



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ  
ESCOLA POLITÉCNICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO E  
SISTEMAS (PPGEPS)

**FRANCESCO SANCHES DOS SANTOS**

MODELOS DE PREVISÃO DE SÉRIES TEMPORAIS APLICADOS A UM CASO  
DE SISTEMA DE ABASTECIMENTO DE ÁGUA

CURITIBA

2023

FRANCESCO SANCHES DOS SANTOS

MODELOS DE PREVISÃO DE SÉRIES TEMPORAIS APLICADOS A UM CASO  
DE SISTEMA DE ABASTECIMENTO DE ÁGUA

Projeto de Pesquisa de Mestrado apresentado ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas (PPGEPS). Área de concentração: Automação e Controle de Sistemas, da Escola Politécnica, da Pontifícia Universidade Católica do Paraná, como requisito parcial à obtenção do título de Mestre em Engenharia de Produção e Sistemas.

Orientador: Dr. Leandro dos Santos Coelho  
Coorientadora: Dra. Viviana Cocco Mariani  
(PPGEM-PUCPR)

CURITIBA

2023

**FRANCESCO SANCHES DOS SANTOS**

**MODELOS DE PREVISÃO DE SÉRIES TEMPORAIS APLICADOS A  
UM CASO DE SISTEMA DE ABASTECIMENTO DE ÁGUA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas (PPGEPS). Área de concentração: Gerência de Produção e Logística, da Escola Politécnica, da Pontifícia Universidade Católica do Paraná, como requisito parcial à obtenção do título de Mestre em Engenharia de Produção e Sistemas.

**COMISSÃO EXAMINADORA**

---

Dr. Leandro dos Santos Coelho

Orientador

Pontifícia Universidade Católica do Paraná

---

Dra. Viviana Cocco Mariani (PPGEM-PUCPR)

Coorientadora

Pontifícia Universidade Católica do Paraná

---

Dr. Helon Vicente Hultmann Ayala

Membro Interno

Pontifícia Universidade Católica do Paraná

Curitiba, 2 de dezembro de 2023

*Dedico essa dissertação de mestrado à Deus, essa força maior, que me guia e ilumina meus  
pensamentos para que eu desenvolva minha luz.*

## **Agradecimentos**

Primeiramente, expresso minha gratidão a Deus por todas as bênçãos recebidas, pois foi Ele quem abriu caminhos e me deu forças para superar esse desafio, tornando-o possível.

À minha família, sou grato pelo apoio incondicional e pelo estímulo constante para seguir em frente com determinação, buscando sempre alcançar novos patamares.

Agradeço ao professor Leandro dos Santos Coelho pela oportunidade de trabalhar ao seu lado e compartilhar seus conhecimentos e experiências ao longo do meu mestrado. Sua orientação contribuiu significativamente para o meu crescimento profissional e pessoal, tornando este trabalho uma realidade.

À professora Viviana Cocco Mariani, agradeço pela disponibilidade e paciência em me auxiliar nas minhas dificuldades, utilizando seu conhecimento para contribuir com o desenvolvimento da pesquisa.

Quero expressar minha gratidão à equipe da Pontifícia Universidade Católica do Paraná (PUCPR) e aos demais professores, especialmente à secretária Denise da Mata Medeiros (PPGEPS), pela paciência, carinho e apoio prestados em diversas ocasiões, sem medir esforços.

Aos meus amigos, que sempre torceram por mim, e aos novos amigos que conquistei ao longo dessa jornada, agradeço por compartilharmos momentos de alegria nessa batalha.

Sou grato ao investimento em bolsas de estudo concedidas pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), que possibilitou a conclusão dessa etapa da minha carreira profissional e acadêmica.

*Se vi mais longe, foi por estar de pé  
sobre ombros de gigantes*

---

- Sir Isaac Newton

## Resumo

O estudo, inserido no contexto do abastecimento de água em Curitiba, concentra-se na eficácia da previsão da demanda no Bairro Alto durante os anos de 2018 a 2020. A questão central investigada é como antecipar a demanda de água para um planejamento mais eficiente no contexto de escassez enfrentada pelos habitantes. O propósito do trabalho é contribuir para o controle eficaz dos recursos hídricos, utilizando modelos de previsão, com ênfase na melhoria do abastecimento d'água em um ambiente competitivo. São explorados modelos como ARIMA (do inglês *Auto-Regressive Integrated Moving Average*), DTR (do inglês *Decision tree regressor*), XGBoost (do inglês *eXtreme Gradient Boosting*) e RNN (do inglês *Recurrent Neural Network*) para a previsão de séries temporais, com uma análise comparativa de eficácia. A necessidade de uma solução nova ou melhor surge da escassez de água enfrentada no Bairro Alto, justificando a busca por métodos eficientes de previsão da demanda. A solução proposta, envolve a aplicação de modelos de aprendizado de máquina, como ARIMA, DTR, XGBoost e especialmente RNN, na previsão da demanda de água. A metodologia básica inclui a aplicação desses modelos aos dados coletados pela SANEPAR (Companhia de Saneamento do Paraná). As características que respondem às questões iniciais são avaliadas por meio de métricas como sMAPE (do inglês *Symmetric Mean Absolute Percentage Error*), MAE (do inglês *Mean Absolute Error*) e RRMSE (do inglês *Root Relative Mean Square Error*), destacando que o modelo RNN demonstrou consistentemente os menores erros em todas as análises. Conclui-se que a abordagem proposta contribui significativamente para a previsão da demanda de água, proporcionando um planejamento eficiente e sustentável do abastecimento hídrico no Bairro Alto.

**Palavras-chave:** Previsão de Séries Temporais, Abastecimento de Água, Aprendizado de Máquina, Redes Neurais Artificiais, Modelos de Previsão.

## Abstract

The study, situated in the context of water supply in Curitiba, focuses on the effectiveness of forecasting water demand in Bairro Alto from 2018 to 2020. The central question investigated is how to anticipate water demand for more efficient planning in the context of scarcity faced by the residents. The purpose of the work is to contribute to the effective control of water resources, utilizing forecasting models, with an emphasis on improving water supply in a competitive environment. Models such as ARIMA (Auto-Regressive Integrated Moving Average), DTR (Decision Tree Regressor), XGBoost (eXtreme Gradient Boosting), and RNN (Recurrent Neural Network) are explored for time series forecasting, with a comparative analysis of effectiveness. The need for a new or better solution arises from the water scarcity in Bairro Alto, justifying the search for more efficient methods of demand forecasting. The proposed solution involves the application of machine learning models such as ARIMA, DTR, XGBoost, and especially RNN in forecasting water demand. The basic methodology includes applying these models to the data collected by SANEPAR (Sanitation Company of Paraná). The features responding to the initial questions are assessed through metrics such as sMAPE (Symmetric Mean Absolute Percentage Error), MAE (Mean Absolute Error), and RRMSE (Root Relative Mean Square Error), highlighting that the RNN model consistently demonstrated the lowest errors in all analyses. It is concluded that the proposed approach significantly contributes to water demand forecasting, providing a more efficient and sustainable planning of water supply in Bairro Alto.

**Keywords:** Time series forecasting, Water supply, Machine learning, Artificial neural networks, Forecasting models.



## Lista de Figuras

1	Etapas para Análise dos Dados. . . . .	22
2	Fluxograma da Revisão Sistemática da Literatura. . . . .	25
3	Modelos de previsão de series temporais na base de dados Scopus e WoS. . . . .	26
4	Elementos do modelo SARIMAX . . . . .	35
5	Fluxograma da árvore de decisão. . . . .	37
6	Fluxograma da floresta aleatória. . . . .	39
7	Fluxograma do XGBoost . . . . .	40
8	Comparação do crescimento em folha com o crescimento em nível . . . . .	41
9	Modelo de uma rede neural artificial MLP . . . . .	42
10	Fluxograma da RNN. . . . .	44
11	Diagrama do funcionamento de uma GRU. . . . .	47
12	Diferenças entre RNN, LSTM, e GRU. . . . .	47
13	Modelo de uma Rede Neural Convolucional. . . . .	49
14	Correlação de Pearson . . . . .	56
15	Relação entre LT01 e PT01 cuja correlação de Pearson é 97%. . . . .	57
16	Autocorrelação . . . . .	58
17	Autocorrelação parcial . . . . .	59
18	Ruído branco . . . . .	59
19	Violino no nível do reservatório . . . . .	60
20	Violino da vazão de recalque . . . . .	60
21	Decomposição STL aditiva dos dados coletados . . . . .	61
22	Comparação dos modelos AR, ARX e MA . . . . .	66
23	Comparação dos modelos ARIMAX, SARIMA e SARIMAX . . . . .	66
24	Comparação dos modelos ARMA e ARIMA . . . . .	66
25	Comparação dos modelos DTR, RFR, XGBoost, Light GBM . . . . .	66
26	Modelo RNN e os vários horizontes . . . . .	67
27	Previsões do modelo Prophet para o reservatório LT01 . . . . .	67
28	Comparação dos modelos ARIMA . . . . .	75
29	Comparação de modelos de regressão . . . . .	75
30	Comparação dos modelos na métrica RRMSE . . . . .	75
31	Comparação dos modelos nas métricas sMAPE, MAE e RRMSE . . . . .	76
32	Demanda média das variáveis de fluxo . . . . .	79

## Lista de Tabelas

1	Combinação de palavras-chave aplicando filtros. . . . .	26
2	Resumo dos artigos obtidos com a RSL nas bases Scopus e WoS. . . . .	27
3	Classificação dos principais periódicos obtidos na RSL. . . . .	27
4	Total de publicações dos principais autores obtidos na RSL. . . . .	29
5	Total de publicações dos principais países obtidos na RSL. . . . .	29
6	Principais modelos de previsão obtidos na RSL. . . . .	31
7	Descrição estatística dos dados de 2018 a 2019 . . . . .	56
8	Teste de Dickey-Fuller Aumentado . . . . .	57
9	Parâmetros utilizados nos modelos ARIMA e seus antecessores obtidos pelo “autoARIMA” do Python. . . . .	63
10	SARIMAX(7, 0, 0) $\times$ (2, 1, [1], 12) Results . . . . .	63
11	Hiperparâmetros dos modelos . . . . .	64
12	Resumo dos Hiperparâmetros dos Modelos de Redes Neurais . . . . .	64
13	Comparação dos modelos de previsão com as métricas de desempenho <b>treino</b>	68
14	Comparação dos modelos de previsão com as métricas de desempenho <b>teste</b>	69
15	Comparação dos modelos de previsão com as métricas de desempenho <b>va-</b> <b>lidação</b> . . . . .	70
16	Comparação dos modelos de previsão com as métricas de desempenho <b>inteiro</b>	71
17	Teste Nemenyi . . . . .	72
18	Métricas de avaliação dos modelos . . . . .	73
19	Comparação dos modelos Ljung Box: Modelos ARIMA com defasagem de 10 para previsão de longo prazo na demanda de água . . . . .	77
20	Demanda de água . . . . .	79

## Lista de Abreviaturas e Siglas

ANN	Rede Neural Artificial (do inglês <i>Artificial Neural Network</i> )
AR	Auto-Regressivo
ARIMA	Média Móvel Integrada Auto-Regressiva (do inglês <i>Auto-Regressive Integrated Moving Average</i> )
ARIMAX	Média Móvel Integrada Auto-Regressiva com entradas eXógenas (do inglês <i>Auto-Regressive Integrated Moving Average with eXogenous inputs</i> )
ARMA	Média Móvel Auto-Regressiva (do inglês <i>Auto-Regressive Moving Average</i> )
ARX	Auto-Regressivo com Variável Exógena (do inglês <i>Auto-Regressive with Exogenous Inputs</i> )
CNN	Rede Neural Convolucional (do inglês <i>Convolutional Neural Networks</i> )
DBN	Rede de Crenças Profundas (do inglês <i>Deep Belief Network</i> )
DTR	Regressor de Árvore de Decisão (do inglês <i>Decision tree regressor</i> )
FT	Flow Transmitter (Transmissor de Fluxo)
GRU	Unidade Recorrente Fechada (do inglês <i>Gated Recurrent Unit</i> )
Hz	Hertz
INMET	Instituto Nacional de Meteorologia
LGBMRegressor	Regressão da Máquina de Impulso de Gradiente Leve
Light GBM	Máquina de Impulso de Gradiente Leve (do inglês <i>Light Gradient Boosting Machine</i> )
LR	Regressão Linear (do inglês <i>Linear Regression</i> )

LSTM	Memória de Longo Curto Prazo (do inglês <i>Long Short-Term Memory</i> )
$m^3$	Metros Cúbicos
$m^3/h$	Metros Cúbicos por Hora
MA	Média Móvel (do inglês <i>Moving Average</i> )
MAE	Erro Médio Absoluto (do inglês <i>Mean Absolute Error</i> )
MAPE	Erro Percentual Médio Absoluto (do inglês <i>Mean Absolute Percentage Error</i> )
mca	Metros Coluna de Água
ML	Aprendizado de Máquina (do inglês <i>Machine Learning</i> )
mm	Milímetros
MSE	Erro Médio Quadrático (do inglês <i>Mean Squared Error</i> )
PR	Estado do Paraná
RBAL	Recalque Bairro Alto
RFR	Regressão de Floresta Aleatória (do inglês <i>Random Forest Regression</i> )
RMSE	Erro de Raiz Média Quadrática (do inglês <i>Root Mean Squared Error</i> )
RNN	Rede Neural Recorrente (do inglês <i>Recurrent Neural Network</i> )
RRMSE	Raiz do Erro Médio Quadrático Relativo (do inglês <i>Root of the Relative Mean Square Error</i> )
SANEPAR	Companhia de Saneamento do Paraná
SARIMA	Auto-Regressivos Integrados de Médias Móveis com Sazonalidade (do inglês <i>Seasonal Auto-Regressive Integrated Moving Averages</i> )
SARIMAX	Média Móvel Sazonal Auto-Regressiva Integrada com Entradas Exógenas (do inglês <i>Seasonal Auto-Regressive Integrated Moving Averages with Exogenous Inputs</i> )
sMAPE	Erro Percentual Absoluto Médio Simétrico (do inglês <i>Symmetric Mean Absolute Percentage Error</i> )
SVM-VAR	Máquinas de Vetor de Suporte - Vetores Auto-Regressivos
TotalBoost	Impulso Total
Transformer	Transformador
XGBRegressor	Regressão XGBoost
XGBoost	Reforço de Gradiente Extremo (do inglês <i>eXtreme Gradient Boosting</i> )

# Sumário

<b>1</b>	<b>Introdução</b>	<b>17</b>
1.1	Contextualização	17
1.2	Motivação	20
1.3	Objetivo Geral	21
1.4	Etapas para Análise dos Dados	21
1.5	Estrutura do Documento	23
<b>2</b>	<b>Revisão da Literatura</b>	<b>24</b>
<b>3</b>	<b>Fundamentos dos Modelos de Previsão</b>	<b>31</b>
3.1	Modelos Clássicos de Séries Temporais	31
3.2	Modelos de Aprendizado de Máquina	35
3.2.1	Prophet	35
3.2.2	Regressão Linear	36
3.2.3	Árvore de Decisão	37
3.2.4	Floresta Aleatória	38
3.2.5	<i>Gradient Boosting</i>	39
3.2.6	<i>LightGBM</i>	40
3.3	Redes Neurais Artificiais	41
3.3.1	MLP	41
3.3.2	Rede Neural Recorrente	43
3.4	Aprendizado Profundo	44
3.4.1	LSTM	44
3.4.2	GRU	45
3.5	Rede Neural Convolucional	48
3.6	Métricas Estatísticas	49
3.7	Correlação de Pearson	50
3.8	Decomposição STL	51
3.9	Teste Dickey-Fuller	52
3.10	Teste de Ljung-Box	52
3.11	Teste de Significância	54
<b>4</b>	<b>Resultados</b>	<b>55</b>
4.1	Análise Exploratória dos Dados	55
4.1.1	Comparação dos Modelos	74
4.2	Aplicação	78
4.2.1	Estudo de Caso 1	78

4.2.2	Estudo de Caso 2 . . . . .	78
<b>5</b>	<b>Conclusões . . . . .</b>	<b>80</b>
5.1	Propostas Futuras . . . . .	80
	<b>Referências . . . . .</b>	<b>81</b>

# 1 Introdução

## 1.1 Contextualização

O acesso à água potável é vital para a saúde e bem-estar das pessoas, sendo um requisito fundamental para a sobrevivência e o desenvolvimento humano. A água potável é essencial para a higiene pessoal, a preparação de alimentos e a prevenção de doenças transmitidas pela água. Além disso, é um componente crucial para o funcionamento adequado de sistemas de saneamento básico. O acesso a água limpa não apenas reduz significativamente a incidência de doenças, mas também promove o crescimento econômico, a educação e a igualdade. Garantir o acesso universal à água potável não apenas salva vidas, mas também contribui para a construção de comunidades saudáveis e sustentáveis em todo o mundo. Os recursos de água potável estão tornando-se mais escassos em algumas comunidades, em parte devido à crescente procura de água potável em regiões urbanas, à pobreza de técnicas de gestão de águas residuais e as secas ou enchentes induzidas pelo clima (KOEBELE et al., 2022).

Dada a crescente escassez de água os órgãos que gerenciam tais recursos recomendam a implementação de novas iniciativas para expandir seus portfólios locais e regionais de água, incluindo a reutilização de águas residuais tratadas para fins potáveis para evitar a escassez de água, questões que poderão piorar drasticamente nas próximas décadas (BARNES; KRISHEN; HU, 2023). A reutilização de águas residuais tratadas (RART) para fins potáveis ajuda a aliviar o uso das águas superficiais e subterrâneas locais que estão relacionadas com preocupações de escassez. Ao reutilizar águas residuais, as comunidades tornam-se menos dependentes dos recursos hídricos disponíveis localmente e operam de forma mais local, girando a economia circular da água centrada (TSATSOU; FRANTZESKAKI; MALAMIS, 2023). Algumas cidades conseguiram implementar este sistema RART para aumentar o abastecimento local de água potável, porém no Brasil ainda não temos isto de forma efetiva, então uma possibilidade é fazer previsão da demanda de tal forma que o abastecimento para a população não seja descontinuado.

O presente estudo envolve dados de abastecimento d'água da cidade de Curitiba no Bairro Alto entre os anos de 2018 e 2020. No entanto, vale destacar, que durante o ano de 2022, os habitantes de Curitiba enfrentaram escassez de água, sendo necessário implementar rodízios, alternando períodos com e sem fornecimento de água potável. Os dados utilizados foram coletados pela companhia de saneamento do Paraná (SANE-PAR). A previsão da demanda de água ao longo do tempo, o que será abordado neste estudo, é essencial para um planejamento sustentável e eficiente do abastecimento hídrico, especialmente no contexto urbano, como é o caso da cidade de Curitiba.

Existem vários modelos que podem ser utilizadas para prever a demanda de água, e a escolha da abordagem dependerá da disponibilidade de dados, da complexidade do sistema e das necessidades específicas da aplicação. Entre os modelos estão: Modelos Estatísticos, tal como Regressão Linear, que pode ser usado quando há uma relação linear entre variáveis como temperatura, população, atividade econômica e consumo de água, Modelos como ARIMA (AutoRegressive Integrated Moving Average) que podem ser eficazes para prever padrões sazonais e tendências ao longo do tempo (OLIVEIRA; STEFFEN; CHEUNG, 2017), as Redes Neurais Artificiais como Redes Neurais Recorrentes (RNN) (ASEERI, 2023) e LSTM (Long Short-Term Memory) (SABZIPOUR et al., 2023) são eficazes para lidar com dados temporais e sequenciais, capturando dependências de longo prazo e os Métodos de Aprendizado de Máquina tais como Máquinas de Vetores de Suporte (SVM) que podem ser usadas para tratar dados não lineares e complexos (CANDELIERI et al., 2019), Random Forest (RF) (ALI et al., 2023) Gradient Boosting (DONG et al., 2023) que são métodos de ensemble (métodos híbridos) que podem ser aplicados para prever padrões complexos, entre outros modelos.

Existem três tipos principais de Aprendizado de Máquina: Aprendizado Supervisionado, Aprendizado Não Supervisionado, e Aprendizado por Reforço. No Aprendizado Supervisionado (LIU; FU, 2023), o algoritmo é treinado em um conjunto de dados rotulado, onde cada exemplo do conjunto de dados possui uma entrada e a saída desejada correspondente. O objetivo é aprender uma função que mapeia as entradas para as saídas, permitindo ao modelo fazer previsões ou tomar decisões em novos dados não vistos anteriormente. No Aprendizado Não Supervisionado (WANG et al., 2022), o algoritmo é treinado em um conjunto de dados não rotulado, e o sistema tenta aprender a estrutura e padrões presentes nos dados. O objetivo principal é explorar a estrutura intrínseca dos dados, identificando agrupamentos, associações ou padrões emergentes sem ter rótulos para orientar o processo. O Aprendizado por Reforço (CHEN et al., 2023) envolve um agente que interage com um ambiente dinâmico. O agente toma decisões sequenciais para alcançar um objetivo específico, recebendo feedback na forma de recompensas ou penalidades. O objetivo é aprender uma política, ou seja, uma estratégia que maximize a recompensa cumulativa ao longo do tempo (SILVA; GOMES, 2021). Os modelos de Aprendizado de Máquina Supervisionados serão usados neste estudo, e optou-se pelo uso de tais modelos porque na literatura da área, geralmente são usados para previsão de séries temporais (UC-CASTILLO et al., 2023).

As séries temporais de abastecimento de água potável geralmente exibem características específicas devido à natureza dinâmica e sazonal do consumo de água, muitas vezes exibem padrões sazonais, com variações regulares ao longo do tempo. Isso pode ser influenciado por fatores como as estações do ano, dias da semana ou horas do dia. Pode



haver tendências a longo prazo nas séries temporais, refletindo mudanças demográficas, crescimento urbano, desenvolvimento industrial ou outros fatores que afetam o consumo de água ao longo do tempo (JI; AHN, 2023). A ocorrência de eventos anômalos, como vazamentos, interrupções no fornecimento devido a situações críticas como secas, enchentes, ou situações de emergência, pode ser evidenciada em picos ou quedas abruptas nas séries temporais. Também podem estar correlacionadas com fatores externos, como eventos climáticos (por exemplo, períodos de seca ou chuvas intensas) e feriados, impactando o comportamento do consumo (BERGLUND; SKARBEEK; KANTA, 2023).

O consumo de água muitas vezes segue padrões diários e semanais previsíveis, como picos de demanda durante o horário de pico diário e variações ao longo da semana (SIEGEL et al., 2020). Flutuações de curto prazo podem ocorrer devido a atividades específicas, eventos locais ou situações temporárias que afetam o consumo de água em um período limitado. Mudanças na infraestrutura, como expansões urbanas, construção de novos empreendimentos ou implementação de políticas de conservação, podem ser refletidas nas séries temporais. Entender essas características é fundamental para o gerenciamento eficiente do abastecimento de água, permitindo a implementação de estratégias proativas, otimização de recursos e resposta adequada a eventos inesperados. O uso de técnicas de previsão e análise de séries temporais, isto é, modelagem preditiva, pode ser valioso para entender as dinâmicas complexas (UC-CASTILLO et al., 2023).

Por meio da utilização de métodos e modelos de séries temporais, neste estudo será realizada a previsão do nível do reservatório na estação de tratamento de água no Bairro Alto em Curitiba, incorporando diversos modelos de previsão nesse processo. Dentre esses modelos, incluem-se os clássicos, como ARIMA e suas variantes, tais como AR (*Auto-Regressive*), ARX (*Auto-Regressive with Exogenous input*), MA (*Moving Average*), ARMA (*Auto-Regressive Moving Average*), SARIMA (*Seasonal Auto-Regressive Integrated Moving Average*), ARIMAX (*Auto-Regressive Integrated Moving Average with Exogenous input*), SARIMAX (*Seasonal Auto-Regressive Integrated Moving Average with Exogenous input*), além dos modelos de aprendizado de máquina, como árvore de decisão, floresta aleatória, Prophet, XGBoost (*Extreme Gradient Boosting*), Light GBM (*Light Gradient Boosting Machine*), e redes neurais como LSTM (*Long Short-Term Memory*), GRU (*Gated Recurrent Unit*), Transformer, CNN (*Convolutional Neural Network*). A diversidade de modelos foi utilizada buscando otimizar a precisão das previsões.

O presente estudo fundamenta-se em modelos de previsão previamente não explorados neste contexto, como GRU, LSTM, XGBOOST, LGBM, Transformer, RNN e CNN. Vários desses modelos não foram aplicados previamente na literatura para a área de demanda de água. Os modelos ARIMA e suas variantes foram aplicados nesta área, como demonstrado por (BUYUKSAHIN; ERTEKIN, 2019a; BHANGU; SANDHU; SAPRA,

2022). Alguns outros modelos, apesar de suas vantagens, ainda não foram devidamente aplicados, como é o caso do modelo RNN (SHIH; SUN; LEE, 2019a), que se mostrará significativamente superior aos demais modelos listados ao longo deste trabalho.

Torna-se evidente que a análise de séries temporais e previsões são ferramentas valiosas para apoiar o processo de tomada de decisão em curto, médio e longo prazo de previsão. Devido às não linearidades, sazonalidades e tendências que podem ocorrer, nos dados temporais de abastecimento de água, o desenvolvimento de modelos de previsão eficientes torna-se uma tarefa desafiadora (RIBEIRO et al., 2021).

## 1.2 Motivação

Estudar e fazer previsão de séries temporais de abastecimento de água é crucial por várias razões, pois isso permite uma gestão mais eficiente e sustentável dos recursos hídricos. Porém pode-se citar alguns motivos importantes. A previsão de séries temporais ajuda a antecipar demandas futuras de água, permitindo que os gestores planejem e desenvolvam infraestruturas adequadas para atender a essas demandas. Isso é vital para garantir que a água esteja disponível em quantidade suficiente para atender às necessidades da população. Fornece *insights* sobre os padrões sazonais e tendências de consumo de água. Com essas informações, os gestores podem tomar decisões informadas sobre a alocação de recursos hídricos e implementar práticas de conservação. Permitem que as autoridades otimizem as operações de abastecimento de água, ajustando a produção e a distribuição com base nas variações de demanda ao longo do tempo. Isso contribui para uma operação mais eficiente do sistema. Podem prever eventos climáticos extremos, como secas prolongadas ou inundações, para trabalhar com situações de emergência. A previsão de séries temporais ajuda a antecipar esses eventos, permitindo que medidas preventivas sejam tomadas para garantir a continuidade do abastecimento de água.

Assim, uma gestão eficiente baseada em previsões precisas pode resultar em economia de recursos financeiros. Isso inclui evitar investimentos desnecessários em infraestrutura e garantir que os recursos sejam alocados de maneira eficaz. Ao compreender os padrões de consumo de água, é possível implementar práticas que promovam a sustentabilidade ambiental, como a redução do desperdício de água e a promoção de fontes alternativas e renováveis. Muitas áreas possuem regulamentações que exigem o monitoramento e relatório regular do abastecimento de água. O estudo e previsão de séries temporais auxiliam as autoridades a cumprir essas normativas de maneira eficaz. A capacidade de prever variações nas condições de abastecimento de água permite uma melhor gestão de riscos, tanto em termos de disponibilidade de água quanto de eventos que possam impactar negativamente a infraestrutura.

Em especial, a situação enfrentada por Curitiba e região metropolitana em 2020, conforme destacado por (VASCONCELOS, 2020) referente ao rodízio de abastecimento de água, com períodos de 36 horas com abastecimento de água, seguidos por 36 horas sem abastecimento mostra a necessidade de previsões mais acuradas do abastecimento de água. Naquele ano a média geral dos reservatórios na região estava em torno de 27,96% de sua capacidade. A crise hídrica teve como principal gatilho a seca meteorológica e estava bastante associada a como era feito o planejamento e a gestão dos recursos hídricos. Deste modo este estudo visa contribuir para a área trazendo algumas percepções que poderão ser usadas na previsão.

