitoramento da qualidade da água é manual, caracterizado por um processo monótono e demorado. Diante disso, é essencial buscar métodos mais eficientes que garantam um monitoramento aprimorado da temperatura da água.

Na análise de regressão da temperatura do rio, a temperatura do ar é comumente utilizada como a variável independente, sendo uma substituta para a troca líquida nos fluxos de calor que afetam a superfície da água. Além disso, a temperatura do ar estima a temperatura de equilíbrio de um curso d'água. (WEBB et al., 2003). Essa abordagem é respaldada pela disponibilidade generalizada de medições da temperatura do ar em comparação com os componentes do orçamento total de energia e condições meteorológicas como velocidade do vento, radiação solar e umidade que desempenham um papel fundamental na determinação das trocas e fluxos de calor que ocorrem na superfície do rio. Portanto, compreender e explicar a relação entre a temperatura do ar e da água é de extrema importância (ZHU et al., 2018).

### **3.7 Estações Meteorológicas**

Uma estação meteorológica desempenha um papel fundamental ao fornecer informações sobre as condições climáticas em nosso ambiente local. Ela é um instrumento ou dispositivo que nos permite obter detalhes sobre a temperatura ambiente, pressão atmosférica, umidade, precipitação, cobertura de nuvens, velocidade e direção do vento, radiação e outros parâmetros relevantes. Por meio da detecção dessas variáveis, como temperatura, pressão, umidade, intensidade da luz e precipitação, as estações meteorológicas desempenham um papel importante na coleta de dados (KODALI; MANDAL, 2016).

Os principais tipos de estações meteorológicas são as convencionais e automáticas. Nas estações convencionais, as medições e observações são realizadas de forma analógica. Por outro lado, nas estações automáticas, são utilizados sensores e softwares capazes de realizar leituras e registrar os dados por meio de tecnologias digitais. Esse método automatizado permite uma coleta mais precisa e eficiente das informações meteorológicas. (MANNRICH; DA ROCHA, 2021).

A estação meteorológica é o dispositivo mais importante para fornecer dados ambientais como temperatura, umidade, precipitação, velocidade do vento, radiação solar, etc. (FOURATI et

al., 2014). Elas estão frequentemente situadas em locais importantes para seus usos principais na aviação ou no monitoramento do clima para grandes centros populacionais (ou seja, cidades e aeroportos) e podem ser mais representativas geograficamente de algumas áreas do que de outras (COLSTON et al., 2018).

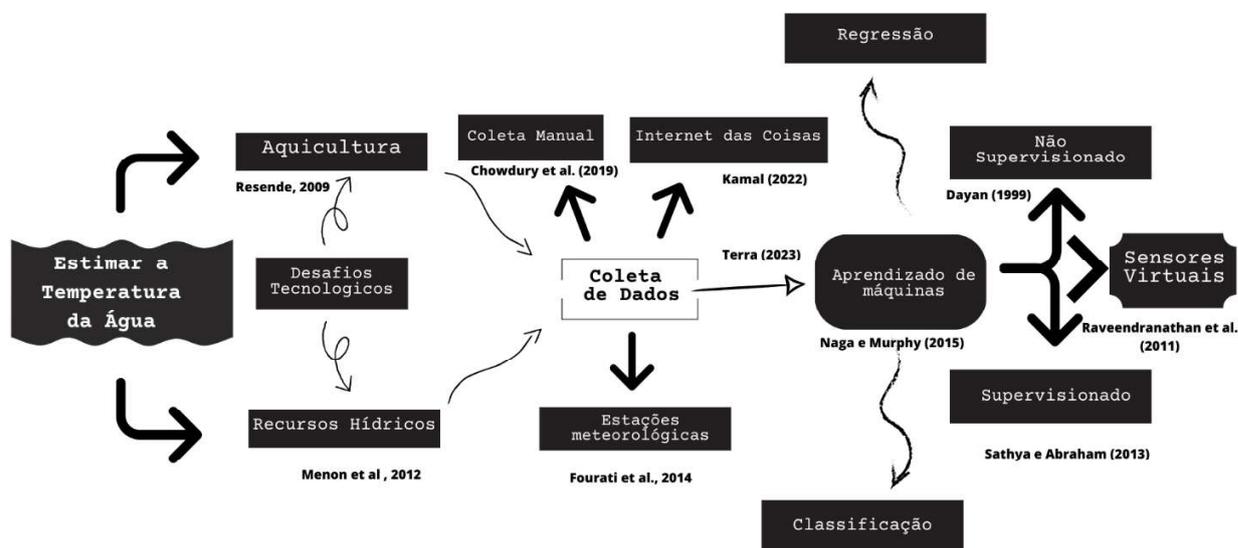
Para obter informações além da temperatura do ar, é possível acessar estações meteorológicas. Essas estações, tanto governamentais quanto privadas, realizam a coleta e o registro de diversos valores meteorológicos (SILVA et al., 2015). Esses dados são de grande importância para a compreensão e monitoramento das condições climáticas em diferentes regiões.

Aprendizado de máquina é amplamente utilizado na previsão do tempo a partir de dados climáticos (YONEKURA et al., 2018). Outros estudos desenvolveram diversas aplicações para estimar a temperatura do solo com aprendizado de máquina e estações meteorológicas (GULERYUZ, 2022). Para prever a temperatura da água de lagos, as estações meteorológicas que mediram as condições meteorológicas locais forneceram uma fonte aprimorada de dados preditivos proporcionando bom desempenho para o modelo (READ et al., 2019).

### **3.8 Considerações Finais**

Neste capítulo, foi apresentado o arcabouço teórico deste trabalho, onde foi possível identificar os conceitos e a importância do monitoramento da temperatura da água para a aquicultura e para a gestão de recursos hídricos. A fim de facilitar a compreensão dos principais achados que contribuíram para o desenvolvimento deste trabalho, foi elaborado um Mapa Mental (Figura 3). Segundo Erdem (2017), os mapas mentais trazem benefícios como relembrar, melhorar a criatividade, resolver problemas, focar em um assunto e organizar os pensamentos.

Figura 3. Mapa Mental dos principais métodos para estimar a temperatura da água



Fonte: Autor (2023).

Os desafios tecnológicos mencionados na literatura, como a infraestrutura limitada para o uso de sensores aquáticos e a necessidade de manutenção constante (REGAN et al., 2009), levaram à formulação da pergunta de pesquisa sobre a busca por uma abordagem alternativa e eficaz para o monitoramento da temperatura da água.

Partindo do estudo de Benyahya et al. (2007), o qual proporcionou uma visão geral dos modelos estatísticos de temperatura da água, foram assimilados conceitos e métodos relacionados aos modelos mais apropriados para prever, simular ou estimar a temperatura da água, levando em consideração o contexto específico em questão.

A aquicultura é a área principal de aplicação dos resultados deste trabalho, uma vez que a produção de pescados desempenha um papel fundamental na economia, contribuindo para o crescimento do setor alimentar (SIQUEIRA, 2017). No entanto, a falta de recursos tecnológicos disponíveis pode limitar a eficiência e a sustentabilidade do setor. Diante disso, pesquisar formas de desenvolver tecnologias acessíveis para os produtores que dispõem de recursos limitados, como sensores virtuais, pode oferecer a eles uma forma de monitorar a qualidade da água, que é essencial para o desenvolvimento da produção.

Com relação aos recursos hídricos, é importante ressaltar que a aquicultura também pode ser realizada em ambientes naturais, como rios e lagos. Nesses casos, os produtores precisam estar atentos à temperatura da água presente nesses corpos hídricos. Assim como em sistemas de

produção em cativeiro, a temperatura da água desempenha um papel crucial na aquicultura em ambientes naturais (VEENSTRA et al., 2003). Variações significativas na temperatura podem afetar o crescimento, a reprodução e a saúde dos organismos aquáticos, impactando diretamente a produção.

Portanto, a temperatura da água desempenha um papel de extrema importância na produção aquícola (LEÓN et al., 2006). Ela tem o potencial de afetar o metabolismo, o crescimento e o desenvolvimento dos organismos aquáticos. Variações significativas na temperatura podem ter impactos diretos na fisiologia dos animais, influenciando suas taxas metabólicas, demandas energéticas, consumo de alimentos e eficiência de conversão alimentar.

Além dos efeitos no desempenho dos organismos cultivados, a temperatura da água também desempenha um papel crucial na ecologia e na gestão da qualidade da água doce. Por exemplo, temperaturas elevadas podem contribuir para o aumento das florações de cianobactérias (CHA et al., 2017), que são proliferações excessivas desses organismos, podendo gerar impactos negativos na qualidade da água, tanto em termos de segurança para a vida aquática quanto para os usos humanos.

#### **4 MODELOS DE APRENDIZADO DE MÁQUINA PARA O DESENVOLVIMENTO DE SENSORES VIRTUAIS NA ESTIMATIVA DA TEMPERATURA DA ÁGUA**

Neste capítulo, desvelamos as contribuições de cunho técnico e científico decorrentes do labor investigativo, as quais foram alcançadas por intermédio da formulação e emprego de métodos científicos, voltados a abordar a questão-chave que norteia esta pesquisa.

A pergunta de pesquisa deste trabalho é: " Qual o modelo estatístico de aprendizagem de máquina mais apropriado para estimar a temperatura da água em diferentes corpos de água, utilizando uma combinação de dados históricos da água e dados meteorológicos?". A resposta a esta questão subsidia a elaboração de sensores virtuais destinados ao monitoramento térmico das massas aquáticas.

## **4.1 Visão Geral**

Esse capítulo apresenta os métodos científicos utilizados no desenvolvimento dos modelos de aprendizagem de máquina visando estimar a temperatura da água, o qual consiste em: materiais e métodos utilizados, além da coleta de dados da área do estudo; apresentação dos modelos de aprendizagem de máquina utilizados; métricas estatísticas utilizadas para avaliar o desempenho dos modelos; e os resultados nos testes de validação apresentados durante a calibração e treinamento.

O resultado do trabalho de pesquisa descrito neste capítulo foi avaliado por especialistas e aceito para publicação (SILVA, SOUZA e NÓBREGA, 2023b) no periódico Open Access e-Journal of the International Measurement Confederation (IMEKO), por meio de um artigo intitulado “*Machine learning models applied to estimate the water temperature of rivers and reservoirs*”. Além disto, foi também submetido e apresentado no I Workshop MATA – Meio Ambiente, Tecnologia e Agro (UFRPE, 2023), por meio do resumo científico intitulado: “Modelos de Aprendizagem de máquina aplicado para estimar a temperatura da água em tanque escavado de Aquicultura 4.0”.

## **4.2 Materiais e Métodos**

A partir da condução do Mapeamento Sistemático da Literatura (Capítulo 2) e do Arcabouço Teórico (Capítulo 3), foi possível compreender acerca dos passos fundamentais a serem empreendidos na aquisição dos dados primordiais para a pesquisa. O primeiro desses passos implica na precisa identificação do local onde o estudo será conduzido, seguido pela coleta minuciosa dos registros históricos da temperatura da água, que posteriormente serão empregados como rótulos para o treinamento dos modelos de aprendizado de máquina.

Adicionalmente, em consonância com as variáveis preditoras, tornou-se imprescindível a obtenção de dados meteorológicos. A preferência recai sobre a utilização de informações provenientes de múltiplas estações meteorológicas, visando a obtenção de uma análise comparativa da precisão. Tal abordagem permite a inferência de conclusões acerca das variáveis que detêm a capacidade de conferir ao modelo um desempenho ótimo, uma vez que essas variáveis poderão ser destacadas por meio da análise comparativa da contribuição de diferentes estações meteorológicas.

### **4.2.1 Coleta de Dados no Tanque de Aquicultura**

No presente estudo, utilizou-se como objeto de investigação um tanque escavado localizado nas dependências da estação de aquicultura do Departamento de Pesca e Aquicultura (DEPAq) da

UFRPE. O mencionado tanque dispõe de um dispositivo IoT para tanques escavados, uma solução completa para aquicultura 4.0, na qual também dispõem de um sensor físico de temperatura mergulhado a 30 centímetros de profundidade (Figura 4), que possibilita o monitoramento da temperatura da água, cujos dados históricos são transmitidos para um servidor. A partir da coleta dos dados disponíveis, relativos à temperatura horária máxima da água durante um período de três meses, estes foram empregados para o treinamento de algoritmos como variáveis de saídas, bem como para a validação e comparação dos dados estimados.

**Figura 4. Boia com sensor de temperatura da água para coleta de dados**



Fonte: Autor (2023).

Os dados de estações meteorológicas foram utilizados como variáveis preditoras para os treinamentos e validação dos modelos. A coleta de dados ocorreu a partir da estação meteorológica Be Weather (B2K) localizada ao lado do tanque escavado (Figura 5). Para coletar dados da estação meteorológica mais distante do tanque e realizar uma comparação de acurácia com a estação próxima ao tanque, foi utilizada a interface de programação de aplicativos (API) Visual Crossing, acessível online. Esta estação remota estava localizada a uma distância de aproximadamente 15 km do tanque. Procedeu-se então o tratamento e sincronização dos dados meteorológicos históricos com os dados de temperatura da água, a partir da respectiva data e hora.

**Figura 5. Estação meteorológica B2K para coleta de dados**



Fonte: Autor (2023).

#### **4.2.2 Coleta dos Dados nos Reservatórios e Rios**

Para subsidiar este trabalho, foram obtidas informações dos rios e reservatórios de Pernambuco, incluindo seus nomes e localizações, fornecidas pela Agência Pernambucana de Águas e Clima (APAC). Os dados referentes à temperatura da água dos rios e do ar foram fornecidos pela mesma fonte. No total, 55 reservatórios e 45 rios foram considerados para análise. Ao analisar os dados, constatou-se que as medições foram realizadas manualmente no período de 2011 a 2022, registrando apenas a temperatura da água e do ar no horário, data e local da coleta. Esses dados são usados como entrada e saída para o treinamento do modelo.

No entanto, os conjuntos de dados obtidos para cada reservatório e rio foram limitados, alguns deles contendo menos de 20 registros. Dada essa limitação, optou-se por treinar o modelo usando dados de todos os rios (para os rios de treinamento) e reservatórios (para os reservatórios de treinamento), a fim de obter um conjunto de dados mais abrangente e representativo.

Como resultado, foram utilizados 1.919 registros, sendo 759 referentes a rios e 1.160 a reservatórios. Essa abordagem permitiu obter um conjunto de dados mais robusto e adequado para o treinamento e validação dos modelos, considerando a disponibilidade limitada de dados individuais para cada corpo d'água.

### 4.2.3 Correlação das Variáveis

A temperatura da água é uma função de muitos fatores ambientais, incluindo a radiação solar, a temperatura do ar, a umidade relativa do ar, a precipitação e a intensidade e a direção do vento, entre outros. (RAHMAN et al., 2015). Além disso, outras variáveis também podem contribuir para o aumento ou diminuição da temperatura da água.

Para identificar as variáveis que possuem uma correlação forte com a temperatura da água, foi utilizado o coeficiente de correlação de Pearson para cada variável observada em relação à temperatura da água. Segundo Lopes (2016), na estatística descritiva este coeficiente tem como objetivo encontrar e mensurar o grau de relação entre duas variáveis, ou seja, um coeficiente de correlação, envolvendo dados contínuos conhecidos como “r de pearson”. Podendo assumir valores positivos ou negativos. A correlação é uma medida que indica o grau de relação linear entre duas variáveis e pode variar de -1 a 1. Valores próximos de -1 indicam uma correlação negativa forte, enquanto valores próximos de 1 indicam uma correlação positiva forte. Uma correlação de 0 indica ausência de correlação. A Tabela 9 fornece um guia para descrever correlações em palavras, com base em seus valores numéricos de acordo com Shimakura (2006).

**Tabela 9. Estrutura do Modelo de Maturidade de Pearson**

<b><i>Coeficiente de correlação</i></b>	<b><i>Interpretação</i></b>
<b><i>0.00 – 0.19 ou 0.00 – -0.19</i></b>	correlação muito fraca
<b><i>0.20 – 0.39 ou -0.20 – -0.39</i></b>	correlação fraca
<b><i>0.40 – 0.69 ou -0.40 – -0.69</i></b>	correlação moderada
<b><i>0.70 – 0.89 ou -0.70 – -0.89</i></b>	correlação forte
<b><i>0.90 – 1.00 ou -0.90 – -1.00</i></b>	correlação muito forte

Fonte: Shimakura (2006).

Para calcular o coeficiente de correlação, é necessário, primeiramente, calcular a covariância entre uma variável preditora e a variável Y. A covariância (Equação 1) é uma medida estatística que quantifica o grau de interdependência linear entre duas variáveis.

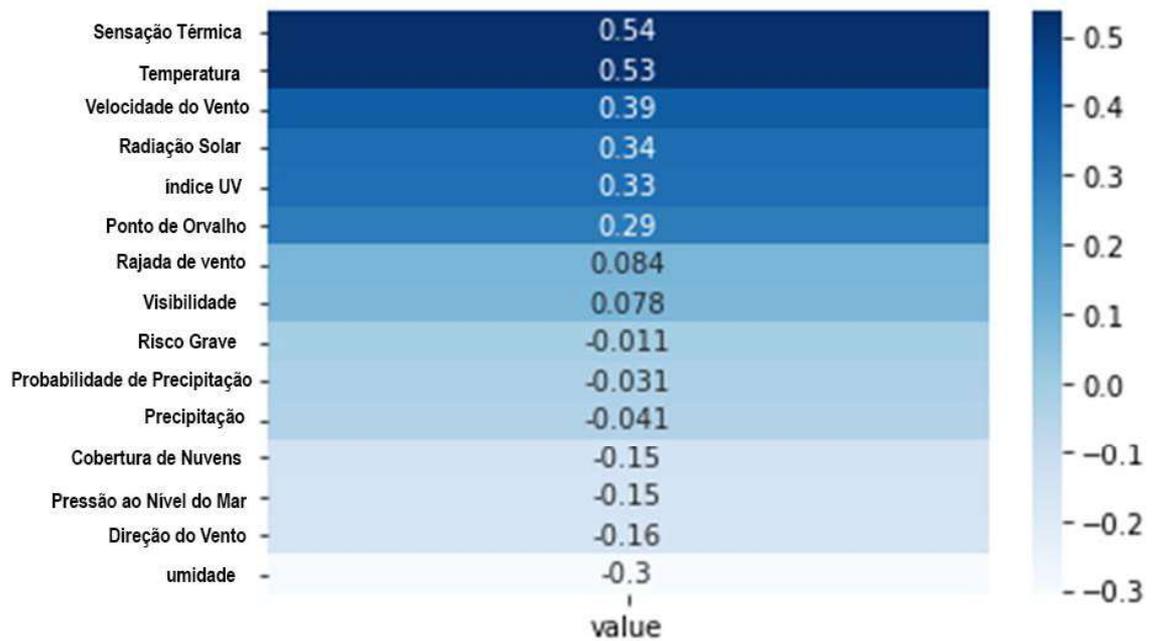
$$Cov(x, y) = \frac{\Sigma(xi - \bar{x}) * (yi - \bar{y})}{n - 1} \quad (1)$$

Onde  $\Sigma$  representa a soma,  $x_i$  é o i-ésimo valor da variável x,  $y_i$  é o i-ésimo valor da variável y,  $\bar{x}$  é a média dos valores da variável x,  $\bar{y}$  é a média dos valores da variável y, n é o número de observações. Finalmente, o valor obtido na Equação (1) é substituído na Equação (2), que representa a fórmula do coeficiente de correlação de Pearson.

$$Cr(x, y) = \frac{Cov(x, y)}{std(x) * std(y)} \quad (2)$$

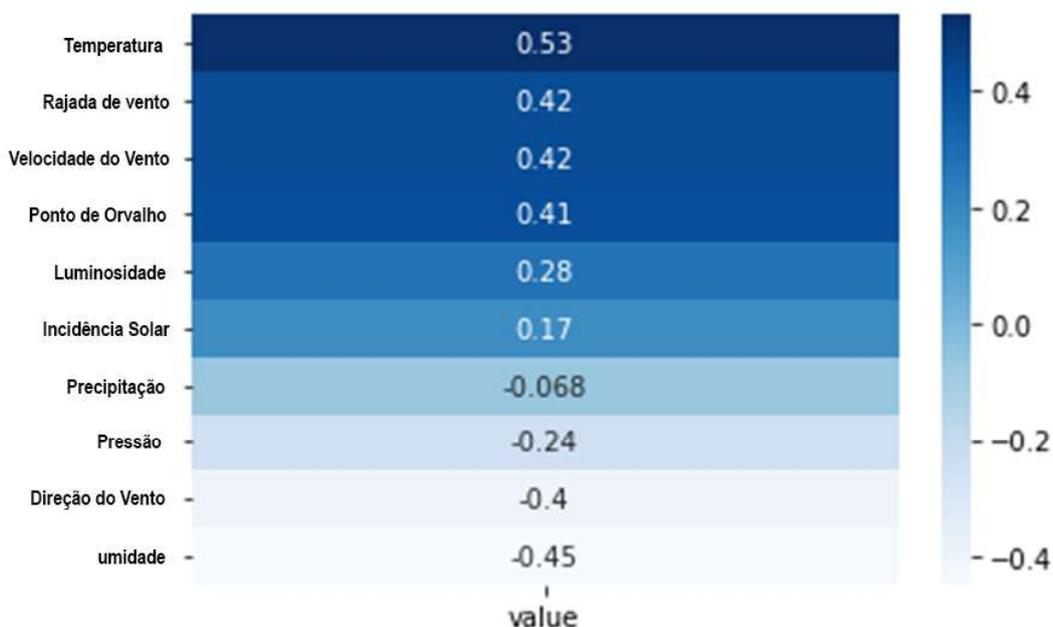
Onde  $Cov(x, y)$  é a covariância entre  $x$  e  $y$ ,  $std(x)$  é o desvio padrão de  $x$  e  $std(y)$  é o desvio padrão de  $y$ . A Figura 6 e a Figura 7 mostram o resultado da correlação em cada estação.

