INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS – *CAMPUS* BAMBUÍ
BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

Juan Pablo Amorim Joanas

ANÁLISE E PREVISÃO DA ENERGIA ARMAZENADA EM RESERVATÓRIOS HIDRELÉTRICOS UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINA

Bambuí - MG
2025

JUAN PABLO AMORIM JOANAS

ANÁLISE E PREVISÃO DA ENERGIA ARMAZENADA EM RESERVATÓRIOS HIDRELÉTRICOS UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINA

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Engenharia de Computação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – *Campus* Bambuí para obtenção do grau de Bacharel em Engenharia de Computação.

Bambuí - MG
2025

Catalogação na Fonte Biblioteca IFMG - *Campus* Bambuí

J62a Joanas, Juan Pablo Amorim

Análise e previsão da energia armazenada em reservatórios hidrelétricos utilizando técnicas de aprendizado de máquina [manuscrito] / Juan Pablo Amorim Joanas – 2025.
54 f. : il.

Orientador: Felipe Lopes de Melo Faria.
Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Computação) – Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais. *Campus* Bambuí, 2025.

1. Aprendizado de máquina. 2. Previsão de energia. 3. Reservatórios hidrelétricos. I. Faria, Felipe Lopes de Melo. II. Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – *Campus* Bambuí. III. Título.

CDD 005.1

Catalogação: João Batista Rodrigues - CRB-6/2022

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS
Campus Bambuí
Diretoria de Ensino
Departamento de Engenharia e Computação
Faz. Varginha - Rodovia Bambuí/Medeiros - Km 05 - Caixa Postal 05 - CEP 38900-000 - Bambuí - MG
37 3431 4900 - www.ifmg.edu.br

Juan Pablo Amorim Joanas

ANÁLISE E PREVISÃO DA ENERGIA ARMAZENADA EM RESERVATÓRIOS HIDRELÉTRICOS UTILIZANDO TÉCNICAS DE APRENDIZADO DE MÁQUINA

Trabalho de conclusão de curso apresentado ao curso de Bacharelado em Engenharia de Computação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – Campus Bambuí para obtenção do grau de bacharel em Engenharia de Computação.

Aprovado em 13/06/2025 pela banca examinadora:

Bambuí,13 de junho de 2025.

Documento assinado eletronicamente por **Calebe Giaculi Junior**, **Professor**, em 25/06/2025, às 15:05, conforme Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Rogerio Amaral Bonatti**, **Professor Substituto**, em 25/06/2025, às 15:08, conforme Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Felipe Lopes de Melo Faria**, **Professor**, em 26/06/2025, às 09:22, conforme Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade do documento pode ser conferida no site https://sei.ifmg.edu.br/consultadocs informando o código verificador **2346786** e o código CRC **E22997F2**.

Aos meus pais e aos meus irmãos.

AGRADECIMENTOS

À minha mãe, Marlene, minha eterna fonte de inspiração e apoio incondicional. Agradeço por acreditar em mim desde o início, por me incentivar a perseguir meus sonhos e por sempre me oferecer um porto seguro nos momentos mais difíceis. Sua força, amor e sabedoria me guiaram durante toda a minha vida e foram essenciais para que eu chegasse até aqui. Aos meus irmãos, Pâmela, Pierre e Ueslei, meus companheiros de todas as horas, obrigado por compartilharem comigo alegrias e tristezas, por me ensinarem tanto sobre a vida e por me proporcionarem momentos inesquecíveis. Agradeço a amizade, por me colocarem sempre para cima e por acreditarem no meu potencial. Agradeço à minha namorada todo o apoio, compreensão e incentivo ao longo desta jornada acadêmica.

Agradeço ao meu orientador, Prof. Me. Felipe Lopes de Melo Faria, sua inestimável orientação, as valiosas contribuições e a constante disponibilidade para auxiliar no desenvolvimento desta pesquisa. Sua expertise e entusiasmo pela área me motivaram a buscar a excelência em meu trabalho. Agradeço a cada um que, direta ou indiretamente, contribuiu para a realização deste trabalho. Agradeço aos amigos, colegas, familiares e demais pessoas que me incentivaram, me ofereceram palavras de apoio e me proporcionaram momentos de alegria e descontração.