### 1.3 **Objetivo Geral**

O objetivo geral deste estudo é desenvolver e aprimorar modelos de previsão de séries temporais para a demanda de água, integrando técnicas estatísticas e de aprendizado de máquina, visando proporcionar uma gestão eficiente e sustentável dos recursos hídricos, em específico na região do Bairro Alto em Curitiba (PR, Brasil), além de contribuir para a otimização do planejamento e operação de sistemas de abastecimento, promovendo a resiliência diante de variações sazonais e eventos imprevistos.

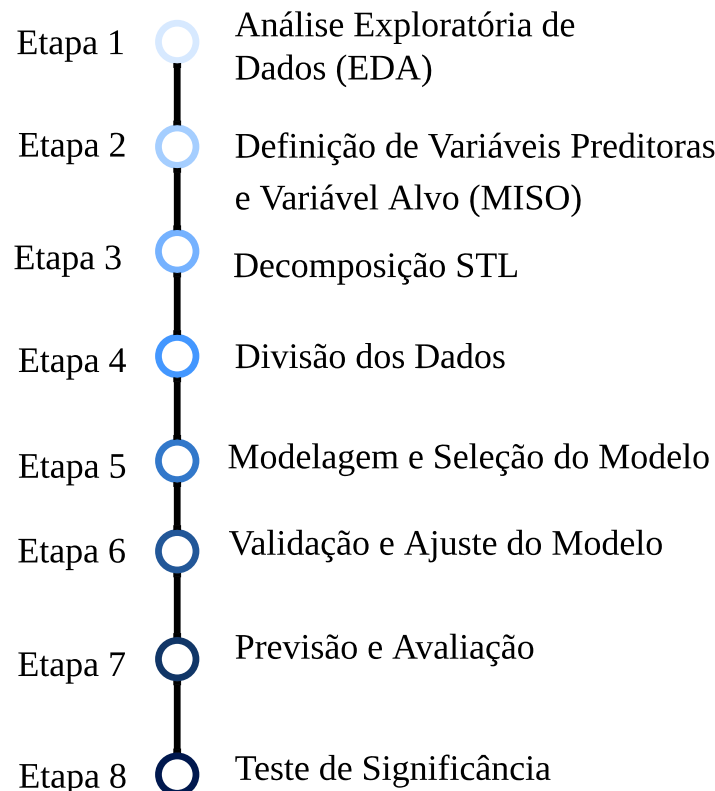
Entre os objetivos específicos do estudo estão:

1. Aplicar diferentes modelos de previsão de séries temporais advindas de dados do Bairro Alto em Curitiba através da SANEPAR.
2. Analisar a precisão, eficiência e capacidade de previsão desses modelos em conjuntos de dados específicos, através de métricas estatísticas.
3. Explorar as estratégias de otimização baseadas em otimização bayesiana utilizando o algoritmo TPE (*Tree-structured Parzen Estimator*) para ajustar os hiperparâmetros dos modelos de previsão de séries temporais.
4. Identificar a combinação mais eficaz de modelo de previsão de séries temporais junto com a configuração de otimização.

### 1.4 **Etapas para Análise dos Dados**

Com o objetivo de realizar as previsões e fazer comparações entre os modelos preditores, a pesquisa adotará um processo bem definido, bem como a seleção dos modelos a serem utilizados na Análise Exploratória de Dados (EDA). A pesquisa foi conduzida seguindo as etapas delineadas, conforme apresentado na Figura 1. As etapas para análise dos dados incluem:

Figura 1: Etapas para Análise dos Dados.



1. Análise Exploratória de Dados (EDA): Nesta etapa tem-se a identificação de valores ausentes, a observação de padrões temporais e a detecção de anomalias. Gráficos de linha são comuns para visualizar a convergência dos dados (ROSTAM et al., 2021).
2. Definição de Variáveis Preditoras e Variável Alvo (Modelo MISO): Na segunda etapa, as variáveis preditoras e a variável alvo para a previsão de Múltiplas Entradas e Uma Saída *Multiple Inputs Single Output* (MISO) são selecionadas. Diferentes modelos, podem incorporar variáveis exógenas na modelagem. Essas variáveis exógenas aprimoram as capacidade de previsão do modelo, especialmente quando o horizonte de previsão se estende além dos dados históricos (PAWŁOWSKI et al., 2022).
3. Decomposição STL: O método de decomposição STL *Seasonal and Trend Decomposition Using locally estimated scatterplot smoothing (Loess)* separa uma série temporal em três componentes: sazonalidade, tendência e resíduo. Essa decomposição permite. Decompor séries temporais em sazonal captura variações periódicas e repetitivas. Decompor séries temporais em tendência reflete a evolução geral dos dados ao longo do tempo. Já a componente de resíduo engloba as variações não explicadas pelas anteriores (BANDARA; HYNDMAN; BERGMEIR, 2021).

4. Divisão dos Dados: É prática comum dividir o conjunto de dados em conjuntos de treinamento, validação e teste para avaliar o desempenho do modelo. Essa divisão permite uma análise da capacidade de generalização dos modelos, evitando problemas de ajuste excessivo ou insuficiente. A proporção de alocação pode variar, mas uma abordagem é alocar 70% para treinamento e validação, e os 30% restantes para o conjunto de testes. A porção de treinamento e validação pode ser subdividida em 80% para treinamento e 20% para validação (TAO et al., 2020).
5. Modelagem e Seleção do Modelo: Nesta etapa, diversos modelos são construídos e avaliados. Alguns modelos comumente usados para previsão de séries temporais incluem ARX, AR, MA, ARIMA, SARIMA, SARIMAX e modelos de aprendizado de máquina como RNN, LSTM, GRU, Transformer, DTR, LR, XGBoost, Light GBM, e Prophet. A escolha do modelo baseia-se em critérios como desempenho na validação, simplicidade do modelo e interpretabilidade dos resultados.
6. Validação e Ajuste do Modelo: Durante a construção do modelo, é importante avaliar seu desempenho usando dados de validação. Neste contexto, métricas de avaliação tais como de avaliação como MAE (Erro Médio Absoluto), sMAPE (Erro Médio Percentual Absoluto Simétrico) e RRMSE (Raiz do Erro Médio Quadrático Relativo) podem ser usadas para comparar e selecionar o melhor modelo. Além disso, técnicas de ajuste como otimização de hiperparâmetros e refinamento do modelo usando dados de treinamento e validação combinados podem melhorar o desempenho de previsão de séries temporais.
7. Previsão e Avaliação: Com o modelo final com os dados de treinamento e validação, é possível fazer previsões para o conjunto de testes, que representa dados futuros não observados. Essas previsões são comparadas com os valores reais correspondentes para avaliar a qualidade e precisão do modelo.
8. Teste de Significância: Aplicar os modelos de previsão e fazer comparativo baseado em testes de significância estatística (*Friedman e Nemenjy*).

Cada uma dessas etapas desempenha um papel crucial na análise dos dados e no processo de modelagem das séries temporais, contribuindo para a compreensão dos dados, construção e validação dos modelos de previsão.

## 1.5 Estrutura do Documento

Esse documento de dissertação está organizado em 6 capítulos. O Capítulo 1, apresentou contextualização, justificativa, objetivos, contribuições, e a organização do

documento. O Capítulo 2 menciona uma visão geral das principais pesquisas na área de previsão de séries temporais aplicadas a demanda de água. No Capítulo 3 são descritos os modelos de previsão que serão utilizados nos dados de séries temporais de abastecimento de água no Bairro Alto em Curitiba, dados estes fornecidos pela SANEPAR através de coletas horárias durante 3 anos consecutivos. O Capítulo 4 apresenta os resultados obtidos ao longo da pesquisa com discussões. Os resultados de previsão obtidos são detalhados e a técnica de otimização e métricas estatísticas de desempenho são aplicadas e analisadas. O Capítulo 5 apresenta os principais resultados da pesquisa, as limitações e delinea possíveis estudos futuros na área.

## 2 Revisão da Literatura

Este capítulo apresenta uma revisão sistemática da literatura (RSL) nos temas relacionados a previsão de séries temporais e aplicações em hidrologia e mais especificamente em abastecimento de água. A revisão bibliográfica realizada consiste em uma análise abrangente e crítica das principais fontes de literatura. As informações extraídas da literatura são fundamentais para embasar a fundamentação teórica, a metodologia e a análise dos resultados deste estudo.

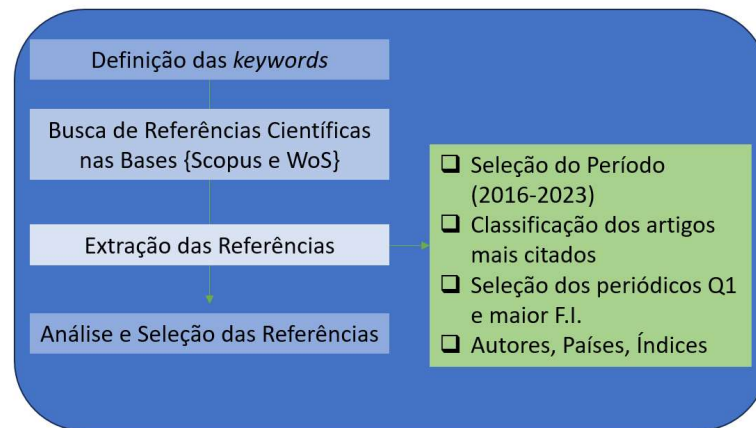
Esta RSL aborda palavras chave como séries temporais, previsão, análise de séries temporais, abastecimento de água, e saneamento, que são palavras relevantes para esta pesquisa.

A seleção das referências foi baseada em critérios específicos tais como. Definição das bases de busca, escolha das keywords, seleção do período de busca (2016 a 2023), seleção do tipo de artigo, organização pela ordem de citações, verificação dos periódicos mais importantes (fator Q1 do Scimago e Fator de Impacto). Embora nem todas as referências obtidas tenham uma relação evidente ou mesmo acentuada com a área de aprendizado de máquina, elas contribuem como material de suporte a implementação de alguns modelos avaliados para previsão neste trabalho e podem servir como base para outros estudos.

A Figura 2 apresenta um fluxograma de como a pesquisa foi realizada, destacando a importância da escolha dos periódicos Q1 e com maior fator de impacto (F.I.), como base para esta RSL. A mesma figura apresenta uma adaptação da metodologia proposta por Martins e Gorschek (2016) para a realização da RSL, onde primeiramente foram realizadas buscas na base Scopus e WoS (*Web of Science*), selecionando referências relevantes para o tema da pesquisa. Para as duas bases de busca utilizadas foram usadas as palavras-chave “*time series forecasting*”, “*time series analysis*”, “*sanitation*” e “*water supply*” .

Na etapa seguinte, foi realizada uma avaliação preliminar de cada artigo obtido,

Figura 2: Fluxograma da Revisão Sistemática da Literatura.



sem aplicar nenhum filtro anual nas buscas. Analisar todos os artigos dessa maneira resultaria em um número elevado, por exemplo, no base de dados Scopus, existiam 831 artigos, enquanto na base de dados WoS, foram encontrados 98 artigos, totalizando 929 artigos, sem a remoção de duplicatas. É importante ressaltar que esses artigos passaram apenas pelo filtro de idioma inglês e pela categoria de serem artigos (*papers*), com o objetivo de aprimorar a busca e a tomada de decisões.

Levando em consideração a diferença entre essa estimativa apresentada na Tabela 2 e a quantidade de artigos restantes após a remoção de duplicatas, tem-se menos de 929 artigos para análise. É válido lembrar que, ao remover as duplicatas, o número diminuiu ainda mais, chegando a 906 artigos.

Na etapa final, foi realizada uma análise dos conteúdos dos artigos selecionados, levando em consideração as áreas de especialização. Como esta revisão está inserida no contexto de um programa de mestrado em Engenharia de Produção e Sistemas, vale a pena analisar a relação dos artigos obtidos com áreas afins como Matemática. Assim, as áreas mais relevantes para a pesquisa foram Informática, Engenharia e Matemática, representando 50% das publicações.

São apresentados os resultados da RSL, utilizando um *software* VOSviewer de cada base de dados utilizada no trabalho. A Figura 3 mostra os modelos de previsão usados com frequência em conjunto com “*time series*” nos artigos obtidos nas bases de dados Scopus e WoS.

Nesse primeiro momento, foram obtidos 2.555 modelos, dos quais 83 modelos são dispostos na Figura 3. É importante destacar que as palavras-chave utilizadas foram *time series forecasting* ou *time series analysis* e *water supply* e *sanitation* em ambas as bases





que muitos deles não se relacionavam diretamente com o objeto de pesquisa deste estudo. Consequentemente, ao longo da condução da RSL artigos discrepantes com o estudo foram excluídos.

Tabela 2: Resumo dos artigos obtidos com a RSL nas bases Scopus e WoS.

Quantidade de artigos obtidos	929
Quantidade de artigos da base WoS	98
Quantidade de artigos da base Scopus	831
Remoção de artigos duplicados	
Porcentagem de artigos duplicados	87%
Quantidade de artigos duplicados	23
Quantidade de artigos sem duplicados	906
Porcentagem de artigos duplicados removidos da base WoS	19,4%
Porcentagem de artigos duplicados removidos da base Scopus	0,5%
Quantidade de artigos duplicados com diferentes citações	3
Porcentagem de artigos duplicados com diferentes citações	13%

A Tabela 3 apresenta os periódicos onde foram publicados o maior número de artigos com as combinações utilizadas na RSL para o tema de estudo em questão. Todos os periódicos são listadas em ordem decrescente pela quantidade de publicações obtidas, incluindo a métrica do *Scimago Journal Rank* (SJR) que avalia a importância relativa de periódicos científicos com base em sua influência e prestígio na comunidade acadêmica. Ele é calculado usando algoritmos complexos que levam em consideração a qualidade das citações recebidas por um periódico. Neste estudo a RSL procurou basear-se em periódicos classificados como Q1 e Q2, bem como o *h-index*.

Tabela 3: Classificação dos principais periódicos obtidos na RSL.

Periódicos	No. de artigos	SJR	<i>h-index</i>
Neurocomputing	27	Q1	143
IEEE Access	18	Q1	127
Applied Soft Computing	12	Q1	143
Energies	11	Q2	93
Energy	11	Q1	343

O software *ScientoPy* obtém os principais tópicos de tendência com base na maior taxa de crescimento médio (AGR) *Average Growth Rate*). A AGR é a diferença média entre o número de documentos publicados em um ano e o número de documentos publicados no ano anterior (RUIZ-ROSETO; RAMIREZ-GONZALEZ; VIVEROS-DELGADO, 2019). Indicando como o número de documentos publicados para um tópico em específico cresceu (número positivo) ou diminuiu (número negativo) em média dentro de um período de tempo. Assim, o AGR é calculado por,

$$\text{AGR} = \frac{\sum_{i=Y_s}^{Y_e} P_i - P_{i-1}}{(Y_e - Y_s) + 1} \quad (2.1)$$

onde AGR é a taxa média de crescimento,  $Y_e$  é o ano final,  $Y_s$  é o ano inicial,  $P_i$  é o número de publicações no ano  $i$ . Para o ano final  $Y_e$ , o *ScientoPy* utiliza o ano final global por defeito configurado nas opções globais ou/em parâmetros do comando *ScientoPy*. O ano de início  $Y_s$  é calculado a partir do ano final  $Y_e$ , conforme calculado por,

$$Y_s = Y_e - (\text{WindowWidth} + 1) \quad (2.2)$$

onde a largura da janela (*Window Width*) é predefinida como 2 anos. Assim, se o ano final for 2018, o AGR é a taxa de crescimento média entre 2017 e 2018 (RUIZ-ROSETO; RAMIREZ-GONZALEZ; VIVEROS-DELGADO, 2019).

A média de documentos por ano (ADY) *average documents per year* é um indicador absoluto que representa o número médio de documentos publicados num período de tempo para um tópico específico. O ADY é calculado por,

$$\text{ADY} = \frac{\sum_{i=Y_s(t)}^{Y_e(t)} P_i}{(Y_e(t) - Y_s(t)) + 1} \quad (2.3)$$

onde *ADY* é a média de documentos por ano,  $Y_e(t)$  é o ano final,  $Y_s(t)$  é o ano inicial, calculado como descrito na equação (2.3),  $P_i$  é o número de publicações no ano  $i$ .

A percentagem de documentos nos últimos anos (PDLY) *Percentage of documents in last years* é um indicador relativo que representa a percentagem do ADY em relação ao número total de documentos para um tópico específico. Desta forma, o PDLY é calculado como,

$$\text{PDLY} = \frac{\sum_{i=Y_s(t)}^{Y_e(t)} P_i}{(Y_e(t) - Y_s(t) + 1) \cdot \text{TND}} \cdot 100\% \quad (2.4)$$

onde *PDLY* é a percentagem de documentos nos últimos anos,  $Y_e(t)$  é o ano final,  $Y_s$  é o ano inicial, calculado como descrito na equação (2.4),  $P_i$  é número de publicações no ano  $i$ , *TND* é o número total de documentos.

Tabela 4 descreve os principais autores obtidos na RSL descrita previamente, sobre o tema em análise. Essa abordagem visa evitar a inclusão de todos os autores e desta-

car aqueles que tiveram uma contribuição significativa na área. Dessa forma, é possível identificar o principal autor que se destacou, fornecendo uma visão geral da distribuição da produção científica entre os pesquisadores. Na Tabela 4 são descritos os valores da taxa de crescimento médio (AGR), documentos médios por ano (ADY), e porcentagem de documentos nos últimos anos (PDLY) no período de 2021 a 2023.

Tabela 4: Total de publicações dos principais autores obtidos na RSL.

Author	No. de artigos	AGR	ADY	PDLY	<i>h-index</i>
Wang et al. (2016)	11	-0,5	2	36,4	8
Shen e Wang (2022)	11	0	3	54,5	5
Xian et al. (2018)	10	1	2,5	50	5
Li et al. (2018)	9	-1,5	2	44,4	4
Sang et al. (2016)	7	1,5	2	57,1	3
Sadaei et al. (2019a)	7	1	2	57,1	3
Hao et al. (2023)	7	1	3	85,7	2
Guo, Pedrycz e Liu (2018)	7	1,5	3	85,7	3
O'Donncha et al. (2022)	6	0	1,5	50	4
Xu et al. (2019b)	6	0	1,5	50	5

A Tabela 5 apresenta os países com maior número de artigos obtidos na RSL usando as palavras-chaves citadas previamente. Tais países estão ordenados de forma decrescente pelo número de publicações obtido. Os principais países que se destacam nessa análise são China, com 179 publicações, Estados Unidos da América com 74 publicações, Índia com 61 publicações, Brasil com 49 publicações, Espanha com 40 publicações, Reino Unido com 40 publicações, Austrália com 31 publicações, Itália com 26 publicações, Canadá com 25, Irã com 20 publicações.

Tabela 5: Total de publicações dos principais países obtidos na RSL.

País	No. de artigos	AGR	ADY	PDLY	<i>h-index</i>
China	179	18,5	48	53,6	31
Estados Unidos da América	74	3	16	43,2	21
Índia	61	0	12	39,3	18
Brasil	49	3,5	12,5	51	17
Espanha	40	1,5	8,5	42,5	12
Reino Unido	40	3	10	50	15
Austrália	31	3,5	7,5	48,4	14
Itália	26	2	7	53,8	10
Canadá	25	1	5,5	44	11
Irã	20	-1	3,5	35	11

Foi realizada uma análise dos artigos obtidos com a RSL. Esses artigos retratam alguns dos modelos de previsão utilizados em Taieb e Atiya (2016), Ursu e Pereau (2016),

Wang et al. (2016), Graff et al. (2017), Tyrallis e Papacharalampous (2017a), Boroojeni et al. (2017), Coelho et al. (2017), Chou e Tran (2018), Bergmeir, Hyndman e Koo (2018), Rossi (2018), Ahmad et al. (2018), Chou e Nguyen (2018), Chen et al. (2018), Sadaei et al. (2019b), Yang et al. (2019), Buyuksahin e Ertekin (2019b), Carvalho Jr. e Costa Jr. (2019), Liu e Chen (2019), Shih, Sun e Lee (2019b), Moon et al. (2019), Yang et al. (2019), Xu et al. (2019a), Golyandina (2020), Martinović, Hunjet e Turcin (2020), Salgotra, Gandomi e Gandomi (2020), Vlachas et al. (2020), Kulshreshtha e Vijayalakshmi (2020a), Samanta et al. (2020), Shen et al. (2020), Sezer, Gudelek e Ozbayoglu (2020a), Du et al. (2020), Li e Bastos (2020), Kumar, Jain e Singh (2021), Lara-Benitez, Carranza-Garcia e Riquelme (2021), Tan et al. (2021), Liu et al. (2021).

Notou-se que em tais artigos diferentes modelos de previsão de séries temporais foram utilizados. Eles representam contribuições significativas para o avanço do conhecimento e aplicação prática das séries temporais, sobre abordagens eficazes nesse campo.

Por exemplo, no estudo conduzido por Xu et al. (2019a), um modelo híbrido foi proposto, combinando o modelo linear AR e LR com o modelo não-linear ARIMA e o modelo (DBN) *Dynamic Bayesian Network*. Essa abordagem permitiu capturar tanto os comportamentos lineares quanto os não-lineares de uma série temporal.

Por outro lado, Li e Bastos (2020) comparou o desempenho de previsão da abordagem MAELS (Modelo Alternativo de Estação Livre Série Temporal) com outros modelos de aprendizado de máquina como CNN, RNN, LSTM, GRU, Transformer, Prophet, ARIMA e (SVM-VAR) *Support Vector Machine Variable Regression*. As abordagens CNN, RNN, GRU, Transformer e LSTM foram capazes de trabalhar com dados multivariados de entrada e saída, enquanto o ARIMA utiliza informações passadas para prever o futuro com base em características como autocorrelação e médias móveis.

Na Tabela 6 é apresentada a quantidade de artigos foram obtidos pela RSL para cada modelo de previsão e também cita-se pelo menos um autor correspondente a cada modelo.

Além dos modelos prévios, também será utilizada a versão atualizada do ARIMA nesta dissertação, bem como os modelos SARIMA e SARIMAX serão comparados para determinar qual deles é o mais adequado. Além disso, serão empregados os modelos Light GBM e XGBoost. Os modelos de aprendizado profundo, como a RNN, ainda é considerada um excelente modelo para previsão de séries temporais no tema de saneamento básico que está estudado.

Embora existam várias ramificações do modelo ARIMA, o modelo desenvolvido pelo Facebook, conhecido como Prophet, sobressai como uma opção superior em comparação com os demais. O Prophet é um modelo mais recente que simplifica significativamente muitas das tarefas que são necessárias ao lidar com o ARIMA. Enquanto o Prophet foi

Tabela 6: Principais modelos de previsão obtidos na RSL.

Modelos	Autores	No.	AGR	ADY	PDLY	<i>h-index</i>
ARX	Gustin, McLeod e Lomas (2018)	3	0	0,7	66,7	2
ARMA	Tyralis e Papacharalampous (2017b)	7	0,3	0,7	28,6	6
ARIMA	Buyuksahin e Ertekin (2019a)	84	1,7	16,7	59,5	27
SARIMA	Kushwah e Wadhvani (2022)	5	1	1,7	100	4
SARIMAX	Bhangu, Sandhu e Sapra (2022)	2	0,3	0,7	100	2
LSTM	Sezer, Gudelek e Ozbayoglu (2020b)	35	3,3	10,7	91,4	16
RNN	Shih, Sun e Lee (2019a)	20	0	4,3	65	11
Árvores de Decisão	Fouilloy et al. (2018)	12	0,7	3	75	7
Transformer	Peimankar et al. (2018)	10	2,3	3	90	5
RF	Yang, Guo e Li (2022)	9	1,7	2,7	88,9	5
CNN	Rostamian e O'Hara (2022)	8	1,3	2,7	100	4
GRU	Yang, Guo e Li (2022)	5	0	1,3	80	4
LR	Mohan et al. (2022)	3	0	0,7	66,7	3
Prophet	Kulshreshtha e Vijayalakshmi (2020b)	3	0,3	1	100	3
XGBoost	Liu et al. (2022)	1	0,3	0,3	100	0

criado em 2017, o modelo ARIMA tem relatos de ter sido desenvolvido na década de 1960. Essa diferença temporal destaca a evolução e a modernização do campo de modelagem de séries temporais ao longo das décadas (RAMOS, 2010).

### 3 Fundamentos dos Modelos de Previsão

Neste capítulo são descritos os modelos de previsão de séries temporais usados neste estudo, bem como as métricas e testes estatísticas de desempenho.

#### 3.1 Modelos Clássicos de Séries Temporais

O modelo ARIMA *AutoRegressive Integrated Moving Average* é um modelo estatístico utilizado para a previsão de séries temporais. Ele é composto por três componentes principais: o componente autoregressivo (AR), o componente de médias móveis (MA) e o componente de diferenciação (I), que é aplicado para tornar a série temporal estacionária.

O componente auto-regressivo do modelo ARIMA é representado por  $AR(p)$ , em que o parâmetro  $p$  determina o número de defasagens ou atrasos (do inglês *lags*) a serem usados. A equação do modelo  $AR(p)$  é expressa da seguinte forma:

$$Y_t = c + \sum_{n=1}^p \alpha_n Y_{t-n} + \varepsilon_t \quad (3.1)$$

na equação (3.1), o termo  $\varepsilon_t$  representa o ruído branco que é caracterizado por um sinal

com média zero e variância sigma. Essa equação pode ser entendida como uma regressão múltipla, em que os valores defasados de  $Y_t$  são utilizados como preditores. Esse modelo é conhecido como modelo autorregressivo de ordem  $p$ , ou  $AR(p)$ .

O modelo ARX é uma extensão do modelo AR, que incorpora variáveis exógenas nos dados para tentar melhorar as previsões. Esse modelo também é multivariado, e foi incluído aqui para fins de comparação com o modelo AR simples, considerando a presença de variáveis exógenas. Pode-se mencionar que de acordo com o valor de  $p$  tem-se alguns aspectos relevantes a citar: Se o parâmetro  $p$  for definido como zero  $AR(0)$ , significa que não há termos autorregressivos no modelo. Nesse caso, a série temporal se comporta como um ruído branco. Com o parâmetro  $p$  definido como 1, o modelo AR leva em consideração o valor anterior da série temporal multiplicado por um coeficiente e, em seguida, adiciona ruído branco. Quando o coeficiente é igual a 0, há apenas ruído branco, resultando em uma série de tempo completamente aleatória, sem padrões previsíveis.

Se o parâmetro  $p$  estiver na faixa  $0 < \alpha < 1$ , ocorre o fenômeno de reversão média. Isso significa que os valores da série tendem a oscilar em torno de uma média central e a regressar em direção a ela após se afastarem. Esse padrão indica uma tendência de retorno à média ao longo do tempo. Aumentar ainda mais o parâmetro  $p$  no modelo AR significa considerar um número crescente de medições de tempo anteriores, cada uma multiplicada pelo seu próprio coeficiente. Isso permite levar em conta uma memória mais longa da série temporal e capturar padrões de dependência complexos ao longo do tempo. No entanto, é importante ter em mente que aumentar excessivamente o valor de  $p$  pode levar a problemas de *overfitting*, onde o modelo se ajusta muito bem aos dados de treinamento, mas tem um desempenho ruim na previsão de novos dados. Portanto, é necessário encontrar um equilíbrio entre a complexidade do modelo e sua capacidade de generalização.

No modelo de média móvel (MA), o componente não é uma média móvel simples, mas sim uma combinação de termos de erro de previsão defasados. O parâmetro  $q$  no modelo MA representa o número de termos de erro de previsão que são levados em consideração na previsão. Este componente não é uma média móvel, mas sim os atrasos no ruído branco (TRENBERTH, 1984). Em um modelo  $MA(1)$ , por exemplo, a previsão é composta por um termo constante, o produto do termo de erro de previsão anterior por um multiplicador, e o termo de erro de previsão atual. Essa abordagem baseia-se em princípios estatísticos e de probabilidade, ajustando a previsão com base em termos anteriores de erro de previsão.