**Figura 6. Correlação das variáveis Estação Meteorológica das API**



Fonte: Autor (2023).

**Figura 7. Correlação das variáveis Estação Meteorológica B2K**



Fonte: Autor (2023).

A Tabela 10 apresenta os resultados derivados do cálculo das correlações entre os rios e reservatórios, considerando as variáveis da temperatura da água e a variável preditiva representando a temperatura do ar.

**Tabela 10. Correlação de temperatura do ar e da água**

Rios	Reservatórios
0.70	0.61

Fonte: (SILVA, SOUZA e NÓBREGA, 2023b).

#### 4.2.4 Normalização dos Dados

Os dados observados foram separados em dois subconjuntos distintos: um subconjunto contendo 80% dos dados, utilizado para a calibração do modelo, e outro subconjunto contendo 20% dos dados, utilizado para a validação do modelo obtido a partir do treinamento. Antes do treinamento e validação do modelo, foi realizada a normalização dos dados de entrada e saída utilizando uma técnica que visa manter as variáveis dentro de um intervalo específico. Essa técnica consiste em escalar os dados de entrada e saída de forma que eles tenham uma distribuição normalizada, ou seja, com média zero e desvio padrão unitário.

A Equação 3 consiste em subtrair a média dos dados e dividir pelo desvio padrão. Onde  $x_{norm}$  é o valor normalizado,  $x$  representa o valor dos dados,  $x_{mean}$  representa a média do conjunto de dados e  $x_{std}$  representa o desvio padrão do conjunto de dados.

$$x_{norm} = (x - x_{mean}) / x_{std} \quad (3)$$

Após concluído o processo de treinamento e validação do modelo, é essencial reverter a normalização aplicada aos dados, a fim de apresentá-los na mesma estrutura do conjunto de dados original. Essa etapa de reversão é extremamente importante para garantir que os resultados obtidos pelo modelo sejam interpretados corretamente e possam ser comparados com os dados originais. Dessa forma, é possível obter uma análise precisa e contextualizada, considerando a escala e as características dos dados originais.

### 4.3 Modelos de Aprendizado de Máquinas

Foram utilizados quatro modelos de aprendizado de máquina de regressão: regressão linear; Descida do Gradiente Estocástico; Árvores Extremamente Aleatórias e Rede Neural Multicamada Perceptron (MLPNN).

#### 4.3.1 Regressão Linear

A regressão linear é um método estatístico usado para modelar a relação entre uma variável dependente (saída) e uma ou mais variáveis independentes (ou preditoras). Considerando que a relação entre essas duas variáveis seja linear, característica típica de um modelo paramétrico. É um dos métodos mais simples e amplamente utilizados na análise de dados. Uma das principais vantagens da regressão linear é sua capacidade de fazer previsões. Com base nos padrões observados nos dados, é possível estimar valores futuros da variável dependente com razoável precisão, fornecendo subsídios valiosos para planejamento e tomada de decisão (MONTGOMERY et al., 2021).

A regressão linear simples envolve apenas uma variável independente e uma variável dependente. A fórmula geral para regressão linear simples é dada pela Equação 4. Onde  $Y$  é variável de previsão dependente,  $\beta_0$  é a constante intercepto,  $\beta_1$  é o coeficiente de regressão (ou inclinação) e  $X_1$  é a variável preditora independente.

$$Y = \beta_0 + \beta_1 * X_1 \quad (4)$$

O objetivo do modelo de regressão linear é determinar os valores de  $\beta_0$  e  $\beta_1$  que minimizam a soma dos quadrados dos erros (SQE)(MSE), ou seja, a diferença entre os valores observados de  $Y$  e os valores previstos pelo modelo. Isso é feito por meio do método dos mínimos quadrados, que é uma técnica de otimização que minimiza a função de custo.

O modelo de regressão linear pode ser expandido para incluir múltiplas variáveis independentes (Equação 5), resultando em um modelo de regressão linear múltipla. Onde  $Y$  é variável de previsão dependente,  $\beta_0$  é a constante intercepto,  $\beta_1, \beta_2, \dots, \beta_n$  é o coeficiente de regressão (ou inclinação) para as respectivas variáveis independentes que são  $X_1, X_2, \dots, X_n$ . Neste estudo será utilizado a regressão linear com múltiplas variáveis de entradas independentes.

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n \quad (5)$$

#### 4.3.2 Descida do Gradiente Estocástico

O *Stochastic Gradient Descent* (SGD) é uma abordagem amplamente utilizada no treinamento de modelos de aprendizado profundo. Este é o algoritmo comumente utilizado para resolver problemas de otimização, ajustando os parâmetros iterativamente através da taxa de aprendizado (KETKAR, 2017). Um modelo estocástico pode ser tanto paramétrico quanto não paramétrico, dependendo de como é utilizado. Nesse contexto, o modelo estocástico é considerado paramétrico. Isso ocorre porque um modelo estocástico paramétrico pressupõe uma forma funcional específica para a relação entre as variáveis independentes (variáveis meteorológicas) e a variável dependente (temperatura da água). Embora algumas das variáveis meteorológicas possam ter uma relação linear com a temperatura da água e possam ser modeladas adequadamente com parâmetros fixos, outras variáveis podem ter relações não lineares e exigir formas funcionais mais flexíveis.

A fórmula SGD, dada pela Equação (6), é uma variação do clássico algoritmo gradiente descendente, que ajusta os parâmetros de um modelo para minimizar uma função de custo. Onde  $W(t+1)$  é o novo valor dos parâmetros no tempo  $t+1$ ,  $W(t)$  é o valor atual dos parâmetros no tempo  $t$ ,  $\alpha$  é a taxa de aprendizado, que controla o tamanho dos passos de atualização dos parâmetros,  $\nabla_w(J(W(t), x(i), y(i)))$  é o gradiente da função de perda  $J$  em relação aos parâmetros  $W(t)$ , calculado com base em uma amostra de dados  $(x(i), y(i))$  específica.

$$W(t+1) = W(t) - \alpha * \nabla_w(J(W(t), x(i), y(i))) \quad (6)$$

Essa fórmula representa a atualização dos parâmetros do modelo usando uma abordagem estocástica, onde a atualização é feita iterativamente para cada amostra de dados individualmente. Em cada iteração, o gradiente da função de perda é calculado com base nessa amostra e usado para ajustar os parâmetros.

### 4.3.3 Árvores Extremamente Aleatórias

O algoritmo de aprendizado de máquina de árvores extremamente aleatórias ou *Extra Tree Regressor* é baseado em árvores de decisão e se destaca pela eficiência e capacidade de lidar com problemas de regressão. Dessa forma, é considerado não paramétrico devido à sua capacidade de ser altamente flexível aos dados. É uma técnica de aprendizado de máquina relativamente recente, que foi proposta como uma extensão do algoritmo *Random Forest* e foi desenvolvida para melhorar o desempenho e a generalização do modelo em relação ao *Random Forest* (AHMAD et al., 2018).

A principal característica do modelo de árvores extremamente aleatórias é que ele introduz aleatoriedade adicional na construção das árvores de decisão, tornando-as ainda mais diversificadas. Enquanto as árvores de decisão tradicionais dividem os dados em cada nó com base em valores de atributos ótimos, as árvores extremamente aleatórias seleciona aleatoriamente os pontos de divisão e escolhe o melhor entre eles. Isso permite que o modelo se adapte de maneira flexível e capture relações complexas e não lineares nos dados.

A fórmula usada para construir cada árvore individual no conjunto é semelhante à fórmula convencional do algoritmo de árvore de decisão. A construção de cada árvore individual envolve a divisão dos dados com base em critérios como a entropia fornecida pela Equação (7) e o ganho de informação calculado pela Equação (8). Entropia é uma medida de impureza ou desordem nos dados, quanto maior a entropia, maior a desordem e a incerteza nos dados.

$$Entropy = \sum_{i=1}^c -p_i \log_2 p_i \quad (7)$$

Onde  $p_i$  é a probabilidade de ocorrência no conjunto de dados  $x$  pertencente à variável  $x$ ,  $\log_2$  (logaritmo na base 2). O ganho de informação mede o quanto a divisão dos dados com base em uma determinada característica reduz a desordem ou incerteza nos dados, ajudando a decidir qual característica é a melhor para separar os dados em grupos mais homogêneos. É usado para determinar a melhor variável para dividir o conjunto de dados e construir a árvore de decisão. Onde  $Entropy(s)$  representa a medida de entropia da variável aleatória  $Y$ , e  $H(S|A)$  é a entropia condicional de  $S$  dada a variável  $A$ .

$$Gain(S, A) = Entropy(s) - H(S|A) \quad (8)$$

Estas fórmulas são fundamentais para a construção de árvores de decisão e determinação das melhores divisões nos dados com base em critérios de ganho de impureza e informação.

#### 4.3.1 Rede Neural Multicamada Perceptron

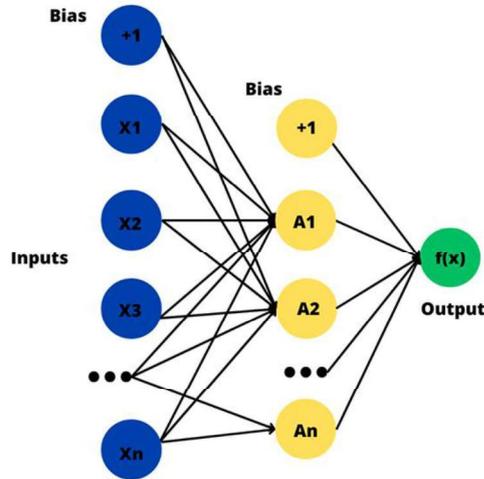
A Rede Neural Perceptron Multicamada (MLPNN) representa uma arquitetura de rede neural do tipo *feed-forward*, baseada na técnica de aprendizado conhecida como *Backpropagation*, estruturada e composta por uma camada de entrada de neurônios que desempenham o papel de receptores, uma ou mais camadas ocultas de neurônios que realizam cálculos iterativos com os dados e, por fim, a camada de saída é responsável por prever os resultados finais da rede (DESAI; SHAH, 2021). Segundo Benyahya al. (2007), um modelo de Rede Neural Artificial, é uma estrutura matemática capaz de descrever relações não lineares complexas entre dados de entrada e saída. O autor considera o modelo que é adaptado à modelagem da temperatura da água como modelos não paramétricos.

A estrutura do MLPNN (Figura 8) consiste em várias camadas de neurônios, incluindo uma camada de entrada, uma ou mais camadas ocultas, e uma camada de saída. Além das conexões ponderadas entre os neurônios, as redes neurais também incorporam bias (ou viés), que são valores ajustáveis adicionados a cada neurônio, permitindo uma maior flexibilidade na modelagem das relações entre as entradas e as saídas. Cada neurônio em uma rede MLP é chamado de perceptron e opera de maneira semelhante a um neurônio biológico. Ele recebe entradas ponderadas, as processa na rede, aplica uma função de ativação e gera uma saída. Este processo é chamado de *feed forward* apresentado pela Equação (9).

$$U = \sum_{i=1}^N x_i * w_i + b \quad (9)$$

Onde  $u$  é a saída gerada pelo perceptron,  $\Sigma$  representa a soma ponderada das entradas multiplicada pelos pesos sinápticos correspondentes,  $x_i$  é o vetor de entradas,  $w_i$  é o vetor de pesos sinápticos,  $b$  é o viés, um termo adicional que permite ajuste a saída do perceptron.

**Figura 8. Multi-layer Perceptron Estrutura**



Fonte: Autor (2023).

Após realizar esta operação, a saída é adicionada a uma função de ativação não linear, neste trabalho foi utilizada a função ReLU (Unidade Linear Retificada) determinada pela Equação (10). ReLU retorna o valor de entrada se for positivo, caso contrário, retorna zero. Onde  $ReLU(x)$  é a saída gerada pela função ReLU,  $x$  é o valor de entrada para a função. Durante o treinamento de uma rede MLP, os pesos e vieses sinápticos são atualizados com base no algoritmo de retro propagação.

$$ReLU(x) = \max(0, x) \quad (10)$$

O objetivo é minimizar uma função de custo, que mede a diferença entre as saídas previstas pela rede e as saídas desejadas. A fórmula para atualizar os pesos e vieses usando o algoritmo de retropropagação envolve o uso do gradiente descendente, que procura a direção de descida mais íngreme na função de custo.

#### **4.4 Desempenho e Avaliação do Modelo**

Para avaliar o desempenho dos modelos estatísticos, quatro diferentes critérios foram utilizados: (1) erro quadrático médio (MSE); (2) raiz do erro quadrático médio (RMSE); (3) erro médio absoluto (MAE); e (4) coeficiente de determinação ( $R^2$ ).

##### **4.4.1 MSE**

O MSE, representado pela Equação (11), é uma medida que quantifica a diferença entre as estimativas do modelo e os valores reais observados, o que indica a magnitude dessa diferença.

Quanto menor o valor do MSE, melhor o desempenho do modelo em termos de precisão de previsão. Onde  $n$  representa o número de observações, o termo soma dos erros quadrados indica a soma das diferenças quadradas entre os valores previstos de  $f_i$  e os valores observados de  $y_i$ .

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 \quad (11)$$

#### 4.4.2 RMSE

O RMSE, calculado pela Equação (12), é uma métrica derivada do MSE em que a raiz quadrada é aplicada ao final, resultando em uma medida de erro expressa na mesma unidade da variável alvo. Quanto menor o valor do RMSE, melhor o desempenho do modelo em termos de precisão de previsão. Onde  $n$  representa o número de observações, o termo soma dos erros quadrados indica a soma das diferenças quadradas entre os valores previstos por  $f_i$  e os valores observados  $O_i$ .

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - O_i)^2} \quad (12)$$

#### 4.4.3 MAE

O MAE, determinado pela Equação (13), representa a média dos valores absolutos dos erros entre as estimativas do modelo e os valores reais da variável de interesse. Um valor de MAE menor indica melhor desempenho do modelo em termos de precisão de previsão. Onde  $n$  representa o número de amostras, o termo  $y_i$  corresponde ao valor verdadeiro da  $i$ -ésima amostra e  $\hat{y}_i$  representa o valor previsto pelo modelo para a  $i$ -ésima amostra.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

#### 4.4.4 R<sup>2</sup>

O coeficiente de determinação ( $R^2$ ), apresentado pela Equação (14), é uma métrica estatística que fornece uma indicação de quão bem o modelo de regressão se ajusta aos dados observados. Em termos gerais, um valor maior de  $R^2$  indica um desempenho superior do modelo, ou seja, uma melhor capacidade de explicar a variação dos dados. Onde SSE é a soma dos erros quadráticos e SST é a soma total dos quadrados.

$$R2 = \frac{SSE}{SST}, \quad (14)$$

$$SSE = \sum_{i=1}^n (f_i - \hat{y})^2, SST = \sum_{i=1}^n (f_i - \bar{y})^2$$

#### 4.5 Método e tecnologias utilizadas para implementação dos algoritmos de aprendizado de máquina

A implementação dos algoritmos de aprendizado de máquina neste estudo foi conduzida por meio da linguagem de programação Python, escolhida devido à sua ampla popularidade em análise de dados e machine learning, beneficiando-se de sua sintaxe acessível e da disponibilidade de bibliotecas especializadas. Para tanto, foi utilizada a biblioteca *scikit-learn*, por já ser otimizada e eficiente em termos de tempo de execução e uso de recursos permitindo trabalhar com conjuntos de dados de tamanhos variados sem comprometer o desempenho. Algumas das principais ferramentas empregadas neste trabalho incluem:

- Quanto à Ambiente de Desenvolvimento Integrado (IDE), optou-se pelo uso do Spyder, uma plataforma científica gratuita e de código aberto, especialmente desenvolvida em Python para cientistas, engenheiros e analistas de dados.
- Todos os datasets (conjunto de dados) e algoritmos desenvolvidos neste trabalho estão disponíveis na plataforma Kaggle<sup>12</sup>. Essa plataforma possibilita o upload e compartilhamento de uma extensa variedade de conjuntos de dados, permitindo a construção e execução de algoritmos utilizando as bibliotecas do Python.

Os resultados das métricas de avaliação apresentam ligeiras variações ao executar os algoritmos de aprendizado de máquina em diferentes iterações. Diversos fatores contribuem para essa variabilidade, incluindo inicializações aleatórias dos pesos, realizadas de maneira não determinística. Além disso, as condições do ambiente de execução desempenham um papel importante. De maneira geral, essas pequenas flutuações não possuem significância substancial no que diz respeito à escolha do algoritmo que alcançou a melhor acurácia. A natureza inerentemente estocástica do treinamento de modelos de aprendizado de máquina pode levar a variações, mas tais

---

<sup>1</sup> Rios e Reservatórios: <https://www.kaggle.com/code/jheklogomesdasilva/rios-e-reservat-rios-dataset-apac>

<sup>2</sup> Tanque (API): <https://www.kaggle.com/code/jheklogomesdasilva/dados-da-api-coletado-na-web-visual-crossing>  
Tanque (B2K): <https://www.kaggle.com/code/jheklogomesdasilva/dados-da-esta-o-b2k-e-base-de-pesca>

nuances não comprometem a robustez das conclusões obtidas. A consistência e relevância das descobertas são mantidas, enfatizando a confiabilidade do algoritmo selecionado para obter os melhores resultados quanto à acurácia no âmbito deste estudo.

## 4.6 Resultados

Nesta seção, serão apresentados os resultados dos testes de validação de cada modelo treinado com as variáveis preditoras das estações meteorológicas para os tanques de aquicultura, bem como, os resultados de validação dos rios e reservatórios utilizando a variável temperatura do ar. Levando em consideração os índices estatísticos utilizados para avaliar o desempenho dos modelos, estes índices fornecem métricas objetivas que permitem uma análise quantitativa e comparativa dos resultados obtidos e identificar qual apresenta o melhor desempenho em termos de precisão e capacidade de previsão.