"A persistência é o caminho do êxito."
(Charles Chaplin)

RESUMO

O presente trabalho propõe uma abordagem para a previsão da energia armazenada em reservatórios hidrelétricos, por meio da aplicação de técnicas de aprendizado de máquina, com destaque para a utilização da rede neural recorrente do tipo LSTM. Utilizou-se como fonte de dados a base Energia Natural Afluente (ENA) Diário por Subsistema, que consiste por valores de energia produzível pelas usinas do Sudeste/Centro-Oeste, disponibilizados pelo Operador Nacional do Sistema Elétrico (2024), realizando um processo de pré-processamento para adequar os dados e em seguida aplicar técnicas de aprendizado de máquina. Esta abordagem permitiu a exploração de padrões históricos de geração e armazenamento de energia, visando o desenvolvimento de modelos de previsão precisos. Posteriormente, foi conduzida uma análise dos resultados obtidos, destacando-se as técnicas que demonstraram melhor desempenho, por meio da avaliação de suas métricas pertinentes. O experimento que mais se destacou foi o Experimento 3, que utilizou médias móveis para suavizar o sinal, obtendo os melhores resultados para as métricas MASE (1,1009) e POCID (93,5094%) com uma janela de suavização de 15 dias.

Palavras-chave: Aprendizado de Máquina, LSTM ,Previsão de Energia, Reservatórios Hidrelétricos

ABSTRACT

This work proposes an approach for forecasting energy stored in hydroelectric reservoirs, through the application of machine learning techniques, with emphasis on the use of the LSTM recurrent neural network. The data source used was the Daily Affluent Natural Energy (ENA) database by Subsystem, which consists of values of energy produced by plants in the Southeast/Central-West, made available by the National Electric System Operator (2024), performing a pre-processing process to adapt the data and then applying machine learning techniques. This approach allowed the exploration of historical patterns of energy generation and storage, aiming at the development of accurate forecasting models. Subsequently, an analysis of the results obtained was conducted, highlighting the techniques that demonstrated the best performance, through the evaluation of their pertinent metrics. The experiment that stood out the most was Experiment 3, which used moving averages to smooth the signal, obtaining the best results for the MASE (1.1009) and POCID (93.5094%) metrics with a 15-day smoothing window.

Keywords: Machine Learning, LSTM, Energy Forecasting, Hydroelectric Reservoirs

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AM – Aprendizado de Máquina

BEN – Balanço Energético Nacional

ENA – Energia Natural Afluente

EPE – Empresa de Pesquisa Energética

LSTM – *Long Short-Term Memory*

MW – Megawatt

ONS – Operador Nacional do Sistema Elétrico

RNA – Redes Neurais Artificiais

RNN – Rede Neural Recorrente

STL – *Seasonal-Trend decomposition using Loess*

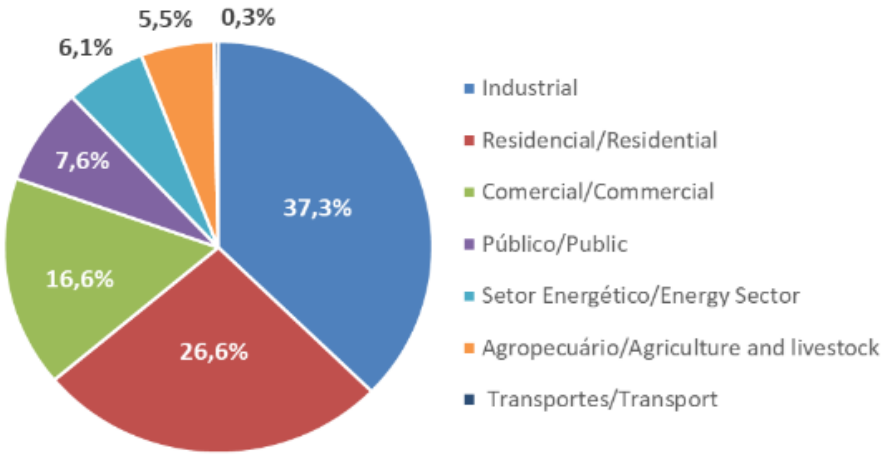
SUMÁRIO

1 INTRODUÇÃO

Nos últimos anos, o Brasil vivencia um acelerado crescimento na demanda por energia elétrica, impulsionado por fatores como a intensificação da urbanização, o avanço da industrialização e o crescente desenvolvimento tecnológico, conforme pode ser visto na Figura 1, onde é possível analisar as parcelas de cada setor no consumo de energia elétrica. Esse cenário impõe a necessidade premente de uma gestão eficiente dos recursos energéticos, visando garantir a estabilidade e a sustentabilidade do sistema elétrico nacional.

De acordo com os dados descritos pela Empresa de Pesquisa Energética (EPE), o Brasil se destaca no cenário mundial como um país com uma matriz energética majoritariamente renovável, em que a geração de energia hidrelétrica assume um papel fundamental, respondendo por uma parcela significativa da eletricidade consumida no país. Em 2022, a geração de energia hidrelétrica participou com 64% da oferta interna de energia elétrica, conforme dados provenientes do Balanço Energético Nacional (BEN) da (Empresa de Pesquisa Energética , 2023).