O modelo MA é uma alternativa ao modelo AR e é usado para capturar padrões de dependência na média móvel, ou seja, a influência de erros passados na previsão atual. Ao combinar o modelo AR e o modelo MA, como no modelo ARMA, é possível obter uma modelagem mais abrangente que considera tanto a dependência autorregressiva quanto a

dependência na média móvel (VIDHYA, 2023), tal que

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3.2)$$

onde  $\varepsilon_t$  representa o ruído branco, esse modelo é conhecido como um modelo de média móvel  $MA(q)$ , em que  $q$  é a ordem da média móvel. É importante ressaltar que não se observam diretamente os valores de  $\varepsilon_t$ , portanto, essa modelagem não se trata de uma regressão no sentido convencional.

Diferentemente de uma regressão comum em que se têm variáveis explicativas observadas, no modelo  $MA(q)$ , são usados os termos de ruído branco defasados para estimar e prever os valores da série temporal. O objetivo é capturar a dependência dos termos de erro passados na previsão atual (VIDHYA, 2023).

O modelo ARMA é uma combinação dos modelos AR e MA, onde o modelo AR é adicionado ao modelo MA. No modelo ARMA, é adicionada uma constante à soma dos termos autorregressivos multiplicados pelos seus coeficientes, juntamente com a soma dos termos de média móvel multiplicados pelos seus coeficientes, além do ruído branco. Essa estrutura é amplamente utilizada em diversos modelos de previsão em diferentes áreas. Esse modelo é bastante semelhante ao modelo ARIMA, pois calcula os termos, mas não inclui a diferenciação presente tanto no modelo ARMA quanto no modelo ARIMA, tal que

$$Y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \dots + \omega_q \varepsilon_{t-q} + \varepsilon_t \quad (3.3)$$

onde  $Y_t$  representa a série temporal que foi diferenciada (possivelmente mais de uma vez). Os preditores no lado direito da equação incluem os valores defasados de  $Y_t$  e os erros defasados. Esse tipo de modelo é conhecido como ARIMA  $(p, d, q)$ .

O modelo ARIMA é uma extensão do modelo ARMA que incorpora uma etapa adicional de pré-processamento chamada de diferenciação. Essa etapa é representada pela notação  $I(d)$ , em que  $d$  denota a ordem de diferenciação, ou seja, o número de transformações necessárias para tornar a série temporal estacionária. Portanto, um modelo ARIMA é simplesmente um modelo ARMA aplicado à série temporal diferenciada. Isso permite lidar com séries temporais que possuem tendências ou padrões não estacionários.

Embora os modelos ARIMA sejam eficazes, incorporar variáveis sazonais e exógenas ao modelo pode potencializar sua capacidade de previsão. No entanto, é importante destacar que o modelo ARIMA pressupõe que a série temporal seja estacionária. Quando

lidamos com séries temporais não estacionárias, é necessário recorrer a outros modelos para a análise e previsão adequadas (VIDHYA, 2023). Um exemplo é o do modelo SARIMA gerado por

$$Y_t = c + \sum_{n=1}^p \alpha_n y_{t-n} + \sum_{n=1}^q \theta_n \epsilon_{t-n} + \sum_{n=1}^P \phi_n y_{t-sn} + \sum_{n=1}^Q \eta_n \epsilon_{t-sn} + \epsilon_t \quad (3.4)$$

O modelo proposto é uma extensão do modelo ARIMA, com a adição de componentes autorregressivos e de média móvel sazonal. Esses componentes extras são ajustados levando em consideração os padrões sazonais presentes nos dados, utilizando atrasos correspondentes à frequência sazonal (por exemplo, 12 para dados mensais). Essa abordagem permite capturar e modelar de forma mais precisa as variações sazonais e melhorar a qualidade das previsões em séries temporais com esse comportamento cíclico (PREDUM, 2021).

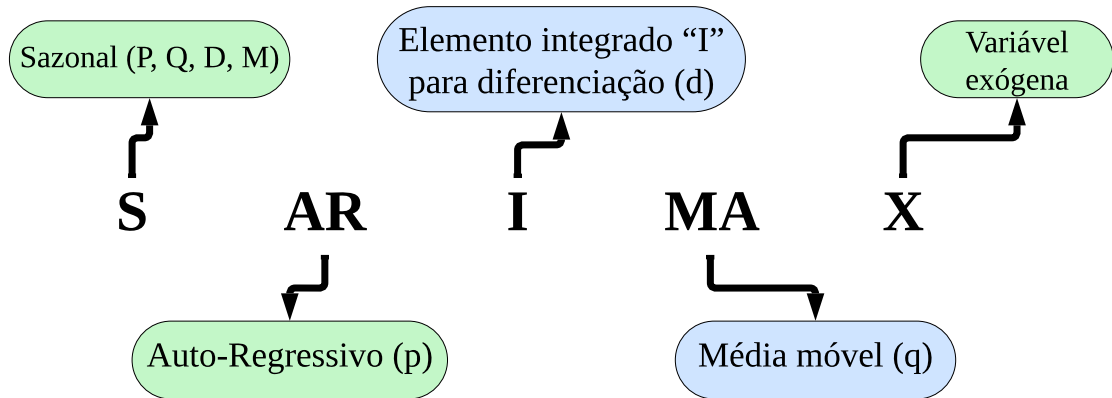
Nos modelos ARIMAX e SARIMAX são consideradas variáveis exógenas, ou seja, são utilizados dados externos para a realização das previsões. É importante ressaltar que mesmo que essas variáveis exógenas sejam indiretamente modeladas no histórico de previsões do modelo, ao incluí-las diretamente, o modelo será capaz de responder de forma ágil aos efeitos dessas variáveis (PREDUM, 2021).

$$d_t = c + \sum_{n=1}^p \alpha_n d_{t-n} + \sum_{n=1}^q \theta_n \epsilon_{t-n} + \sum_{n=1}^r \beta_n x_{n_t} + \sum_{n=1}^P \phi_n d_{t-sn} + \sum_{n=1}^Q \eta_n \epsilon_{t-sn} + \epsilon_t \quad (3.5)$$

onde  $p$  é a ordem de autorregressão de tendência (ACF), é o número de termos autorregressivos (parte AR), permitindo incorporar o efeito de valores passados no modelo,  $d$  é diferença de tendência, o número de diferenças não sazonais necessárias para estacionariedade,  $q$  é a ordem da média móvel de tendência (PCAF), é o número de erros de previsão defasados na equação de previsão (parte MA). Enquanto os elementos sazonais em SARIMAX são,  $P$  é a ordem autorregressiva sazonal,  $D$  é a Ordem das diferenças sazonais,  $M$  é o número de etapas de tempo para um único período sazonal,  $M$  é igual à defasagem ACF com o valor mais alto (normalmente em uma defasagem alta).  $D = 1$  se a série tiver um padrão sazonal estável ao longo do tempo,  $D = 0$  se a série tiver um padrão sazonal instável ao longo do tempo,  $P \geq 1$  se a FAC for positiva na defasagem  $M$ , senão  $P = 0$ ,  $Q \geq 1$  se a ACF for negativa na defasagem  $M$ , caso contrário  $Q = 0$ , e  $X$  é a variável exógena. Na Figura 4 é mostrado detalhes do modelo SARIMAX.



Figura 4: Elementos do modelo SARIMAX



## 3.2 Modelos de Aprendizado de Máquina

Os modelos de aprendizado de máquina para séries temporais têm sido amplamente reconhecidos e utilizados na literatura atual, especialmente por não serem baseados em métodos de gradiente. Esses modelos são valorizados por sua capacidade de capturar relações complexas e não lineares nos dados, permitindo previsões eficientes. Sua popularidade reflete o reconhecimento da eficácia desses modelos em abordar uma ampla gama de problemas de previsão de séries temporais em diferentes áreas de estudo (AL-SHABI, 2021; SEN et al., 2022; KHEIRI; KARIMI, 2023). A seguir são mencionados alguns dos modelos de aprendizado de máquina utilizados e analisados neste estudo.

### 3.2.1 Prophet

O Prophet é um modelo de previsão de séries temporais desenvolvido pelo Facebook. Foi projetado para simplificar a previsão de séries temporais que apresentam padrões sazonais, tendências e feriados. O Prophet é útil para usuários que desejam realizar previsões sem requerer um profundo conhecimento em estatística ou aprendizado de máquina (STEFENON et al., 2023). O modelo se baseia em uma abordagem aditiva que desagrega a série temporal em vários componentes individuais, como tendência de longo prazo, sazonalidade semanal e anual, e efeitos de feriados. Esses componentes são combinados para formar uma previsão geral. A equação básica do modelo Prophet pode ser representada da seguinte forma:

$$p(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (3.6)$$

onde  $p(t)$  é o valor da série temporal no tempo  $t$ , que se deseja prever,  $g(t)$  representa a tendência de longo prazo da série,  $s(t)$  representa os componentes sazonais, que podem incluir padrões semanais e anuais,  $h(t)$  é a representação dos efeitos de feriados ou eventos especiais.

O modelo Prophet ajusta esses componentes aos dados históricos de séries temporais para criar uma previsão futura. Ele utiliza um procedimento de ajuste automático para estimar os parâmetros desses componentes com base nos dados fornecidos. A abordagem aditiva do Prophet permite que os padrões sazonais, tendências e feriados sejam capturados separadamente e, em seguida, somados para gerar uma previsão global (KULSHRESHTHA; VIJAYALAKSHMI, 2020b).

### 3.2.2 Regressão Linear

A regressão linear é definida da seguinte forma:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon \quad (3.7)$$

onde há  $p$  variáveis explicativas, denotadas por  $x$ . Existe uma variável alvo, denotada por  $y$ . O valor de  $y$  é calculado como uma constante  $\beta_0$ , somada aos valores das variáveis  $x$  multiplicados por seus coeficientes  $\beta_1$  a  $\beta_p$ .

Para utilizar a regressão linear (KORSTANJE, 2021), é necessário estimar os coeficientes beta com base em um conjunto de dados de treinamento. Esses coeficientes podem ser estimados por meio de,

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (3.8)$$

onde,  $\hat{\beta}$  é um vetor de coeficientes estimados que minimiza a soma dos quadrados dos resíduos no método de mínimos quadrados ordinários (OLS, *Ordinary Least Squares method*). Cada  $\hat{\beta}_i$  representa o coeficiente estimado para a variável independente  $X_i$ ;  $X$  é a matriz de dados independentes, onde cada coluna representa uma variável independente diferente e cada linha representa uma observação separada;  $X^T$  denota a transposição da matriz  $X$ , ou seja, as linhas de  $X$  tornam-se colunas de  $X^T$  e vice-versa;  $y$  é o vetor de variável dependente, que contém os valores observados que estão sendo modelados ou previstos;  $(X^T X)^{-1}$  representa a inversa da matriz resultante da multiplicação da transposta de  $X$  por  $X$ . Esta é a matriz inversa de  $X^T X$ ;  $X^T y$  denota o produto matricial de  $X^T$  com  $y$ , resultando em um vetor;  $\hat{\beta} = (X^T X)^{-1} X^T y$  representa a multiplicação da inversa de  $X^T X$  com  $X^T y$ , resultando no vetor de coeficientes estimados  $\hat{\beta}$  que minimiza a soma dos

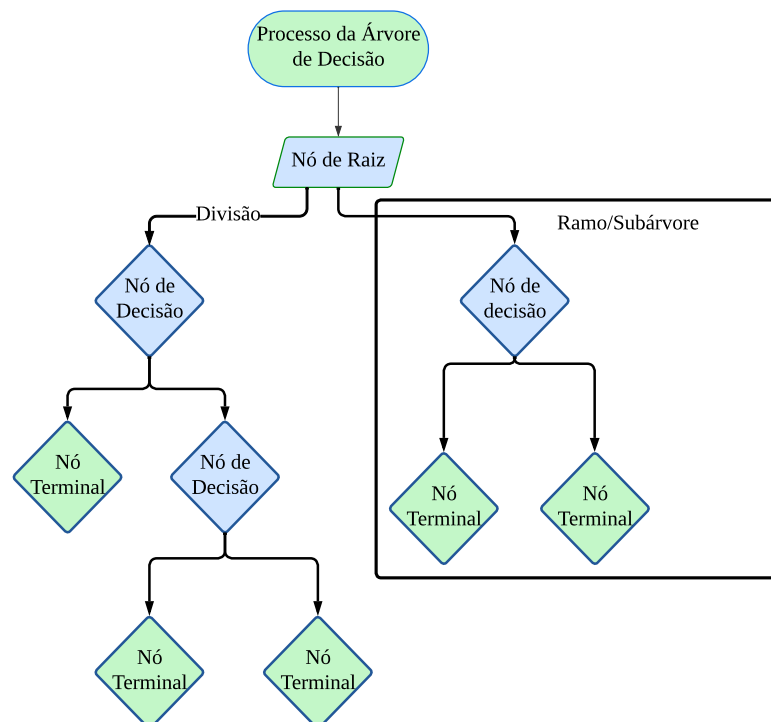
quadrados dos resíduos.

### 3.2.3 Árvore de Decisão

Uma árvore de decisão é um dos modelos de aprendizado de máquina frequentemente utilizados para resolver problemas de regressão e classificação. Como o nome sugere, o algoritmo utiliza um modelo de decisões semelhante a uma árvore para prever o valor de destino (regressão) ou determinar a classe de destino (classificação) (SHI et al., 2023). É importante se familiarizar com as terminologias básicas associadas a uma árvore de decisão (Singh Kushwah et al., 2022).

Na Figura 5 tem-se o nó raiz que é o nó mais alto da árvore que representa todos os pontos de dados. A divisão refere-se à criação de um nó em dois ou mais sub-nós. O nó de decisão são os nós que são divididos em sub-nós, ou seja, esse nó que é dividido é chamado de nó de decisão. O nó folha/terminal são os nós que não se dividem. Esses nós estão geralmente no final da árvore. O ramo/subárvore é uma subseção de toda a árvore e é chamada de galho ou subárvore. O nó pai e filho é um nó, que é dividido em sub-nós e é chamado de um nó pai de sub-nós, enquanto sub-nós são os filhos do nó pai. Na Figura 5, o nó de decisão é o pai dos nós terminais (filhos). A poda é a remoção de sub-nós de um nó de decisão. A poda costuma ser feita em árvores de decisão para evitar o *overfitting* (READER, 2023).

Figura 5: Fluxograma da árvore de decisão.



A árvore de decisão pode ser mais robusta que modelo de regressão linear, tendo a capacidade de otimizar os parâmetros para trabalhar com horizontes de tempo mais longos. Vale destacar novamente, que a árvore de decisão representa um conjunto de regras de decisão hierárquicas, organizadas na forma de uma árvore. Cada nó na árvore representa uma decisão ou teste sobre um atributo, e cada ramo representa um possível resultado dessa decisão (GIFFORD; BAYRAK, 2023). Observa-se analisando a Figura 5 que repetir exatamente a mesma árvore de decisão várias vezes não adiciona valor significativo em comparação com o uso dessa árvore de decisão apenas uma vez. Em modelos de conjunto, é crucial que cada modelo individual apresente pequenas variações em relação aos demais.

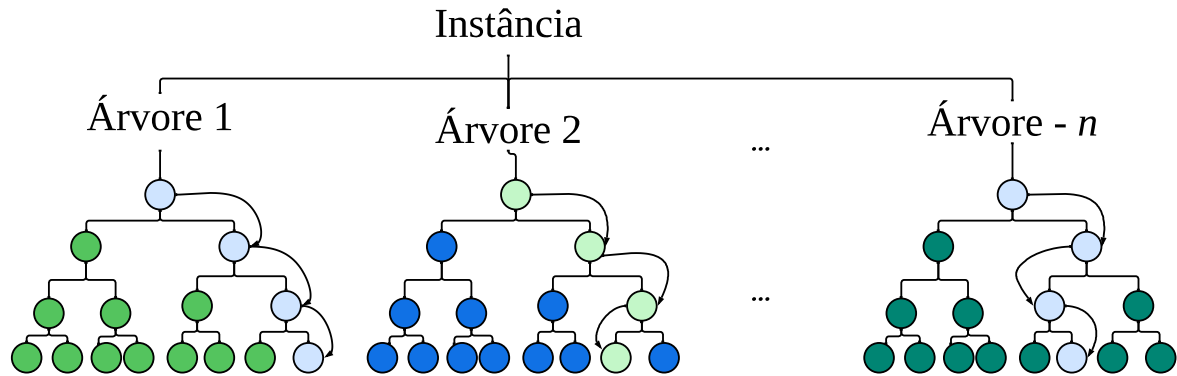
### 3.2.4 Floresta Aleatória

Dois modelos de previsão amplamente reconhecidos para criar conjuntos são *bagging* e *boosting*. A floresta aleatória (do inglês *Random Forest*) (RF) utiliza o ensacamento para criar um conjunto de árvores de decisão, onde cada árvore é construída com uma amostra aleatória do conjunto de dados original. Isso assegura que as árvores sejam distintas e diversificadas, contribuindo para a robustez e eficácia do modelo (SEMAN et al., 2023).

Cada árvore da RF é construída por meio de um algoritmo de aprendizado individual que divide o conjunto de variáveis de entrada em subconjuntos, com base em um teste de valor de atributo, como o coeficiente de Gini. Ao contrário das árvores de decisão clássicas, as árvores do modelo RF são construídas sem poda e selecionam aleatoriamente um subconjunto de variáveis de entrada em cada nó. Atualmente, o número de variáveis utilizadas para dividir um nó em uma RF, denotado por  $m$ , corresponde à raiz quadrada do número total de variáveis de entrada.

Essa abordagem ajuda a aumentar a diversidade das árvores e aprimorar o desempenho do modelo (PELLETIER et al., 2016). Na Figura 6, um fluxograma do modelo RF ilustra como as árvores funcionam. Na construção da próxima árvore, os dois processos anteriores se repetirão, levando à criação de uma nova árvore. Provavelmente, essa árvore será diferente da primeira, pois tanto na seleção das amostras quanto na seleção das variáveis, o processo ocorre de maneira aleatória.

Figura 6: Fluxograma da floresta aleatória.



A utilização do modelo RF em vez do modelo árvore de decisão tem várias vantagens, e essa escolha é frequentemente motivada por considerações de desempenho, robustez e generalização.

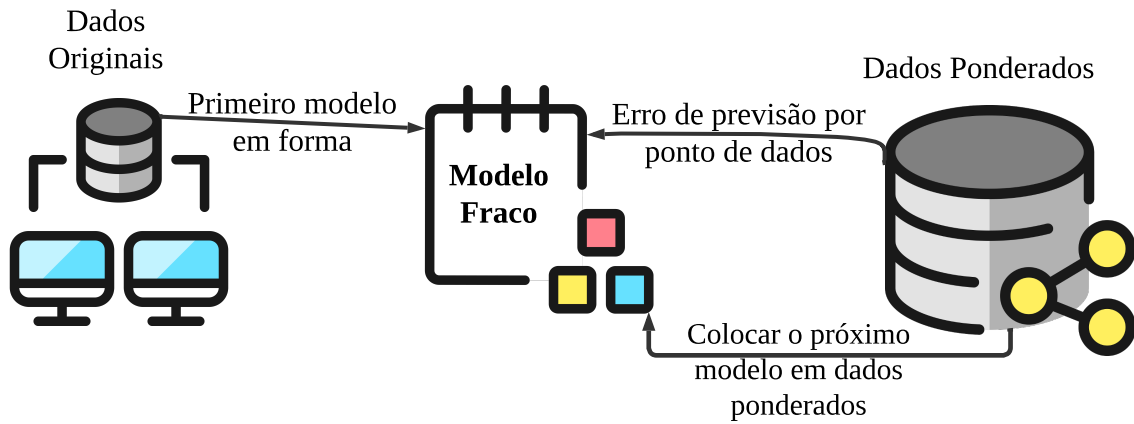
### 3.2.5 Gradient Boosting

O *Gradient Boosting* é um modelo que combina vários modelos de árvore de decisão para realizar previsões. Cada uma dessas árvores de decisão é única, pois a diversidade é um elemento importante nesse processo. A diversidade é alcançada através de um processo chamado *boosting*, que é uma abordagem iterativa, que adiciona modelos fracos ao conjunto de forma inteligente, dando peso maior aos pontos de dados que ainda não foram previstos de forma adequada (BUEECHI et al., 2023).

A Figura 7 apresenta uma visão esquemática do modelo XGBoost. À medida que novos modelos fracos são adicionados, todos os modelos fracos intermediários são mantidos. O modelo final é uma combinação de todos esses modelos fracos, resultando em um ensemble que oferece uma melhor capacidade de previsão do que um único modelo.

O *boosting* é apenas um dos métodos de ensemble utilizados em conjunto com o *bagging*. O *bagging* também é um método que utiliza múltiplos modelos de árvore de decisão, porém, em vez de adicionar os modelos de forma iterativa, cada modelo é treinado independentemente em subconjuntos aleatórios dos dados de treinamento. Ambos os métodos, *boosting* e *bagging*, têm como objetivo melhorar o desempenho do modelo combinando as previsões de múltiplos modelos individuais.

Figura 7: Fluxograma do XGBoost



### 3.2.6 *LightGBM*

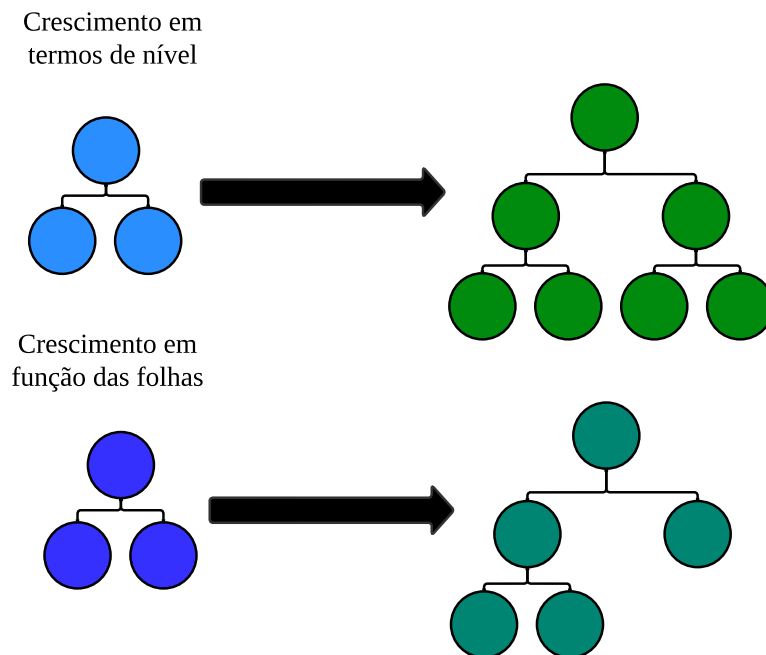
Uma alternativa proposta pelo XGBoost é a segmentação baseada em histograma. Nesse caso, em vez de iterar por todas as partições possíveis, o modelo constrói um histograma para cada variável e utiliza-os para encontrar a melhor divisão geral entre as variáveis. O LightGBM, desenvolvido pela Microsoft, adota uma abordagem mais eficiente para a definição das divisões. Essa abordagem é conhecida como amostragem *Gradient-Based One-Side Sample* (GOSS). O GOSS calcula o gradiente para cada ponto de dados e utiliza-o para filtrar os pontos de dados com gradientes baixos (SUN; LIU; SIMA, 2020).

O LightGBM também utiliza uma abordagem chamada *Exclusive Feature Bundling* (EFB), que acelera a seleção de variáveis correlacionadas. Outra diferença é que o modelo LightGBM é adequado para o crescimento de folhas, enquanto o XGBoost cultiva as árvores em níveis. Essa diferença pode ser visualizada na Figura 8 (YE; ZHAO; DENG, 2023). Essa diferença teoricamente favorece o LightGBM em termos de precisão, mas também apresenta um maior risco de sobre-ajuste quando há poucos dados disponíveis.

Na Figura 8, é possível visualizar como cada modelo é ajustado durante o processo de crescimento de árvore em folhas e em níveis. Essa representação gráfica oferece uma compreensão visual das diferenças entre os modelos. A Figura 8, apresenta um diagrama que ilustra o crescimento de uma árvore em termos de níveis, modelo XGBoost, e crescimento de uma árvore em termos de folhas, modelo LightGBM.

No crescimento de árvore em folhas, no modelo LightGBM, novas folhas são adicionadas à árvore de forma iterativa, visando maximizar a redução do erro de treinamento. Isso significa que as árvores são expandidas adicionando folhas, uma a uma, até que o critério de parada seja alcançado. No crescimento em níveis, no modelo XGBoost, as árvores são expandidas em profundidade de forma simultânea em todos os níveis. Ou

Figura 8: Comparação do crescimento em folha com o crescimento em nível



seja, em cada nível, todas as folhas são expandidas ao mesmo tempo, resultando em um crescimento mais uniforme da árvore. Essa distinção no modo de crescimento das árvores pode afetar o comportamento e o desempenho do modelo.

### 3.3 Redes Neurais Artificiais

Uma rede neural é um modelo de processamento de informações inspirado pelo funcionamento do cérebro humano. Consiste em um conjunto interconectado de unidades de processamento, conhecidas como neurônios artificiais, que trabalham em conjunto para realizar tarefas de aprendizado a partir de dados (XIANG et al., 2018). Assim como os neurônios no cérebro estão interligados por sinapses, os neurônios artificiais são conectados por conexões ponderadas. Essas conexões permitem que a rede neural analise padrões complexos nos dados, reconhecendo relações e características importantes para executar tarefas como classificação, previsão, reconhecimento de padrões (BABU; REDDY, 2014). Conforme a rede é exposta a exemplos e informações, ela ajusta suas conexões para melhorar seu desempenho, tornando-a capaz de generalizar e lidar com novos dados (RAO; ZHAO; DENG, 2020).

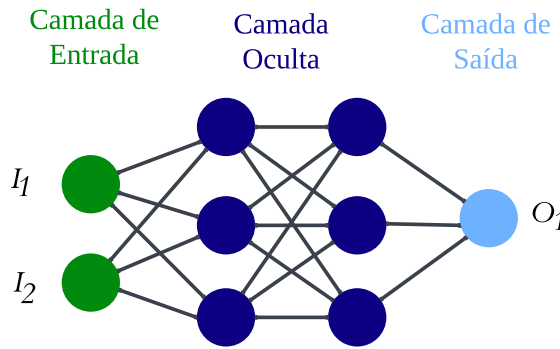
#### 3.3.1 MLP

Uma Rede Neural *Multilayer Perceptron* (MLP) é um tipo de arquitetura de rede neural artificial composta por várias camadas de neurônios (ou unidades) organizados em

uma estrutura de múltiplas camadas. Cada neurônio em uma camada está conectado a todos os neurônios da camada seguinte, sem formar ciclos (rede feedforward). Essa arquitetura é projetada para realizar tarefas de aprendizado supervisionado, como classificação e regressão (QIN et al., 2023).

A topologia da MLP funciona como uma rede *feedforward*, rede progressiva, a saída de um neurônio se conecta com outro neurônio da próxima camada, no sentido esquerda/direita, formada por um conjunto de neurônios denominados nós, como na Figura 9. A rede possui uma camada de entrada, sem função de ativação, uma ou mais camadas ocultas e uma camada de saída. A complexidade da rede neural MLP se dá pela quantidade de camadas ocultas que houver e a quantidade de neurônios que essas camadas possuam (GRÜBLER, 2018).