### 4.6.1 Resultados com dados da API coletados na Web utilizando todas as variáveis

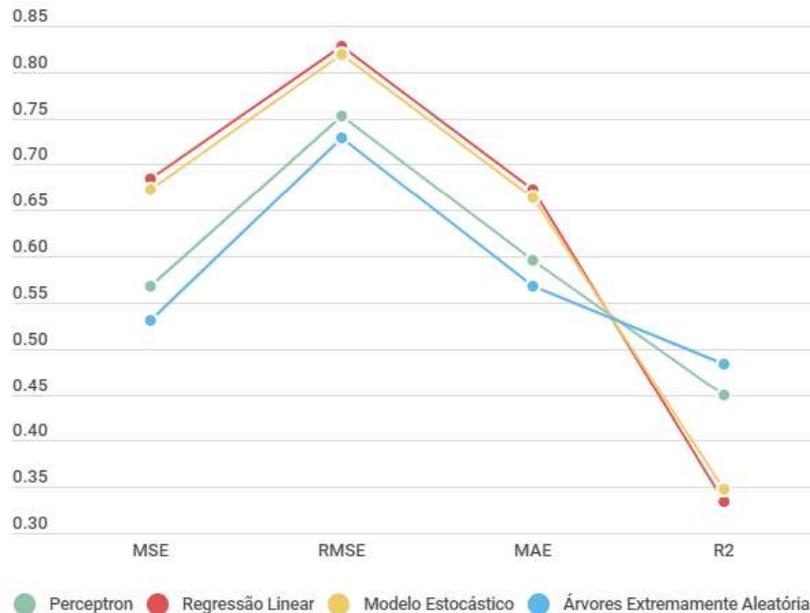
O primeiro treinamento e validação foi com os dados do taque escavado de aquicultura, nesse contexto consideramos como sistema lântico. No entanto, alguns tanques escavados podem ter sistemas de aeração ou circulação artificial da água, o que pode afetar sua classificação como sistemas lânticos. O modelo de árvores extremamente aleatórias, demonstrou maior eficácia quando treinado com os dados provenientes da estação da API coletada na web, seguido pela rede neural (MLPNN), que também apresentou um desempenho satisfatório em relação aos demais modelos. As variáveis preditoras dessa estação que foi utilizada foram: temperatura do ar, sensação térmica do ar, ponto do orvalho, umidade precipitação, probabilidade de precipitação, rajada de vento, velocidade do vento, direção do vento, pressão do nível do mar, cobertura de nuvens, visibilidade radiação solar, índice UV, risco grave. A Tabela 11 exibe o desempenho de cada modelo e os índices estatísticos ao utilizar todas as variáveis preditoras disponíveis, incluindo aquelas com correlação forte e fraca. A Figura 9 ilustra gráficos decorrentes dos dados apresentados na Tabela 11.

**Tabela 11. Performance dos modelos com dados API coletados na web**

<b>Modelos:</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>R<sup>2</sup></b>
<b>Perceptron (MLPNN):</b>	0.567	0.753	0.595	0.449
<b>Regressão Linear:</b>	0.685	0.828	0.672	0.334
<b>Modelo Estocástico:</b>	0.672	0.819	0.663	0.347
<b>Árvores Extremamente Aleatórias:</b>	0.531	0.729	0.568	0.483

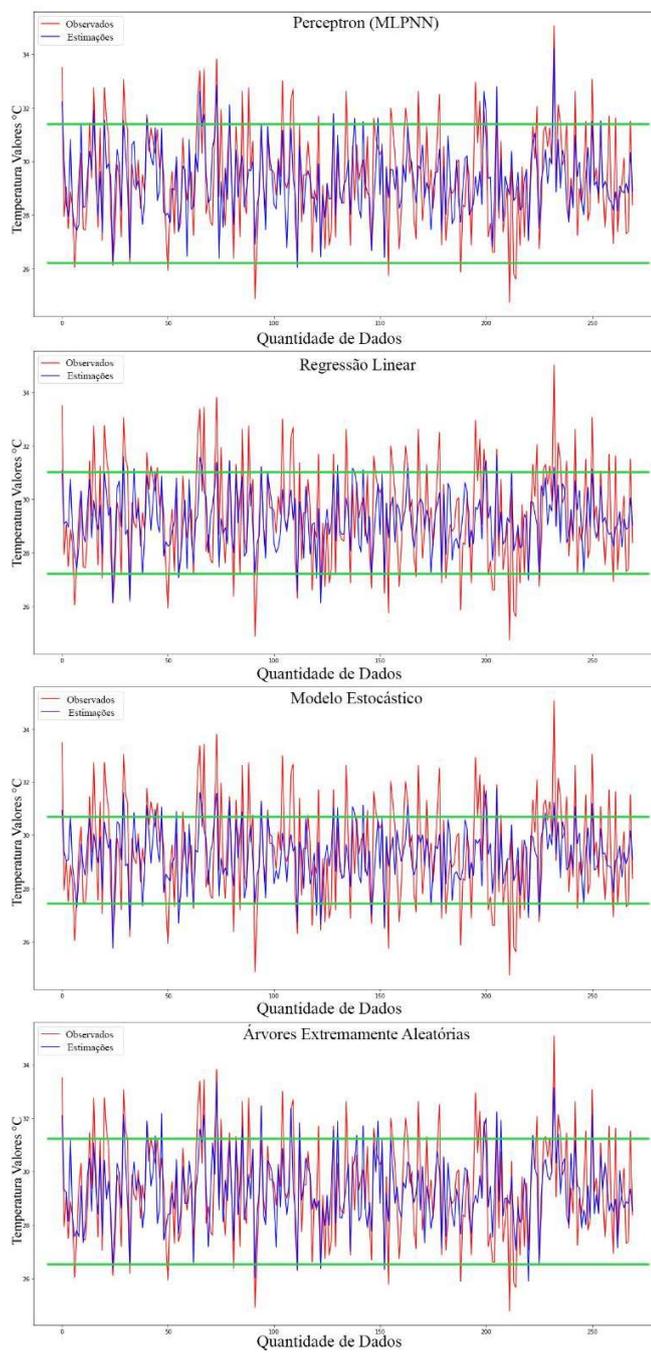
Fonte: Autor (2023).

**Figura 9. Gráficos das métricas estatísticas dos dados da API coletados na web**

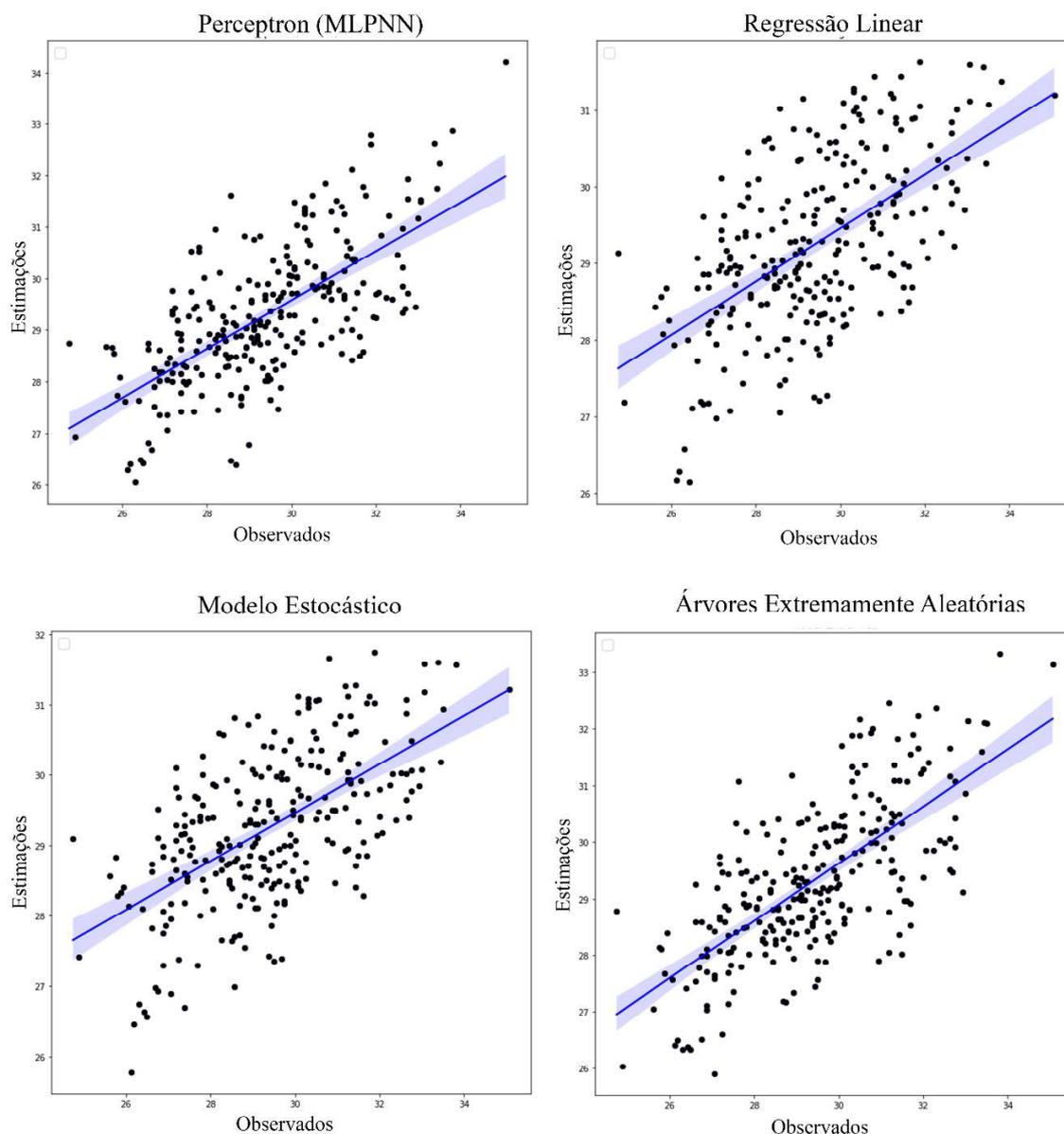


O modelo de *árvores* extremamente aleatórias, obteve o menor MSE (0.531) e RMSE (0.729), indicando que suas estimativas apresentaram menor erro médio em comparação aos outros modelos. Apresentou ainda um MAE de 0.568, que representa uma aproximação moderada entre as estimativas e os valores reais dos dados de entrada. Em relação ao R<sup>2</sup>, que mede a capacidade do modelo de explicar a variabilidade dos dados, o modelo de árvores extremamente aleatórias obteve um valor de 0.483, indicando um desempenho moderado na explicação da variação dos dados. A Figura 10 apresenta os resultados dos testes de validação, destacando a comparação entre os dados observados e os estimados pelos modelos. Observa-se que a linha azul, em certos casos, segue o padrão de curva dos dados observados, variando entre 26°C e 31°C nas estimativas. Entretanto, em algumas observações onde os dados ultrapassaram os 32°C, as estimativas não conseguiram acompanhar, resultando em uma margem de erro entre 2 a 3°C. Na Figura 11, exibimos um gráfico de dispersão, onde os pontos próximos à linha diagonal indicam uma correspondência adequada entre os valores observados e os previstos. É notável que, nos modelos de árvores extremamente aleatórias e perceptron, os pontos se aproximam mais da linha diagonal. Por outro lado, nos modelos estocástico e de regressão linear, os pontos estão mais distantes da linha, sugerindo um desempenho inferior.

**Figura 10. Gráfico do teste de validação da estação API todas variáveis de entrada**



**Figura 11. Dispersão do teste de validação da estação API todas variáveis de entrada**



Fonte: Autor (2023).

#### **4.6.2 Resultados com dados da estação B2K utilizando todas as variáveis**

A análise dos resultados do treinamento dos modelos utilizando os dados da estação B2K revelou um desempenho superior em relação aos modelos treinados com os dados da estação anteriormente mencionada. Esses resultados indicam que a estação meteorológica B2K forneceu variáveis com uma correlação mais significativa com a variável de saída temperatura da água, o que resultou em uma maior precisão nos modelos de aprendizado de máquina. As variáveis predictoras

dessa estação que foi utilizada foram: temperatura do ar, humidade, pressão, velocidade do vento e direção do vento. A rede neural (MLPNN) demonstrou a maior eficácia entre os modelos, seguida pelo modelo de árvores extremamente aleatórias. A Tabela 12 apresenta os resultados detalhados de cada modelo, incluindo as métricas estatísticas.

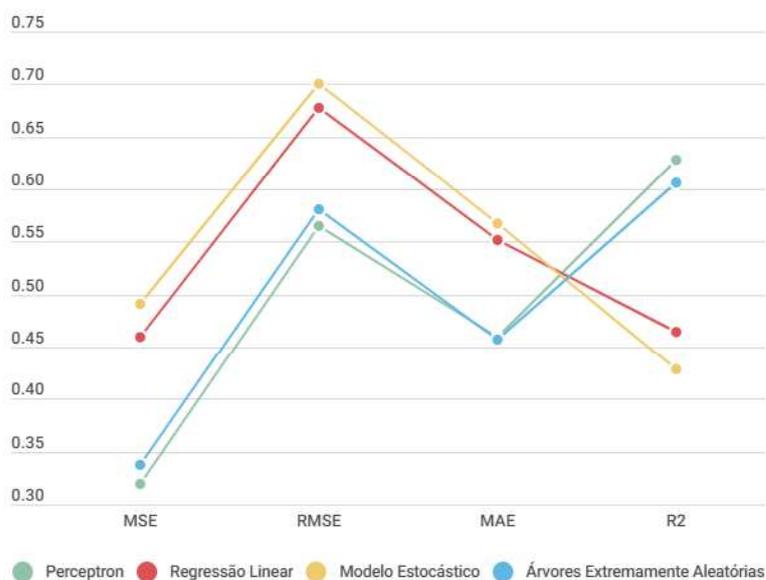
**Tabela 12. Performance dos modelos com dados da estação meteorológica B2K**

<i>Models:</i>	<i>MSE</i>	<i>RMSE</i>	<i>MAE</i>	<i>R<sup>2</sup></i>
<i>Perceptron (MLPNN):</i>	0.319	0.565	0.460	0.628
<i>Regressão Linear:</i>	0.460	0.678	0.552	0.464
<i>Modelo Estocástico:</i>	0.491	0.701	0.567	0.428
<i>Árvores Extremamente Aleatórias:</i>	0.338	0.581	0.457	0.606

Fonte: Autor (2023).

Após analisar os valores da Tabela 3, podemos observar que o modelo Perceptron (MLPNN) obteve o menor MSE (0.319) e RMSE (0.565), indicando um menor erro médio nas previsões. Além disso, apresentou um MAE de 0.460, o que indica uma boa aproximação entre as previsões e os valores reais. O modelo MLPNN também teve um valor elevado para o coeficiente de determinação  $R^2$ , alcançando 0.628, demonstrando uma maior capacidade de explicar a variação dos dados em relação aos outros modelos. Esta observação fica evidente na Figura 12, que ilustra gráficos decorrentes dos dados apresentados na Tabela 12.

**Figura 12. Gráfico das métricas estatísticas dos dados da estação b2k coletados**

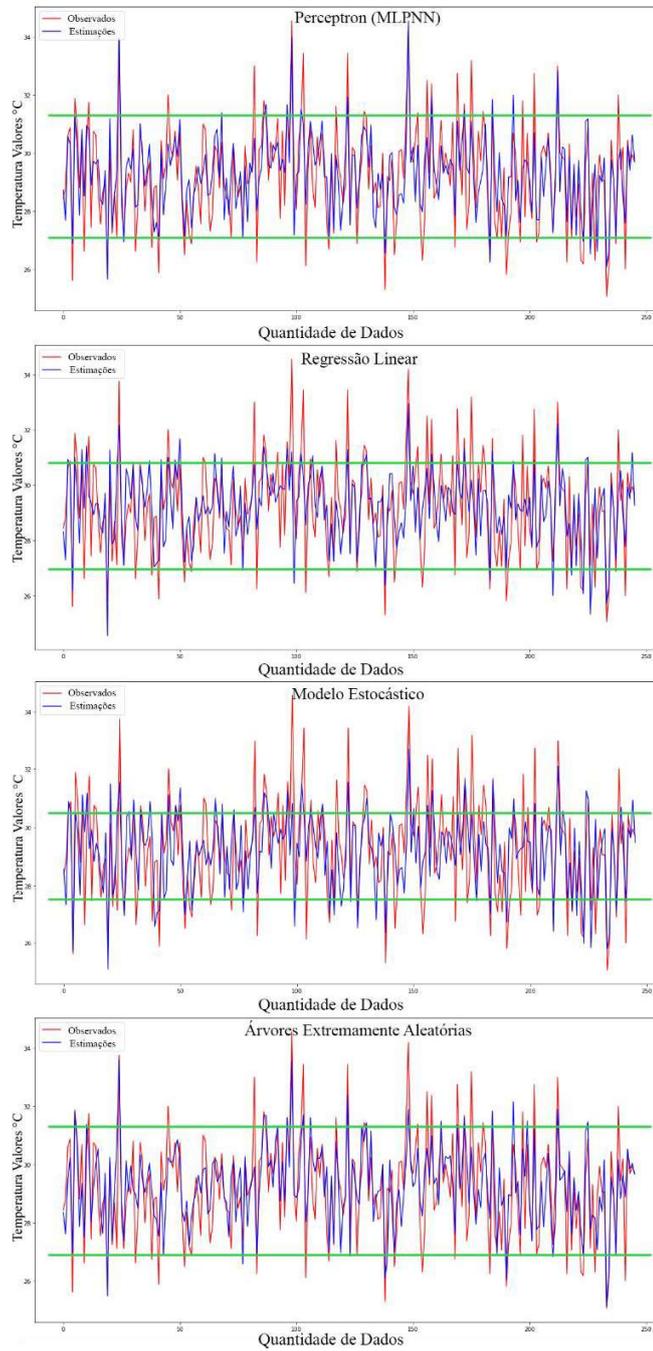


Fonte: Autor (2023).

Ao analisar os gráficos presentes na Figura 13, é possível notar que a linha vermelha representa o conjunto de teste observado, enquanto a linha azul representa os dados estimados. Nesta estação, nos modelos de perceptron e árvores, a semelhança entre o padrão de curva dos dados observados e estimados é ainda mais evidente em comparação com os gráficos da estação anterior. Notavelmente, no modelo de perceptron, em alguns casos, as estimativas conseguiram acompanhar os dados que ultrapassaram os 32°C. Porém, para os dados observados próximos a 26°C, em sua maioria, as estimativas não conseguiram acompanhar, resultando em uma margem de erro de 2°C. No caso do modelo de árvore, as estimativas, na maioria das vezes, não conseguiram acompanhar bem os dados acima de 31°C, resultando em uma margem de erro de 3°C. Os outros modelos também apresentaram uma diferença média de 3°C na maioria dos casos.