Figura 9: Modelo de uma rede neural artificial MLP



$$I = [I_1, I_2] = \text{Vetor de Entrada}$$

$$O = [O_1] = \text{Vetor de Saída}$$

O modelo de rede neural artificial MLP é dado por,

$$v_j = \sum_{i=0}^m w_i y_i + b \quad (3.9)$$

o funcionamento geral de uma rede MLP está representada na Figura 9. Cada neurônio recebe todos os valores das entradas, representadas pelo símbolo  $y$ , que são multiplicadas pelos pesos sinápticos simbolizados pelo  $w$  e somadas entre si junto com uma constante chamada de polarização ou bias, representada pelo símbolo  $b$ .



### 3.3.2 Rede Neural Recorrente

Uma Rede Neural Recorrente (*Recurrent Neural Network*, RNN) é um tipo de arquitetura de rede neural que pode ser utilizada para usar dados sequenciais ou temporais (NASIRI; EBADZADEH, 2023). Ao contrário das redes neurais convencionais, onde as entradas e saídas são tratadas como dados independentes, as RNNs levam em consideração a ordem e a relação entre os elementos em uma sequência, tornando-as ideais para lidar com dados como séries temporais.

A característica principal das RNNs é que elas contêm laços em sua estrutura, permitindo que informações anteriores influenciem o processamento de informações subsequentes. Isso significa que a saída em um determinado passo de tempo não depende apenas da entrada atual, mas também das entradas anteriores na sequência.

$$h_t = f(W_{hh} \cdot h_{t-1} + W_{xh} \cdot x_t + b_h) \quad (3.10)$$

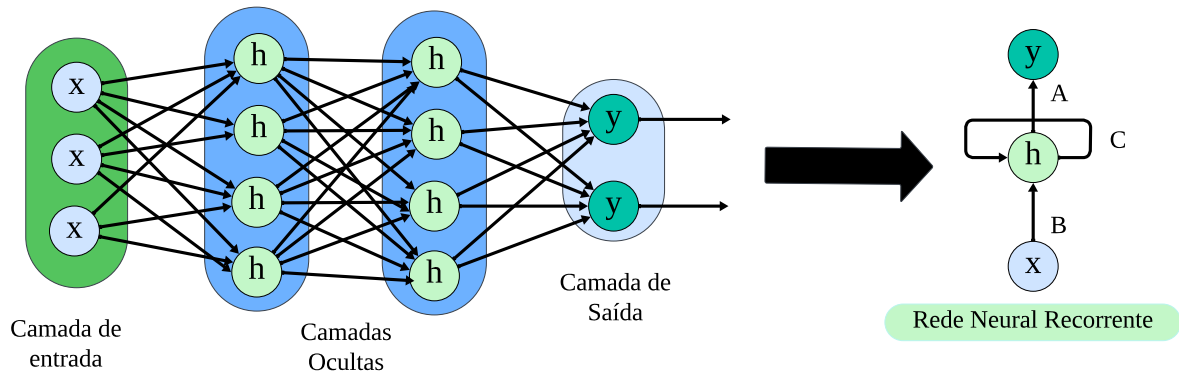
onde  $h_t$  é o estado oculto (ou saída) no tempo  $t$ ,  $h_{t-1}$  é o estado oculto anterior no tempo  $t - 1$ ,  $x_t$  é a entrada no tempo  $t$ ,  $W_{hh}$  é a matriz de pesos que controla a influência do estado oculto anterior,  $W_{xh}$  é a matriz de pesos que controla a influência da entrada,  $b_h$  é o vetor de viés,  $f$  é uma função de ativação, frequentemente a função tangente hiperbólica (tanh) ou a função sigmoide (TAM, 2023).

A equação representa a propagação do estado oculto ao longo do tempo em uma RNN. A cada novo passo de tempo, a RNN considera a entrada atual  $x_t$  e o estado oculto anterior  $h_{t-1}$ , calculando o novo estado oculto  $h_t$  usando as matrizes de pesos e a função de ativação.

No entanto, as RNNs tradicionais podem enfrentar dificuldades em capturar dependências de longo prazo, devido ao problema de dissipação do gradiente. Para entender isso, surgiram variações avançadas, como *Long Short-Term Memory* (LSTM) e *Gated Recurrent Units* (GRU), que incorporam mecanismos de aprendizado de esquecimento e controle de informação, permitindo que informações relevantes sejam mantidas por períodos mais longos de tempo (WANG; YING, 2023), (ZHAO et al., 2023).

Como pode ser visto na Figura 10, a grande diferença da RNN é que há um laço de reações. Enquanto cada entrada de uma rede totalmente conectada é completamente independente, as entradas de uma RNN têm uma relação de realimentação entre si. Isso faz com que ela seja capaz de capturar padrões em dados sequenciais de uma maneira que redes neurais tradicionais talvez não consigam.

Figura 10: Fluxograma da RNN.



### 3.4 Aprendizado Profundo

Aprendizado Profundo ou *Deep Learning* (DL) refere-se a um subcampo do aprendizado de máquina que envolve a construção e o treinamento de modelos de rede neural profunda. O termo profundo se refere ao uso de arquiteturas de modelos que consistem em várias camadas, conhecidas como redes neurais profundas (KOTHONA et al., 2023).

#### 3.4.1 LSTM

As redes neurais LSTMs são uma evolução das RNNs, projetadas para superar desafios na captura de dependências de longo prazo em sequências de dados. Diferentemente das RNNs convencionais, as LSTMs têm a capacidade de manter informações relevantes por longos períodos, tornando-as especialmente eficazes em tarefas que envolvem padrões complexos e dependências temporais distantes (ZHANG, 2021).

Uma das principais inovações das LSTMs é a introdução de unidades de memória chamadas células, que possuem três componentes principais: uma porta de entrada (*input gate*), uma porta de esquecimento (*forget gate*) e uma porta de saída (*output gate*). Essas portas permitem que as LSTMs controlem o fluxo de informações através da célula, decidindo quais informações devem ser mantidas, esquecidas ou passadas para a saída (ZHANG, 2021).

$$f_t = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + b_f) \quad (3.11)$$

$$i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + b_i) \quad (3.12)$$

$$\tilde{C}_t = \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_c) \quad (3.13)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (3.14)$$

$$o_t = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + b_o) \quad (3.15)$$

$$h_t = o_t \odot \tanh(C_t) \quad (3.16)$$

onde  $x_t$  é a entrada no tempo  $t$ ,  $h_{t-1}$  é o estado oculto anterior no tempo  $t - 1$ ,  $f_t$  é o valor da porta de esquecimento,  $i_t$  é o valor da porta de entrada,  $\tilde{C}_t$  é o candidato a novo estado de memória,  $C_t$  é o novo estado de memória,  $o_t$  é o valor da porta de saída,  $h_t$  é o novo estado oculto (saída) no tempo  $t$ ,  $\sigma$  é a função de ativação sigmoide,  $\odot$  representa a multiplicação elemento a elemento.

Essa estrutura permite que as LSTMs controlem o fluxo de informações e aprendam a armazenar ou descartar informações relevantes para diferentes tarefas. As portas de entrada, esquecimento e saída funcionam como mecanismos de controle, permitindo que as LSTMs aprendam a manter informações importantes, esquecer informações desnecessárias e gerar saídas precisas ao longo de sequências temporais.

### 3.4.2 GRU

Uma rede neural GRU é um tipo de arquitetura de RNN que foi projetado para trabalhar com o problema de dissipação de gradiente e captura de dependências de longo prazo em sequências de dados. Essa variação das RNNs tradicionais introduz mecanismos para controlar o fluxo de informação por meio das unidades de tempo. A GRU é uma alternativa vantajosa para a análise de séries temporais, devido à sua habilidade de lidar com sequências de dados de extensões variáveis e de capturar dependências de longo prazo presentes em informações sequenciais. Além disso, a GRU apresenta uma estrutura de simplicidade superior à LSTM, permitindo um processo de treinamento ágil (MIGLIATO; PONTI, 2021).

A estrutura da GRU inclui dois portões principais: o portão de atualização (*update gate*) e o portão de reinicialização (*reset gate*). Esses portões permitem que o GRU decida quais informações serão transmitidas para a próxima etapa de tempo e quais informações serão descartadas, nessas equações

(3.17), (3.18), (3.19) e (3.20):

$h_t$  representa o estado oculto na etapa de tempo  $t$ ,  $h_{t-1}$  é o estado oculto na etapa

de tempo anterior  $t - 1$ ,  $x_t$  é a entrada na etapa de tempo  $t$ ,  $r_t$  é o valor do portão de reinicialização na etapa  $t$ ,  $z_t$  é o valor do portão de atualização na etapa  $t$ ,  $\odot$  denota a multiplicação elemento a elemento,  $\sigma$  é a função sigmoid, que retorna valores entre 0 e 1,  $\tanh$  é a função tangente hiperbólica, que retorna valores entre  $-1$  e  $1$ ,  $W_r, W_z$  e  $W_h$  são matrizes de pesos que o modelo aprende durante o treinamento.

O portão de reinicialização ( $r_t$ ) controla a quantidade de informação do passado a ser esquecida, dada por:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (3.17)$$

O portão de atualização ( $z_t$ ) controla a quantidade de informação do passado a ser passada para o próximo estado, como:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (3.18)$$

A ativação do candidato ( $\tilde{h}_t$ ), candidato a novo estado oculto calculado por:

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t]) \quad (3.19)$$

O novo estado oculto ( $h_t$ ) é a combinação ponderada do estado anterior e do novo candidato, dado por:

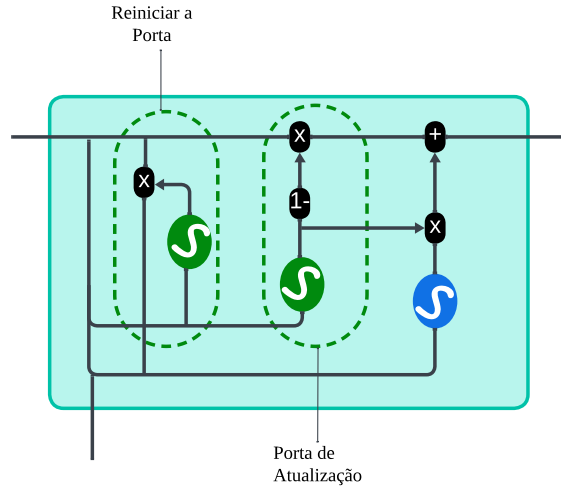
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (3.20)$$

Na Figura 11 é representado um diagrama de um modelo de GRU para análise de séries temporais. O modelo GRU é um tipo de RNN que possui dois portões: um portão de atualização e um portão de reinicialização. Esses portões controlam como a informação é armazenada e atualizada na memória oculta da rede. Um modelo GRU é capaz de aprender padrões temporais complexos e dependências de longo prazo nos dados sequenciais.

A Figura 11 apresenta uma representação simplificada do modelo com três portões: o portão de reinicialização, o portão de atualização e o portão de saída. Os portões são interconectados por linhas tracejadas, representando o fluxo de informação entre eles. O diagrama está rotulado em português, com Porta de Reinicialização, Porta de Atualização, e Porta de Saída (SARANYA; SIVAKUMAR, 2020; JORDAN; SOKÓŁ; PARK, 2021;

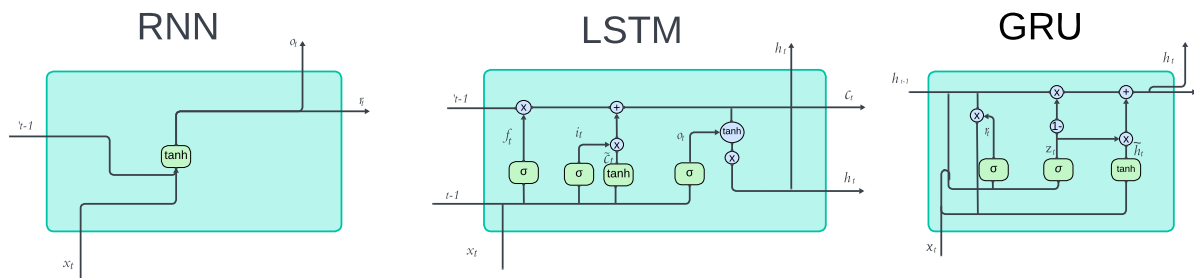
KHAN et al., 2022).

Figura 11: Diagrama do funcionamento de uma GRU.



As redes neurais GRUs, LSTMs, e RNNs apresentam variações em suas arquiteturas, todas projetadas para abordar a dificuldade de capturar dependências temporais em sequências de dados. Enquanto as RNNs tradicionais têm uma tendência a sofrer com o desvanecimento do gradiente ao longo do tempo, as LSTMs e GRUs foram desenvolvidas para superar essa limitação. As LSTMs introduzem células de memória e portas de controle que permitem armazenar e atualizar informações relevantes ao longo das etapas temporais, sendo especialmente adequadas para capturar relações de dependência de longo prazo. As GRUs, por sua vez, simplificam a arquitetura das LSTMs, utilizando portas de atualização para permitir o fluxo de informações e controle sobre o estado oculto. A Figura 12 ilustra as RNNs, LSTMs, e GRUs, permitindo uma visualização das diferenças entre essas redes neurais. Cada seção possui um diagrama da arquitetura da rede, com nós representando neurônios e arestas representando conexões entre neurônios. A RNN tem um único neurônio recorrente, a LSTM tem vários neurônios recorrentes com conexões adicionais que formam portões e células de memória, e a GRU tem dois portões que controlam o fluxo de informação na memória oculta da rede.

Figura 12: Diferenças entre RNN, LSTM, e GRU.



As LSTMs e GRUs oferecem soluções mais sofisticadas em relação às RNNs tradicionais, apresentando mecanismos que permitem capturar dependências de longo prazo de maneira mais eficaz.

### 3.5 Rede Neural Convolucional

As Redes Neurais Convolucionais (*Convolutional Neural Networks*, CNN) são um tipo de rede neural que utiliza a operação de convolução em vez da multiplicação por matrizes em ao menos uma de suas camadas. Esse tipo de rede é efetiva em aplicações em que os dados são dispostos de forma que a relação de vizinhança entre os elementos é relevante, no caso de séries temporais, que são sequências unidimensionais de dados amostrados em intervalos de tempo regulares tem-se este tipo de característica (SILVA, 2021), (REICHMAN; MALOF; COLLINS, 2016) .

A camada convolucional tem como objetivo extrair as características mais importantes da entrada. Dessa forma, sua saída é um mapa de características obtido a partir da convolução da entrada com um *kernel* aprendido, seguido da aplicação de uma função de ativação não linear (LUCAS, 2019). Os mapas de características completos são obtidos por,

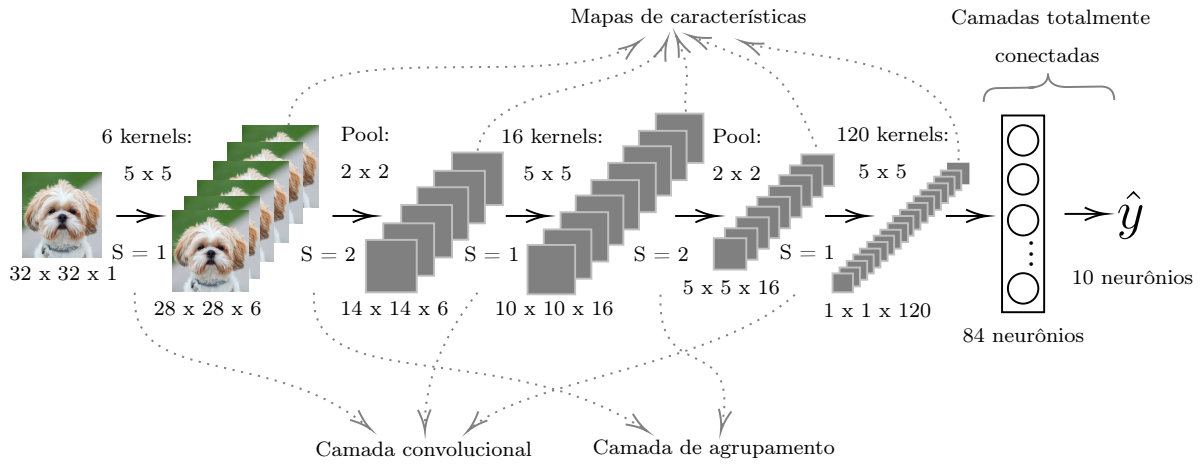
$$Z_{i,j,k}^L = W_k^L \cdot X_{i,j}^L + b_k^L \quad (3.21)$$

onde  $Z_{i,j,k}^L$  é o mapa de características obtido pela convolução do  $k$ -ésimo filtro da  $L$ -ésima camada com a célula de entrada centrada na localização  $(i, j)$ ,  $W_k^L$  vetor de pesos do  $k$ -ésimo filtro da  $L$ -ésima camada,  $b_k^L$  termo de polarização do  $k$ -ésimo filtro da  $L$ -ésima camada,  $X_{i,j}^L$  é a célula de entrada centrada na localização  $(i, j)$  da  $L$ -ésima camada.

A profundidade dos mapas de características é dada pelo número de *kernels* (ou filtros) de convolução. Observe na Figura 13 que a 1ª camada de convolução com 6 *kernels* gera uma saída de profundidade 6. Isso porque, cada *kernel* possui pesos diferentes para extrair diferentes características da entrada (LUCAS, 2019).

Uma vantagem das camadas de convolução é o compartilhamento do vetor de pesos para toda a circunvolução na construção de um mapa de características, pois reduz o número de parâmetros na rede, resultando em treinamento e previsões mais eficientes (LUCAS, 2019). A largura e a altura desses mapas são definidas pelo tamanho do *kernel* e do *stride* (passo da circunvolução) dado por,

Figura 13: Modelo de uma Rede Neural Convolucional.



$$T_{\text{map}} = \left( \frac{I - F}{S + 1} \right) \quad (3.22)$$

onde  $T_{\text{map}}$  é a altura ou largura do mapa de características,  $I$  é a altura ou largura da entrada,  $F$  é a altura ou largura do *kernel* de convolução,  $S$  é o tamanho do *stride*.

### 3.6 Métricas Estatísticas

As métricas estatísticas são utilizadas na análise de previsão de séries temporais para avaliar se o modelo preditor possui um desempenho adequado e/ou superior. Quanto menor for o erro da maior parte das métricas, melhor será o desempenho do modelo aplicados.

A Raiz do Erro Médio Quadrático Relativo (RRMSE) é uma variante do erro quadrático médio (RMSE) sendo calculado por,

$$RRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sum_{i=1}^n (\hat{y}_i)^2} \quad (3.23)$$

onde  $n$  número total de observações ou amostras no conjunto de dados,  $y_i$  valor real da observação  $i$ ,  $\hat{y}_i$  valor previsto ou estimado da observação  $i$  pelo modelo,  $\sum_{i=1}^n$  soma sobre todas as observações no conjunto de dados.

O Erro Absoluto Médio (MAE) é utilizado como uma métrica para avaliar o desempenho de modelos de previsão. Em vez de calcular a média das diferenças entre os

valores reais e previstos, o MAE calcula a média dos valores absolutos dessas diferenças, garantindo que os erros positivos e negativos não se anulem, calculada por,

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (3.24)$$

sua interpretação é similar ao RRMSE, em que o erro é expresso na mesma escala ou ordem de grandeza da variável estudada.

O Erro Percentual Absoluto Médio Simétrico (sMAPE) é outra métrica comumente utilizada para avaliar a precisão de modelos de previsão. O sMAPE é expresso como uma porcentagem, facilitando a compreensão da precisão relativa do modelo. O sMAPE é adequado para trabalhar com valores nulos nos dados, pois a divisão por zero é evitada no cálculo da métrica. O sMAPE é sensível a valores extremos nos dados. Se houver valores discrepantes que não representem a tendência geral, eles podem influenciar significativamente a métrica. Seu cálculo é dado por,

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{2|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)} \times 100 \quad (3.25)$$

### 3.7 Correlação de Pearson

A correlação de Pearson é uma medida estatística que avalia a relação linear entre duas variáveis. No contexto de séries temporais, a correlação de Pearson pode ser útil para entender se existe uma relação linear entre duas séries temporais, ou entre diferentes variáveis de uma mesma série temporal (Cesar de Lima Nogueira et al., 2023). No entanto, é importante ter em mente algumas considerações ao aplicar a correlação de Pearson a séries temporais: (i) a correlação de Pearson é sensível a tendências e sazonalidades nas séries temporais. Se houver uma tendência ou sazonalidade em ambas as séries, a correlação pode indicar uma relação mesmo que a relação real seja mais complexa. (ii) a correlação de Pearson pressupõe uma relação linear entre as variáveis. Se a relação entre as séries temporais for não linear, a correlação de Pearson pode não capturar essa relação de maneira adequada, e (iii) a correlação de Pearson pode ser sensível a valores extremos (outliers) (dos Santos Coelho; Hultmann Ayala; Cocco Mariani, 2024). Valores extremos podem distorcer a medida de correlação, tornando-a menos representativa da relação geral entre as séries.

Ao usar a correlação de Pearson para análise de séries temporais, é importante interpretar os resultados com cautela e considerar outros métodos estatísticos e gráficos



para uma compreensão mais completa da relação entre as séries. A equação do coeficiente de correlação de Pearson é dada por:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum (x_i - \bar{x})^2)(\sum (y_i - \bar{y})^2)}} \quad (3.26)$$

onde  $x_i$  e  $y_i$  representam os valores das variáveis  $X$  e  $Y$ , respectivamente.  $\bar{x}$  e  $\bar{y}$  são as médias dos valores  $x_i$  e  $y_i$ .

O coeficiente de correlação de Pearson mede a força e a direção da relação linear entre as variáveis  $X$  e  $Y$ . Valores próximos a 1 indicam uma correlação positiva forte, valores próximos a  $-1$  indicam uma correlação negativa forte, e valores próximos a 0 indicam ausência de correlação entre as variáveis.

### 3.8 Decomposição STL

A decomposição sazonal e de tendência utilizando o procedimento de suavização de diagramas de dispersão estimada localmente (Loess) (STL) é uma técnica amplamente utilizada para decompor séries temporais em seus componentes sazonais, de tendência e restantes (RIBEIRO et al., 2023). O método STL realiza a decomposição aditiva dos dados por meio de uma sequência de aplicações do Loess mais suave, onde regressões polinomiais ponderadas localmente são aplicadas em cada amostra do conjunto de dados, tendo como variáveis explicativas os valores próximos do amostra cuja resposta está sendo estimada (THEODOSIOU, 2011).

Ao aplicar a decomposição STL, a série temporal pode ser expressa como a soma dos componentes sazonais, de tendência e restantes. Essa técnica é útil para análise e modelagem de séries temporais, pois proporciona uma compreensão clara dos padrões de variação presentes nos dados.

A decomposição STL é formalmente definida como:

$$y_t = f(S_t, T_t, R_t) = \begin{cases} y_t = S_t + T_t + R_t & \text{modelo aditivo} \\ y_t = S_t T_t R_t & \text{modelo multiplicativo} \end{cases} \quad (3.27)$$

onde  $y_t$  é o valor da série temporal no tempo  $t$ ,  $T_t$  é a componente de tendência no tempo  $t$ ,  $S_t$  é a componente de sazonalidade no tempo  $t$ ,  $R_t$  é a componente de resíduo no tempo  $t$ .

### 3.9 Teste Dickey-Fuller

O teste Dickey-Fuller (DF) segue uma regressão linear autoregressiva (AR) de primeira ordem e testa a hipótese nula de que uma raiz unitária está presente em uma série temporal, indicando a não estacionariedade. A forma mais comum do teste Dickey-Fuller é conhecida como ADF (Augmented Dickey-Fuller), que inclui termos adicionais na regressão para trabalhar com possíveis problemas de autocorrelação (AGIAKLOGLOU; NEWBOLD, 1992). O procedimento geral do teste Dickey-Fuller envolve as seguintes etapas:

**Formulação da Hipótese Nula (H0):** A hipótese nula assume a presença de raízes unitárias, o que indica não estacionariedade na série temporal. **Formulação da Hipótese Alternativa (H1):** A hipótese alternativa busca rejeitar a hipótese nula, sugerindo a estacionariedade da série temporal. **Realização do Teste:** O teste Dickey-Fuller é realizado calculando uma estatística de teste. Se o valor-p associado à estatística de teste for menor que um determinado nível de significância pré-definido, a hipótese nula é rejeitada. **Interpretação dos Resultados:** Se a hipótese nula for rejeitada, isso sugere que a série temporal é estacionária. Caso contrário, a não estacionariedade da série temporal não pode ser descartada.

De acordo com o Reisen et al. (2017), o teste DF é representado por,

$$\Delta y(t) = \alpha + \beta \cdot t + \gamma \cdot y(t-1) + \delta_1 \cdot \Delta y(t-1) + \delta_2 \cdot \Delta y(t-2) + \dots + \delta_p \quad (3.28)$$

$$\Delta y(t-p) + \varepsilon(t) \quad (3.29)$$

onde  $\Delta y(t)$  é a diferença entre os valores consecutivos da série temporal no tempo  $t$ ,  $y(t-1)$  é o valor da série temporal no tempo anterior,  $t$  é uma variável de tendência temporal,  $\alpha, \beta, \gamma$  são parâmetros a serem estimados,  $\delta_1, \delta_2, \dots, \delta_p$  são os coeficientes associados às diferenças defasadas,  $\varepsilon(t)$  é o termo de erro.

### 3.10 Teste de Ljung-Box

A ideia do teste de Ljung-Box é verificar se as autocorrelações dos resíduos em diferentes defasagens são estatisticamente diferentes de zero. Se a estatística de teste do Ljung-Box indicar significância estatística, isso sugere que há autocorrelações remanescentes nos resíduos, indicando que o modelo não está capturando completamente a estrutura da série temporal (BOX; PIERCE, 1970), (LJUNG; BOX, 1978), (DAVIES; NEWBOLD, 1979). As etapas básicas do teste de Ljung-Box:

**Formulação da Hipótese Nula (H0):** A hipótese nula assume que não há autocor-

relação nos resíduos. Formulação da Hipótese Alternativa (H1): A hipótese alternativa sugere que há autocorrelação nos resíduos. Cálculo da Estatística de Teste: A estatística de teste de Ljung-Box é calculada usando as autocorrelações dos resíduos em várias defasagens. A fórmula é baseada na soma dos quadrados dessas autocorrelações. Distribuição da Estatística de Teste: A estatística de teste é comparada a uma distribuição qui-quadrado com um número apropriado de graus de liberdade. Isso depende do número de defasagens considerado no teste. Decisão Estatística: Se o valor-p associado à estatística de teste for menor que um nível de significância escolhido (por exemplo, 0,05), a hipótese nula é rejeitada, indicando a presença de autocorrelação nos resíduos.

A estatística de teste Box-Pierce é uma versão simplificada da estatística de Ljung-Box para a qual estudos de simulação subsequentes mostraram baixo desempenho (DAVIES; NEWBOLD, 1979). A estatística de teste é (LJUNG; BOX, 1978):

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k} \quad (3.30)$$

onde  $n$  é o tamanho da amostra,  $\hat{\rho}_k$  é a autocorrelação da amostra no lag  $k$  e  $h$  é o número de defasagens que estão sendo testadas. Debaixo  $H_0$  a estatística  $Q$  segue assintoticamente um  $\chi^2_{(h)}$ . Para o nível de significância  $\alpha$ , a região crítica para rejeição da hipótese de aleatoriedade é:

$$Q > \chi^2_{1-\alpha, h} \quad (3.31)$$

onde  $\chi^2_{1-\alpha, h}$  é o  $(1 - \alpha)$ -quantil (BROCKWELL; DAVIS, 2002) da distribuição qui-quadrada com graus  $h$  de liberdade.

O teste de Ljung-Box é comumente usado na modelagem de média móvel integrada ARIMA. Note que ele é aplicado aos resíduos de um modelo ARIMA ajustado, não à série original, e em tais aplicações a hipótese que está sendo testada é que os resíduos do modelo ARIMA não têm autocorrelação. Ao testar os resíduos de um modelo ARIMA estimado, os graus de liberdade precisam ser ajustados para refletir a estimação do parâmetro. Por exemplo, para um modelo ARIMA  $(p, 0, q)$ , os graus de liberdade devem ser definidos como  $h - p - q$  (DAVIDSON, 2000). O teste Box-Pierce utiliza a estatística do teste, na notação descrita previamente, dada por (BOX; PIERCE, 1970)

$$Q_{BP} = n \sum_{k=1}^h \hat{\rho}_k^2 \quad (3.32)$$

e usa a mesma região crítica definida previamente. Estudos de simulação mostraram que a distribuição para a estatística Ljung-Box é mais próxima de um  $\chi_{(h)}^2$  6 distribuição do que é a distribuição para a estatística Box-Pierce para todos os tamanhos de amostra, incluindo os pequenos.

### 3.11 Teste de Significância

O teste de Friedman classifica os modelos  $K$  em cada conjunto de dados em relação ao valor absoluto dos resultados dados por esses algoritmos. A classificação do algoritmo com maior desempenho é 1, e o com menor desempenho é classificado como  $K$ . Em seguida, o valor da estatística com base em todas as classificações é calculado como mostrado em equações (3.33) e (3.34) com  $r_{eu}^j$  sendo a classificação do desempenho do  $j$ -ésimo algoritmo no  $i$ -ésimo conjunto de dados. Essa estatística obedece à distribuição do quiquadrado com  $K - 1$  graus de liberdade (LIU; XU, 2022).

$$\chi_F^2 = \frac{12N}{K(K+1)} \left[ \sum_{j=1}^K R_j^2 - \frac{K(K+1)^2}{4} \right] \quad (3.33)$$

$$R_j = \frac{1}{N} \sum_{i=1}^N r_{ei}^j \quad (3.34)$$

$$F_F = \frac{(N-1)\chi_F^2}{N(K-1)\chi_F^2} \quad (3.35)$$

As estatísticas  $F_F$  mostrados na equação (3.35) obedecem à distribuição  $F$  com graus de liberdade  $K - 1$  e  $(K - 1)(N - 1)$ . Pode-se obter o valor crítico abaixo do nível de significância especificado (geralmente  $\alpha = 0,05$  ou  $0,01$ ). Ao comparar esse valor crítico com o valor calculado com a equação (3.35), a hipótese nula é rejeitada se o valor estatístico  $F_F$  é maior que o valor crítico, indicando que há diferenças significativas entre os algoritmos  $K$ . Em seguida, pode-se realizar um procedimento *post hoc* para analisar melhor se o algoritmo de controle é significativamente melhor do que cada algoritmo de referência nos experimentos. Ao contrário, se o valor for menor ou igual ao valor crítico, a hipótese nula é aceita, indicando que não há diferenças significativas entre os algoritmos  $K$ .

Adicionalmente, utilizou-se a *Critical Difference* (CD) para determinar se dois

classificadores eram significativamente diferentes entre si. O CD foi calculado conforme a fórmula mencionada anteriormente:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \quad (3.36)$$

na equação do CD,  $q_\alpha$  representa o valor crítico obtido da Tabela 17 de teste de Nemenyi,  $k$  é o número de classificadores e  $N$  é o número total de amostras (LIU; XU, 2022).

## 4 Resultados

Neste capítulo, é fornecida uma síntese dos resultados obtidos até o momento. Apresenta-se um resumo sucinto da análise de resultados das principais realizações e descobertas alcançadas.

### 4.1 Análise Exploratória dos Dados

Os dados foram processados usando a análise exploratória de dados, em inglês *Exploratory Data Analysis* (EDA) (BANDARA; HYNDMAN; BERGMEIR, 2021) resumindo suas principais características, e formulando hipóteses que possam direcionar a coleta adicional de dados, se necessária.

Existem dados anômalos com NaN, do inglês *Not a Number* que representam a falta de dados coletados, logo tais dados foram interpolados usando os valores existentes e vizinhos a ele.

Assim como em qualquer empresa de saneamento básico e tratamento de água, é utilizado um mecanismo de acionamento automático denominado trava de segurança para evitar que o nível do tanque se aproxime de zero e haja falta de água nos locais abastecidos por esse tanque. O nível máximo que o tanque pode alcançar é de  $5,26m^3$  (equivalente a 5264.56 litros). As bombas são ativadas em sua potência máxima para evitar que sejam acionadas quando o nível do tanque estiver dentro dessa faixa. No entanto, a bomba 1 ainda estaria operando para completar o nível do tanque caso esteja dentro dessa faixa.

Na Tabela 7, o desvio padrão é representado pela sigla STD, que corresponde à expressão em inglês *STandard Deviation*. A quantidade de dados medidos hora a hora pela SANEPAR de 2018 a 2019 são de 17.522.

No contexto de análises de dados várias técnicas de EDA têm sido adotadas. Neste estudo serão abordadas várias análises entre elas correlação de Pearson, para verificar quais variáveis podem ser excluídas tais como ruídos (correlação baixa) ou altamente correlacionadas com a variável de saída do nosso estudo, que será a variável LT01, nível

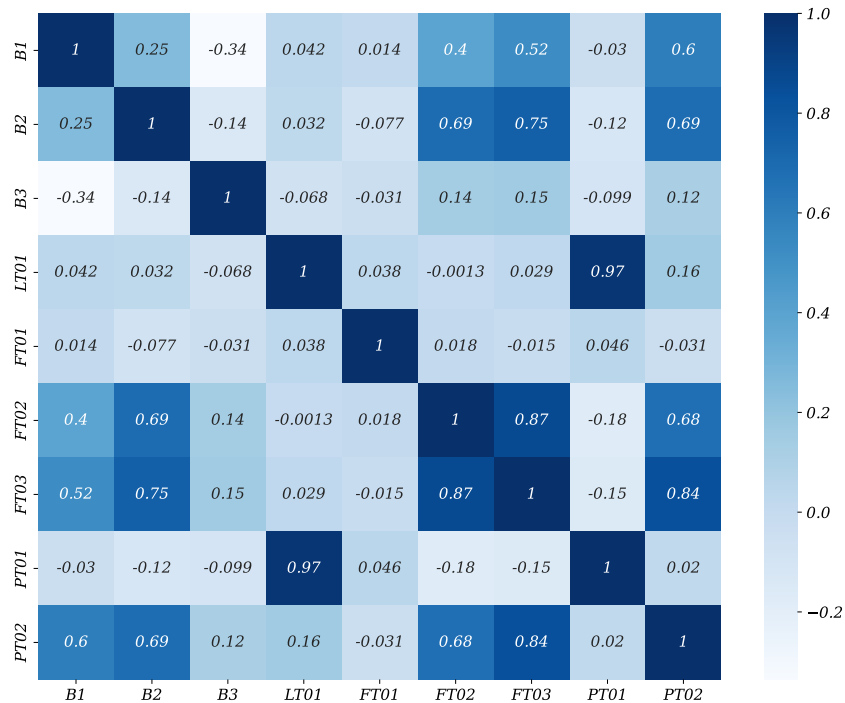
Tabela 7: Descrição estatística dos dados de 2018 a 2019

Métricas	B1	B2	B3	LT01	FT01	FT02	FT03	PT01	PT02
<b>Média</b>	52,289	18,421	3,338	3,513	215,699	114,832	104,195	4,448	20,724
<b>STD</b>	11,421	19,742	12,624	0,670	110,223	43,604	25,636	0,700	3,610
<b>Min.</b>	0	0	0	0,294	0	0	0	0,842	0
<b>25%</b>	49,519	0	0	3,077	255,454	74,912	81,430	4,015	18,072
<b>50%</b>	57,925	0,050	0	3,715	265,357	122,149	109,911	4,602	21,791
<b>75%</b>	57,989	36,796	0	4,047	272,609	145,865	123,189	4,990	23,051
<b>Max.</b>	59,988	59,992	59,988	4,445	390,683	400,415	183,900	5,639	29,008

do tanque de armazenamento de água pela SANEPAR no Bairro Alto. Nesse caso, as variáveis removidas tem pouca correlação com a LT01, as que apresentam correlação baixa são B3 e FT02 e serão removidas.

A Figura 14 mostra a correlação de Pearson entre as variáveis do conjunto de dados deste estudo. Essa figura representa a matriz (simétrica) da dependência/correlação entre as variáveis. Analisamos as correlações das variáveis de entrada com a variável de saída. Serão descartadas as variáveis com correlação menor que 5% e maior que 90%.

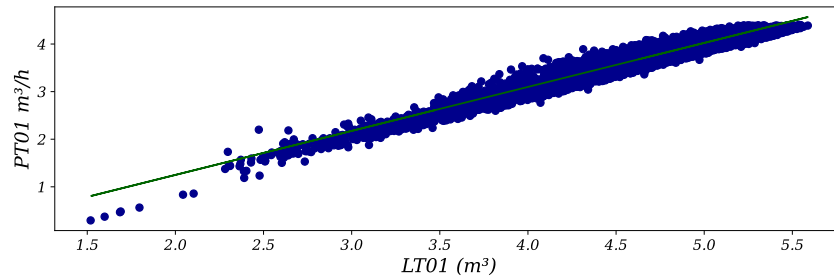
Figura 14: Correlação de Pearson



Nesse conjunto de dados que está sendo trabalhado, há uma forte correlação da variável PT01 com o LT01 conforme visto na Figura 15 que fornece uma representação dos coeficientes  $\beta_0$  e  $\beta_1$ , que são os coeficientes da correlação linear entre as variáveis. Um aumento de 1 na variável  $x$  está associado a um aumento proporcional de  $\beta_1$  na variável

$y$ . O valor de  $\beta_0$  representa o valor de  $y$  quando  $x$  é igual a 0.

Figura 15: Relação entre LT01 e PT01 cuja correlação de Pearson é 97%



A estatística ADF (do inglês *Augmented Dickey-Fuller*) de  $-12,515$  indica a evidência de estacionariedade na série temporal. Quanto mais negativo for o valor da estatística ADF, maior é a evidência de estacionariedade nos dados.

O valor de  $p$ , aproximadamente de  $0,000000000000000000000000262$ , expresso de forma mais concisa como  $2,62 \times 10^{-23}$  usando a notação científica, está associado ao teste ADF. Este valor- $p$  representa a probabilidade de obter um resultado igual ou mais extremo do que o observado, sob a suposição de que a hipótese nula seja verdadeira. No contexto do teste ADF, a hipótese nula é a presença de raiz unitária na série temporal, indicando não estacionariedade. Portanto, um valor de  $p$  baixo, geralmente abaixo de um nível de significância predefinido, como 0,05, sugere que a série temporal é estacionária, enquanto um valor de  $p$  alto sugere não estacionariedade. Dado o valor de  $p$  de  $2,62 \times 10^{-23}$ , evidencia-se uma probabilidade muito baixa, indicando forte suporte contra a hipótese nula e sugerindo que a série temporal é estacionária. Na Tabela 8, são apresentados todos os dados do teste para estacionalidade. Os resultados indicam fortes evidências contra a hipótese nula. Com um teste ADF de  $-12,515$  e um valor de  $p$  extremamente baixo de  $2,62 \times 10^{-23}$ , rejeita-se a hipótese nula de presença de raiz unitária. Os 44 atrasos utilizados e as 17.477 observações corroboram a análise estatística.

Tabela 8: Teste de Dickey-Fuller Aumentado

Teste ADF	$-12,515$
Valor de $p$	$2,62 \times 10^{-23}$
Atrasos utilizados	44
Número de observações	17.477
Valor crítico (1%)	$-3,431$
Valor crítico (5%)	$-2,862$
Valor crítico (10%)	$-2,567$

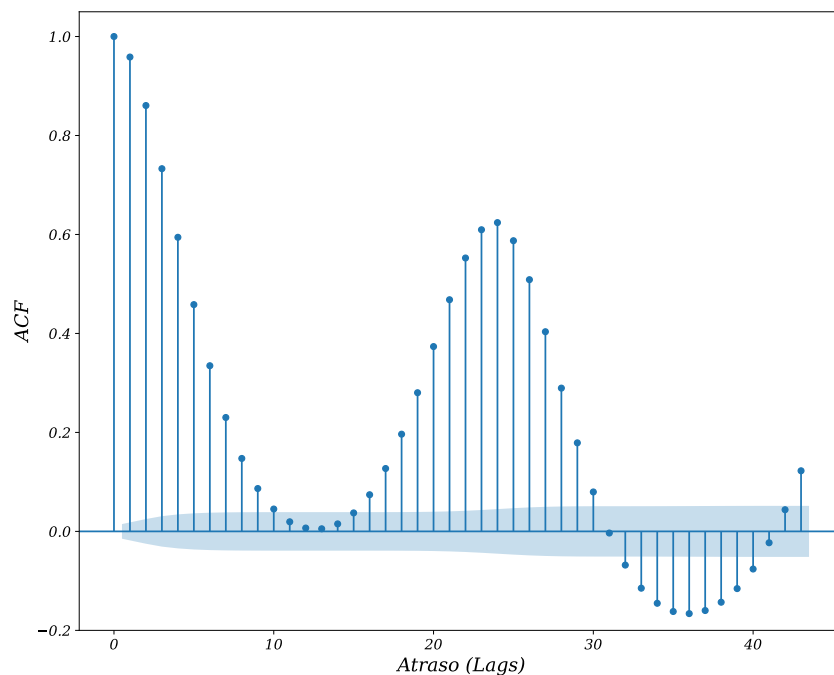
Ao comparar a estatística de teste ADF com os valores críticos, observa-se que está significativamente abaixo deles em todos os níveis de significância (1%, 5%, 10%).

Portanto, a conclusão é de que os dados não possuem raiz unitária, indicando que são estacionários.

Na Figura 16, pode-se observar a diferença entre a autocorrelação (ACF) exibida na Figura 16 e a autocorrelação parcial (PACF) exibida na Figura 17. A autocorrelação é uma medida da correlação entre os valores da série temporal em diferentes defasagens, levando em consideração tanto a correlação direta quanto a correlação indireta. Por outro lado, a autocorrelação parcial mede apenas a correlação direta entre os valores, desconsiderando a influência das defasagens intermediárias. Essas análises são úteis para identificar padrões e relações de dependência entre os valores da série temporal, fornecendo informações importantes para a modelagem e previsão desses dados. O intervalo de confiança padrão de 95% é representado pela marca azul nas Figuras 16 e 17. As observações que estão fora desse intervalo são consideradas estatisticamente correlacionadas, indicando a presença de padrões ou estrutura na série temporal.

A correlação visualizada na Figura 16 é fundamental para a interpretação do teste ADF. Em uma série de ruído branco, os valores são completamente aleatórios e não apresentam correlação significativa. Portanto, quando há correlação presente na série, isso indica a existência de padrões ou dependências entre os valores, o que pode ser explorado para a modelagem e previsão da série temporal.

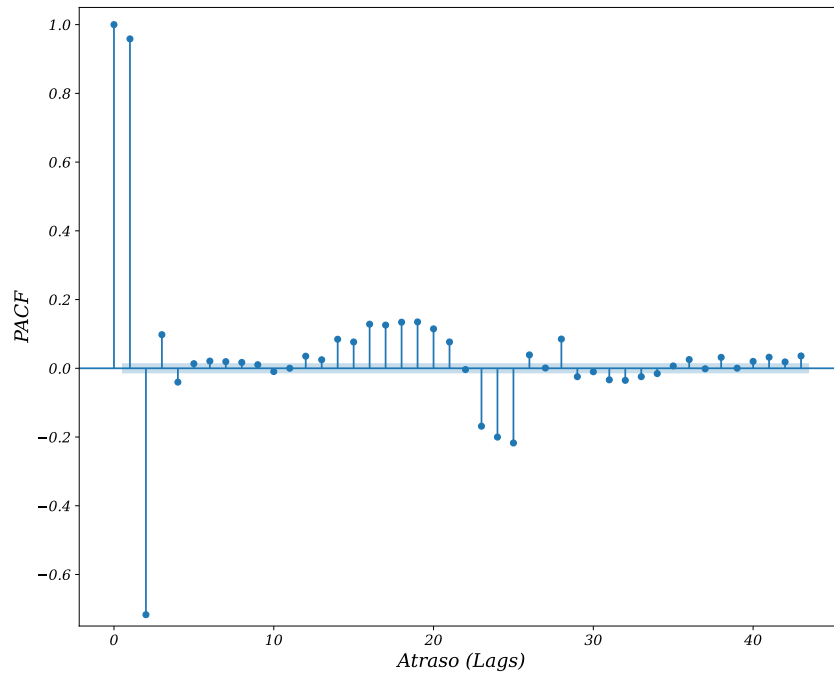
Figura 16: Autocorrelação



Demonstrar que uma série temporal tem ou pode ter um ruído branco também é conveniente para a análise da EDA. Na Figura 18, é possível observar uma série temporal que pode ser caracterizada como ruído branco. Uma série temporal é considerada ruído

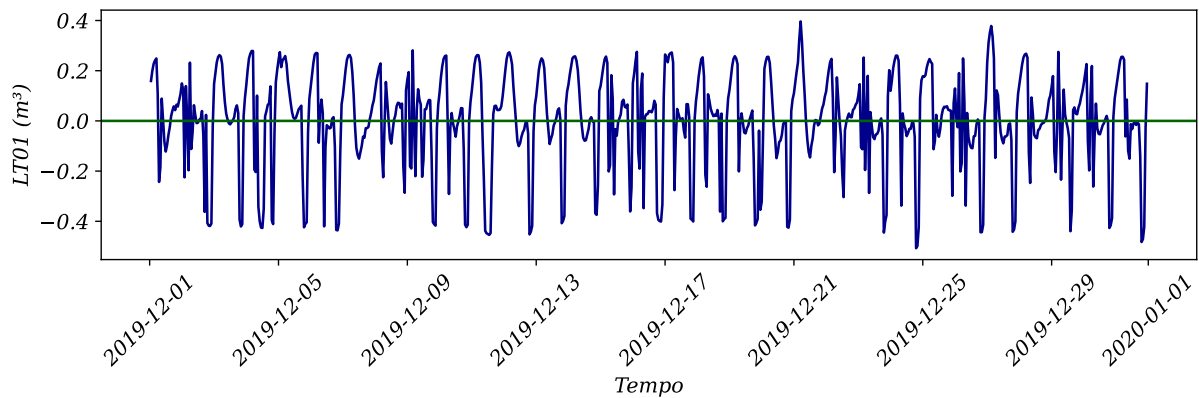


Figura 17: Autocorrelação parcial



branco se suas variáveis forem independentes e distribuídas de forma idêntica, com média zero. Isso implica que todas as variáveis possuem a mesma variância ( $\sigma^2$ ) e que cada valor não possui correlação com os demais valores da série.

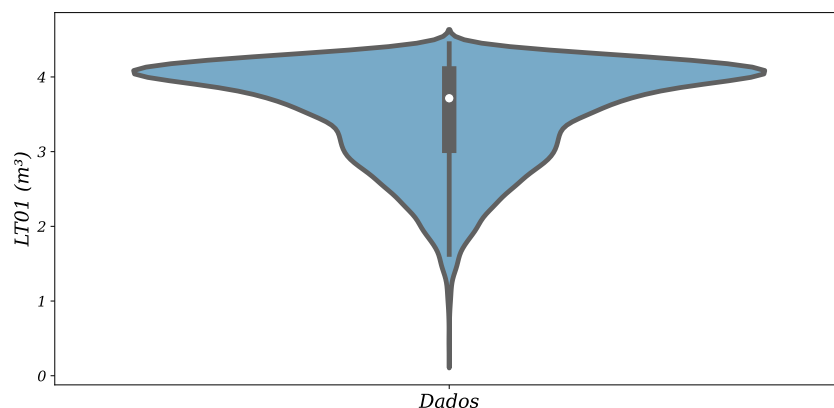
Figura 18: Ruído branco



Nesse exemplo, ao utilizar os dados da SANEPAR, a série temporal trabalhada é estacionária e também apresenta ruído branco (do inglês *white noise*).

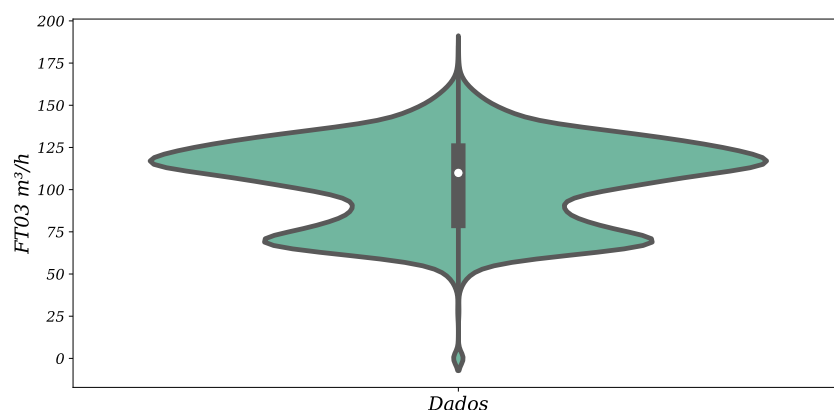
Com base na forte evidência contra a hipótese nula, podemos rejeitar a hipótese nula. A Figura 19, podemos notar um aumento na demanda durante essas horas durante o ano de 2019. Conforme mencionado na subseção 1.2, as anomalias climáticas ocorridas em 2020, especialmente a falta de chuvas e devido ao COVID-19, tiveram um impacto significativo nos resultados. Isso contribuiu para as mudanças observadas na demanda de água ao longo desse período.

Figura 19: Violino no nível do reservatório



A Figura 20 mostra como a vazão pode ser afetada pelo nível do tanque. É interessante observar que a vazão de recalque tem um impacto mais significativo no nível do tanque em comparação com as outras vazões. Isso ocorre porque a vazão de recalque está associada à injeção de água diretamente no tanque por meio da bomba localizada próxima à base do tanque. Por outro lado, as demais vazões apresentam alguns valores ausentes, o que limita sua influência na análise geral.

Figura 20: Violino da vazão de recalque

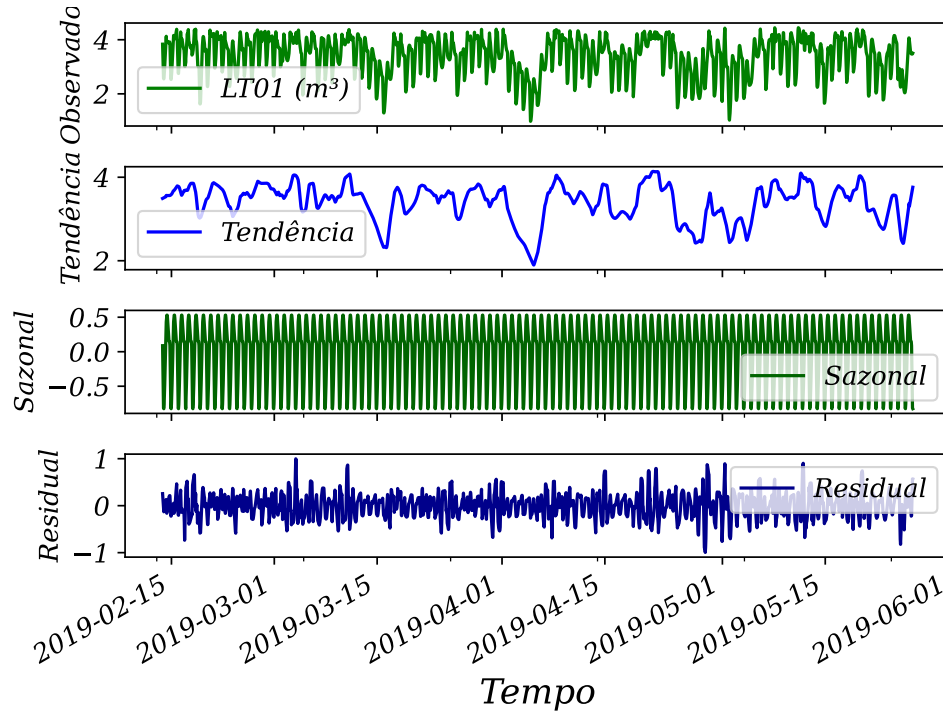


**Múltiplas entradas e saída única (MISO)** (do inglês *Multiple Inputs, Single Output*) na dissertação. O modelo ARIMA, juntamente com suas variantes e extensões, foi amplamente estudado durante a pesquisa, assim como modelos regressivos que envolvem múltiplas variáveis de entrada e uma variável de saída, neste caso, a LT01. As demais variáveis foram utilizadas como suporte para melhorar o modelo do tipo ARIMAX ou modelos com variáveis exógenas. Quando aplicado sem o uso de variáveis exógenas, o modelo ARIMA apresenta apenas uma entrada, semelhante ao modelo de LR. No entanto, ao incluir variáveis exógenas, o modelo se torna MISO, permitindo uma modelagem abrangente e considerando a interação de várias variáveis para prever a variável de interesse.

**Decomposição STL** através da decomposição, é possível analisar se a série apre-

senta tendência, sazonalidade e resíduos. Ao observar a Figura 21, é evidente que os dados exibem ambos os padrões. Isso indica que a série é estacionária, como confirmado pelo seguinte teste ADF anterior.

Figura 21: Decomposição STL aditiva dos dados coletados



**Separação dos Dados** os dados foram divididos em conjuntos de treinamento, teste e validação. Essa prática é comum entre profissionais de aprendizado de máquina, pois permite avaliar o desempenho do modelo em conjuntos de dados diferentes (RASHKA, 2015; GÉRON, 2017).

Quanto à divisão dos dados, foi adotada uma estratégia básica em que 70% dos dados foram destinados ao conjunto de treinamento e os 30% restantes foram reservados para o conjunto de teste. Dentro dos 70% de treinamento, foi realizada uma subdivisão em que 80% desses dados foram usados novamente para treinamento e os 20% restantes foram utilizados para validação. Essa abordagem foi implementada em linguagem de programação para facilitar o processo e evitar a necessidade de recalculá-la a cada modificação do modelo.

**Modelagem e seleção do modelo** a estratégia recursiva é mencionada por Petropoulos et al. (2022) como uma abordagem eficaz na previsão de séries temporais de múltiplos passos. De acordo com o autor, essa estratégia envolve o uso de previsões anteriores como entradas para prever os próximos passos da série temporal. A abordagem recursiva tem demonstrado potencial para melhorar a acurácia das previsões de séries temporais de longo prazo.

A previsão dos dados em uma janela de horizonte de previsão estendida, abrangendo diferentes períodos de tempo, como uma hora, seis horas, doze horas e um dia. Essa estratégia de previsão recorrente permite a comparação entre modelos de regressão e modelos ARIMA em diferentes horizontes temporais.

Essa abordagem é vantajosa, pois cada modelo possui suas próprias características e desempenho ao lidar com previsões de curto prazo, como um dia, e previsões de prazo mais longo, como um dia. Ao utilizar uma janela de previsão mais ampla, é possível observar e avaliar melhor as diferenças entre os modelos e analisar seu desempenho em horizontes de tempo variados.

Além desses modelos, vários outros foram implementados no documento, tais como DTR, RFR, XGBRegressor, LGBMRegressor, LSTM, GRU, Prophet, RNN, Transformer, CNN e ANN, a fim de obter o melhor resultado para a previsão de séries temporais de abastecimento de água.

**Validação e ajuste do modelo** o horizonte de previsão foi personalizado com base no método recursivo de previsão de série temporal e na previsão do nível do tanque LT01. Foram selecionados os seguintes passos para a previsão à frente: uma hora, seis horas, doze horas e um dia. Essa escolha do horizonte de previsão foi feita levando em consideração a estratégia recursiva e os objetivos específicos do estudo. Identifica-se que essa janela de tempo proporciona uma análise mais adequada e comparável entre os modelos utilizados.