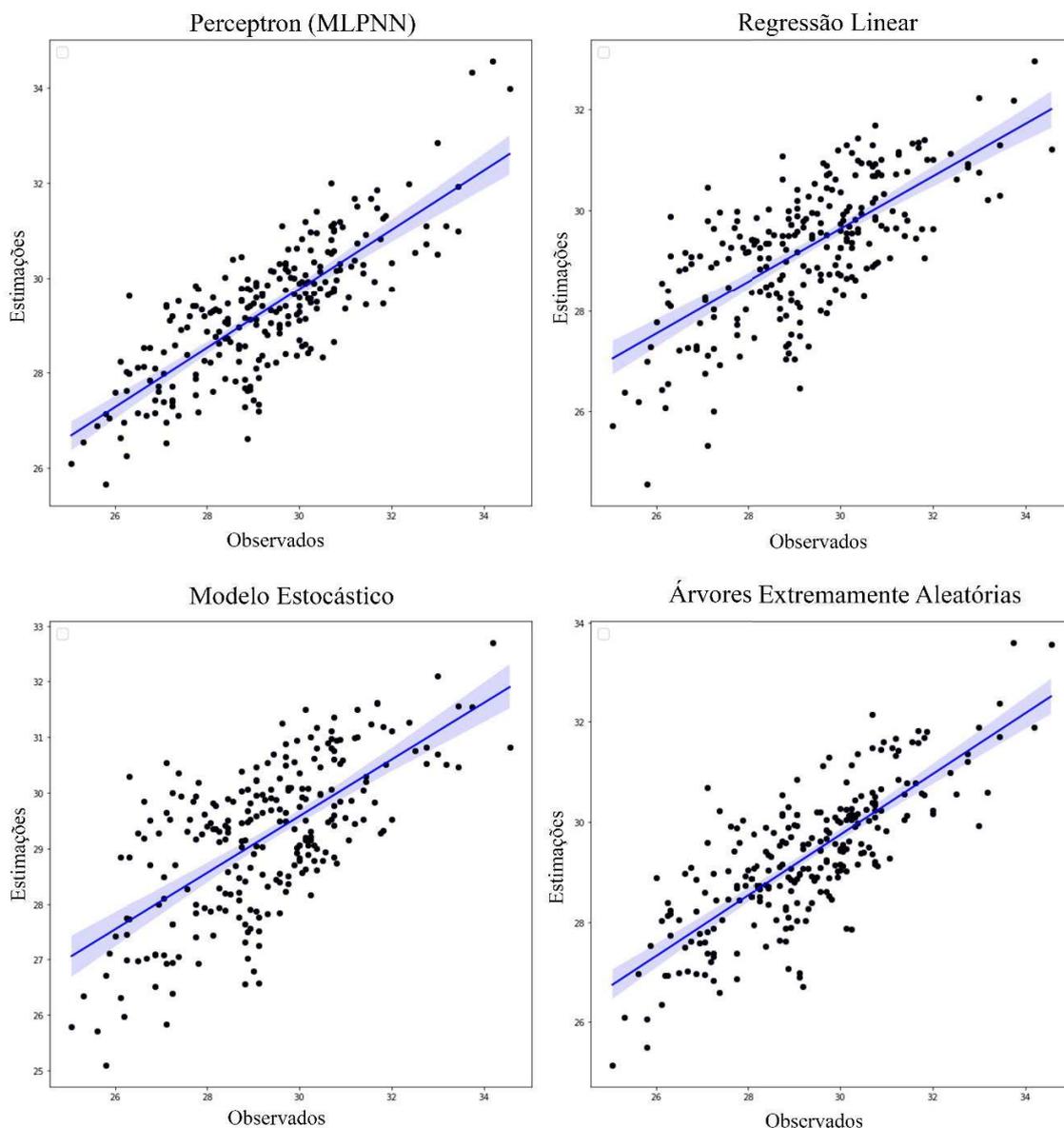
Na Figura 14, os gráficos de dispersão revelam pontos próximos à linha diagonal, indicando uma correspondência adequada entre os valores observados e previstos. Notavelmente, tanto o perceptron quanto o modelo de árvore apresentam pontos mais próximos à linha diagonal, demonstrando um desempenho mais preciso em relação aos outros modelos.

**Figura 13. Gráfico do teste de validação da estação B2K com todas as variáveis de entrada**



Fonte: Autor (2023).

**Figura 14. Dispersão do teste de validação da estação B2K com todas variáveis de entrada**



Fonte: Autor (2023).

Em comparação com os resultados anteriores, podemos destacar que tanto a regressão linear quanto o modelo estocástico tiveram desempenho inferior, apresentando valores mais elevados para MSE, RMSE e MAE, além de um coeficiente de determinação  $R^2$  menor. O modelo de árvores extremamente aleatórias, também mostrou uma queda em seu desempenho, embora ainda tenha obtido resultados razoáveis.

#### 4.6.3 Resultados com dados da API coletados na Web utilizando variáveis com correlações moderadas

Tendo em vista que a estação que apresentou variáveis com correlação mais forte foi aquela que obteve um melhor resultado, foram selecionadas as variáveis com correlação mais moderadas para o treinamento: temperatura do ar, sensação térmica, umidade, velocidade do vento, radiação solar, uvindex. O objetivo foi fornecer aos modelos sinais mais claros e relevantes para aprender a relação entre os dados de entrada e a variável de saída. A correlação entre variáveis pode indicar a presença de informações redundantes. Ao selecionar variáveis altamente correlacionadas, podemos simplificar o modelo e capturar as informações essenciais para a tarefa de aprendizado (SAEYS et al., 2007).

Considerando a estação da API disponível na Web, observou-se que a maioria das variáveis apresentava uma correlação fraca com a variável de saída. Portanto, para o treinamento dos modelos, foram selecionadas apenas as variáveis que possuíam uma correlação acima de 0,30, incluindo temperatura, sensação térmica, umidade, velocidade do vento, radiação solar e índice UV. Entre os modelos avaliados, o MLPNN demonstrou um melhor desempenho, apresentando um menor MSE (0.684) e RMSE (0.827), além de um MAE de 0.685 (Tabela 13). No entanto, todos os modelos tiveram valores relativamente baixos para o coeficiente de determinação  $R^2$ , indicando uma capacidade limitada de explicar a variabilidade dos dados.

**Tabela 13. Desempenho do modelo com dados da API coletados na Web usando variáveis de correlação moderada**

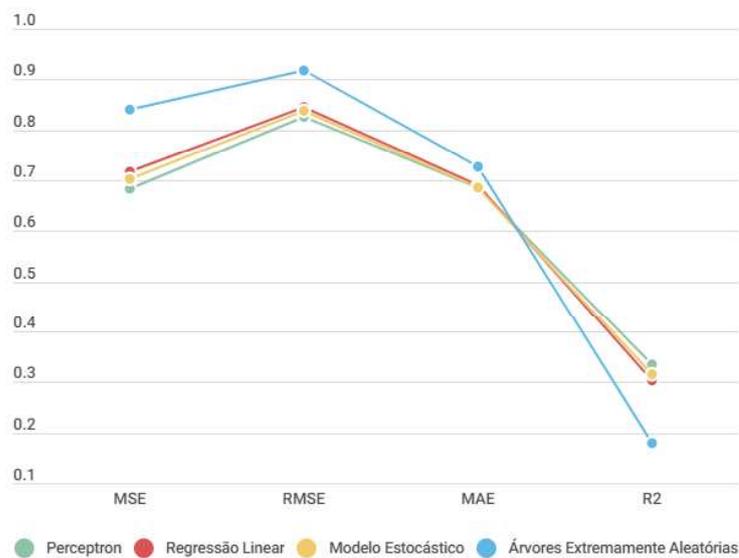
<i>Models:</i>	<i>MSE</i>	<i>RMSE</i>	<i>MAE</i>	<i>R<sup>2</sup></i>
<i>Perceptron (MLPNN):</i>	0.684	0.827	0.685	0.334
<i>Regressão Linear:</i>	0.716	0.846	0.690	0.304
<i>Modelo Estocástico:</i>	0.703	0.839	0.685	0.316
<i>Árvores Extremamente Aleatórias:</i>	0.841	0.917	0.727	0.181

Fonte: Autor (2023).

Na Figura 15 é possível visualizar a representação gráfica dos dados apresentados na Tabela 13. Nos gráficos apresentados na Figura 16, torna-se evidente a presença de uma discrepância ampliada entre os conjuntos de dados observados e os correspondentes conjuntos de dados estimados, em contraposição aos desfechos obtidos nos procedimentos de validação, os quais mostraram-se melhores quando todas as variáveis de entrada foram empregadas.

Observa-se que no modelo de perceptron, a margem de erro pode aumentar até 4°C em alguns casos, em comparação com os dados observados. Na regressão linear, esses erros parecem ser ainda maiores, seguindo a tendência do modelo estocástico. Em relação à árvore de decisão, a perda de precisão torna-se evidente. Essa análise indica que, neste conjunto de dados, a acurácia alcançada foi a mais baixa em comparação com os outros conjuntos testados.

**Figura 15. Gráficos das métricas estatísticas da correlação moderada da API**



#### **4.6.4 Resultados com Dados da Estação B2K utilizando Variáveis com Correlações Moderadas**

Ao considerar a estação B2K, que mostrou variáveis com fortes correlações com a variável de saída, foram selecionadas as variáveis com correlações acima de 0,40, incluindo temperatura, umidade, velocidade e direção do vento, rajada de vento e ponto de orvalho.

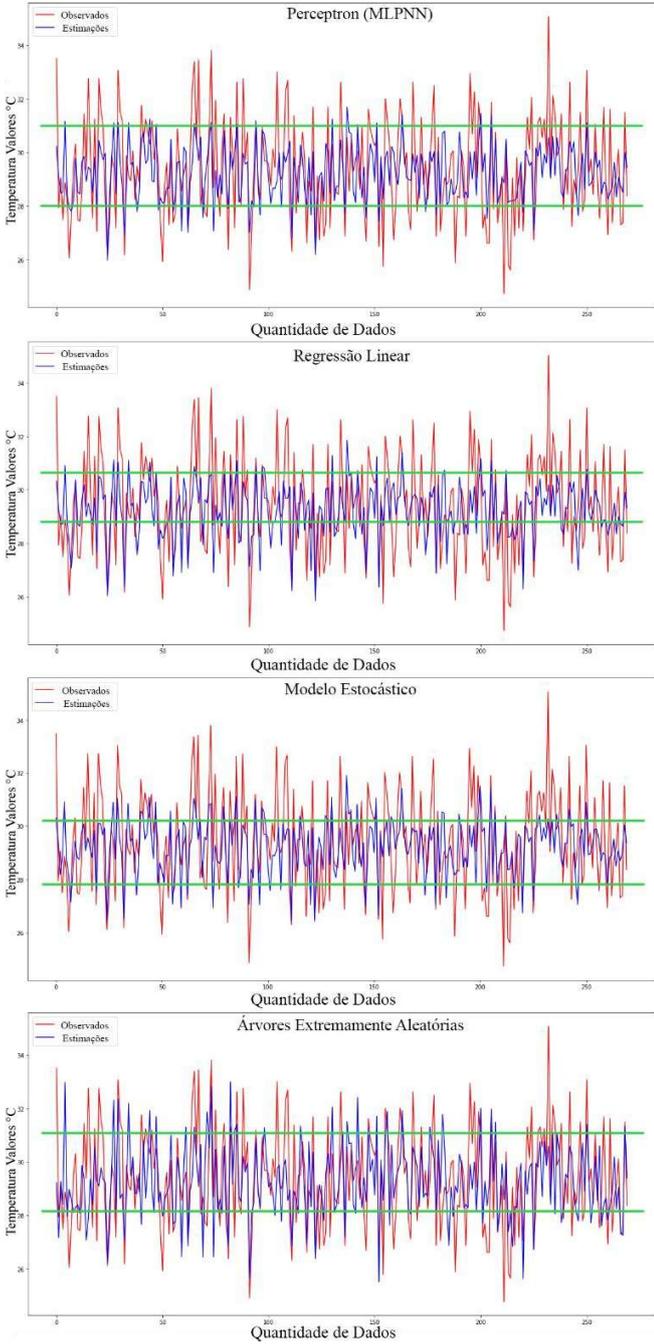
Nesse caso, o modelo MLPNN continuou demonstrando o melhor desempenho, com um MSE de 0.514 e RMSE de 0.717, seguido pelo Linear Regression e Model Stochastic, que obtiveram resultados semelhantes (Tabela 14). No entanto, todos os modelos tiveram um R<sup>2</sup> relativamente baixo, indicando que ainda há variabilidade nos dados que não foi capturada pelos modelos. Esta observação fica mais evidente quando se analisa a Figura 17, que apresenta gráficos decorrentes dos dados apresentados na Tabela 14.

A representação gráfica fornecida pela Figura 18 revela que mesmo na estação B2K, ao reduzir a amplitude das variáveis de entrada, ocorre uma diminuição na correspondência entre os

dados estimados e os dados observados, em que na maioria dos casos os dados estimados não coincidem muito bem demonstrando um baixo desempenho em todos os modelos. Contudo, uma análise detalhada da Tabela 14 demonstra que, ao adotar tal redução de variáveis, a estação B2K manifestou uma melhora em sua capacidade de precisão quando comparada às estações API.

É importante notar que a consideração da correlação não são os únicos fatores a serem considerados. Embora as variáveis com forte correlação tenham demonstrado uma tendência a melhorar o desempenho dos modelos. Nesse contexto os resultados mostraram que outras variáveis menos correlacionadas podem conter informações complementares que são relevantes para a tarefa e conseqüentemente trazem resultados mais acurados.

**Figura 16. Gráfico do teste de validação da estação API variáveis com correlação moderada**



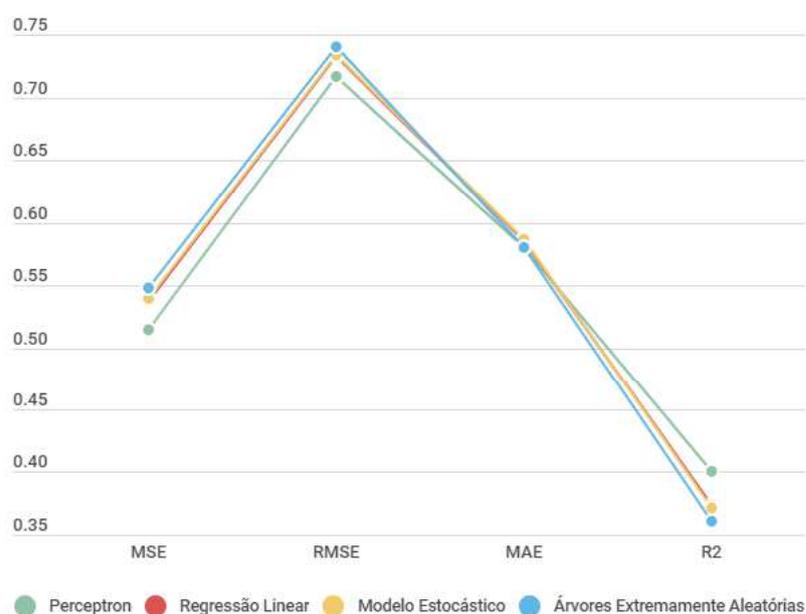
Fonte: Autor (2023).

**Tabela 14. Desempenho do modelo com dados da estação B2K usando variáveis de correlação moderada**

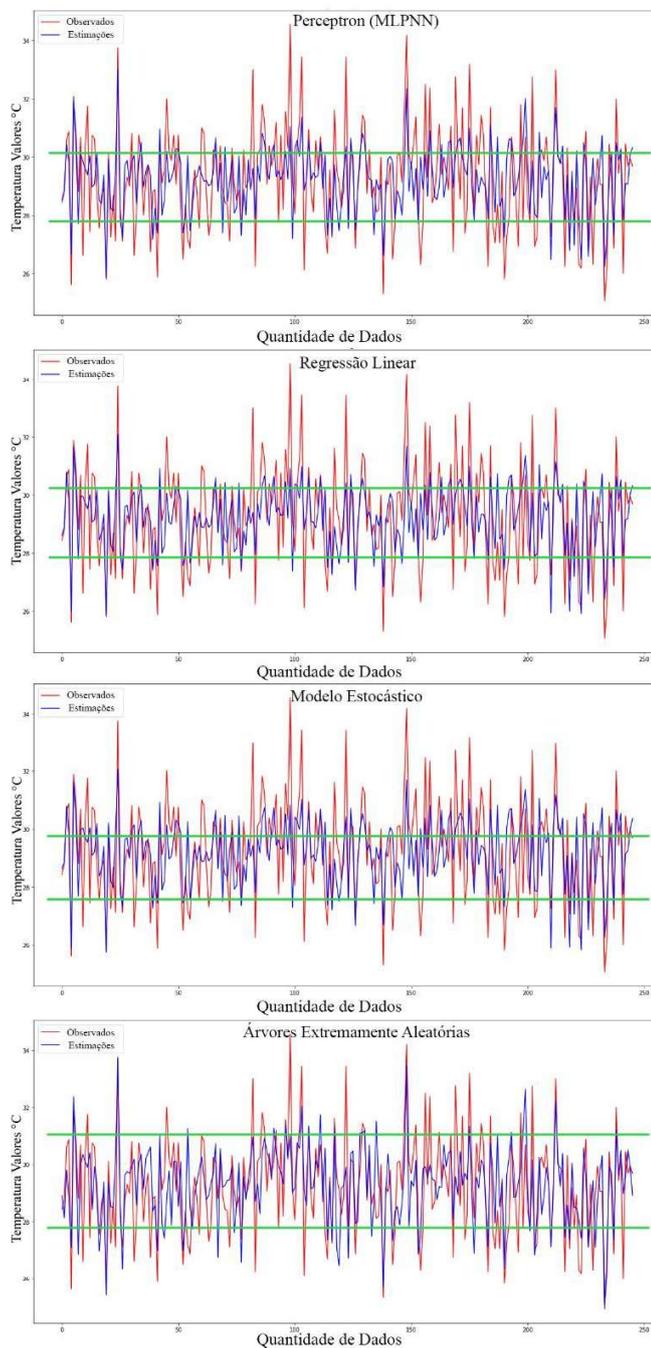
<i>Models:</i>	<i>MSE</i>	<i>RMSE</i>	<i>MAE</i>	<i>R<sup>2</sup></i>
<b>Perceptron (MLPNN):</b>	0.514	0.717	0.580	0.401
<b>Regressão Linear:</b>	0.537	0.733	0.586	0.374
<b>Modelo Estocástico:</b>	0.539	0.734	0.587	0.372
<b>Árvores Extremamente Aleatórias:</b>	0.548	0.740	0.580	0.361

Fonte: Autor (2023).

**Figura 17. Gráfico das métricas estatísticas da correlação moderada B2k**



**Figura 18. Gráfico do teste de validação da estação B2K variáveis com correlação moderada**



Fonte: Autor (2023).

#### 4.6.5 Resultados dos testes de validação dos rios

No processo de treinamento com os dados coletados nos rios, serão apresentados os resultados referentes aos sistemas lógicos, os quais se caracterizam pelo fluxo contínuo e direcionado das águas. O modelo de rede neural MLPNN apresentou um desempenho superior em comparação aos outros algoritmos, seguido pelo modelo de regressão linear e o Stochastic. Essa constatação é respaldada pelos resultados obtidos nos testes de validação (Tabela 15) e baseados em índices estatísticos relevantes.

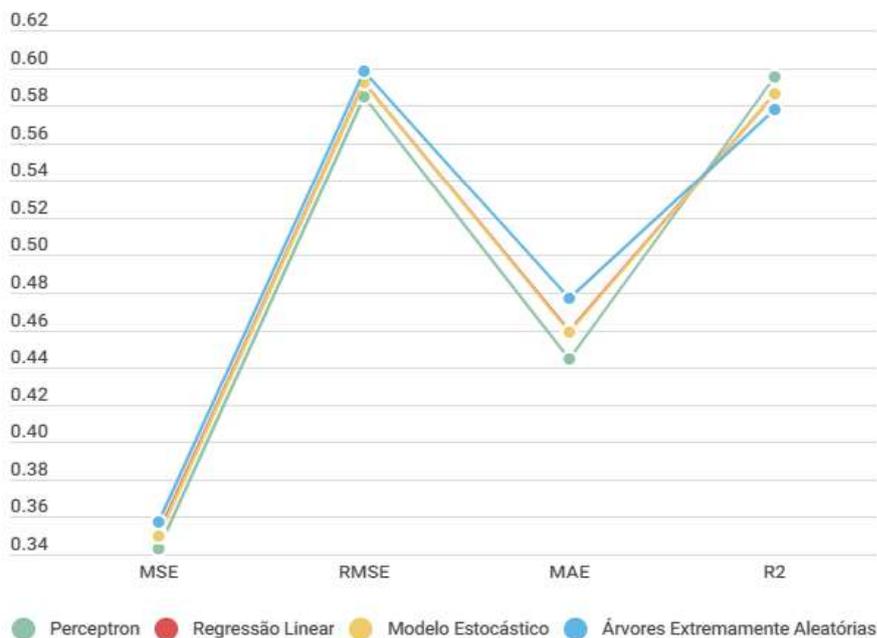
**Tabela 15. Desempenho dos modelos com os dados coletados nos Rios**

<b>Models:</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>R<sup>2</sup></b>
<i>Perceptron (MLPNN):</i>	0.343	0.585	0.445	0.595
<i>Regressão Linear:</i>	0.351	0.592	0.460	0.586
<i>Modelo Estocástico:</i>	0.350	0.592	0.459	0.586
<i>Árvores Extremamente Aleatórias:</i>	0.357	0.598	0.477	0.578

Fonte: (SILVA, SOUZA e NÓBREGA, 2023b).