Foram utilizados os parâmetros obtidos pelo autoARIMA, que são  $(p = 7, d = 0, q = 0)(P = 2, D = 1, Q = 1)_{M=12}$ , mas foram ajustados para obter um melhor resultado, sendo  $(p = 7, d = 1, q = 7)(P = 2, D = 1, Q = 1)_{M=12}$ . Na Tabela 9, são exibidos todos os modelos obtidos por esse método do “autoARIMA” e ajustados para que obtenham o melhor resultado.  $p$ : Ordem do componente AR (*Auto-Regressivo*),  $d$ : Número de diferenciações não sazonais,  $q$ : Ordem do componente MA (*Média Móvel*),  $P$ : Ordem do componente AR sazonal,  $D$ : Número de diferenciações sazonais,  $Q$ : Ordem do componente MA sazonal,  $M$ : Período sazonal (número de observações em um ciclo sazonal). Na Tabela 10 mostra como a biblioteca do Python autoARIMA obteve os resultados dos parâmetros, exibindo o STD e os intervalos de confiança nos quais o modelo alcançou o melhor desempenho. O leve ajuste realizado não altera significativamente os parâmetros obtidos nesta biblioteca, permitindo que cada modelo seja trabalhado de maneira eficiente.

Para os modelos de gradiente *boosting* e redes neurais artificiais, os hiperparâmetros foram otimizados usando a biblioteca Optuna do Python. Nesse contexto, são empregadas técnicas bayesianas, especificamente o algoritmo TPE, visando uma otimização mais eficiente.

Tabela 9: Parâmetros utilizados nos modelos ARIMA e seus antecessores obtidos pelo “autoARIMA” do Python.

Modelo	Parâmetros Utilizados	Método de Estimação
AR(p)	$p = 7$	AutoARIMA
ARX(p)	$p = 7$	AutoARIMA
MA(q)	$q = 7$	AutoARIMA
ARMA(p, q)	$p = 7, q = 7$	AutoARIMA
ARIMA(p, d, q)	$p = 7, d = 1, q = 7$	AutoARIMA
ARIMAX(p, d, q)	$p = 7, d = 1, q = 7$	AutoARIMA
SARIMA(p, d, q)(P, D, Q)	$p = 7, d = 1, q = 7, P = 2, D = 1, Q = 1, M = 12$	AutoARIMA
SARIMAX(p, d, q)(P, D, Q, M)	$p = 7, d = 1, q = 7, P = 2, D = 1, Q = 1, M = 12$	AutoARIMA

Tabela 10: SARIMAX(7, 0, 0)  $\times$  (2, 1, [1], 12) Results

AutoARIMA	Coef	STD Err	z	P >  z	[0,025	0,975]
Intercept	0,0003	0	1,053	0,292	0	0,001
ar.L1	1,6149	0,011	141,865	0	1,593	1,637
ar.L2	-0,8879	0,021	-42,045	0	-0,929	-0,847
ar.L3	0,3167	0,024	13,033	0	0,269	0,364
ar.L4	-0,1056	0,027	-3,961	0	-0,158	-0,053
ar.L5	-0,1099	0,028	-3,928	0	-0,165	-0,055
ar.L6	0,1431	0,027	5,368	0	0,091	0,195
ar.L7	-0,0673	0,015	-4,583	0	-0,096	-0,039
ar.S.L12	-0,1222	0,016	-7,705	0	-0,153	-0,091
ar.S.L24	0,1692	0,014	12,244	0	0,142	0,196
ma.S.L12	-0,8728	0,012	-74,569	0	-0,896	-0,850
sigma2	0,0157	0	60,022	0	0,015	0,016

Os modelos XGBoost e LightGBM tem como parâmetros e hiperparâmetros mostrado na Tabela 11 a otimização dos parâmetros dos modelos XGBoost, LightGBM, RFR e DTR. Esses modelos, devido à sua semelhança, exibem tempos de desempenho próximos um do outro.

Os modelos de rede neural artificial, como RNN, ANN, CNN, GRU, LSTM e Transformer, obtidos na otimização do Optuna do Python, tiveram seus hiperparâmetros melhorados, conforme exibido na Tabela 12. Esses modelos, por serem modelos de rede neural artificial, são melhores para otimizar do que os outros.

**Previsão e avaliação** foram empregadas três métricas amplamente utilizadas na literatura para avaliar e comparar os modelos ARIMA e os modelos de regressão, conforme detalhado na seção 3.6.

Na análise dos modelos desenvolvidos, observou-se que o modelo DTR obteve o me-

Tabela 11: Hiperparâmetros dos modelos

Modelo	Estimadores	Profund. Máxima	Min. Amos- tras Divisão	Min. Amos- tras por Folha	Máx. Recursos	Taxa de Aprendi- zado
XGB Regressor	503	5	7	2	“sqrt”	0,034
LGBM Regressor	820	10	3	5	“auto”	0,014
Random Forest Regressor	135	10	4	2	None	N/A
Decision Tree Regressor	N/A	229	32	20	None	N/A

Tabela 12: Resumo dos Hiperparâmetros dos Modelos de Redes Neurais

Modelo	Unidades/ Layers	Heads/ Dimen- sões	Tamanho do Batch	Épocas	Dropout/ Lear- ning Rate	Outros Parâmetros
LSTM	128	–	32	77	–	–
GRU	–	–	32	50	–	–
Transformers	–	8 heads, 217; 433	–	50	–	2 camadas
RNN	79	–	16	50	0,0008612	–
CNN	–	–	61	10	0,2799; 0,00052	Kernel: 7, Densas: 1, Verbosidade: 1
ANN	125	–	27	96	0,4135, 0,0004057	Densas: 1, Verbosidade: 0

lhor desempenho, tanto para previsões de curto prazo, durante as horas de pico entre 18h e 21h, quanto para outros períodos. Além disso, os modelos MA, AR, SARIMA, ARIMA, SARIMAX, ARIMAX, ARX, LGBMRegressor, XGBRegressor, RFR, RNN, ANN, CNN, GRU, LSTM, Prophet e Transformer também apresentaram resultados satisfatórios, seguindo uma ordem decrescente de desempenho.

No âmbito das previsões de longo prazo, abrangendo casos de um dia, os modelos ARMA, AR, MA, ARIMA, ARIMAX, ARX, SARIMA, SARIMA, XGBRegressor, RFR, LGBMRegressor, DTR, RNN, ANN, CNN, GRU, LSTM, Prophet e Transformer foram avaliados. Uma observação recorrente foi a superioridade dos modelos que incorporam variáveis exógenas em termos de capacidade de previsão, evidenciada nas Figuras de 22 a 27 e nas Tabelas de 13 a 16, onde os valores menores foram destacados em **negrito** para

facilitar a análise. O modelo RNN destacou-se tanto nos conjuntos de treinamento quanto na avaliação global, consolidando-se como o modelo mais eficaz nas previsões realizadas.

Cada figura, desde a 22 até a 27, ilustra cenários distintos de previsão e comparação entre modelos semelhantes. Os modelos Prophet e RNN, sendo este último a escolha superior, são apresentados de forma isolada. A decisão de não incluir o modelo LR na comparação baseou-se na constância observada em suas previsões a longo prazo.

Ao avaliar os modelos de previsão, tanto nas tabelas quanto nas imagens, o modelo RNN destaca-se como a opção mais eficaz. No caso específico da SANEPAR, esse modelo demonstra um desempenho superior em comparação com os demais modelos de previsão adotados. Três métricas amplamente utilizadas na literatura foram empregadas para avaliar e comparar os modelos ARIMA e os modelos de regressão, conforme detalhado na seção 3.6.

Na análise dos modelos desenvolvidos, verificou-se que o modelo DTR alcançou o melhor desempenho, tanto para previsões de curto prazo, durante as horas de pico entre 18h e 21h, quanto para outros períodos. Adicionalmente, os modelos MA, AR, SARIMA, ARIMA, SARIMAX, ARIMAX, ARX, LGBMRegressor, XGBRegressor, RFR, RNN, ANN, CNN, GRU, LSTM, Prophet e Transformer também apresentaram resultados satisfatórios, seguindo uma ordem decrescente de desempenho.

No contexto das previsões de longo prazo, abrangendo períodos de um dia, os modelos ARMA, AR, MA, ARIMA, ARIMAX, ARX, SARIMA, SARIMA, XGBRegressor, RFR, LGBMRegressor, DTR, RNN, ANN, CNN, GRU, LSTM, Prophet e Transformer foram avaliados. Destacou-se a superioridade dos modelos que incorporam variáveis exógenas em termos de capacidade de previsão, evidenciada nas Figuras de 22 a 27 e nas Tabelas de 13 a 16, onde os valores menores foram destacados em **negrito** e *itálico* para facilitar a análise. O modelo RNN destacou-se tanto nos conjuntos de treinamento quanto na avaliação global, consolidando-se como o modelo mais eficaz nas previsões realizadas.

Cada Figura, desde a 22 até a 27, ilustra cenários distintos de previsão e comparação entre modelos semelhantes. Os modelos Prophet e RNN, sendo este último a escolha superior, são apresentados de forma isolada. A decisão de não incluir o modelo LR na comparação baseou-se na constância observada em suas previsões a longo prazo.

Ao avaliar os modelos de previsão, tanto nas tabelas quanto nas imagens, o modelo RNN destaca-se como a opção mais eficaz. No caso específico da SANEPAR, esse modelo demonstra um desempenho superior em comparação com os demais modelos de previsão adotados.

Figura 22: Comparação dos modelos AR, ARX e MA

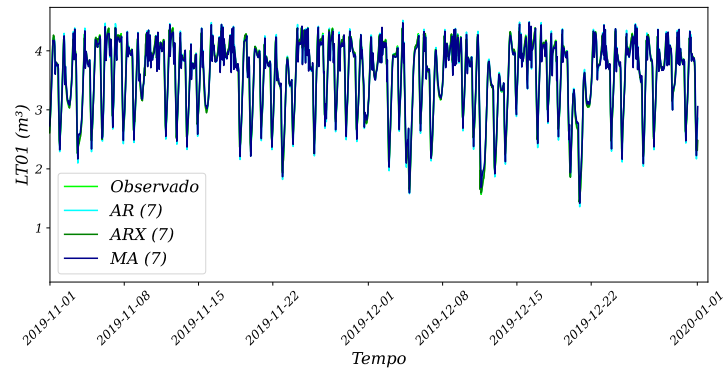


Figura 23: Comparação do modelos ARIMAX, SARIMA e SARIMAX

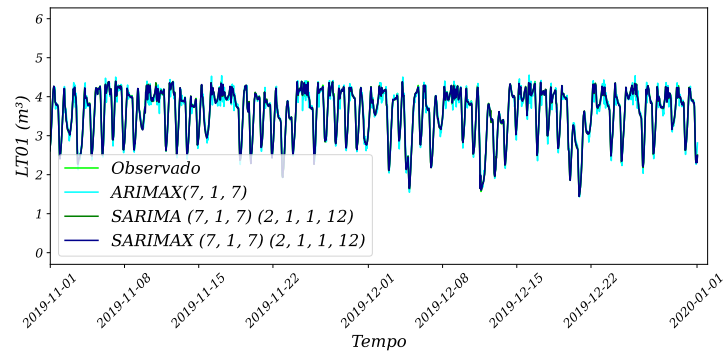


Figura 24: Comparação dos modelos ARMA e ARIMA

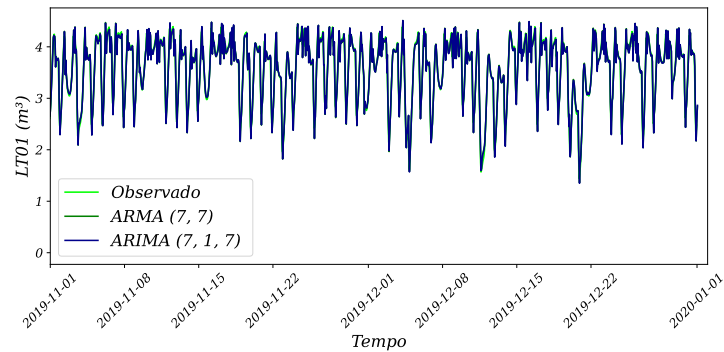


Figura 25: Comparação dos modelos DTR, RFR, XGBoost, Light GBM

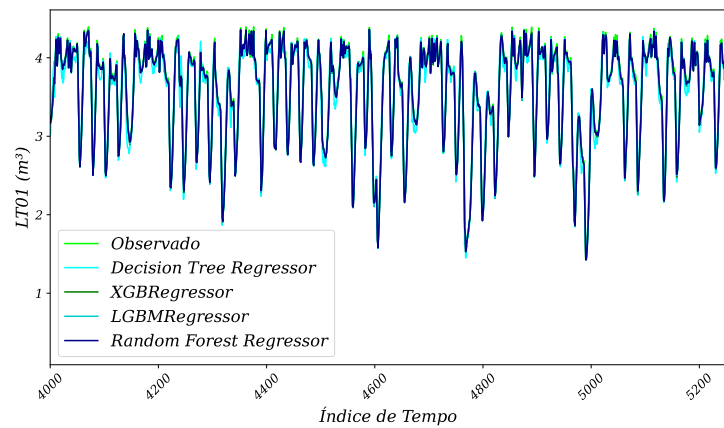




Figura 26: Modelo RNN e os vários horizontes

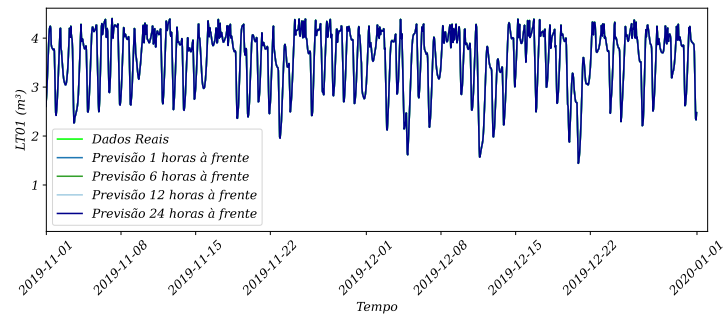


Figura 27: Previsões do modelo Prophet para o reservatório LT01

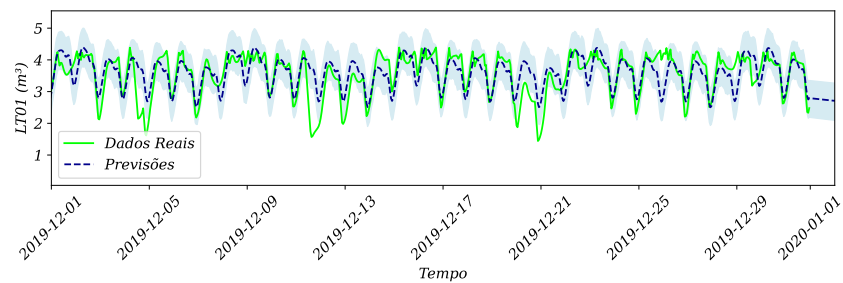


Tabela 13: Comparação dos modelos de previsão com as métricas de desempenho **treino**

		Modelos Treino																		
Horizontes	Métricas	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1 hora à frente	sMAPE	3,91	4,01	4,03	3,91	3,92	3,89	3,82	3,86	8,85	9,31	9,52	9,37	35,4	35,8	9	<b>0,0665</b>	16,8	23	23
	MAE	<b>0,25</b>	<b>0,25</b>	<b>0,26</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,24</b>	<b>0,25</b>	0,36	0,65	0,67	0,65	1,42	1,44	<b>0,2</b>	<i>0,0023</i>	0,55	0,83	0,83
	RRMSE	<b>0,09</b>	<b>0,10</b>	<b>0,09</b>	<b>0,09</b>	<b>0,09</b>	<b>0,09</b>	<b>0,09</b>	<b>0,09</b>	<b>0,21</b>	<b>0,21</b>	<b>0,21</b>	<b>0,21</b>	2,3	0,65	<b>0,2</b>	<i>0,0008</i>	0,31	0,48	0,48
6 horas à frente	sMAPE	9,97	10,1	9,7	9,98	9,97	10	10,1	9,99	6,99	12,4	12,7	9,369	66,2	83,9	20	<i>0,0230</i>	16,7	20,6	20,6
	MAE	0,64	0,65	0,62	0,64	0,64	0,64	0,65	0,64	0,59	0,9	0,93	0,651	3,37	4,95	0,6	<i>0,0007</i>	0,55	0,72	0,72
	RRMSE	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,16</b>	0,32	0,33	<b>0,209</b>	5,02	1,71	0,6	<i>0,0006</i>	0,31	0,45	0,45
12 horas à frente	sMAPE	11,6	11,6	11,3	11,6	11,5	11,7	11,8	11,6	6,99	12,4	12,7	9,369	72	98,6	25	<b>0,0683</b>	16,8	29,2	29,2
	MAE	0,75	0,75	0,74	0,75	0,75	0,76	0,77	0,75	0,59	0,9	0,93	0,651	3,83	6,69	0,8	<i>0,0022</i>	0,55	1,11	1,11
	RRMSE	<b>0,27</b>	<b>0,27</b>	<b>0,26</b>	<b>0,27</b>	<b>0,26</b>	<b>0,27</b>	<b>0,27</b>	<b>0,27</b>	<b>0,16</b>	0,32	0,33	<b>0,209</b>	5,69	2,25	0,9	<i>0,0009</i>	0,31	0,55	0,55
24 horas à frente	sMAPE	6,77	6,85	6,67	6,77	6,69	6,82	6,86	6,82	6,99	12,4	12,7	9,369	74,4	104	26	<b>0,2328</b>	16,8	26,8	26,8
	MAE	0,43	0,44	0,43	0,43	0,43	0,44	0,44	0,43	0,59	0,9	0,93	0,651	4,04	7,5	0,8	<i>0,0079</i>	0,55	1	1
	RRMSE	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,16</b>	0,32	0,33	<b>0,209</b>	5,99	2,5	1	<i>0,0024</i>	0,31	0,52	0,52

Legenda para os Modelos de Previsão: A - AR, B - ARX, C - MA, D - ARMA, E - ARIMA, F - SARIMA, G - ARIMAX, H - SARIMAX, I - Decision Tree Regressor, J - Random Forest Regressor, K - XGBRegressor, L - LGBMRegressor, M - LSTM, N - GRU, O - Prophet, P - RNN, Q - Transformer, R - CNN, S - ANN.

Tabela 14: Comparação dos modelos de previsão com as métricas de desempenho teste

Modelos Teste																				
Horizontes	Métricas	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1 hora à frente	sMAPE	3,93	4,15	3,99	3,93	3,92	3,91	4,16	4,16	7,76	8,46	8,68	8,45	15,6	15,9	9	<b>0,0744</b>	15,1	20,6	20,6
	MAE	<b>0,26</b>	<b>0,27</b>	<b>0,26</b>	<b>0,26</b>	<b>0,26</b>	<b>0,26</b>	<b>0,27</b>	<b>0,27</b>	0,40	0,61	0,63	0,61	0,53	0,54	<b>0,2</b>	<i>0,0024</i>	0,52	0,76	0,76
	RRMSE	<b>0,09</b>	<b>0,10</b>	<b>0,09</b>	<b>0,09</b>	<b>0,09</b>	<b>0,09</b>	<b>0,10</b>	<b>0,10</b>	<b>0,18</b>	<b>0,19</b>	<b>0,20</b>	<b>0,19</b>	1,01	0,33	<b>0,2</b>	<i>0,0029</i>	0,34	0,5	0,5
6 horas à frente	sMAPE	9,74	9,94	9,44	9,74	9,71	9,76	9,96	9,96	6,36	10,7	11	8,446	59,5	72,7	20	<i>0,0308</i>	15,1	17,3	17,3
	MAE	0,65	0,66	0,63	0,65	0,65	0,65	0,66	0,66	0,56	0,8	0,82	0,609	2,97	4,04	0,6	<i>0,0007</i>	0,51	0,62	0,62
	RRMSE	<b>0,23</b>	<b>0,23</b>	<b>0,22</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,14</b>	<b>0,28</b>	<b>0,29</b>	<b>0,191</b>	4,9	1,42	0,6	<i>0,0033</i>	0,34	0,46	0,46
12 horas à frente	sMAPE	11,1	11,2	10,9	11,1	11,1	11,2	11,2	11,3	6,36	10,8	11	8,446	68,4	94,1	25	<b>0,0745</b>	15,1	18,8	18,8
	MAE	0,74	0,75	0,73	0,74	0,74	0,75	0,75	0,75	0,56	0,8	0,82	0,609	3,67	6,31	0,8	<i>0,0023</i>	0,52	0,68	0,68
	RRMSE	<b>0,26</b>	<b>0,26</b>	<b>0,25</b>	<b>0,26</b>	<b>0,26</b>	<b>0,26</b>	<b>0,26</b>	<b>0,26</b>	<b>0,14</b>	<b>0,28</b>	<b>0,29</b>	<b>0,191</b>	6,01	2,11	0,9	<i>0,0032</i>	0,34	0,48	0,48
24 horas à frente	sMAPE	6,15	6,34	6,08	6,15	6,14	6,24	6,36	6,37	6,36	10,7	11	8,446	71,5	102	26	<b>0,2385</b>	15,1	18,1	18,1
	MAE	0,4	0,41	0,4	0,4	0,4	0,41	0,42	0,42	0,56	0,8	0,83	0,609	3,92	7,36	0,8	<i>0,0081</i>	0,52	0,65	0,65
	RRMSE	<b>0,16</b>	<b>0,16</b>	<b>0,16</b>	<b>0,16</b>	<b>0,16</b>	<b>0,16</b>	<b>0,16</b>	<b>0,16</b>	<b>0,16</b>	<b>0,14</b>	<b>0,28</b>	<b>0,29</b>	<b>0,191</b>	6,42	2,43	1	<i>0,0041</i>	0,34	0,47

Legenda para os Modelos de Previsão: A - AR, B - ARX, C - MA, D - ARMA, E - ARIMA, F - SARIMA, G - ARIMAX, H - SARIMAX, I - Decision Tree Regressor, J - Random Forest Regressor, K - XGBRegressor, L - LGBMRegressor, M - LSTM, N - GRU, O - Prophet, P - RNN, Q - Transformer, R - CNN, S - ANN.

Tabela 15: Comparação dos modelos de previsão com as métricas de desempenho **validação**

		Modelos Validação																		
Horizontes	Métricas	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1 hora à frente	sMAPE	4,08	4,28	4,20	4,09	4,10	4,20	4,26	4,29	8,54	10,47	10,66	10,45	29,8	29,4	9	<b>0,0675</b>	17,4	18,3	18,3
	MAE	<b>0,25</b>	<b>0,26</b>	<b>0,26</b>	<b>0,25</b>	<b>0,25</b>	<b>0,26</b>	<b>0,26</b>	<b>0,26</b>	0,32	0,72	0,74	0,72	1,1	1,08	<b>0,2</b>	<i>0,0023</i>	0,56	0,6	0,6
	RRMSE	<b>0,10</b>	<b>0,10</b>	<b>0,10</b>	<b>0,10</b>	<b>0,10</b>	<b>0,10</b>	<b>0,10</b>	<b>0,10</b>	<b>0,20</b>	<b>0,23</b>	<b>0,24</b>	<b>0,23</b>	1,87	0,56	<b>0,2</b>	<i>0,0008</i>	0,33	0,39	0,39
6 horas à frente	sMAPE	10,9	11,1	10,6	10,9	10,9	11	11,1	11,1	6,8	13,9	14,2	10,45	67,9	84	20	<i>0,0229</i>	17,4	20,5	20,5
	MAE	0,68	0,69	0,66	0,68	0,68	0,69	0,69	0,69	0,57	1,01	1,04	0,721	3,39	4,81	0,6	<i>0,0007</i>	0,56	0,69	0,69
	RRMSE	<b>0,25</b>	<b>0,25</b>	<b>0,24</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,16</b>	<b>0,36</b>	<b>0,37</b>	<b>0,233</b>	4,98	1,72	0,6	<i>0,0005</i>	0,34	0,44	0,44
12 horas à frente	sMAPE	12,7	12,8	12,4	12,7	12,6	12,8	12,8	12,8	6,8	13,9	14,2	10,45	74,4	100	25	<b>0,0689</b>	17,4	22,9	22,9
	MAE	0,8	0,81	0,79	0,8	0,8	0,81	0,81	0,81	0,57	1,01	1,04	0,721	3,92	6,71	0,8	<i>0,0022</i>	0,56	0,79	0,79
	RRMSE	<b>0,29</b>	<b>0,29</b>	<b>0,28</b>	<b>0,29</b>	<b>0,29</b>	<b>0,29</b>	<b>0,29</b>	<b>0,29</b>	<b>0,16</b>	<b>0,36</b>	<b>0,37</b>	<b>0,233</b>	5,73	2,33	0,9	<i>0,0008</i>	0,33	0,48	0,48
24 horas à frente	sMAPE	7,3	7,45	7,19	7,3	7,27	7,37	7,43	7,46	6,8	13,9	14,2	10,45	76,9	106	26	<b>0,2342</b>	17,4	22,9	22,9
	MAE	0,46	0,46	0,45	0,46	0,45	0,46	0,46	0,46	0,57	1,01	1,04	0,721	4,14	7,59	0,8	<i>0,0077</i>	0,56	0,79	0,79
	RRMSE	<b>0,18</b>	<b>0,18</b>	<b>0,18</b>	<b>0,18</b>	<b>0,18</b>	<b>0,18</b>	<b>0,18</b>	<b>0,18</b>	<b>0,16</b>	<b>0,36</b>	<b>0,37</b>	<b>0,233</b>	6,04	2,61	1	<i>0,0024</i>	0,33	0,48	0,48

Legenda para os Modelos de Previsão: A - AR, B - ARX, C - MA, D - ARMA, E - ARIMA, F - SARIMA, G - ARIMAX, H - SARIMAX, I - Decision Tree Regressor, J - Random Forest Regressor, K - XGBRegressor, L - LGBMRegressor, M - LSTM, N - GRU, O - Prophet, P - RNN, Q - Transformer, R - CNN, S - ANN.

Tabela 16: Comparação dos modelos de previsão com as métricas de desempenho **inteiro**

Modelos Inteiros																				
Horizontes	Métricas	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1 hora à frente	sMAPE	3,94	4,08	4,05	3,93	3,95	3,91	4,05	4,05	8,51	9,22	9,43	9,244	17,1	17,4	9	<b>0,0690</b>	16,4	22,5	22,5
	MAE	<b>0,25</b>	<b>0,26</b>	<b>0,26</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,26</b>	<b>0,26</b>	0,36	0,65	0,67	0,648	0,57	0,58	<b>0,2</b>	<i>0,0023</i>	0,54	0,81	0,81
	RRMSE	<b>0,09</b>	<b>0,1</b>	<b>0,09</b>	<b>0,09</b>	<b>0,09</b>	<b>0,09</b>	<b>0,1</b>	<b>0,1</b>	<b>0,2</b>	<b>0,21</b>	<b>0,21</b>	<b>0,207</b>	1,01	0,31	<b>0,2</b>	<i>0,0017</i>	0,32	0,49	0,49
6 horas à frente	sMAPE	10	10,2	9,75	10	10	10,1	10,2	10,1	6,77	12,1	12,4	12,07	61,7	74,6	20	<i>0,0253</i>	16,3	20	20
	MAE	0,65	0,66	0,63	0,65	0,65	0,65	0,66	0,65	0,58	0,89	0,91	0,885	3,04	4,08	0,6	<i>0,0007</i>	0,54	0,7	0,7
	RRMSE	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,23</b>	<b>0,16</b>	<b>0,32</b>	<b>0,32</b>	<b>0,316</b>	4,65	1,45	0,6	<i>0,0019</i>	0,33	0,46	0,46
12 horas à frente	sMAPE	11,6	11,7	11,4	11,6	11,6	11,7	11,8	11,7	6,77	12,1	12,4	12,12	70,7	96	25	<b>0,0703</b>	16,4	28,7	28,7
	MAE	0,76	0,76	0,74	0,76	0,75	0,76	0,77	0,76	0,58	0,89	0,91	0,889	3,75	6,38	0,8	<i>0,0023</i>	0,54	1,09	1,09
	RRMSE	<b>0,27</b>	<b>0,27</b>	<b>0,26</b>	<b>0,27</b>	<b>0,27</b>	<b>0,27</b>	<b>0,27</b>	<b>0,27</b>	<b>0,16</b>	<b>0,32</b>	<b>0,32</b>	<b>0,317</b>	5,69	2,16	0,9	<i>0,0019</i>	0,32	0,56	0,56
24 horas à frente	sMAPE	6,66	6,79	6,57	6,66	6,6	6,71	6,82	6,8	6,77	12,1	12,4	12,21	73,8	104	26	<b>0,2347</b>	16,4	26,2	26,2
	MAE	0,43	0,43	0,42	0,43	0,42	0,43	0,44	0,43	0,58	0,89	0,92	0,897	4,01	7,44	0,8	<i>0,0080</i>	0,54	0,98	0,98
	RRMSE	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,17</b>	<b>0,16</b>	<b>0,32</b>	<b>0,32</b>	<b>0,319</b>	6,07	2,49	1	<i>0,0030</i>	0,32	0,53	0,53

Legenda para os Modelos de Previsão: A - AR, B - ARX, C - MA, D - ARMA, E - ARIMA, F - SARIMA, G - ARIMAX, H - SARIMAX, I - Decision Tree Regressor, J - Random Forest Regressor, K - XGBRegressor, L - LGBMRegressor, M - LSTM, N - GRU, O - Prophet, P - RNN, Q - Transformer, R - CNN, S - ANN.

**Teste de significância** o teste de Friedman e o teste de Nemenyi para comparar as classificações médias entre os diversos classificadores. O teste de Nemenyi é uma ferramenta de comparação múltipla frequentemente empregada após a aplicação de testes não paramétricos com três ou mais fatores.

A matriz de comparação entre os classificadores, apresentada na Tabela 17, exibe os valores de comparação múltipla de Nemenyi, onde as entradas evidenciam as diferenças significativas entre os pares de classificadores.

A Tabela 17 apresenta os resultados do teste de Nemenyi, um método utilizado para comparar as classificações médias entre diferentes classificadores após a aplicação de testes não paramétricos com três ou mais fatores. Cada célula da tabela mostra os valores de comparação múltipla de Nemenyi, que indicam as diferenças significativas entre os pares de classificadores. O valor na interseção da linha  $i$  e da coluna  $j$  representa a diferença significativa entre os classificadores  $i$  e  $j$ .

Tabela 17: Teste Nemenyi

Nemenyi	B1	B2	B3	LT01	FT01	FT02	FT03	PT01	PT02
<b>B1</b>	1	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
<b>B2</b>	0,001	1	0,001	0,001	0,001	0,001	0,001	0,001	0,001
<b>B3</b>	0,001	0,001	1	0,001	0,001	0,001	0,001	0,001	0,001
<b>LT01</b>	0,001	0,001	0,001	1	0,001	0,001	0,001	0,001	0,001
<b>FT01</b>	0,001	0,001	0,001	0,001	1	0,001	0,131	0,001	0,001
<b>FT02</b>	0,001	0,001	0,001	0,001	0,001	1	0,001	0,001	0,001
<b>FT03</b>	0,001	0,001	0,001	0,001	0,131	0,001	1	0,001	0,001
<b>PT01</b>	0,001	0,001	0,001	0,001	0,001	0,001	0,001	1	0,001
<b>PT02</b>	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	1

No contexto do estudo, os resultados da análise comparativa revelaram diferenças estatisticamente significativas entre vários pares de classificadores, como indicado pelas entradas da tabela. Isso sugere que pelo menos um modelo é considerado estatisticamente superior aos demais, com base nas comparações realizadas.

O valor crítico CD foi utilizado para determinar se dois classificadores eram significativamente diferentes entre si. Esse valor é calculado com base no valor crítico obtido da Tabela 17 de teste de Nemenyi, o número de classificadores e o número total de amostras. O valor CD é uma métrica que auxilia na interpretação das diferenças entre os classificadores, ajudando a identificar quais pares de classificadores apresentam diferenças estatisticamente significativas.

Os resultados da pesquisa indicaram a existência de evidências estatísticas que sugerem a superioridade de pelo menos um modelo em relação aos demais. Além disso, a análise de comparação significativa entre os modelos revelou pares de classificadores que

apresentam diferenças estatisticamente significativas em seus desempenhos. Essas informações são valiosas para a seleção e avaliação dos modelos de classificação, permitindo uma compreensão mais precisa das diferenças de desempenho entre os classificadores avaliados no estudo. Na Tabela 18 é mostrado como cada modelo foi comparado entre si em 24 passos à frente.

Tabela 18: Métricas de avaliação dos modelos

Modelo	sMAPE	MAE	RRMSE
Prophet	25,67	0,844	0,975
Transformer	16,39	0,544	0,324
ANN	26,22	0,980	0,531
CNN	26,22	0,980	0,531
<b>RNN</b>	<b>0,235</b>	<b>0,008</b>	<b>0,003</b>
LSTM	73,75	4,010	6,068
GRU	103,57	7,443	2,485
AR	6,66	0,428	0,169
ARX	6,79	0,434	0,173
MA	6,57	0,423	0,166
ARMA	6,66	0,428	0,169
ARIMA	6,60	0,424	0,167
SARIMA	6,71	0,432	0,170
ARIMAX	6,82	0,436	0,173
SARIMAX	6,80	0,435	0,173
DTR	6,77	0,577	0,158
RFR	12,09	0,886	0,316
XGBRegressor	12,41	0,916	0,323
LGBMRegressor	12,21	0,897	0,319

**Modelo com menor valor em cada métrica:** Primeiramente, os diversos modelos de previsão de séries temporais foram avaliados para um horizonte de previsão de um dia. Para cada métrica (sMAPE, MAE e RRMSE), identificou-se o modelo que apresentou o menor valor. A métrica sMAPE apontou que o modelo **RNN** obteve o menor valor. Quanto à métrica MAE, novamente o modelo **RNN** demonstrou o menor valor. A métrica RRMSE também indicou que o modelo **RNN** teve o menor valor.

**Evidências estatísticas de que pelo menos um modelo é superior:** Para validar estatisticamente as diferenças entre os modelos, foi realizado um teste estatístico denominado **Teste de Friedman**. Esse teste avalia o desempenho dos modelos em todas as métricas simultaneamente. O resultado do teste de Friedman revelou **evidências estatísticas** que pelo menos um dos modelos apresenta superioridade estatística em relação aos demais, considerando um nível de significância de 0.05.

**Comparação significativa entre modelos - Teste de Nemenyi:** A fim de determinar quais modelos apresentam diferenças estatisticamente significativas entre si, foi conduzido o **teste de comparações múltiplas de Nemenyi**. Esse teste avalia todos os

pares possíveis de modelos e identifica quais deles possuem diferenças estatisticamente significativas. Os resultados indicaram **diferenças estatisticamente significativas** entre vários pares de modelos. Especificamente:

O modelo **RNN** apresentou diferenças significativas em relação aos modelos **LSTM** e **GRU**. O modelo **LSTM** apresentou diferenças significativas em relação ao modelo **RNN**. O modelo **GRU** exibiu diferenças significativas em relação ao modelo **RNN**. Com base na análise estatística de Friedman e no teste de comparações múltiplas de Nemenyi, conclui-se que o modelo **RNN** apresenta o melhor desempenho geral em relação às métricas consideradas (**sMAPE**, **MAE** e **RRMSE**) para um horizonte de previsão de um dia, utilizando os dados completos.

#### 4.1.1 Comparação dos Modelos

Com o objetivo de obter uma análise mais aprofundada do desempenho de cada modelo, foi realizada uma comparação por meio de um gráfico de violino e de barra. Dessa forma, pôde-se observar qual dos modelos apresentava o melhor desempenho.

Ao examinar os modelos representados nas Figuras 28 e 29, identifiquei os modelos que se destacam em relação à natureza dos dados. Na Figura 31, que compara os modelos ARIMA e XGBoost com outros, torna-se evidente que os modelos ARIMA como AR, ARX, MA, ARMA, ARIMAX e SARIMAX demonstram um desempenho sólido. Além disso, os modelos baseados em gradientes e regressão, como o XGBoost, exibem resultados comparáveis, beneficiando-se da otimização por meio do Optuna, uma abordagem de bayesiana usando o método TPE.

Na Figura 30, que contrasta as redes neurais com o modelo Prophet, é importante destacar que os modelos de redes neurais, incluindo RNN, LSTM, GRU, ANN, CNN e Transformer, foram avaliados em conjunto com o modelo Prophet. A análise estatística também demonstrou que o modelo RNN se sobressai como o vencedor entre as métricas avaliadas. Essa conclusão é respaldada pelas evidências de que pelo menos um modelo é superior aos demais. Os modelos com valores de p-valor abaixo de 0,05 foram realçados em *itálico* para enfatizar sua significância.

Na Figura 29, é feita uma comparação entre os modelos de gradiente e regressor. Esses modelos, por serem mais robustos e utilizar técnicas de otimização mais avançadas, mostram-se superiores aos modelos comparados. O modelo XGBoost, em particular, é identificado como superior em relação aos outros modelos na análise.

Na Figura 30, nota-se que todos os modelos trabalhados aqui, exceto o modelo LR, foram comparados em relação às métricas de desempenho. Mesmo sendo muito robustos, esses modelos não conseguiram obter um resultado tão bom quanto o RNN.



Figura 28: Comparação dos modelos ARIMA

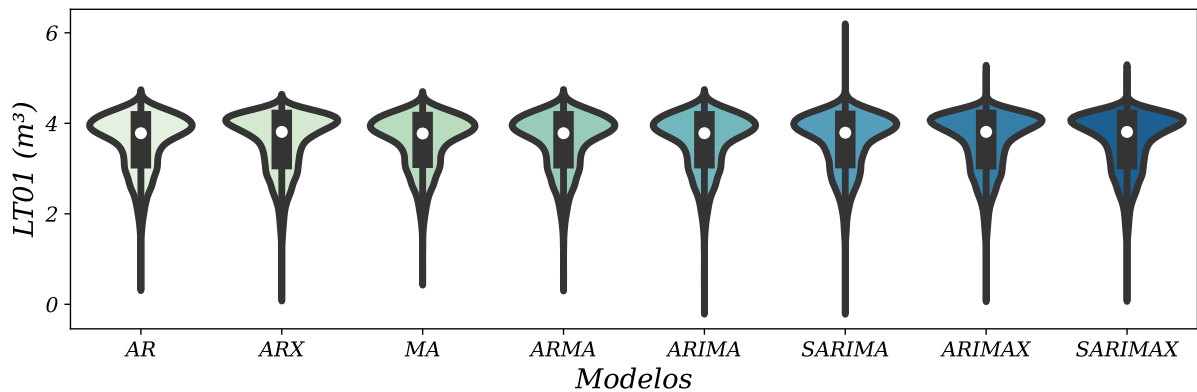


Figura 29: Comparação de modelos de regressão

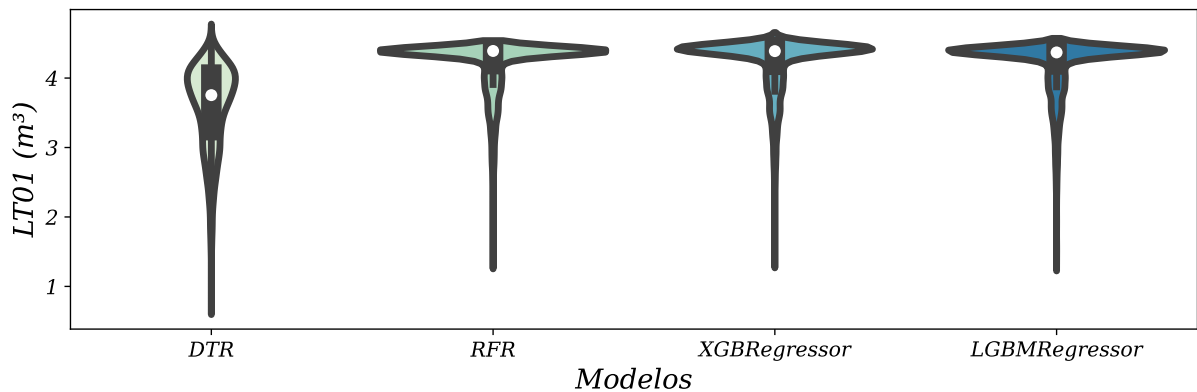
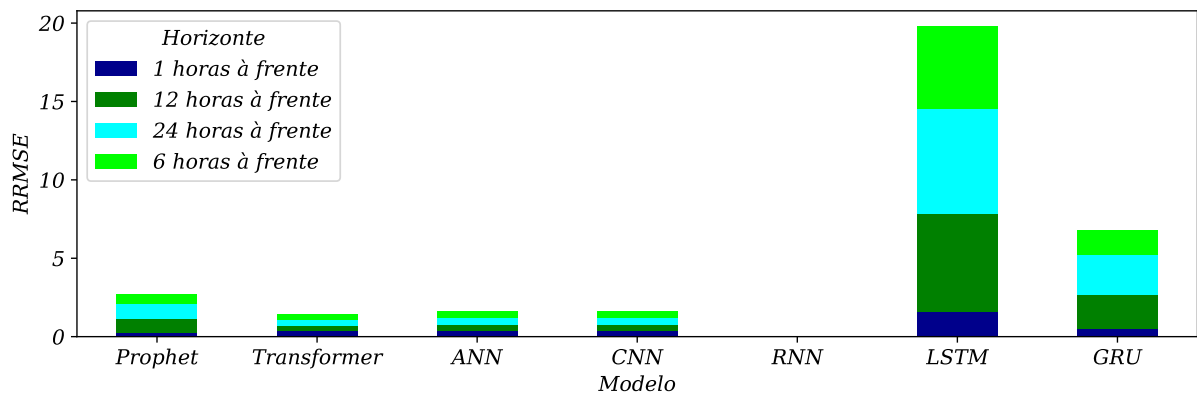
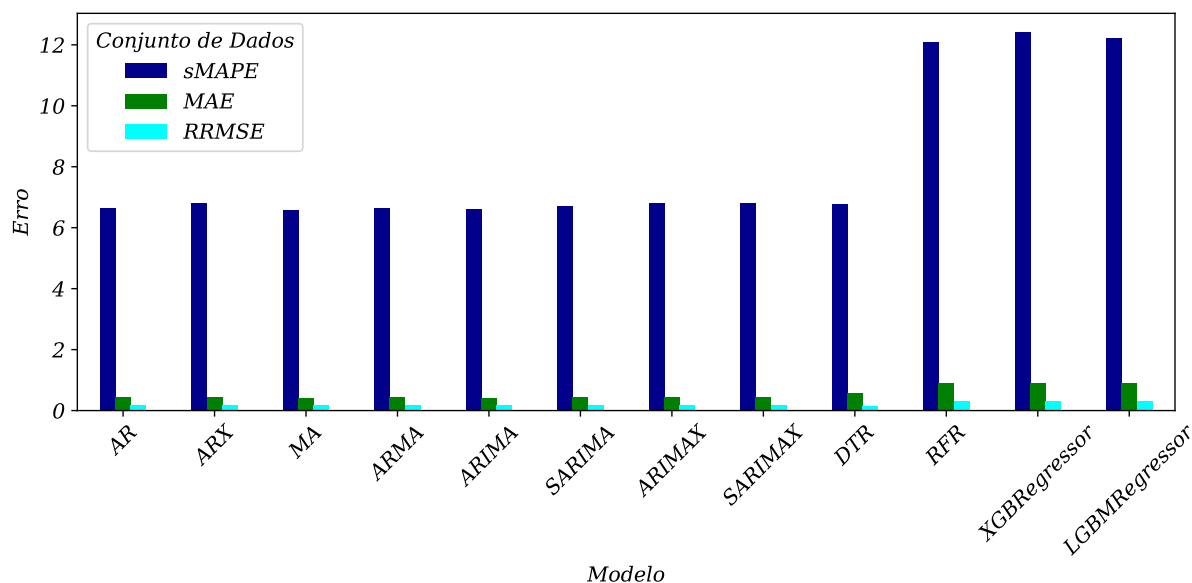


Figura 30: Comparação dos modelos na métrica RRMSE



A avaliação da eficácia dos modelos ARIMA em previsões de longo prazo emprega o teste de Ljung-Box, conforme detalhado nas Tabelas 19a a 19d ilustram a acurácia dos modelos ARIMA ao longo do tempo, com valores menores sendo destacados em **negrito** para facilitar a interpretação. Modelos como ARX, ARIMAX e SARIMAX, que incorporam variáveis exógenas, demonstram um desempenho superior nesse contexto. Esses modelos não lineares apresentam uma capacidade de previsão robusta em horizontes

Figura 31: Comparação dos modelos nas métricas sMAPE, MAE e RRMSE



temporais mais longos, diferenciando-se positivamente dos outros modelos ARIMA. Na Figura 28, são selecionados os modelos ARIMA e seus antecessores. Esses modelos têm suas limitações, tanto para horizontes de previsão de curto prazo quanto para horizontes de longo prazo. Nessa comparação no gráfico de violino, são combinados vários outros gráficos em um só, como o gráfico de barras e o *boxplot*. Esse gráfico pode fornecer várias informações, mas o objetivo aqui é identificar apenas o melhor modelo entre os modelos ARIMA.

Como essa série não apresentou uma estacionariedade bem definida e os dados não a tornaram estacionária, os modelos que não têm sazonalidade mostraram-se superiores, tais como AR, MA, ARX, ARMA, ARIMA e ARIMAX. O modelo ARIMAX demonstrou ser bastante robusto para este caso, mas mesmo assim, modelos mais básicos como AR e MA ainda apresentaram resultados melhores.

Tabela 19: Comparação dos modelos Ljung Box: Modelos ARIMA com defasagem de 10 para previsão de longo prazo na demanda de água

(a) Treinamento			(b) Teste		
Ljung Box	Estatística de Teste	Valor De p	Ljung Box	Estatística de Teste	Valor De p
ARX	59,677	<b>0</b>	ARX	47,177	<b>0</b>
AR	52,312	<b>0,265</b>	AR	49,965	0,444
MA	57,268	<b>0</b>	MA	77,884	<b>01</b>
ARMA	<b>6,945</b>	<b>0,731</b>	ARMA	<b>1,545</b>	0,999
ARIMA	16,724	0,081	ARIMA	<b>5,354</b>	0,866
SARIMA	48,505	<b>0</b>	SARIMA	24,663	<b>0,006</b>
ARIMAX	89,931	<b>0</b>	ARIMAX	36,738	<b>0</b>
SARIMAX	29,093	<b>0</b>	SARIMAX	21,236	<b>0,020</b>
(c) Validação			(d) Inteiro		
Ljung Box	Estatística de Teste	Valor De p	Ljung Box	Estatística de Teste	Valor De p
ARX	<b>5,108</b>	0,884	ARX	48,870	<b>0</b>
AR	4,360	0,930	AR	49,432	<b>0,035</b>
MA	46,252	<b>0</b>	MA	57,629	<b>0</b>
ARMA	<b>7,515</b>	0,676	ARMA	<b>10,053</b>	<b>0,436</b>
ARIMA	<b>7,738</b>	0,654	ARIMA	<b>10,053</b>	<b>0,436</b>
SARIMA	28,998	<b>0,001</b>	SARIMA	<b>10,053</b>	<b>0,436</b>
ARIMAX	<b>6,115</b>	<b>0</b>	ARIMAX	70,458	<b>0</b>
SARIMAX	<b>4,443</b>	0,925	SARIMAX	<b>2,897</b>	<b>0</b>

## 4.2 Aplicação

A previsão da demanda d'água é uma preocupação fundamental para muitas organizações e autoridades responsáveis pelo abastecimento de água. Neste estudo de caso, explorou-se como a análise de séries temporais pode ser aplicada para prever a demanda d'água ao longo do tempo.

### 4.2.1 Estudo de Caso 1

Confirmou-se que a ativação das bombas de sucção durante o período de 18h às 21h resulta em um maior custo energético para a SANEPAR. Portanto, é recomendado evitar o acionamento das bombas durante esse período, utilizando estratégias de armazenamento e gerenciamento eficientes.

Com base nos resultados obtidos, conclui-se que as pressões atuais das variáveis **PRESSÃO DE SUCÇÃO - PT01** e **PRESSÃO DE RECALQUE - PT02** são adequadas para atender à demanda diária. O percentil 10 das pressões de sucção (3,48 mca) indica que apenas 10% dos valores estão abaixo desse limite, o que sugere que a pressão de sucção geralmente se mantém em níveis adequados para o funcionamento adequado do sistema. Da mesma forma, o percentil 90 das pressões de recalque (24,02 mca) indica que apenas 10% dos valores estão acima desse limite, evidenciando que a pressão de recalque também se mantém dentro dos padrões necessários para atender à demanda diária.

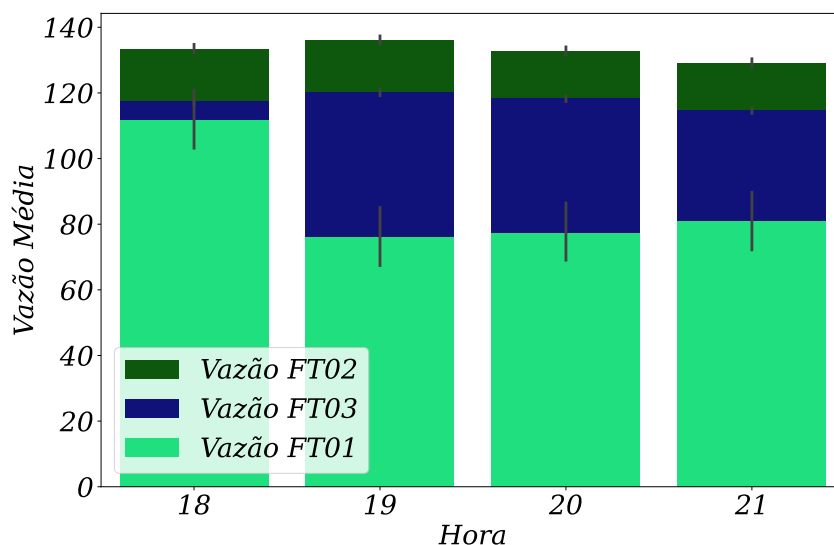
Com base na frequência de funcionamento das bombas e na demanda durante o horário de pico, determinou-se que é necessário manter um volume máximo d'água no reservatório, correspondente a 5285,90 litros, para evitar o acionamento das bombas nesse período.

### 4.2.2 Estudo de Caso 2

Ao analisar os dados dos últimos 3 anos do Bairro Alto, identificou-se a presença de tendências sazonais e padrões de consumo de água. Essas informações são valiosas para compreender os padrões de demanda e planejar o abastecimento de forma eficiente.

O gráfico de barras apresentado na Figura 32 mostra a demanda média das variáveis de fluxo (Vazão de Entrada – FT01, Vazão de Gravidade – FT02 e Vazão de Recalque – FT03) durante o intervalo das 18h às 21h. Cada barra representa a média da demanda para cada variável em um horário específico dentro desse intervalo. A altura de cada barra indica a magnitude da demanda média para a respectiva variável. Essa visualização permite que sejam identificados os horários em que as variáveis de fluxo apresentaram maior demanda, o que é útil para o planejamento e gerenciamento adequado do sistema.

Figura 32: Demanda média das variáveis de fluxo



A Tabela 20 apresenta os resultados para as três variáveis estudadas: vazão de entrada – FT01, vazão de gravidade – FT02 e vazão de recalque – FT03. Os resultados destacam os horários específicos em que cada variável apresentou maior demanda dentro do intervalo das 18h às 21h, fornecendo importantes para o planejamento e gerenciamento adequado do sistema. A Tabela 20 resume essas informações.

Tabela 20: Demanda de água

Vazões	Horário de Maior Demanda	Demanda
Entrada – FT01	2020/10/08 21:00:00	383,87 $m^3/h$
Gravidade – FT02	2020/10/20 18:00:00	326,17 $m^3/h$
Recalque – FT03	2020/11/26 19:00:00	194,35 $m^3/h$

Durante as horas de pico, é necessário que o nível do reservatório esteja mantido dentro na média de  $3.9005 m^3$  para evitar o acionamento das bombas. Manter o nível do reservatório dentro dessa faixa permitirá que o sistema opere de forma eficiente, atendendo à demanda de água sem a necessidade de acionar as bombas.

É importante destacar que a vazão de recalque exerce um impacto significativo no nível do reservatório em comparação com as outras vazões. Essa diferença se deve ao fato de que a vazão de recalque está diretamente relacionada à injeção de água no reservatório por meio da bomba localizada próxima à sua base. Em contraste, as demais vazões possuem alguns valores ausentes, o que limita sua influência na análise geral do sistema.

## 5 Conclusões

Na dissertação realizada, foi conduzido um estudo abrangente sobre a previsão da demanda d'água por meio da análise de séries temporais. Através da análise exploratória dos dados e da aplicação da decomposição STL, foram identificados padrões sazonais e tendências na demanda de água. Ao longo do estudo, foram empregados os modelos ARIMA, DTR e XGBoost para validar o estudo de caso da SANEPAR.

No segundo estudo de caso, que tratou do impacto do acionamento das bombas durante o horário de pico em uma rede de distribuição de água, a análise se concentrou nos horários em que as pessoas estão em casa e consomem mais água. O objetivo geral do trabalho foi desenvolver modelos de previsão de séries temporais específicos para o abastecimento de água. Embora a literatura aborde diversos modelos de séries temporais, apenas alguns deles são aplicados ao contexto de abastecimento d'água. Nesse sentido, foram comparados 19 tipos diferentes de modelos.

Com base nos resultados obtidos, conclui-se que a abordagem de séries temporais é uma ferramenta eficaz para prever a demanda futura d'água. Os resultados também indicaram a importância de considerar as flutuações sazonais e as diferentes partes do dia ao determinar a vazão e o volume mínimo de reserva no reservatório. Apesar dos progressos obtidos nesta pesquisa, é crucial destacar algumas limitações a serem consideradas. Primeiramente, a análise fundamentou-se em dados históricos de demanda d'água de uma única região, especificamente o maior bairro de Curitiba. O estudo não considerou fatores externos, como mudanças climáticas ou eventos imprevistos, que poderiam impactar a demanda d'água.

### 5.1 Propostas Futuras

Apesar dos resultados promissores evidenciados por esta pesquisa, é essencial que se reconheçam suas limitações e que se instigue a exploração de novos horizontes em pesquisas subsequentes. Uma análise mais profunda e abrangente pode ser realizada, investigando modelos de redes neurais mais avançados. Além disso, a implementação de técnicas de otimização matemática mais refinadas, como o uso do método CMAES (do inglês *Covariance Matrix Adaptation Evolution Strategy*), pode ser considerada. Seria prudente incluir cuidadosamente variáveis exógenas em todos os modelos pertinentes, como o uso de variáveis climáticas e dados de precipitação do tempo. Implementa modelos que utilizam lógica *fuzzy* para aprimorar a previsão do tanque. Usa essa previsão juntamente com modelos existentes na literatura, como a otimização BOA, que não foi abordada neste contexto.