Observa-se que, em relação ao erro quadrático médio (MSE), todos os modelos apresentaram resultados próximos, variando de 0.343 a 0.357. Isso indica que os modelos têm uma boa capacidade de estimar os valores corretos em relação aos valores observados. O mesmo padrão é observado para a raiz do erro quadrático médio (RMSE), onde as diferenças entre os modelos são mínimas, variando de 0.585 a 0.598. A Figura 19 ilustra gráficos decorrentes dos dados apresentados na Tabela 15.

**Figura 19. Gráficos das métricas estatísticas do teste de validação nos rios**



Fonte: (SILVA, SOUZA e NÓBREGA, 2023b).

Quanto à média absoluta do erro (MAE), o modelo de Perceptron (MLPNN) obteve o menor valor (0.445), indicando uma melhor precisão na estimativa dos valores em comparação aos outros modelos. Em relação ao coeficiente de determinação ( $R^2$ ), que mede a proporção da variabilidade dos dados explicada pelo modelo, o modelo Perceptron (MLPNN) apresentou o valor mais alto (0.595), indicando uma melhor capacidade de previsão em relação aos outros modelos. Com base nos resultados do teste de validação, podemos concluir que o modelo Perceptron (MLPNN) demonstrou o melhor desempenho geral, apresentando menor erro médio absoluto e maior coeficiente de determinação. No entanto, é importante destacar que todos os modelos avaliados apresentaram resultados semelhantes e mostraram uma capacidade satisfatória de estimar os valores observados.

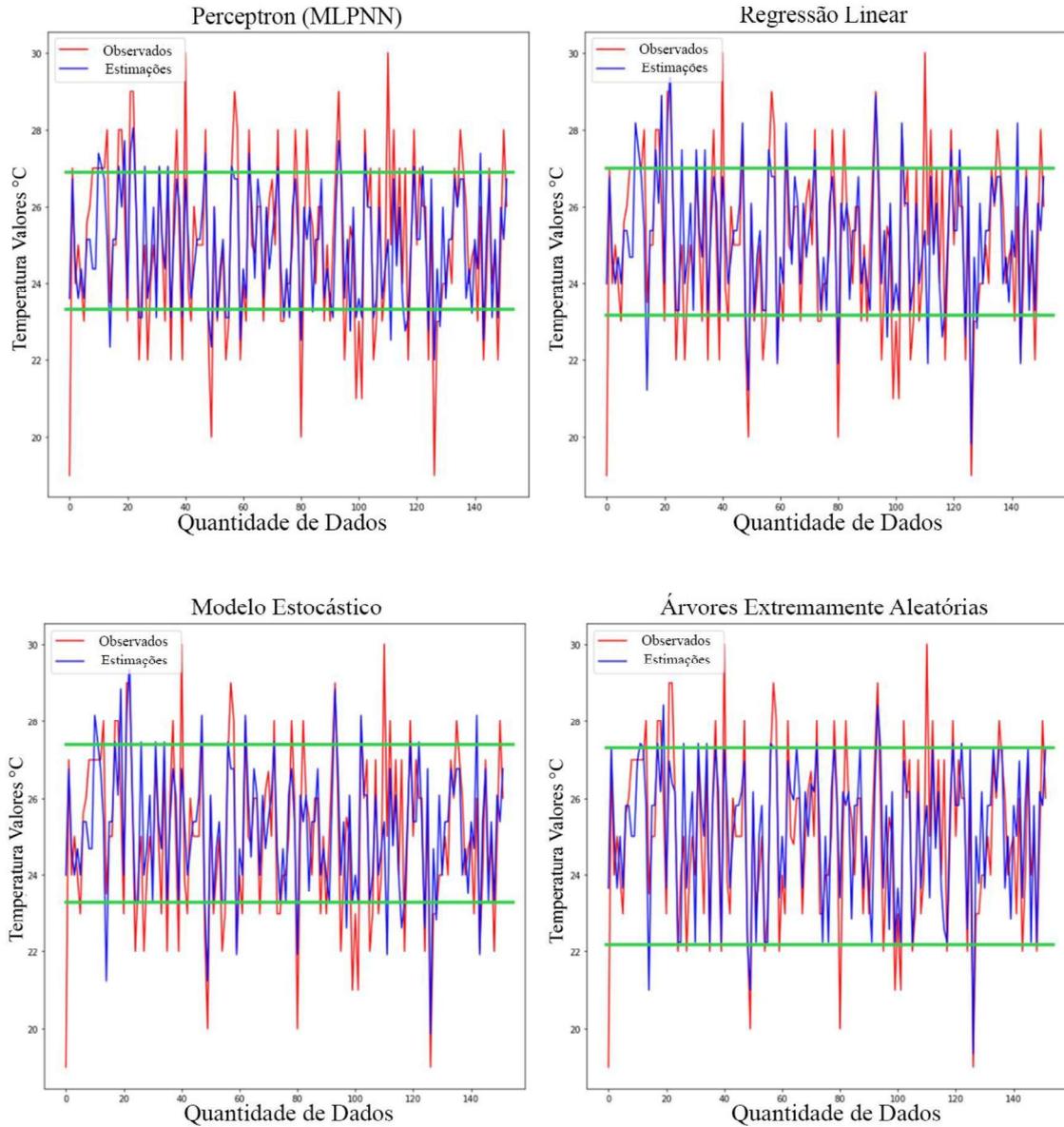
A Figura 20 exibe os dados observados e estimados no teste de validação, destacando que as estimativas seguem de perto os dados observados em temperaturas mais elevadas, com margens de erro em torno de 2°C. No entanto, para temperaturas mais baixas, abaixo de 23°C, as estimativas geralmente não conseguem acompanhar os dados observados na maioria dos casos. O modelo de árvore, no entanto, consegue acompanhar melhor essas temperaturas mais baixas.

#### **4.6.6 Resultado dos testes de validação dos reservatórios**

Os reservatórios, por sua vez, representam corpos de água de grandes proporções. A classificação precisa de um reservatório como sistema lótico ou lêntico depende da sua dinâmica hídrica específica, bem como da presença ou ausência de um fluxo contínuo de água. Durante o treinamento dos dados coletados nos reservatórios, mais uma vez a rede neural MLPNN demonstrou uma ligeira superioridade em relação aos outros modelos, seguida pelo modelo de regressão linear e o Stochastic. No entanto, desta vez, mesmo com uma amostra de dados maior nos reservatórios, os resultados apresentaram uma menor acurácia em comparação aos dados coletados nos rios.

Uma possível explicação para essa diferença de desempenho pode ser encontrada na seção anterior a análise dos cálculos de correlação, entre a temperatura do ar e a temperatura da água nos dois tipos de corpos d'água. Foi observado que a correlação positiva entre essas variáveis nos rios é mais forte do que nos reservatórios, o que pode ter contribuído para um melhor desempenho dos modelos nos dados dos rios. Isso acontece porque, em casos com correlações fortes, os padrões e relações entre os dados se tornam mais evidentes e, conseqüentemente, são mais facilmente aprendidos pelo modelo. Além disso, no caso dos dados dos reservatórios, a presença de ruídos, como outliers, teve um efeito negativo no desempenho do modelo. Por outro lado, nos dados dos rios, o conjunto de dados foi menor e continha menos ruídos, o que permitiu que o modelo se concentrasse nos padrões mais importantes. A Tabela 16 apresenta os resultados dos índices estatísticos obtidos nos reservatórios, fornecendo uma visão geral do desempenho dos modelos nesse contexto específico.

**Figura 20. Estimativas e observações para o conjunto de validação do rio**



Fonte: (SILVA, SOUZA e NÓBREGA, 2023b).

**Tabela 16. Desempenho dos modelos com os dados coletados nos Reservatórios**

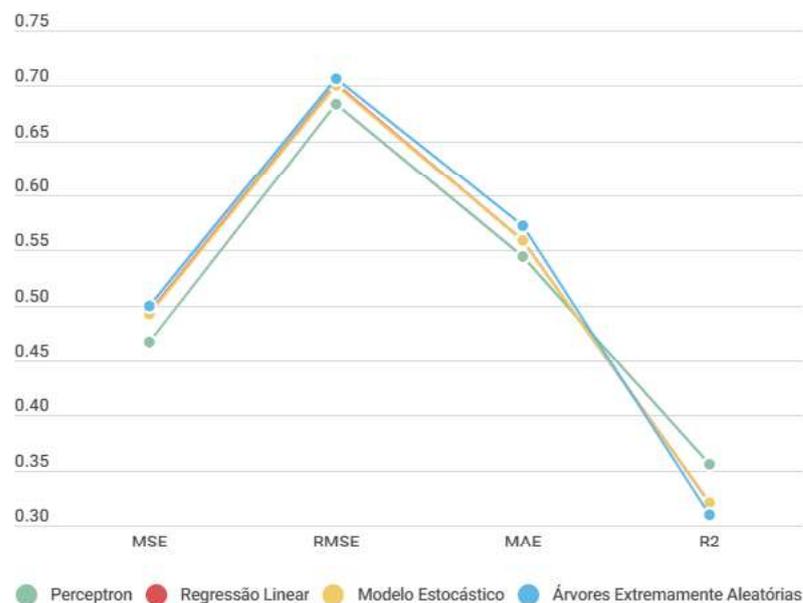
<b>Modelos:</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>R<sup>2</sup></b>
<b><i>Perceptron (MLPNN):</i></b>	0.467	0.683	0.544	0.356
<b><i>Regressão Linear:</i></b>	0.493	0.702	0.559	0.320
<b><i>Modelo Estocástico:</i></b>	0.492	0.701	0.559	0.321
<b><i>Árvores Extremamente Aleatórias:</i></b>	0.499	0.707	0.572	0.310

Fonte: (SILVA, SOUZA e NÓBREGA, 2023b).

Os resultados apresentados na Tabela 16 mostra que em termos de erro quadrático médio (MSE) e raiz do erro quadrático médio (RMSE), os modelos mostraram resultados próximos entre si, variando de 0.467 a 0.499 para o MSE e de 0.683 a 0.707 para o RMSE. Isso indica que os modelos apresentaram uma capacidade moderada de estimar os valores corretos em relação aos valores observados nos reservatórios.

No que diz respeito à média absoluta do erro (MAE), o modelo Perceptron (MLPNN) obteve o menor valor (0.544), indicando uma melhor precisão na estimativa dos valores em comparação aos outros modelos. Em relação ao coeficiente de determinação ( $R^2$ ), que mede a proporção da variabilidade dos dados explicada pelo modelo, todos os modelos apresentaram valores baixos, variando de 0.310 a 0.356. Esta observação fica evidente na Figura 21, que ilustra gráficos decorrentes dos dados apresentados na Tabela 16. Isso indica que os modelos tiveram dificuldade em capturar a variação dos dados nos reservatórios.

**Figura 21. Gráficos das métricas estatísticas do teste de validação nos rios**

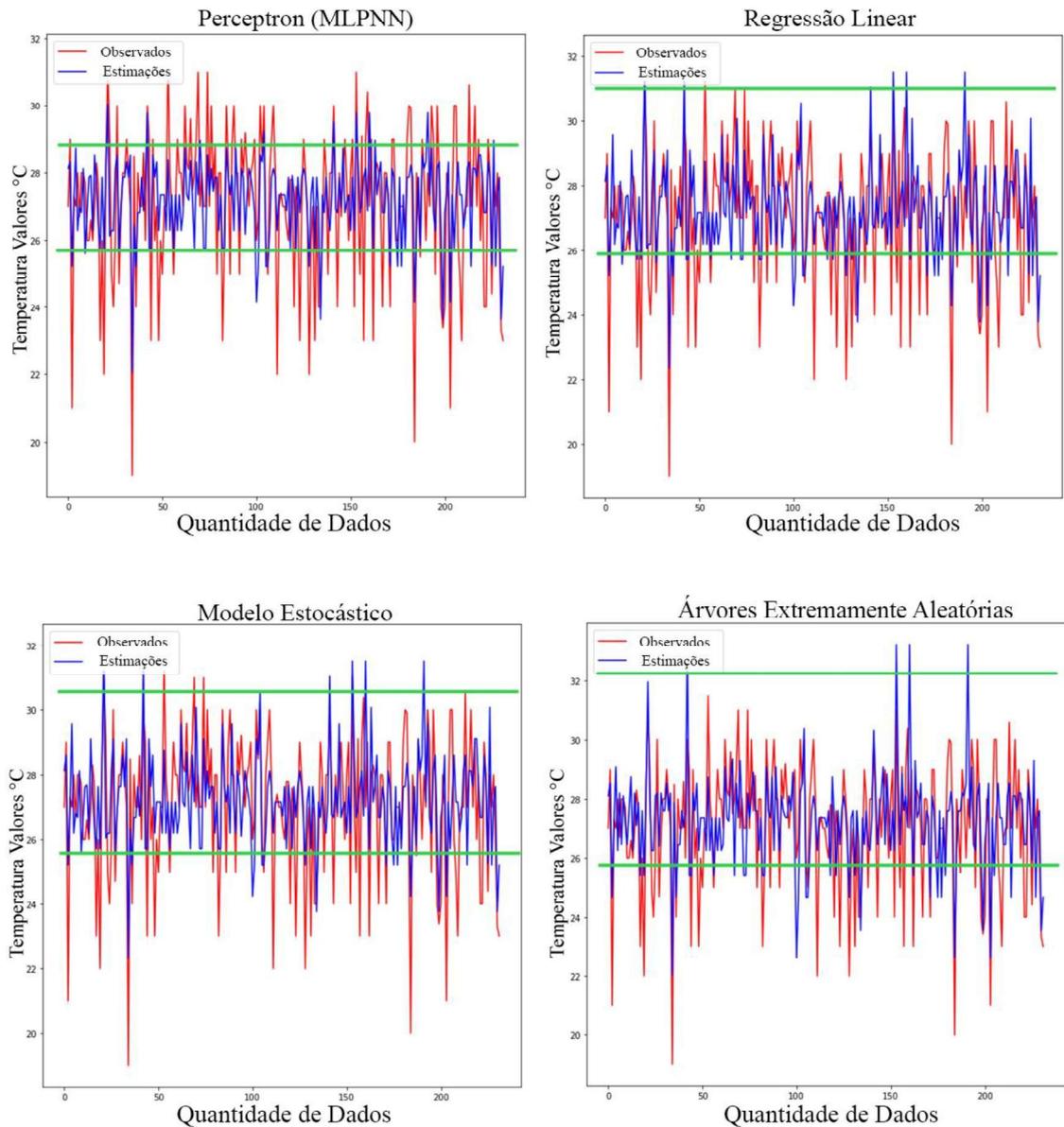


Fonte: (SILVA, SOUZA e NÓBREGA, 2023b).

Com base nos resultados do teste de validação, podemos concluir que os modelos tiveram um desempenho inferior ao lidar com os dados dos reservatórios em comparação aos dados dos rios. O modelo Perceptron (MLPNN) demonstrou uma ligeira superioridade em termos de menor erro médio absoluto.

No entanto, é importante ressaltar que todos os modelos avaliados apresentaram um desempenho moderado e mostraram uma capacidade razoável de estimar os valores observados nos reservatórios. A Figura 22 apresenta os dados observados e as estimativas obtidas durante o teste de validação. Ao analisar esta figura, torna-se evidente que os dados estimados geralmente não acompanham de forma precisa os dados observados durante a validação dos modelos. Em grande parte dos casos, especialmente em temperaturas inferiores a 23°C, a margem de erro pode chegar a 4°C, ou até mesmo ultrapassar esse valor em algumas instâncias. Por outro lado, em temperaturas mais altas, com exceção do perceptron, as estimativas geralmente excedem os dados observados, apresentando uma diferença média de cerca de 1°C e até 3°C no modelo de árvore. Esses resultados sugerem que os modelos tendem a ter um erro significativo, o que pode ser atribuído ao baixo volume de dados disponíveis para o treinamento, juntamente com a fraca correlação entre as variáveis envolvidas.

**Figura 22. Estimativas e observações para o conjunto de validação de reservatórios**



Fonte: (SILVA, SOUZA e NÓBREGA, 2023b).

#### 4.7 Considerações Finais

Este capítulo abordou a mais relevante contribuição científica deste trabalho, que se concentrou na realização de estimativas da temperatura da água através da aplicação de modelos de aprendizado de máquina com vistas ao desenvolvimento de sensores virtuais. Dentre as principais contribuições científicas observadas, destaca-se a superior performance demonstrada pelo modelo MLPNN no âmbito tanto dos tanques de aquicultura quanto dos reservatórios.

Adicionalmente, no escopo da análise dos dados concernentes aos cálculos de correlação, esta pesquisa apresentou resultados que evidenciam como correlações robustas entre os dados de entrada e os dados de saída são prontamente assimiladas pelos modelos, culminando em um notável aprimoramento da precisão na estimativa da temperatura da água, como atestado pelos resultados obtidos nos estágios de treinamento e validação.

## **5 UTILIZAÇÃO DE SENSORES VIRTUAIS PARA APOIAR A AQUICULTURA 4.0 E A GESTÃO DOS RECURSOS HÍDRICOS**

Neste capítulo é apresentado como foram realizadas as aplicações do sensoriamento virtual desenvolvido por meio do aprendizado de máquina para apoiar a Aquicultura 4.0, principalmente quando não se dispõe de sensores físicos para medir a temperatura da água ou como alternativa aos sensores defeituosos, além de se mostrar como uma forma alternativa de monitoramento da temperatura dos rios e reservatórios.

Este capítulo dedica-se à exploração de aplicações tecnológicas em um contexto do mundo real, com o intuito de oferecer contribuições técnicas para áreas que têm como objetivo primordial a produção de peixe e o monitoramento dos recursos hídricos.

## 5.1 Visão Geral

Os sensores virtuais são uma alternativa viável quando os sensores convencionais não estão disponíveis ou estão inoperantes devido a falhas ou defeitos. Neste trabalho, os sensores virtuais foram desenvolvidos com base em dados coletados e interfaces foram criadas para simplificar o monitoramento realizado pelos usuários. Essas interfaces foram disponibilizadas para dois parceiros: uma startup e uma agência governamental estadual:

1. **Startup Smart Rural:** dedicada à Aquicultura 4.0, está desempenhando um papel crucial ao oferecer tecnologias acessíveis e de baixo custo para aprimorar a indústria da produção de pescado.
2. **Agência Pernambucana de Águas e Clima (APAC):** desempenha um importante papel governamental, concentrando seus esforços no gerenciamento do monitoramento hidrometeorológico de rios e reservatórios do estado de Pernambuco.

## 5.2 Tecnologias utilizadas para a interface dos Sensores Virtuais

Nesta seção, serão delineadas as tecnologias empregadas na construção do algoritmo de aprendizado de máquina, uma vez já treinado com seus respectivos pesos, para viabilizar a aplicação dos sensores virtuais. Adicionalmente, será abordada a estrutura da interface do usuário destinada ao processo de monitoramento.

### 5.2.1 Tecnologias da interface Web para rios e reservatórios

Com o objetivo de disponibilizar os resultados da pesquisa por meio de um mapa interativo que permita ao usuário obter, com apenas um clique, a temperatura estimada para um reservatório ou rio em qualquer hora e local, fez-se necessário desenvolver um aplicativo Web utilizando marcação HTML.

O foco do desenvolvimento é a criação de uma interface de mapa que apresente pontos de localização dos corpos d'água. Para facilitar esse processo, foi utilizada a API MapTiler, uma plataforma de mapeamento e geocodificação que oferece acesso a serviços de mapas, imagens de satélite e outros recursos relacionados por meio de uma Interface de Programação de Aplicativos (API). Essa API permite que os desenvolvedores integrem facilmente mapas interativos e recursos de geolocalização em seus aplicativos e websites.

Dentro do código HTML, foi incorporado o algoritmo treinado com a melhor acurácia para estimar a temperatura da água. A linguagem de programação JavaScript foi utilizada para essa

finalidade. O dado de entrada, temperatura do ar, foi adquirido por meio de uma API web chamada Visual Crossing, pois a APAC ainda não disponibiliza uma API que permita obter a temperatura do ar local no momento do acesso.

Dessa forma, ao combinar essas tecnologias e serviços, será possível criar uma interface interativa que ofereça aos usuários informações precisas e em tempo real sobre a temperatura da água em diferentes locais de interesse.

### **5.2.2 Tecnologias da interface Web para tanques**

Um dos objetivos primordiais da indagação investigativa reside na elaboração de uma tecnologia de acessibilidade universal, apropriada para diversos produtores, considerando que as aplicações Web atualmente detêm um amplo alcance junto à maioria da população, em virtude de sua facilidade de acesso e usabilidade. Nesse contexto, a concepção de um Web site para o monitoramento em tanques de aquicultura emerge como uma possibilidade viável.

A interface proposta centra-se na estimativa da temperatura da água, baseada na localização geográfica do usuário ou no CEP (Código de Endereçamento Postal) fornecido, além de pontos de localização em um mapa utilizando a API Mapbox. Para viabilizar tal funcionalidade, optou-se pela utilização do kit de desenvolvimento Flutter, lançado pela Google em 2015 e implementado por meio da linguagem de programação Dart. Este kit capacita a criação de aplicativos compilados nativamente, abrangendo os sistemas operacionais Android, iOS, Windows, Mac e Linux e distribuição Web. No contexto deste trabalho, a ênfase recai sobre a plataforma Web, dado que uma parcela substancial dos pequenos produtores utiliza essa distribuição em diversos dispositivos móveis.

O algoritmo com a melhor acurácia obtida no treinamento foi implementado em Dart. Os dados de entrada foram adquiridos por meio da API Web Visual Crossing, selecionada por oferecer uma abordagem de baixo custo para obtenção das variáveis preditoras.

A meta final consiste na elaboração de uma interface que apresente simplicidade, permitindo ao produtor monitorar a temperatura da água em seus reservatórios com base em dados meteorológicos que possuem potencial relação com a temperatura da água.

### **5.3 Desenvolvimento das Interfaces para suporte tecnológico**

Esta seção apresenta os resultados alcançados no âmbito dos suportes tecnológicos desenvolvidos, que englobam um mapa interativo web destinado à representação dos rios e

reservatórios situados em Pernambuco, bem como um aplicativo móvel concebido para a utilização por parte dos produtores nos seus tanques de aquicultura.

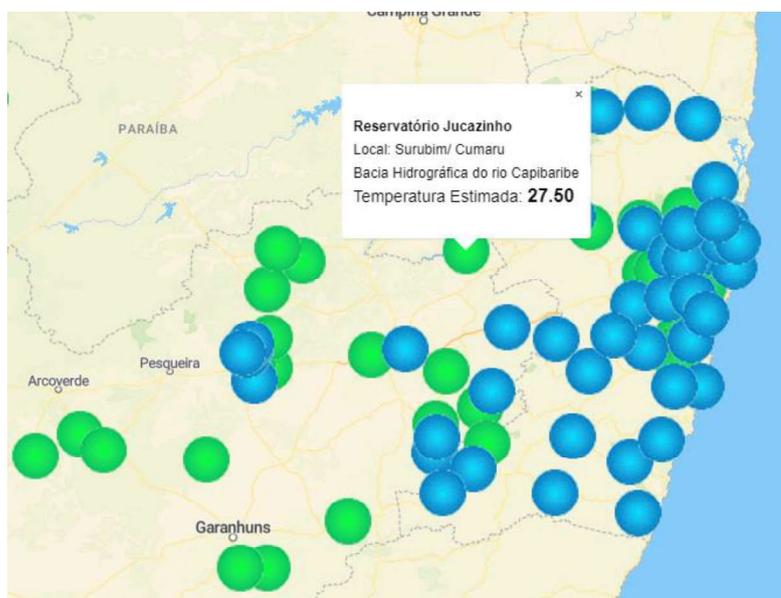
### 5.3.1 Aplicativo Web Para os Sensores Virtuais nos Rios e Reservatórios

Após a realização dos testes de validação dos modelos, foi possível identificar o MLPNN como o modelo de melhor desempenho. Com base nessa informação, foram obtidos os pesos da rede neural e implementado o feedforward (predict) para o cálculo da temperatura estimada da água em reservatórios e rios.

O sistema desenvolvido permite ao usuário navegar pelo mapa e, ao clicar em uma determinada localização (longitude e latitude), obter informações em tempo real sobre a temperatura da água estimada para o reservatório ou rio selecionado. Essa funcionalidade proporciona ao usuário acesso imediato às informações desejadas, independentemente de sua localização.

No protótipo da interface do mapa (Figura 23), os pontos azuis representam os rios, enquanto os pontos verdes representam os reservatórios. Essa diferenciação visual dos pontos permite uma identificação clara dos corpos d'água em análise, facilitando a interação e a compreensão dos resultados pelos usuários.

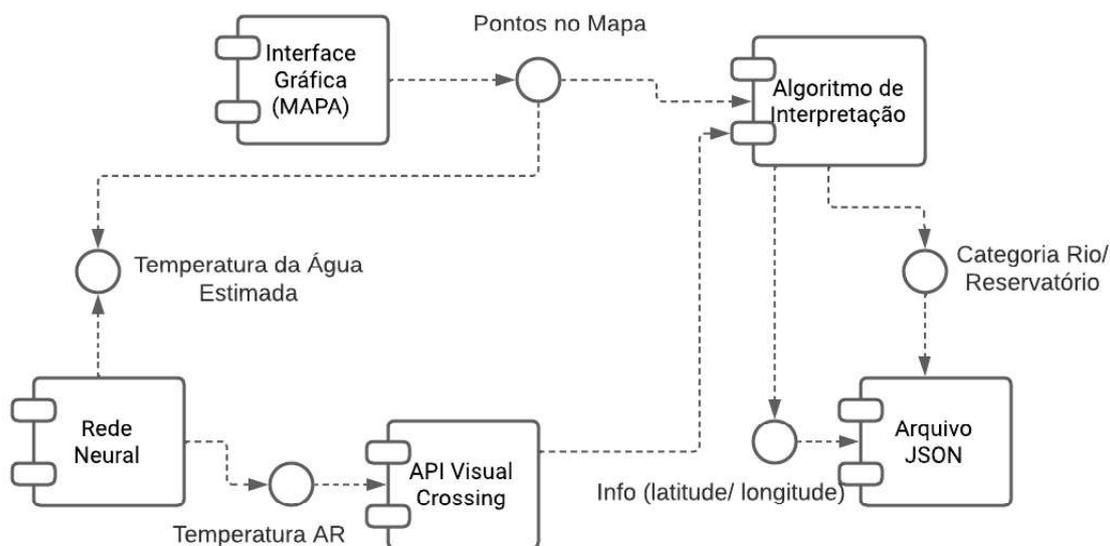
**Figura 23. Interface com o Usuário do mapa interativo**



Fonte: (SILVA, SOUZA e NÓBREGA, 2023b).

Uma perspectiva arquitetural da interface pode ser apreciada na Figura 24 por meio do Diagrama de Componentes da UML, o qual ilustra a sequência de interação entre os componentes do sistema e os relacionamentos de dependência entre eles.

**Figura 24. Diagrama de Componentes do Sensor Virtual para Rios e Reservatórios**



Fonte: Autor (2023).

Dessa forma, o protótipo desenvolvido é um aplicativo web interativo que permite aos usuários obter informações sobre a temperatura da água de corpos d'água específicos de forma rápida e precisa, utilizando o melhor modelo (MLPNN) treinado e os dados disponíveis. Isso contribui para uma melhor compreensão e monitoramento da temperatura da água em tempo real em diversas localidades.

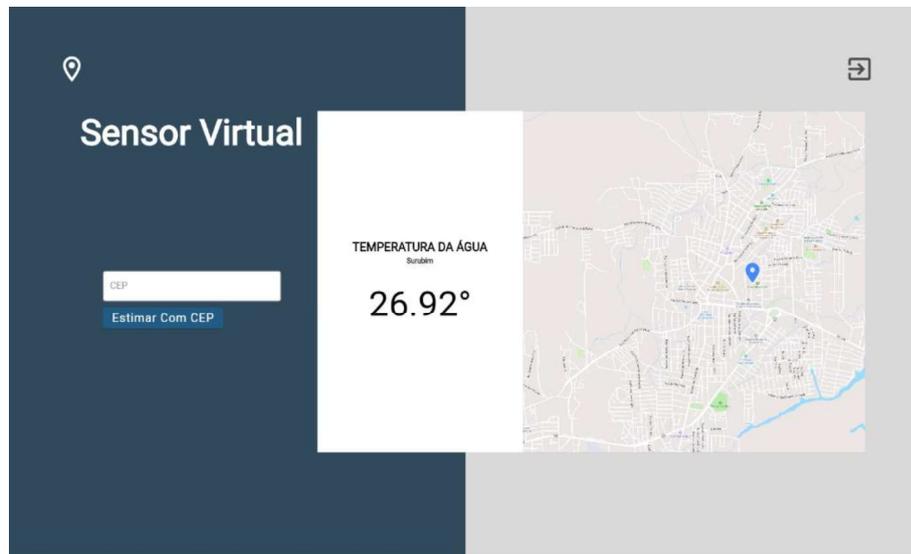
### 5.3.2 Aplicação Web Para os Sensores Virtuais nos Tanques de Aquicultura

Os procedimentos de validação conduzidos nos tanques de aquicultura evidenciaram que a rede neural de perceptron multicamadas (MLPNN) manifestou um desempenho superior em relação aos demais modelos investigados. Conseqüentemente, esse algoritmo foi escolhido para prosseguir, culminando na obtenção dos pesos da rede neural e na implementação do mecanismo de avanço (feedforward) para a estimativa da temperatura da água.

A aplicação desenvolvida, conforme ilustrado na Figura 25, confere ao usuário a capacidade de acessar a temperatura da água em concordância com sua geolocalização. Além disso, possibilita

a inserção de um CEP no caso de o usuário desejar obter informações sobre a temperatura da água em um tanque situado em localidade distinta.

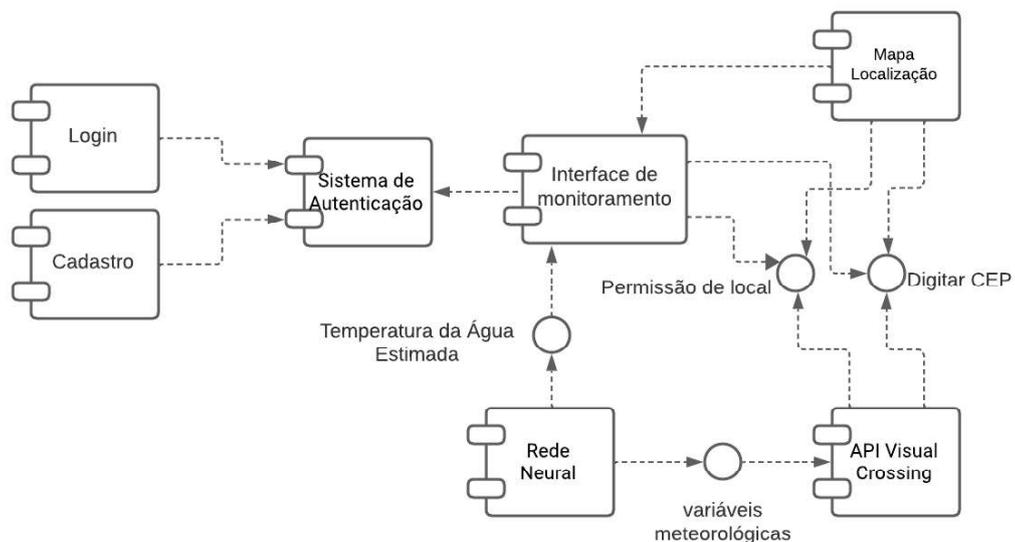
**Figura 25. Interface com o usuário do aplicativo Web**



Fonte: Autor (2023).

Para elucidar uma perspectiva da arquitetura da interface desenvolvida, a Figura 26 exibe, por meio de um diagrama UML, a sequência de interações entre os componentes do sistema e os relacionamentos de dependência entre eles.

**Figura 26. Diagrama de Componentes do Sensor Virtual para Aquicultura 4.0**



Fonte: Autor (2023).

## **5.4 Sensores Virtuais aplicados no Governo e na Iniciativa Privada**

Essa seção apresenta como o sensoriamento virtual pode ser utilizado pelo governo e pela iniciativa privada que necessitam realizar o monitoramento da temperatura da água. Considerando que em alguns momentos há indisponibilidade de sensores físicos, a aplicação de sensores virtuais pode ser uma alternativa ou complementar às soluções existentes.

### **5.4.1 A startup Smart Rural**

A Smart Rural é uma startup incubada na UFRPE dedicada à área da aquicultura inteligente, com o objetivo de aumentar a produtividade e reduzir os custos de produção de peixes e camarões por meio da implementação de tecnologias inovadoras e acessíveis, como a Internet das Coisas (IoT). Com o foco na acessibilidade tecnológica para um amplo público-alvo, a Smart Rural busca atender diferentes tipos de produtores, incluindo os pequenos produtores que necessitam de soluções acessíveis para melhorar sua produção. As soluções da Smart Rural são resultantes de pesquisas acadêmicas e tecnológicas desenvolvidas por pesquisadores vinculados aos departamentos de Computação (DC) e de Pesca e Aquicultura (DEPAQ) da UFRPE.

Diante da necessidade do monitoramento da temperatura da água nos tanques escavados de aquicultura, a aplicação em questão contribui para um serviço que está disponível no website da Smart Rural, acessível em <<https://www.smartrural.com.br>>.

A solução permite que os produtores que não possuam sensores físicos ou que tenham sensores inoperantes estimem a temperatura da água por meio do sensoriamento virtual, garantindo assim um monitoramento contínuo em seus tanques. A utilização da aplicação é gratuita e ilimitada. Os usuários interessados em obter informações sobre a temperatura da água estimada podem simplesmente acessar a seção de "Serviços" e clicar no botão "Começar Monitoramento". Em seguida só precisam realizar um cadastro para que a empresa possa gerenciar essas informações, compreender as necessidades e o perfil dos produtores e, desse modo, desenvolver soluções tecnológicas ainda mais inovadoras para contribuir com a Aquicultura 4.0. os usuários podem acessar a aplicação diretamente por meio do seguinte link: <<https://sensorvirtual.smartrural.com.br/>>.

#### **5.4.2 A Agência Pernambucana de Águas e Clima**

A Agência Pernambucana de Águas e Clima (APAC) foi instituída por meio da Lei Estadual nº14.028, datada de 26 de março de 2010, com o propósito de fortalecer o planejamento e a regulamentação dos diversos usos dos Recursos Hídricos do Estado e o Sistema Integrado de Gerenciamento dos Recursos Hídricos (SIGRH). O fortalecimento institucional visa à consolidação da Política Estadual de Recursos Hídricos (Lei nº 12.984/2005) e do Sistema Integrado de Gerenciamento dos Recursos Hídricos (SIGRH).

A missão da APAC consiste em executar a Política Estadual de Recursos Hídricos, planejar e regulamentar o uso da água em Pernambuco, gerenciar o monitoramento hidrometeorológico dos rios e reservatórios, além de realizar previsões de tempo e temperatura no Estado, como estabelecido na lei de sua criação. A visão da APAC é se tornar uma referência no Nordeste em relação ao planejamento e regulação de Recursos Hídricos, bem como na gestão integrada e participativa e na prevenção de enchentes e eventos climáticos extremos.

A temperatura da água desempenha um papel crucial nos processos hidrometeorológicos, impactando diretamente o ciclo da água e a interação entre a superfície terrestre e a atmosfera. A temperatura da água é fundamental para o ciclo hidrológico, que envolve a evaporação da água da superfície, sua ascensão na atmosfera, a formação de nuvens, a ocorrência de precipitação e o retorno da água para os corpos d'água. O ciclo hidrológico desempenha um papel vital na hidrometeorologia.

Em Pernambuco, os rios e reservatórios carecem de sensores físicos para monitorar a temperatura da água, porém várias barreiras surgem, incluindo falta de infraestrutura, desafios no monitoramento desses sensores e áreas remotas. Como alternativa, o monitoramento manual é utilizado, no entanto, essa abordagem não garante uma observação constante dos dados, resultando em lacunas nas informações.

A aplicação desenvolvida neste trabalho, por meio do uso de sensores virtuais, contribui significativamente para o monitoramento contínuo dos rios e reservatórios. Essa aplicação poderá ser uma das capacidades de monitoramento disponibilizadas no website da APAC, atualmente a aplicação se encontra acessível em: <<https://estimatetempapac.web.app/>>.

Os usuários interessados em obter informações sobre a temperatura da água estimada podem simplesmente interagir com o mapa, escolher uma localização e realizar um monitoramento em tempo real.

## **5.5 Considerações Finais**

As contribuições tecnológicas deste trabalho podem ter um impacto significativo na Aquicultura 4.0 e na gestão de recursos hídricos em Pernambuco. A introdução de sensores virtuais representa uma inovação importante no monitoramento da temperatura da água em rios e reservatórios. Isso é particularmente relevante em áreas onde sensores físicos são escassos devido às limitações geográficas ou restrições financeiras. A aplicação desenvolvida oferece uma solução acessível e confiável para preencher as lacunas de informações, permitindo o monitoramento contínuo em tempo real da temperatura da água.

O entendimento da temperatura da água desempenha um papel vital na hidrometeorologia, afetando o ciclo da água e os padrões climáticos. A capacidade de monitorar a temperatura da água de forma eficaz é essencial para previsões meteorológicas precisas e para a gestão sustentável dos recursos hídricos.

## 6 CONCLUSÃO

Este trabalho descreveu o desenvolvimento de sensores virtuais destinados à estimativa da temperatura da água em ambientes de aquicultura, rios e reservatórios, utilizando abordagens de aprendizado de máquina. Os dados utilizados para treinamento dos algoritmos de aprendizado de máquina foram obtidos a partir de sensores físicos de Internet das Coisas (IoT) instalados em tanques de aquicultura, juntamente com variáveis meteorológicas coletadas em estações de monitoramento. Além disso, foram incorporados dados de medições manuais da temperatura da água de rios e reservatórios do estado de Pernambuco, bem como da temperatura do ar.

O objetivo desse estudo foi o desenvolvimento de uma tecnologia acessível destinada a auxiliar na implementação da Aquicultura 4.0 e na gestão de recursos hídricos em rios e reservatórios. O capítulo final deste trabalho oferece um resumo conciso das atividades realizadas, destacando as contribuições científicas e técnicas alcançadas em conformidade com os resultados obtidos. Também são identificadas as principais limitações do estudo e delineadas as possíveis oportunidades de pesquisas futuras.

## 6.1 Visão Geral

Considerando a relevância do monitoramento contínuo da temperatura da água, bem como as limitações associadas aos sensores físicos convencionais e suas dificuldades de implantação em áreas remotas, torna-se imperativo a busca por soluções inteligentes acessíveis que possibilitem um monitoramento ininterrupto.