## Referências

- AGIAKLOGLOU, C.; NEWBOLD, P. Empirical evidence on dickey-fuller-type tests. **Journal of Time Series Analysis**, v. 13, n. 6, p. 471–483, 1992.
- AHMAD, T. et al. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. **ENERGY AND BUILDINGS**, v. 165, p. 301–320, 2018. ISSN 0378-7788.
- AL-SHABI, M. Q. Machine learning: Algorithms, real-world applications and research directions. **SN Computer Science**, Springer, v. 2, n. 3, p. 1–12, 2021.
- ALI, M. et al. Ensemble robust local mean decomposition integrated with random forest for short-term significant wave height forecasting. **Renewable Energy**, v. 205, p. 731–746, 2023.
- ASEERI, A. O. Effective rnn-based forecasting methodology design for improving short-term power load forecasts: Application to large-scale power-grid time series. **Journal of Computational Science**, v. 68, p. 101984, 2023.
- BABU, C. N.; REDDY, B. E. A moving-average filter based hybrid arima-ann model for forecasting time series data. **Applied Soft Computing**, v. 23, p. 27–38, 2014.
- BANDARA, K.; HYNDMAN, R. J.; BERGMEIR, C. Mstl: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. **arXiv preprint arXiv:2107.13462**, 2021. Disponível em: <<https://arxiv.org/abs/2107.13462>>.
- BARNES, J. L.; KRISHEN, A. S.; HU, H. fen. Public tap water perceptions and potable reuse acceptance: A cognitive dissonance theoretical understanding. **Journal of Cleaner Production**, v. 429, p. 139587, 2023.
- BERGLUND, E. Z.; SKARBEEK, M.; KANTA, L. A sociotechnical framework to characterize tipping points in water supply systems. **Sustainable Cities and Society**, v. 97, p. 104739, 2023.
- BERGMEIR, C.; HYNDMAN, R.; KOO, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. **Computational Statistics and Data Analysis**, v. 120, p. 70–83, 2018.
- BHANGU, K.; SANDHU, J.; SAPRA, L. Time series analysis of covid-19 cases. **World Journal of Engineering**, v. 19, p. 40–48, 2022. ISSN 17085284.
- BOROOJENI, K. et al. A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon. **Electric Power Systems Research**, v. 142, p. 58–73, 2017.
- BOX, G. E. P.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American Statistical Association**, v. 65, n. 332, p. 1509–1526, 1970. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/01621459.1970.10481180>>.
- BROCKWELL, P. J.; DAVIS, R. A. **Introduction to Time Series and Forecasting**. [S.l.]: Springer New York, 2002. ISBN 978-0-387-95351-9.

BUEECHI, E. et al. Crop yield anomaly forecasting in the pannonian basin using gradient boosting and its performance in years of severe drought. **Agricultural and Forest Meteorology**, v. 340, p. 109596, 2023.

BUYUKSAHIN, U.; ERTEKIN. Improving forecasting accuracy of time series data using a new arima-ann hybrid method and empirical mode decomposition. **Neurocomputing**, v. 361, p. 151–163, 2019. ISSN 09252312.

BUYUKSAHIN, U.; ERTEKIN. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. **Neurocomputing**, v. 361, p. 151–163, 2019.

CANDELIERI, A. et al. Tuning hyperparameters of a svm-based water demand forecasting system through parallel global optimization. **Computers & Operations Research**, v. 106, p. 202–209, 2019.

Carvalho Jr., J. G.; Costa Jr., C. T. Non-iterative procedure incorporated into the fuzzy identification on a hybrid method of functional randomization for time series forecasting models. **Applied Soft Computing Journal**, Elsevier Ltd, Postgraduate Program in Electrical Engineering, Federal University of Pará, Brazil, v. 80, p. 226–242, 2019. ISSN 15684946 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85064441622&doi=10.1016%2Fj.asoc.2019.03.059&partnerID=40&md5=84d0bd291cc451de280dc9ed77524736>>.

Cesar de Lima Nogueira, S. et al. Prediction of the nox and co2 emissions from an experimental dual fuel engine using optimized random forest combined with feature engineering. **Energy**, v. 280, p. 128066, 2023.

CHEN, M. et al. The predictive management in campus heating system based on deep reinforcement learning and probabilistic heat demands forecasting. **Applied Energy**, v. 350, p. 121710, 2023.

CHEN, Y. Y. et al. Applications of Recurrent Neural Networks in Environmental Factor Forecasting: A Review. **NEURAL COMPUTATION**, v. 30, n. 11, p. 2855–2881, 2018. ISSN 0899-7667.

CHOU, J.-S.; NGUYEN, T.-K. Forward Forecast of Stock Price Using Sliding-Window Metaheuristic-Optimized Machine-Learning Regression. **IEEE Transactions on Industrial Informatics**, v. 14, n. 7, p. 3132–3142, 2018.

CHOU, J.-S.; TRAN, D.-S. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. **Energy**, v. 165, p. 709–726, 2018.

COELHO, I. et al. A GPU deep learning metaheuristic based model for time series forecasting. **Applied Energy**, v. 201, p. 412–418, 2017.

DAVIDSON, J. **Econometric Theory**. Wiley-Blackwell, 2000. ISBN 978-0-631-21584-4. Disponível em: <<https://www.wiley.com/en-cn/Econometric+Theory-p-9780631215844>>.



- DAVIES, N.; NEWBOLD, P. Some power studies of a portmanteau test of time series model specification. **Biometrika**, v. 66, n. 1, p. 153–155, 04 1979. ISSN 0006-3444. Disponível em: <<https://doi.org/10.1093/biomet/66.1.153>>.
- DONG, J. et al. Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with xgboost in different regions of china. **Engineering Applications of Artificial Intelligence**, v. 117, p. 105579, 2023.
- dos Santos Coelho, L.; Hultmann Ayala, H. V.; Cocco Mariani, V. Co and nox emissions prediction in gas turbine using a novel modeling pipeline based on the combination of deep forest regressor and feature engineering. **Fuel**, v. 355, p. 129366, 2024.
- DU, S. et al. Multivariate time series forecasting via attention-based encoder–decoder framework. **Neurocomputing**, v. 388, p. 269–279, 2020.
- FOUILLOY, A. et al. Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability. **Energy**, v. 165, p. 620–629, 2018. ISSN 03605442.
- GIFFORD, M.; BAYRAK, T. A predictive analytics model for forecasting outcomes in the national football league games using decision tree and logistic regression. **Decision Analytics Journal**, v. 8, p. 100296, 2023.
- GOLYANDINA, N. Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing. **WILEY INTERDISCIPLINARY REVIEWS-COMPUTATIONAL STATISTICS**, v. 12, n. 4, 2020. ISSN 1939-0068.
- GRAFF, M. et al. Time series forecasting with genetic programming. **Natural Computing**, v. 16, n. 1, p. 165–174, 2017.
- GRÜBLER, M. Entendendo o funcionamento de uma rede neural artificial. **Medium**, Jun 2018. Disponível em: <<https://medium.com/brasil-ai/entendendo-o-funcionamento-de-uma-rede-neural-artificial-4463fcf44dd0>>.
- GUO, H.; PEDRYCZ, W.; LIU, X. Hidden markov models based approaches to long-term prediction for granular time series. **IEEE Transactions on Fuzzy Systems**, v. 26, p. 2807–2817, 2018. ISSN 10636706.
- GUSTIN, M.; MCLEOD, R.; LOMAS, K. Forecasting indoor temperatures during heatwaves using time series models. **Building and Environment**, v. 143, p. 727–739, 2018. ISSN 03601323.
- GÉRON, A. **Hands-On Machine Learning with Scikit-Learn and TensorFlow**. 2nd. ed. O'Reilly Media, 2017. Disponível em: <<https://www.oreilly.com/library/view/hands-on-machine-learning/9781491962282/>>.
- HAO, J. et al. A bi-level ensemble learning approach to complex time series forecasting: Taking exchange rates as an example. **Journal of Forecasting**, v. 42, p. 1385–1406, 2023. ISSN 02776693.

- JI, S.; AHN, K.-H. Temperature change-informed future multisite streamflow generation to support water supply vulnerability assessments under climate change. **Journal of Hydrology**, v. 624, p. 129928, 2023.
- JORDAN, I. D.; SOKÓŁ, P. A.; PARK, I. M. Gated recurrent units viewed through the lens of continuous time dynamical systems. **Frontiers in Computational Neuroscience**, v. 15, p. 678158, 2021. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fncom.2021.678158/full>>.
- KHAN, M. et al. Cyclic gate recurrent neural networks for time series classification with missing values. **Neural Processing Letters**, v. 55, n. 1, p. 1–32, 2022. Disponível em: <<https://link.springer.com/article/10.1007/s11063-022-10950-2>>.
- KHEIRI, K.; KARIMI, H. Sentimentgpt: Exploiting gpt for advanced sentiment analysis and its departure from current machine learning. **arXiv preprint arXiv:2307.10234**, 2023.
- KOEBELE, E. A. et al. A role for water markets in enhancing water security in the western United States?: Lessons from the Walker River Basin. **Water Policy**, v. 24, n. 11, p. 1757–1771, 2022.
- KORSTANJE, J. **Advanced Forecasting with Python**. [S.l.]: Springer, 2021.
- KOTHONA, D. et al. Deep learning forecasting tool facilitating the participation of photovoltaic systems into day-ahead and intra-day electricity markets. **Sustainable Energy, Grids and Networks**, v. 36, p. 101149, 2023.
- KULSHRESHTHA, S.; VIJAYALAKSHMI, A. An ARIMA-LSTM hybrid model for stock market prediction using live data. **Journal of Engineering Science and Technology Review**, v. 13, n. 4, p. 117–123, 2020.
- KULSHRESHTHA, S.; VIJAYALAKSHMI, A. An arima-lstm hybrid model for stock market prediction using live data. **Journal of Engineering Science and Technology Review**, v. 13, p. 117–123, 2020. ISSN 17919320.
- KUMAR, G.; JAIN, S.; SINGH, U. P. Stock Market Forecasting Using Computational Intelligence: A Survey. **ARCHIVES OF COMPUTATIONAL METHODS IN ENGINEERING**, v. 28, n. 3, p. 1069–1101, 2021. ISSN 1134-3060.
- KUSHWAH, A.; WADHVANI, R. Trend triplet based data clustering for eliminating nonlinear trend components of wind time series to improve the performance of statistical forecasting models. **Multimedia Tools and Applications**, v. 81, p. 33927–33953, 2022. ISSN 13807501.
- LARA-BENITEZ, P.; CARRANZA-GARCIA, M.; RIQUELME, J. C. An Experimental Review on Deep Learning Architectures for Time Series Forecasting. **INTERNATIONAL JOURNAL OF NEURAL SYSTEMS**, v. 31, n. 3, 2021. ISSN 0129-0657.
- LI, A. W.; BASTOS, G. S. Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review. **IEEE ACCESS**, v. 8, p. 185232–185242, 2020. ISSN 2169-3536.

LI, P. et al. Dynamic similar sub-series selection method for time series forecasting. **IEEE Access**, v. 6, p. 32532–32542, 2018. ISSN 21693536.

LIU, H.; CHEN, C. Data processing strategies in wind energy forecasting models and applications: A comprehensive review. **APPLIED ENERGY**, v. 249, p. 392–408, 2019. ISSN 0306-2619.

LIU, H. et al. Dual-stage time series analysis on multifeature adaptive frequency domain modeling. **International Journal of Intelligent Systems**, v. 37, p. 7837–7856, 2022. ISSN 08848173.

LIU, J.; FU, Y. Renewable energy forecasting: A self-supervised learning-based transformer variant. **Energy**, v. 284, p. 128730, 2023.

LIU, J.; XU, Y. T-friedman test: A new statistical test for multiple comparison with an adjustable conservativeness measure. **International Journal of Computational Intelligence Systems**, v. 15, p. 29–43, 2022. Disponível em: <<https://doi.org/10.1007/s44196-022-00083-8>>.

LIU, Z. Y. et al. Forecast Methods for Time Series Data: A Survey. **IEEE ACCESS**, v. 9, p. 91896–91912, 2021. ISSN 2169-3536 J9 - IEEE ACCESS JI - IEEE Access.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297–303, 08 1978. ISSN 0006-3444. Disponível em: <<https://doi.org/10.1093/biomet/65.2.297>>.

LUCAS, P. d. O. e. **Previsão de Séries Temporais de Evapotranspiração de Referência com Redes Neurais Convolucionais**. Tese (Doutorado) — Universidade Federal de Minas Gerais, 2019. Disponível em: <<https://www.ppgee.ufmg.br/defesas/1748M.PDF>>.

MARTINOVIĆ, M.; HUNJET, A.; TURCIN, I. Time series forecasting of the austrian traded index (Atx) using artificial neural network model. **Tehnicki Vjesnik**, v. 27, n. 6, p. 2053–2061, 2020.

MARTINS, L. E. G.; GORSCHKE, T. Requirements engineering for safety-critical systems: A systematic literature review. **Information and Software Technology**, v. 75, p. 71–89, 2016. ISSN 0950-5849. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950584916300568>>.

MIGLIATO, A. L. T.; PONTI, M. A. **Detecção de Outliers em Dados não Vistos de Séries Temporais por meio de Erros de Predição com SARIMA e Redes Neurais Recorrentes LSTM e GRU**. Dissertação (Mestrado) — Universidade de São Paulo, 2021.

MOHAN, S. et al. Predicting the impact of the third wave of covid-19 in india using hybrid statistical machine learning models: A time series forecasting and sentiment analysis approach. **Computers in Biology and Medicine**, v. 144, 2022. ISSN 00104825.

MOON, J. et al. Temporal data classification and forecasting using a memristor-based reservoir computing system. **Nature Electronics**, v. 2, n. 10, p. 480–487, 2019.

NASIRI, H.; EBADZADEH, M. M. Multi-step-ahead stock price prediction using recurrent fuzzy neural network and variational mode decomposition. **Applied Soft Computing**, v. 148, p. 110867, 2023.

O'DONNCHA, F. et al. A spatio-temporal lstm model to forecast across multiple temporal and spatial scales. **Ecological Informatics**, v. 69, 2022. ISSN 15749541.

OLIVEIRA, P. J.; STEFFEN, J. L.; CHEUNG, P. Parameter estimation of seasonal arima models for water demand forecasting using the harmony search algorithm. **Procedia Engineering**, v. 186, p. 177–185, 2017. XVIII International Conference on Water Distribution Systems, WDSA2016.

PAWŁOWSKI, A. et al. Model predictive control using miso approach for drug co-administration in anesthesia. **Journal of Process Control**, v. 117, p. 98–111, 2022. ISSN 0959-1524. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959152422001287>>.

PEIMANKAR, A. et al. Multi-objective ensemble forecasting with an application to power transformers. **Applied Soft Computing Journal**, v. 68, p. 233–248, 2018. ISSN 15684946.

PELLETIER, C. et al. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. **Remote Sensing of Environment**, v. 187, p. 156–168, 2016. Cited By 296. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84992151859&doi=10.1016/j.rse.2016.10.010&partnerID=40&md5=09efc79bab8e893b97fd21cb4844b98d>>.

PETROPOULOS, F. et al. Forecasting: theory and practice. **International Journal of Forecasting**, v. 38, n. 3, p. 705–871, 2022. ISSN 0169-2070. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169207021001758>>.

PREDUM, R. **Time Series Forecasting with ARIMA, SARIMA, and SARIMAX**. 2021. Disponível em: <<https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6>>.

QIN, Y. et al. Spatio-temporal hierarchical mlp network for traffic forecasting. **Information Sciences**, v. 632, p. 543–554, 2023.

RAMOS, A. S. **Previsões de Séries Temporais combinando modelos ARMA e Redes Neurais Artificiais**. Tese (Doutorado) — Universidade Federal de Pernambuco, 2010.

RAO, X.; ZHAO, H.; DENG, Q. Artificial-neural-network (ann) based proxy model for performances forecast and inverse project design of water huff-n-puff technology. **Journal of Petroleum Science and Engineering**, v. 195, p. 107851, 2020.

RASCHKA, S. A practical guide to machine learning in python. **Machine Learning with Python**, 2015. Disponível em: <[https://sebastianraschka.com/pdf/books/machine\\_learning\\_with\\_python/mlwp.pdf](https://sebastianraschka.com/pdf/books/machine_learning_with_python/mlwp.pdf)>.

READER, T. C. Decision tree regression explained with implementation in python. **Medium**, 2023. Disponível em: <<https://medium.com/@theclickreader/decision-tree-regression-explained-with-implementation-in-python-1e6e48aa7a47>>.

REICHMAN, D.; MALOF, J. M.; COLLINS, L. M. Leveraging seed dictionaries to improve dictionary learning. In: **2016 IEEE International Conference on Image Processing (ICIP)**. [S.l.: s.n.], 2016. p. 3723–3727.

REISEN, V. et al. Robust dickey–fuller tests based on ranks for time series with additive outliers. **Metrika**, v. 80, n. 1, p. 115–131, 2017. Cited By 1. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84986317325&doi=10.1007%2fs00184-016-0594-8&partnerID=40&md5=c83f82d0c372e22d5970aff448f05411>>.

RIBEIRO, M. H. D. M. et al. Cooperative ensemble learning model improves electric short-term load forecasting. **Chaos, Solitons & Fractals**, v. 166, p. 112982, 2023.

RIBEIRO, M. H. D. M. et al. Time series forecasting based on ensemble learning methods applied to agribusiness, epidemiology, energy demand, and renewable energy. Pontificia Universidade Católica do Paraná, 2021.

ROSSI, R. Relational time series forecasting. **Knowledge Engineering Review**, v. 33, 2018.

ROSTAM, N. A. P. et al. A complete proposed framework for coastal water quality monitoring system with algae predictive model. **IEEE Access**, Institute of Electrical and Electronics Engineers Inc., v. 9, p. 108249 – 108265, 2021. ISSN 21693536. Cited by: 12; All Open Access, Gold Open Access. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85112659996&doi=10.1109%2fACCESS.2021.3102044&partnerID=40&md5=a078d7fe0d04f37177f1ae6f798ff71b>>.

ROSTAMIAN, A.; O'HARA, J. G. Event prediction within directional change framework using a cnn-lstm model. **NEURAL COMPUTING & APPLICATIONS**, v. 34, p. 17193–17205, 2022. ISSN 0941-0643.

RUIZ-ROSETO, J.; RAMIREZ-GONZALEZ, G.; VIVEROS-DELGADO, J. Software survey: Scientopy, a scientometric tool for topics trend analysis in scientific publications. **Scientometrics**, v. 121, n. 2, p. 1165–1188, Nov 2019. ISSN 1588-2861. Disponível em: <<https://doi.org/10.1007/s11192-019-03213-w>>.

SABZIPOUR, B. et al. Comparing a long short-term memory (lstm) neural network with a physically-based hydrological model for streamflow forecasting over a canadian catchment. **Journal of Hydrology**, v. 627, p. 130380, 2023.

SADAEI, H. et al. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. **Energy**, v. 175, p. 365–377, 2019. ISSN 03605442.

SADAEI, H. et al. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. **Energy**, v. 175, p. 365–377, 2019.

SALGOTRA, R.; GANDOMI, M.; GANDOMI, A. Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming. **Chaos, Solitons and Fractals**, v. 138, 2020.

SAMANTA, S. et al. Learning elastic memory online for fast time series forecasting. **Neurocomputing**, v. 390, p. 315–326, 2020.

SANG, Y.-F. et al. Wavelet-based hydrological time series forecasting. **Journal of Hydrologic Engineering**, v. 21, 2016. ISSN 10840699.

SARANYA, S.; SIVAKUMAR, R. Gated recurrent units (gru) for time series forecasting in higher education. **International Journal of Engineering Research and Technology**, v. 13, n. 7, p. 1809–1813, 2020. Disponível em: <<https://www.ijert.org/gated-recurrent-units-gru-for-time-series-forecasting-in-higher-education>>.

SEMAN, L. O. et al. Ensemble learning methods using the hodrick–prescott filter for fault forecasting in insulators of the electrical power grids. **International Journal of Electrical Power & Energy Systems**, v. 152, p. 109269, 2023.

SEN, J. et al. Machine learning: Algorithms, models, and applications. **arXiv preprint arXiv:2201.01943**, 2022.

SEZER, O. B.; GUDELEK, M. U.; OZBAYOGLU, A. M. Financial time series forecasting with deep learning : A systematic literature review: 2005-2019. **APPLIED SOFT COMPUTING**, v. 90, 2020. ISSN 1568-4946.

SEZER, O. B.; GUDELEK, M. U.; OZBAYOGLU, A. M. Financial time series forecasting with deep learning : A systematic literature review: 2005-2019. **APPLIED SOFT COMPUTING**, v. 90, 2020. ISSN 1568-4946.

SHEN, L.; WANG, Y. Tcct: Tightly-coupled convolutional transformer on time series forecasting. **Neurocomputing**, v. 480, p. 131–145, 2022. ISSN 09252312.

SHEN, Z. et al. A novel time series forecasting model with deep learning. **Neurocomputing**, v. 396, p. 302–313, 2020.

SHI, M. et al. Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine. **Mechanical Systems and Signal Processing**, v. 188, p. 110022, 2023.

SHIH, S.-Y.; SUN, F.-K.; LEE, H.-Y. Temporal pattern attention for multivariate time series forecasting. **Machine Learning**, v. 108, p. 1421–1441, 2019. ISSN 08856125.

SHIH, S.-Y.; SUN, F.-K.; LEE, H.-Y. Temporal pattern attention for multivariate time series forecasting. **Machine Learning**, v. 108, n. 8-9, p. 1421–1441, 2019.

SIEGEL, J. E. et al. Safe energy savings through context-aware hot water demand prediction. **Engineering Applications of Artificial Intelligence**, v. 90, p. 103481, 2020.

SILVA, A. C.; GOMES, L. F. A. M. Inteligência artificial: estado atual, desafios e oportunidades de pesquisa. **Estudos Avançados**, v. 35, n. 101, p. 7–26, 2021. Disponível em: <<https://www.scielo.br/j/ea/a/wXBdv8yHBV9xHz8qG5RCgZd/>>.

SILVA, J. P. **Como funcionam as Redes Neurais Convolucionais (CNNs)**. 2021. <<https://medium.com/data-hackers/como-funcionam-as-redes-neurais-convolucionais-cnns-71978185c1>>. Acessado em 04/05/2023.

Singh Kushwah, J. et al. Comparative study of regressor and classifier with decision tree using modern tools. **Materials Today: Proceedings**, v. 56, p. 3571–3576, 2022. First International Conference on Design and Materials.

STEFENON, S. F. et al. Aggregating prophet and seasonal trend decomposition for time series forecasting of italian electricity spot prices. **Energies**, v. 16, n. 3, 2023.

SUN, X.; LIU, M.; SIMA, Z. A novel cryptocurrency price trend forecasting model based on lightgbm. **Finance Research Letters**, v. 32, p. 101084, 2020.

TAIEB, S. B.; ATIYA, A. F. A Bias and Variance Analysis for Multistep-Ahead Time Series Forecasting. **IEEE Transactions on Neural Networks and Learning Systems**, Institute of Electrical and Electronics Engineers Inc., Department of Computer Science, Université Libre de Bruxelles, Brussels, 1050, Belgium, v. 27, n. 1, p. 62–76, 2016. ISSN 2162237X (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84925431469&doi=10.1109%2FTNNLS.2015.2411629&partnerID=40&md5=e1c7f3c7a1136a0e0e4d2aff817b4008>>.

TAM, A. **LSTM for Time Series Prediction in PyTorch**. Machine Learning Mastery, 2023. Disponível em: <<https://machinelearningmastery.com/lstm-for-time-series-prediction-in-pytorch/>>.

TAN, Y. F. et al. Exploring Time-Series Forecasting Models for Dynamic Pricing in Digital Signage Advertising. **FUTURE INTERNET**, v. 13, n. 10, 2021. ISSN 1999-5903.

TAO, H. et al. Training and testing data division influence on hybrid machine learning model process: Application of river flow forecasting. **Complexity**, Hindawi, Oct 2020.

THEODOSIOU, M. Forecasting monthly and quarterly time series using stl decomposition. **International Journal of Forecasting**, v. 27, n. 4, p. 1178–1195, 2011. Cited By 86. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-80052160927&doi=10.1016%2fj.ijforecast.2010.11.002&partnerID=40&md5=e8242471ba1ec14ada46ab567f3a364d>>.

TRENBERTH, K. E. Signal versus noise in the southern oscillation. **Monthly Weather Review**, v. 112, n. 2, p. 326–332, 1984.

TSATSOU, A.; FRANTZESKAKI, N.; MALAMIS, S. Nature-based solutions for circular urban water systems: A scoping literature review and a proposal for urban design and planning. **Journal of Cleaner Production**, v. 394, p. 136325, 2023.

TYRALIS, H.; PAPACHARALAMPOUS, G. Variable selection in time series forecasting using random forests. **Algorithms**, v. 10, n. 4, 2017.

TYRALIS, H.; PAPACHARALAMPOUS, G. Variable selection in time series forecasting using random forests. **Algorithms**, v. 10, 2017. ISSN 19994893.

UC-CASTILLO, J. L. et al. A systematic review and meta-analysis of groundwater level forecasting with machine learning techniques: Current status and future directions. **Environmental Modelling & Software**, v. 168, p. 105788, 2023.

URSU, E.; PEREAU, J. C. Application of periodic autoregressive process to the modeling of the Garonne river flows. **STOCHASTIC ENVIRONMENTAL RESEARCH AND RISK ASSESSMENT**, v. 30, n. 7, p. 1785–1795, 2016. ISSN 1436-3240.

VASCONCELOS, F. **Falta d'água em curitiba e região metropolitana não É culpa só da estiagem**. 2020. Disponível em: <<https://www.brasildefato.com.br/2020/11/03/falta-d-agua-em-curitiba-e-regiao-metropolitana-nao-e-culpa-so-da-estiagem>>.

VIDHYA, A. **Time Series Forecasting and Analysis — ARIMA and Seasonal ARIMA**. Medium, 2023. Disponível em: <<https://medium.com/analytics-vidhya/time-series-forecasting-and-analysis-arima-and-seasonal-arima-caca61ae863>>.

VLACHAS, P. et al. Backpropagation algorithms and Reservoir Computing in Recurrent Neural Networks for the forecasting of complex spatiotemporal dynamics. **Neural Networks**, v. 126, p. 191–217, 2020.

WANG, J. et al. Financial time series prediction using elman recurrent random neural networks. **Computational Intelligence and Neuroscience**, v. 2016, 2016. ISSN 16875265.

WANG, L. et al. Hybrid application of unsupervised and supervised learning in forecasting absolute open flow potential for shale gas reservoirs. **Energy**, v. 243, p. 122747, 2022.

WANG, M.; YING, F. Point and interval prediction for significant wave height based on lstm-gru and kde. **Ocean Engineering**, v. 289, p. 116247, 2023.

WANG, Y. et al. Recycling combustion ash for sustainable cement production: A critical review with data-mining and time-series predictive models. **CONSTRUCTION AND BUILDING MATERIALS**, v. 123, p. 673–689, 2016. ISSN 0950-0618.

XIAN, S. et al. A novel fuzzy time series forecasting method based on the improved artificial fish swarm optimization algorithm. **Soft Computing**, v. 22, p. 3907–3917, 2018. ISSN 14327643.

XIANG, Y. et al. A svr-ann combined model based on ensemble emd for rainfall prediction. **Applied Soft Computing**, v. 73, p. 874–883, 2018.

XU, W. et al. Deep belief network-based AR model for nonlinear time series forecasting. **Applied Soft Computing Journal**, v. 77, p. 605–621, 2019.

XU, W. et al. A hybrid modelling method for time series forecasting based on a linear regression model and deep learning. **Applied Intelligence**, v. 49, p. 3002–3015, 2019. ISSN 0924669X.

YANG, S.; GUO, H.; LI, J. Cnn-grua-fc stock price forecast model based on multi-factor analysis. **Journal of Advanced Computational Intelligence and Intelligent Informatics**, v. 26, p. 600–608, 2022. ISSN 13430130.

YANG, W. et al. Hybrid wind energy forecasting and analysis system based on divide and conquer scheme: A case study in China. **Journal of Cleaner Production**, v. 222, p. 942–959, 2019.



YE, J.; ZHAO, B.; DENG, H. Photovoltaic power prediction model using pre-train and fine-tune paradigm based on lightgbm and xgboost. **Procedia Computer Science**, v. 224, p. 407–412, 2023.

ZHANG, E. **Recurrent Neural Network is All You Need**. 2021. <<https://medium.com/mcgill-mma-intro-to-ai/recurrent-neural-network-is-all-you-need-f576782c5d2>>. Acessado em: 22 de Março de 2023.

ZHAO, L. et al. A hybrid vmd-lstm/gru model to predict non-stationary and irregular waves on the east coast of china. **Ocean Engineering**, v. 276, p. 114136, 2023.