O emprego de algoritmos de aprendizado de máquina em diversas áreas de aplicação tem se tornado cada vez mais comum para a realização de estimativas. Este estudo teve como objetivo desenvolver, avaliar e aplicar um método que utiliza tais algoritmos para a criação de sensores virtuais. Para alcançar tal propósito, as seguintes etapas foram conduzidas, conforme delineado no capítulo introdutório e em conformidade com a metodologia de pesquisa adotada:

1. **Desenvolver Mapeamento Sistemático da Literatura:** conforme descrito no Capítulo 2, foi conduzida uma revisão abrangente da literatura para analisar as abordagens empregadas na estimativa da temperatura da água por meio de modelos estatísticos. Foi observado que muitos autores utilizaram mais de um modelo estatístico em seus estudos para comparar o desempenho e obter resultados mais robustos, no entanto, nenhum dos estudos descreviam a seleção desses modelos para o desenvolvimento de sensores virtuais na estimação da temperatura da água em tempo real, além de aplicar em outras áreas como aquicultura, possibilitando assim, uma oportunidade de pesquisa;
2. **Analisar as tecnologias para implementação dos algoritmos de aprendizagem de máquina:** conforme apresentado no Capítulo 3, foi conduzida uma análise dos procedimentos e tecnologias empregados na aquisição de dados, nos tipos e as técnicas de avaliação dos algoritmos, com a finalidade de subsidiar a seleção das tecnologias empregadas e identificar os passos necessários para a estimativa da temperatura da água;
3. **Implementar os algoritmos de aprendizagem de máquina:** conforme descrito no Capítulo 4, a partir dos resultados dos passos anteriores, foram implementados quatro modelos de aprendizado de máquina: Paramétricos - (1) regressão linear e (2) descida do gradiente estocástico; Não Paramétricos - (3) *árvores extremamente aleatórias* e (4) rede neural multicamada perceptron.
4. **Coletar dados das estações meteorológicas** - conforme detalhado no Capítulo 4, a aquisição das variáveis independentes foi executada por meio da coleta de dados na estação meteorológica mais próxima do tanque escavado, Be Weather (B2K), localizado

na base de aquicultura da UFRPE. Além disso, foi utilizada a API Visual Crossing disponível na Internet, especificamente as relacionadas às informações climáticas. Embora em comparação com a estação B2K, a API tenha resultando em uma acurácia inferior, ela se mostrou uma alternativa viável e de baixo custo;

5. **Coletar dados históricos em tanque:** conforme descrito no Capítulo 4, os dados das variáveis dependentes foram obtidos por meio dos sensores IoT disponíveis no tanque escavado da base de aquicultura da UFRPE, os quais registraram essas informações no servidor ThingsBoard.
6. **Coletar dados históricos dos rios e reservatórios:** conforme descrito no Capítulo 4, a partir dos dados fornecidos pela Agência Pernambucana de Águas e Clima (APAC), foram obtidas informações dos rios e reservatórios de Pernambuco, incluindo seus nomes e localizações, histórico da temperatura da água e do ar obtido em coletas manuais realizadas no período de 2011 a 2022;
7. **Treinar e avaliar a acurácia dos algoritmos de aprendizagem de máquina:** conforme descrito no Capítulo 4, foi alocada uma parcela correspondente a 20% dos dados para ser utilizada no processo de validação dos modelos. A fim de apoiar essa prática, quatro métricas estatísticas foram considerados;
8. **Analisar e selecionar algoritmo que apresentou melhor desempenho:** a Rede Neural Multicamada Perceptron demonstrou o desempenho mais satisfatório em relação aos outros algoritmos utilizados, inclusive quando comparada às variáveis provenientes das estações meteorológicas;
9. **Desenvolver mapa web interativo para reservatórios e rios:** foi desenvolvida uma interface web que possibilita a navegação por um mapa contendo a localização dos rios e reservatórios. Essa interface permite ao usuário obter informações em tempo real sobre a estimativa da temperatura da água nesses ambientes aquáticos;
10. **Desenvolver aplicativo de software para aquicultura:** desenvolveu-se um aplicativo web destinado que pode ser utilizado por soluções da Aquicultura 4.0. Essa aplicação permite a estimativa da temperatura da água em tempo real com base na localização atual do usuário ou em uma localização de interesse selecionada por ele;
11. **Aplicar sensoriamento virtual da temperatura da água em reservatórios e rios:** conforme descrito no Capítulo 5, o desenvolvimento de sensores virtuais apresenta-se

como uma alternativa para o monitoramento da temperatura da água de reservatórios e rios, podendo ser adotado por agências governamentais como a APAC;

12. **Aplicar sensoriamento virtual da temperatura da água em solução da Aquicultura 4.0:** conforme descrito no Capítulo 5, os resultados dos modelos tornaram possível o desenvolvimento de uma aplicação web que se tornou um serviço no website da startup Smart Rural que provê soluções de Aquicultura 4.0;
13. **Comunicar os resultados da pesquisa:** publicados resultados parciais deste trabalho em periódicos e eventos científicos - artigo “Estimação da temperatura da água: um levantamento dos modelos estatísticos para a aplicação em IOT e Tanques de Aquicultura” na revista Research, Society and Development (SILVA, SOUZA e NÓBREGA, 2023a); artigo “*Machine learning models applied to estimate the water temperature of rivers and reservoirs*” no journal ACTA of IMEKO - the International Measurement Confederation (SILVA, SOUZA e NÓBREGA, 2023b); resumo “Modelos de aprendizagem de máquina aplicado para estimar a temperatura da água em tanque escavado de Aquicultura 4.0” no I Workshop MATA – Meio Ambiente, Tecnologia e Agro (UFRPE, 2023). Além disso, foi elaborada a dissertação de mestrado com a descrição de todo o trabalho de pesquisa realizado.

## 6.2 Dificuldades e Lições Aprendidas

Por meio da condução deste trabalho, foi possível identificar uma série de desafios que se apresentaram durante a busca pelos resultados desejados. Entre as dificuldades identificadas, merecem destaque:

- A disponibilidade dos dados usados no processo de treinamento dos algoritmos de aprendizado de máquina. As falhas nos sensores resultaram em lacunas nos dados, tanto na estação meteorológica quanto nos sensores IoT. Isso exigiu a necessidade de aguardar um período de tempo para que os sensores pudessem continuar a coletar informações até que se acumulasse uma amostra de dados significativa;
- Os trabalhos encontrados na literatura que abordam o emprego de modelos de aprendizado de máquina para a estimativa ou previsão da temperatura da água focam predominantemente na avaliação desses modelos. Não foi identificado nenhum trabalho que se dedicasse ao desenvolvimento de interfaces capazes de transformar os resultados do treinamento desses modelos em sensores virtuais específicos para a tarefa em questão. Diante dessa ausência,

foi necessário recorrer a trabalhos em outras áreas de aplicação como referência para conceber a abordagem proposta; e

No que tange às lições aprendidas, este estudo permitiu a aquisição de conhecimento por meio das experiências vivenciadas, da observação sistemática, do registro metódico, da análise crítica e da disseminação do conhecimento adquirido. De forma particular, essas lições proporcionaram uma compreensão mais aprofundada de que:

- Apesar do notável crescimento na adoção de tecnologias digitais na área da aquicultura nas últimas décadas, ainda persiste uma carência considerável em termos de recursos tecnológicos, sobretudo para enfrentar os desafios presentes em regiões com escassez de recursos. Nestes contextos, a pesquisa e a inovação desempenham um papel fundamental para suprir essa deficiência e promover avanços significativos;
- A gestão hídrica de rios e reservatórios no governo carece de tecnologias inteligentes e representa um desafio, mas também oferece oportunidades para melhorias significativas na eficiência e sustentabilidade das soluções disponibilizadas para os interessados;
- O uso de modelos de aprendizado de máquina na ciência de dados para previsão e estimação tem sido comumente eficaz e bem-sucedido em uma ampla variedade de aplicações;
- Variáveis com correlações mais robustas têm o potencial de aprimorar a eficiência do aprendizado dos modelos, contudo, mesmo a inclusão de variáveis com correlações mais fracas pode contribuir para o aumento da acurácia dos modelos; e
- Foi possível enxergar várias questões no que diz respeito a aplicar métodos alternativos e acessíveis para monitorar a temperatura da água, bem como de sua importância.

### **6.3 Contribuições do Trabalho**

A partir dos resultados desse trabalho, possibilita-se a descrição das contribuições científicas e técnicas previstas no Capítulo 1.

#### **Contribuições Científicas:**

- Importância de estimar a temperatura da água na ecologia aquática e na aquicultura.
  - Conforme descrito no Capítulo 1, a temperatura da água desempenha um papel crucial na ecologia aquática e na aquicultura, portanto, deve-se realizar o

monitoramento de forma contínua através de sensores físicos estimando a temperatura da água caso esses sensores se tornem inoperáveis.

- Desenvolvimento de uma abordagem inteligente para estimar a temperatura da água de forma contínua e em tempo real por meio de modelos de aprendizado de máquina.
  - Conforme descrito no Capítulo 4, são apresentados os materiais e métodos que podem ser empregados como referência em pesquisas abrangendo diversas áreas de estudo. Este capítulo consolida os principais modelos utilizados na estimação de variáveis de regressão, bem como as métricas estatísticas empregadas para a avaliação desses modelos além das análises das correlações das variáveis;
- A importância das variáveis meteorológicas como preditoras na estimativa da temperatura da água em corpos de água usando modelos de aprendizado de máquina.
  - Ao incorporar essas variáveis como preditoras em modelos de aprendizado de máquina, podemos entender melhor os padrões e as tendências climáticas que afetam a temperatura da água. Isso permite fazer previsões eficazes da temperatura da água em diferentes momentos e locais.

#### **Contribuição Técnica:**

- Desenvolvimento de uma aplicação que torne possível estimar a temperatura da água dos tanques escavados de aquicultura, utilizando dados meteorológico de acordo com a localização.
  - Conforme delineado no início do Capítulo 5, foi concebida uma aplicação Web com o propósito de servir como uma alternativa aos sensores físicos tradicionais. Importante destacar que esta versão da aplicação é passível de evolução, de modo a atender a novos requisitos à medida que forem demandados.
- Desenvolvimento de um aplicativo web de suporte tecnológico por meio de um mapa interativo que permite ao usuário obter a estimação da temperatura da água dos rios e reservatórios de Pernambuco.
  - Conforme também descrito no Capítulo 5, esta abordagem tem como objetivo contribuir como suporte à Agência Pernambucana de Águas e Clima (APAC) para a

estimativa da temperatura da água em rios e reservatórios. Disponível em <https://estimatetempapac.web.app/>

## 6.4 Limitações e Trabalhos Futuros

### As principais limitações do trabalho são:

- Baixo volume na amostra de dados coletados;
- Falha na transmissão dos sensores tanto na estação meteorológica quanto nos sensores do tanque;
- Coleta de dados manual nos rios e reservatórios realizada em datas e horários específicos resultando em baixo volume de dados; e
- Disponibilidade de apenas uma variável preditora (temperatura do ar) coletada no rio e reservatórios.

Como trabalhos futuros, é crucial abordar e mitigar as limitações identificadas anteriormente, refinando métodos ou explorando abordagens alternativas para melhorar a eficácia e robustez do estudo. Buscar um conjunto de dados mais abrangente pode aprimorar a confiabilidade dos resultados e melhorar a acurácia dos modelos, enquanto a inclusão de variáveis meteorológicas adicionais, como radiação solar e vento, nos rios e reservatórios enriqueceria os modelos. A automação do processo de treinamento do modelo, enquanto os sensores físicos estiverem em operação pode ser uma melhoria significativa. Isso não apenas otimizaria o uso do recurso, mas também garantiria que o modelo esteja sempre atualizado e adaptado às condições em evolução. Além disso, estender a aplicação do modelo a outros tipos de tanques de aquicultura, como tanques de rede com sensores submersos em profundidade entre 2 a 4 metros, seria valioso para ampliar a generalização dos modelos e proporcionar insights específicos sobre diferentes configurações de tanques, contribuindo para uma compreensão mais abrangente e aplicável no contexto da aquicultura 4.0.

## REFERÊNCIAS

- ADI, E.; ANWAR, A.; BAIG, Z.; ZEADALLY, S. **Machine learning and data analytics for the IoT**. Neural Computing and Applications, v. 32, 2020. p. 16205-16233.
- ADI, E.; ANWAR, A.; BAIG, Z.; ZEADALLY, S. **Machine learning and data analytics for the IoT**. Neural Computing and Applications, v. 32, 2020. p. 16205-16233.
- AHMAD, M. W., REYNOLDS, J., & REZGUI, Y. **predictive modelling for solar thermal energy systems: a comparison of support vector regression, random forest, extra trees and regression trees**. Journal of cleaner production, 203, 2018, pp.810-821.
- AHMADI-NEDUSHAN, B.; ST-HILAIRE, A.; OUARDA, T. B.; BILODEAU, L.; ROBICHAUD, E.; THIÉMONGE, N.; BOBÉE, B. **Predicting river water temperatures using stochastic models: case study of the Moisie River (Québec, Canada)**. In: Hydrological Processes: An International Journal, v. 21, n. 1, 2007. p. 21-34.
- ALI, S.; MISHRA, P. K.; ISLAM, A.; ALAM, N. M. **Simulation of water temperature in a small pond using parametric statistical models: implications of climate warming**. In: Journal of Environmental Engineering, v. 142, n. 3, 2016. 04015085.
- ANTONOPOULOS, V. Z.; GIANNIOU, S. K. **Simulation of water temperature and dissolved oxygen distribution in Lake Vegoritis, Greece**. In: Ecological modelling, v. 160, n. 1-2, 2003. p. 39-53.
- B2K. **Technology Solutions**. Disponível em: <https://www.b2ktech.com.br/>. Acessado em: 22 de julho. 2023.
- BALTAR, R. **Roteiro para Elaboração do Projeto de Pesquisa**. Texto de orientação, 2000.
- BENYAHYA, L.; CAISSIE, D.; ST-HILAIRE, A.; OUARDA, T. B.; BOBÉE, B. **A review of statistical water temperature models**. In: Canadian Water Resources Journal, v. 32, n. 3, 2007. p. 179-192.
- BOKINGKITO JR, P. B.; LLANTOS, O. E. **Design and implementation of real-time mobile-based water temperature monitoring system**. In: Procedia Computer Science, vol. 124, 2017. p. 698-705.
- BUZELLI, G. M.; CUNHA-SANTINO, M. B. D. **Análise e diagnóstico da qualidade da água e estado trófico do reservatório de Barra Bonita, SP**. In: Revista Ambiente & Água, v. 8, 2013.p. 186-205.
- CAISSIE, D. **The thermal regime of rivers: a review**. In: Freshwater Biology, v. 51, n. 8, 2006. p. 1389-1406. DOI: 10.1111/j.1365-2427.2006.01597.x.
- CALDWELL, J.; RAJAGOPALAN, B.; DANNER, E. **Statistical modeling of daily water temperature attributes on the Sacramento River**. In: Journal of Hydrologic Engineering, v. 20, n. 5, 2015. 04014065.

- CALDWELL, R. J.; GANGOPADHYAY, S.; BOUNTRY, J.; LAI, Y.; ELSNER, M. M. **Statistical modeling of daily and subdaily stream temperatures: Application to the Methow River Basin, Washington**. Water Resources Research, v. 49, n. 7, 2013. p. 4346-4361.
- CHA, YoonKyung et al. **The relative importance of water temperature and residence time in predicting cyanobacteria abundance in regulated rivers**. In: Water research, v. 124, 2017. p. 11-19.
- CHAUDHURI, A. **Internet of Things, for Things, and by Things**. In: CRC Press, 2018.
- CHENARD J.F. and CAISSIE D. **Stream temperature modelling using artificial neural networks: application on Catamaran Brook, New Brunswick, Canada**. In: Hydrol. Process., v. 22, 2008. p. 336.
- CHOWDURY, M. S. U. et al. **IoT based real-time river water quality monitoring system**. In: Procedia Computer Science, v. 155, 2019. p. 161-168.
- CHRIST, R. D.; WERNLI, R. L. Chapter 2 - **The Ocean Environment**. In: The ROV Manual (Second Edition), 2014. p. 21-52. Butterworth-Heinemann, ISBN 9780080982885.
- COLSTON, J. M. et al. **Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study**. In: Environmental research, v. 165, 2018. p. 91-109.
- CORDOVIL, V. R. D. S.; FRANCELIN, M. M. **Organização e representações: uso de mapa mental e mapa conceitual**. Xix encontro nacional de pesquisa em ciência da informação (xix enancib); xix encontro nacional de pesquisa em ciência da informação (xix enancib), v. 24, n. 2. Disponível em: <https://brapci.inf.br/index.php/res/v/103035>, 2018.
- COSTA, A. F. S. et al. **Recursos hídricos**. In: Caderno de Graduação-Ciências Exatas e Tecnológicas-UNIT-SERGIPE, v. 1, n. 1, 2012. p. 67-73.
- DART programming language**. Disponível em: <https://dart.dev/>. Accessed at 06/30/2023.
- DAYAN, P.; SAHANI, M.; DEBACK, G. **Unsupervised learning**. In: The MIT Encyclopedia of the Cognitive Sciences, 1999. p. 857-859.
- DE CARA, B. E. D.; LUIZ, A. J. B.; NEVES, M. C. **Método para expansão de uma série temporal de temperatura da água a partir de dados do sistema automático de monitoramento de variáveis ambientais (SIMA): aplicação em aquicultura no reservatório de Furnas**. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO-SBSR 16, 2013. p.5491-5498.
- DE OLIVEIRA SANTOS, D.; DE FREITAS, E. B. **A Internet das Coisas e o Big Data inovando os negócios**. Refas-Revista Fatec Zona Sul, v. 3, n. 1, 2016. p. 1-18.
- DESAI, M., & SHAH, M. **An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)**. Clinical eHealth, 4, 2021, pp.1-11.

- DIMITROV, N.; GÖÇMEN, T. **Virtual sensors for wind turbines with machine learning-based time series models**. In: Wind Energy, v. 25, n. 9, 2022. p. 1626-1645,
- DOGAN, A.; BIRANT, D. **Machine learning and data mining in manufacturing**. In: Expert Systems with Applications, v. 166, 2021. p. 114060.
- DRAKOULELIS, M. et al. **Virtual sensors: an industrial application for illumination attributes based on machine learning techniques**. In: Annals of Telecommunications, v. 76, n. 7-8, 2021. p. 529-535.
- EL NAQA, I.; MURPHY, M. J. **What is machine learning?** Springer International Publishing, 2015.
- EMBRAPA. **Trajatória / Desempenho Recente do Agro Nacional**. In: Embrapa. Disponível em: <https://www.embrapa.br/visao-de-futuro/trajetoria-do-agro/desempenho-recente-do-agro/aquicultura>. Acessado em: 17 de Mar. 2023.
- ERDEM, A. **Mind Maps as a Lifelong Learning Tool**. In: Universal Journal of Educational Research, v. 5, n. 12A, 2017. p. 1-7.
- FERCHICHI, H.; ST-HILAIRE, A.; OUARDA, T. B.; LÉVESQUE, B. **Prediction of coastal water temperature using statistical models**. In: Estuaries and Coasts, v. 45, n. 7, 2022. p. 1909-1927.
- FLUTTER. **Build apps for any screen**. Disponível em: <https://flutter.dev/>. Acessado em: 30 de julho. 2023.
- FOURATI, M. A.; CHEBBI, W.; KAMOUN, A. **Development of a web-based weather station for irrigation scheduling**. In: 2014 third IEEE international colloquium in information science and technology (CIST), 2014. p. 37-42.
- FUJIWARA, K. et al. **Soft-sensor development using correlation-based just-in-time modeling**. In: AIChE Journal, v. 55, n. 7, 2009. p. 1754-1765.
- FUNARI, E.; MANGANELLI, M.; SINISI, L. **Impact of climate change on waterborne diseases**. In: Annali dell'Istituto superiore di sanità, v. 48, 2012. p. 473-487.
- GARBOSSA, L. H. P., NOVAES, A. L. T., & LAPA, K. R. **Automação de baixo custo como alternativa para monitoramento hidrológico**. Agropecuária Catarinense, v. 33, n. 3, p. 72-76, 2020.
- GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. In: Editora Plageder, 2009.
- GOTHREAUX, C. T.; GREEN, C. C. **Effects of shading on the reproductive output and embryo viability of Gulf Killifish**. North American Journal of Aquaculture, v. 74, n. 2, p. 266-272, 2012.
- GULERYUZ, D. **Estimation of soil temperatures with machine learning algorithms—Giresun and Bayburt stations in Turkey**. In: Theoretical and Applied Climatology, v. 147, n. 1-2, 2022. p. 109-125.

- HAGUE, M. J.; PATTERSON, D. A. **Evaluation of statistical river temperature forecast models for fisheries management**. North American Journal of Fisheries Management, v. 34, n. 1, 2014. p. 132-146.
- HARVEY, R.; LYE, L.; KHAN, A.; PATTERSON, R. **The influence of air temperature on water temperature and the concentration of dissolved oxygen in Newfoundland Rivers**. Canadian Water Resources Journal, v. 36, n. 2, 2011. p. 171-192.
- HEDDAM, S.; PTAK, M.; ZHU, S. **Modelling of daily lake surface water temperature from air temperature: Extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN**. In: Journal of Hydrology, v. 588, 2020. p. 125130.
- ILYAS, E. B. et al. **Virtual sensor creation to replace faulty sensors using automated machine learning techniques**. In: 2020 Global Internet of Things Summit (GloTS), IEEE, 2020, p. 1-6.
- JAVASCRIPT. Disponível em: <https://developer.mozilla.org/en-US/docs/Web/JavaScript>. Acessado em 06/30/2023.
- JEPPESEN, E.; IVERSEN, T. M. **Two simple models for estimating daily mean water temperatures and diel variations in a Danish low gradient stream**. In: Oikos, 1987. p. 149-155.
- JIANG, D.; XU, Y.; LU, Y.; GAO, J.; WANG, K. **Forecasting Water Temperature in Cascade Reservoir Operation-Influenced River with Machine Learning Models**. In: Water, v. 14, n. 14, 2022. p. 2146.
- JOSHI, S. **What Is Statistical Modeling? When and Where to Use It**. In: G2: Business Software and Services Reviews. Disponível em: <https://www.g2.com/articles/statistical-modeling>. Acessado em: 13 de fev. 2023.
- KAGGLE. **Yor Machine Learning and Data Science Community**. Disponível em: <https://www.kaggle.com/>. Acessado em: 19 de Dezembro. 2023.
- KAMAL, R. **A Short Guide On IoT Data Collection**. In: Intuz, 2022. Disponível em: <https://www.intuz.com/blog/guide-on-iot-data-collection>. Acessado em: 15 de Mar. 2023.
- KANO, M.; FUJIWARA, K. **Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications**. Journal of Chemical Engineering of Japan, v. 46, n. 1, 2013. p. 1-17.
- KARVONEN, A. et al. **Increasing water temperature and disease risks in aquatic systems: climate change increases the risk of some, but not all, diseases**. International journal for parasitology, v. 40, n. 13, p. 1483-1488, 2010.
- KEELE, S. **Guidelines for performing systematic literature reviews in software engineering**, 2007.
- KETKAR, N., & KETKAR, N. **Stochastic gradient descent**. Deep learning with Python: A hands-on introduction. 2017, ISBN 978-1-4842-2766-4, pp.113-132.

- KODALI, R. K.; MANDAL, S. **IoT based weather station**. In: 2016 international conference on control, instrumentation, communication and computational technologies (ICCICCT), 2016. p. 680-683. IEEE.
- KODITALA, N. K.; PANDEY, P. S. **Water quality monitoring system using IoT and machine learning**. In: 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), 2018. pp. 1-5. IEEE.
- KUBITZA, F.; CAMPOS, J. **Aquicultura no Brasil. Conquistas e Desafios. Panorama da Aquicultura**, v. 25, n. 150, 2015. p. 11-13.
- LAANAYA, F.; ST-HILAIRE, A.; GLOAGUEN, E. **Water temperature modelling: comparison between the generalized additive model, logistic, residuals regression and linear regression models**. In: Hydrological Sciences Journal, v. 62, n. 7, 2017. p. 1078-1093.
- LARNIER, K.; ROUX, H.; DARTUS, D.; CROZE, O. **Water temperature modeling in the Garonne River (France)**. In: Knowledge and Management of Aquatic Ecosystems, n. 398, 2010. p. 04.
- LEÓN, C. J.; HERNÁNDEZ, J. M.; LEÓN-SANTANA, M. **The effects of water temperature in aquaculture management**. In: Applied Economics, v. 38, n. 18, 2006. p. 2159-2168.
- LETCHER, B. H.; HOCKING, D. J.; O'NEIL, K.; WHITELEY, A. R.; NISLOW, K. H.; O'DONNELL, M. J. **A hierarchical model of daily stream temperature using air-water temperature synchronization**, autocorrelation, and time lags. PeerJ, v. 4, 2016. e1727.
- LIMA, E. L. **Módulo de sensores para monitoramento da qualidade da água com transmissão sem fio utilizando plataforma de prototipagem**. Disponível em: <http://200.128.81.65:8080/handle/tede/863>. Acessado em: 11 de fev. 2023.
- LIU, W. C.; CHEN, W. B. **Prediction of water temperature in a subtropical subalpine lake using an artificial neural network and three-dimensional circulation models**. In: Computers & Geosciences, v. 45, 2012. p. 13-25.
- LOURENÇO, J. D. P.; MALTA, J. D. O.; DE SOUSA, F. N. **A importância de monitorar a qualidade da água na piscicultura**. In: Embrapa. Disponível em: <https://www.infoteca.cnptia.embrapa.br/bitstream/doc/668748/1/IT599.pdf>. Acessado em: 10 de fev. 2023.
- LUO, M. et al. **The effects of water temperature on the survival, feeding, and growth of the juveniles of *Anguilla marmorata* and *A. bicolor pacifica***. Aquaculture, v. 400, p. 61-64, 2013.
- LYONETTE, D. R. M. et al. **A new directional comparison technique for the protection of teed transmission circuits**. In: 2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No. 00CH37077). IEEE, 2000. p. 1979-1984.
- MAHESH, B. **Machine learning algorithms-a review**. In: International Journal of Science and Research (IJSR), v. 9, n. 1, 2020. p. 381-386.

- MANNRICH, J. P.; da ROCHA NEVES Jr, O. **Proposta de estação meteorológica baseada em internet das coisas de código aberto e baixo custo como ferramenta para articulação entre novas tecnologias digitais da informação e comunicação no ensino de viências/física na educação básica.** In: RENOTE, v. 19, n. 2, 2021. p. 323-334.
- MAPBOX. **Location intelligence for business.** Disponível em: <https://www.mapbox.com/>. Acessado em: 28 de julho. 2023.
- MAPBOX. **Location intelligence for business.** Disponível em: <https://www.mapbox.com/>. Accessed at 07/30/2023.
- MAPTILER. **Maps for developers.** Disponível em: <https://www.mapbox.com/>. Acessado em: 29 de julho. 2023.
- MARTIN, D. et al. **Virtual sensors.** In: Business & Information Systems Engineering, v. 63, p. 315-323, 2021.
- MASTI, D. et al. **A machine-learning approach to synthesize virtual sensors for parameter-varying systems.** In: European Journal of Control, v. 61, 2021.p. 40-49.
- MCGRATH, E. O.; NEUMANN, N. N.; NICHOL, C. F. **A statistical model for managing water temperature in streams with anthropogenic influences.** River Research and Applications, v. 33, n. 1, 2017. p. 123-134.
- MCGRATH, E. O.; NEUMANN, N. N.; NICHOL, C. F. **A statistical model for managing water temperature in streams with anthropogenic influences.** In: River Research and Applications, v. 33, n. 1, 2017. p. 123-134.
- MENON, K. U.; DIVYA, P.; RAMESH, M. V. **Wireless sensor network for river water quality monitoring in India.** In: 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), 2012. p. 1-7. IEEE.
- MOHR, S.; DRAINAS, K.; GEIST, J. **Assessment of Neural Networks for Stream-Water-Temperature Prediction.** In: 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 2021. p. 891-896.
- MOHSENI, O.; STEFAN, H. G.; ERICKSON, T. R. **A nonlinear regression model for weekly stream temperature.** In: Water Resour. Res., v. 34, n. 10, 1998. p. 2685–2692.
- MONTGOMERY, D. C., PECK, E. A., & VINING, G. G. **Introduction to linear regression analysis.** John Wiley & Sons, 2021, ISBN 9781119578741.
- MORRILL, J. C.; BALES, R. C.; CONKLIN, M. H. **Estimating stream temperature from air temperature: implications for future water quality.** In: Journal of Environmental Engineering, v. 131, n. 1, 2005. p. 139-146.
- MOURA, G. D. S. et al. **Desempenho e atividade de amilase em tilápias-do-nylo submetidas a diferentes temperaturas.** In: Pesquisa Agropecuária Brasileira, v. 42, 2007. p. 1609-1615.

Nações Unidas Brasil. **Objetivos de Desenvolvimento Sustentável no Brasil**. Disponível em: <https://brasil.un.org/pt-br/sdgs> . Acessado em: 11 de fev. 2023.

NGUEKU, B. **Water monitoring in fish ponds**. In: International Journal of Fisheries and Aquatic Studies, vol. 2, 2014. p. 32-32.

NOBRE, P. **Aquecimento global, oceanos & sociedade**. In: InterfacEHS-Revista de Saúde, Meio Ambiente e Sustentabilidade, v. 3, n. 1. Disponível em: <http://www3.sp.senac.br/hotsites/blogs/InterfacEHS/wp-content/uploads/2013/07/art-1-2008-6.pdf>. Acessado em: 11 de fev. 2023.

OBIERO, K. et al. **Predicting uptake of aquaculture technologies among smallholder fish farmers in Kenya**. Aquaculture International, v. 27, p. 1689-1707, 2019.

PARRA, L. et al. **Design and deployment of low-cost sensors for monitoring the water quality and fish behavior in aquaculture tanks during the feeding process**. Sensors, v. 18, n. 3, p. 750, 2018.

PEIXEBR. **Associação Brasileira da Piscicultura**. Disponível em: <https://www.peixebr.com.br/>.

PETERSEN, K.; FELDT, R.; MUJTABA, S.; MATTSSON, M. **Systematic Mapping Studies in Software Engineering**. In: 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), University of Bari, Italy, 2008.

PICCOLROAZ, S.; CALAMITA, E.; MAJONE, B.; GALLICE, A.; SIVIGLIA, A.; TOFFOLON, M. **Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches**. In: Hydrological Processes, v. 30, n. 21, 2016. p. 3901-3917.

PICCOLROAZ, S.; TOFFOLON, M.; MAJONE, B. **A simple lumped model to convert air temperature into surface water temperature in lakes**. In: Hydrology and earth system sciences, v. 17, n. 8, 2013. p. 3323-3338.

PIKE, A.; DANNER, E.; BOUGHTON, D.; MELTON, F.; NEMANI, R.; RAJAGOPALAN, B.; LINDLEY, S. **Forecasting river temperatures in real time using a stochastic dynamics approach**. In: Water Resources Research, v. 49, n. 9, 2013. p. 5168-5182.

POLANSKI, D. M.; ANGELÓPOULOS, C. M. **Modelling virtual sensors for indoor environments with machine learning**. In: 18th International Conference on Distributed Computing in Sensor Systems (DCOSS), 2022, p. 222-228.

PYTHON. **Programming Language**. Disponível em: <https://www.python.org/>. Acessado em: 20 de janeiro. 2023.

QUARTIERI, R. **Termometria**. PROFESSORA RENATA QUARTIERI. Disponível em: <https://renataquartieri.com/vestibular-2/exercicios/termometria/>. Acessado em: 11 de fev. 2023.

RABI, A.; HADZIMA-NYARKO, M.; ŠPERAC, M. **Modelling river temperature from air temperature: case of the River Drava (Croatia)**. In: Hydrological sciences journal, v. 60, n. 9, 2015. p. 1490-1507.

- RAND, J. M.; NANKO, M. O.; LYKKEGAARD, M. B.; WAIN, D.; KING, W.; BRYANT, L. D.; HUNTER, A. **The human factor: Weather bias in manual lake water quality monitoring.** In: *Limnology and Oceanography: Methods*, vol. 20, n. 5, 2022. p. 288-303.
- RASTEGARI, H.; NADI, F.; LAM, S. S.; ABDULLAH, M. I.; KASAN, N. A.; RAHMAT, R. F.; MAHARI, W. A. W. **Internet of Things in aquaculture: A review of the challenges and potential solutions based on current and future trends.** *Smart Agricultural Technology*, 2023. 100187.
- RAVEENDRANATHAN, N. et al. **From modeling to implementation of virtual sensors in body sensor networks.** *IEEE Sensors Journal*, v. 12, n. 3, 2011. p. 583-593.
- READ, J. S. et al. **Process-guided deep learning predictions of lake water temperature.** *Water Resources Research*, v. 55, n. 11, 2019. p. 9173-9190.
- REGAN, F. et al. **A demonstration of wireless sensing for long term monitoring of water quality.** In: 2009 IEEE 34th Conference on Local Computer Networks, 2009, p. 819-825.
- RESENDE, E. K. D. **Pesquisa em rede em aquicultura: bases tecnológicas para o desenvolvimento sustentável da aqüicultura no Brasil. Aquabrasil.** In: *Revista Brasileira de Zootecnia*, vol. 38, 2009. p. 52-57.
- RODRIGUES, A. S. de L.; MALAFAIA, G.; CASTRO, P. D. T. A. **A importância da avaliação do habitat no monitoramento da qualidade dos recursos hídricos: uma revisão.** In: *SaBios-Revista de Saúde e Biologia*, v. 5, 2010. n. 1.
- RODRIGUES, L. S. et al. **Panorama da aquicultura no Brasil: desafios e oportunidades.** *BNDES Setorial*, n. 35, mar. 2012, p. 421-463.
- ROSE, K.; ELDRIDGE, S.; CHAPIN, L. **The internet of things: An overview.** In: *The Internet Society (ISOC)*, 2015, p. 1-50.
- SAEED, S.; HONEYEH, K.; OZGUR, K.; WEN-CHENG, L. **Water temperature prediction in a subtropical subalpine lake using soft computing techniques.** In: *Earth Sciences Research Journal*, v. 20, n. 2, 2016. p. 1-11.
- SAHOO, G. B.; SCHLADOW, S. G.; REUTER, J. E. **Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models.** In: *Journal of hydrology*, v. 378, n. 3-4, 2009. p. 325-342.
- SALEEM, T. J.; CHISHTI, M. A. **Deep learning for Internet of Things data analytics.** *Procedia Computer Science*, v. 163, 2019. p. 381-390.
- SANTOS, J. A.; PARRA FILHO, D. **Metodologia científica.** Disponível em: <http://ava.institutoalfa.com.br/tcc/apostila-de-metodologia-cient%C3%ADfica.pdf>, 14 de fev. 2023.

- SATHYA, R.; ABRAHAM, A. **Comparison of supervised and unsupervised learning algorithms for pattern classification.** International Journal of Advanced Research in Artificial Intelligence, v. 2, n. 2, 2013. p. 34-38.
- SILVA, R. B. D. et al. **Estações meteorológicas de código aberto: Um projeto de pesquisa e desenvolvimento tecnológico.** In: Revista brasileira de ensino de Física, v. 37, 2015.
- SILVA, SOUZA e NÓBREGA. **Estimation of water temperature: a survey of statistical models for application in IOT and Aquaculture Tanks.** Research, Society and Development, v. 12, n. 4, p. e26912441142. 2023a
- SILVA, SOUZA e NÓBREGA. **Machine learning models applied to estimate the water temperature of rivers and reservoirs.** The International Measurement Confederation , Acta Imeko. 2023b
- SILVERIO-FERNÁNDEZ, M.; RENUKAPPA, S.; SURESH, S. **What is a smart device? - A conceptualization within the paradigm of the internet of things.** In: Visualization in Engineering, v. 6, n. 1, 2018. p. 1-10.
- SIMON, M. K.; GOES, J. **Developing a theoretical framework.** In: Seattle, WA: Dissertation Success, LLC, 2011.
- SIQUEIRA, T. V. D. **Aquicultura: a nova fronteira para aumentar a produção mundial de alimentos de forma sustentável,** 2017.
- SPYDER. **The Scientific python Development Environment.** Disponível em: <https://www.spyder-ide.org/>. Acessado em: 20 de janeiro. 2023.
- ST-HILAIRE, A.; OUARDA, T. B.; BARGAOU, Z.; DAIGLE, A.; BILODEAU, L. **Daily river water temperature forecast model with ak-nearest neighbour approach.** In: Hydrological Processes, v. 26, n. 9, 2012. p. 1302-1310.
- SUN, M.; CHEN, J.; LI, D. **Water temperature prediction in sea cucumber aquaculture ponds by RBF neural network model.** In: 2012 International Conference on Systems and Informatics (ICSAI2012), 2012. p. 1154-1159.
- TASNIM, B.; JAMILY, J. A.; FANG, X.; ZHOU, Y.; HAYWORTH, J. S. **Simulating diurnal variations of water temperature and dissolved oxygen in shallow Minnesota lakes.** In: Water, v. 13, n. 14, 2021. p. 1980.
- TERRA, J. **Regression vs. Classification in Machine Learning for Beginners.** In: Simplilearn, 2023. Disponível em: [https://www.simplilearn.com/regression-vs-classification-in-machine-learning-article#:~:text=The%20regression%20algorithm's%20task%20is%20mapping%20input%20value%20\(x\)%20with,are%20used%20with%20continuous%20data](https://www.simplilearn.com/regression-vs-classification-in-machine-learning-article#:~:text=The%20regression%20algorithm's%20task%20is%20mapping%20input%20value%20(x)%20with,are%20used%20with%20continuous%20data). 17 de Mar. 2023.
- TOFFOLON, M.; PICCOLROAZ, S. **A hybrid model for river water temperature as a function of air temperature and discharge.** In: Environmental Research Letters, v. 10, n. 11, 2015.114011.

- VEENSTRA, J. et al. **Impact of net pen aquaculture on lake water quality**. In: Water Science and Technology, v. 47, n. 12, 2003. p. 293-300.
- VISUALCROSSING. **weaheer data and API**. Disponível em: <https://www.visualcrossing.com/>. Acessado em: 20 de julho. 2023.
- WEBB, B. W.; CLACK, P. D.; WALLING, D. E. **Water–air temperature relationships in a Devon river system and the role of flow**. In: Hydrological processes, v. 17, n. 15, 2003. p. 3069-3084.
- WENXIAN, Guo et al. **PSO-BP neural network model for predicting water temperature in the middle of the Yangtze river**. In: 2010 International Conference on Intelligent Computation Technology and Automation. IEEE, 2010. p. 951-954.
- YANG, H. et al. **A review of remote sensing for water quality retrieval: progress and challenges**. Remote Sensing, v. 14, n. 8, p. 1770, 2022.
- YONEKURA, K. et al. **Short-term local weather forecast using dense weather station by deep neural network**. In: 2018 IEEE international conference on big data (big data), 2018. p. 1683-1690. IEEE.
- ZAIDAN, M. A. et al. **Virtual SenSorS: toward High-reSolution air Pollution Monitoring uSing ai and iot**. In: IEEE Internet of Things Magazine, v. 6, n. 1, 2023. p. 76-81.
- ZHU, S.; HEDDAM, S.; NYARKO, E. K.; HADZIMA-NYARKO, M.; PICCOLROAZ, S.; WU, S. **Modeling daily water temperature for rivers: comparison between adaptive neuro-fuzzy inference systems and artificial neural networks models**. In: Environmental Science and Pollution Research, v. 26, 2019. p. 402-420..
- ZHU, S.; HEDDAM, S.; WU, S.; DAI, J.; JIA, B. **Extreme learning machine-based prediction of daily water temperature for rivers**. In: Environmental Earth Sciences, v. 78, 2019. p. 1-17,
- ZHU, S.; NYARKO, E. K.; HADZIMA-NYARKO, M. **Modelling daily water temperature from air temperature for the Missouri River**. In: PeerJ, v. 6, 2018. e4894,
- ZHU, S.; NYARKO, E. K.; HADZIMA-NYARKO, M. **Modelling daily water temperature from air temperature for the Missouri River**. In: PeerJ, v. 6, 2018.
- ZHU, S.; PTAK, M.; YASEEN, Z. M.; DAI, J.; SIVAKUMAR, B. **Forecasting surface water temperature in lakes: A comparison of approaches**. In: Journal of Hydrology, v. 585, 2020. p. 124809.