Figura 1 – Participação setorial no consumo de eletricidade



Fonte: (Empresa de Pesquisa Energética , 2023).

De acordo com (Barros, 2005), o Brasil enfrentou uma grande preocupação em relação ao setor energético devido ao racionamento de energia elétrica em 2001, que afetou os Estados do Sudeste e Nordeste, evidenciando falhas no modelo de operação adotado nos anos 1990.

De acordo com (Agência Nacional de Energia Elétrica (ANEEL), 2005), a energia firme de uma hidrelétrica é a quantidade máxima de energia que a usina consegue gerar de forma constante, levando em consideração o período mais seco já registrado no histórico de vazões do rio onde ela está situada.

As usinas hidrelétricas utilizam dados históricos de vazões que cobrem um período de 70 anos. Com base nesses dados, é possível realizar estudos aplicando técnicas estatísticas para simular diferentes cenários de vazão das usinas. A energia assegurada

representa a quantidade máxima de energia que as usinas podem oferecer de forma contínua ao longo dos anos. Esse valor é estimado por meio de simulações que utilizam dados estatísticos, considerando o risco de não conseguir atender totalmente à demanda. Em termos práticos, as simulações indicam que, em 5% dos cenários, pode ocorrer racionamento (Agência Nacional de Energia Elétrica (ANEEL), 2005).

Desta forma, o presente trabalho propôs uma abordagem para usar técnicas de Aprendizado de Máquina (AM) para prever a energia armazenada no reservatório hidrelétrico do Sudeste/Centro-Oeste. Para isso, empregaram-se Redes Neurais Recorrentes do tipo LSTM em um conjunto de dados diário de Energia Natural Afluente (ENA) fornecido pelo Operador Nacional do Sistema Elétrico (ONS. . . , 2024).

1.1 Objetivo Geral

O objetivo geral deste trabalho foi desenvolver e avaliar modelos de aprendizado de máquina para a análise e previsão da energia armazenada em reservatórios hidrelétricos do subsistema Sudeste/Centro-Oeste. Por meio da base de dados, fundamentada em valores diários coletados nos reservatórios, em que a base compôs valores da ENA bruta, ou seja, a energia produzível pela usina em Megawatt (MW) médio, informações relevantes foram extraídas para alimentar os modelos de previsão. O objetivo é aprimorar a precisão das previsões de geração hidrelétrica, utilizando técnicas computacionais avançadas, o que contribuirá para a otimização da gestão dos recursos hídricos e energéticos.

1.2 Objetivos específicos

Os objetivos específicos do presente trabalho foram os seguintes:

- Pré-processar a base de dados ENA Diário por Subsistema;
- Desenvolver modelo de aprendizado de máquina;
- Avaliar o modelo de aprendizado de máquina.

1.3 Justificativa

De acordo com a (Empresa de Pesquisa Energética , 2023), o cenário brasileiro apresenta uma matriz energética predominantemente renovável, com a produção de energia proveniente principalmente de fontes hídricas. Ressaltam-se a importância e o impacto, tanto econômico quanto social, que a escassez de energia pode ocasionar ao país.

Conforme a (Empresa de Pesquisa Energética , 2023), em 2022, a produção de energia elétrica proveniente de hidrelétricas foi responsável por 64% da matriz energética brasileira. Isso levanta preocupações sobre a possibilidade de racionamento de energia. De acordo com (Barros, 2005), o Brasil enfrentou um racionamento de energia no ano de

2001, que afetou os Estados do Nordeste e do Sudeste, evidenciando falhas no modelo operacional. Esses acontecimentos destacam a importância de estudar medidas que assegurem a antecipação e a eficácia na operação do sistema.

Segundo a (Agência Nacional de Energia Elétrica (ANEEL), 2005), a projeção operacional das hidrelétricas é realizada com base em dados históricos das vazões dos rios onde essas instalações estão situadas. Esse método tem como objetivo prever diferentes cenários operacionais e analisar a possibilidade de racionamento em períodos futuros. De acordo com (Haykin, 1998), o AM envolve a extração de regras e padrões utilizando algoritmos baseados em estatísticas.

Portanto, a integração de técnicas de aprendizado de máquina na previsão de geração energética será fundamental para o gerenciamento eficaz, a previsão precisa e a antecipação da tomada de decisões no futuro, otimizando o processo de geração de energia. Dessa forma, o trabalho está voltado para a aplicação de técnicas de AM na previsão da geração de energia nos reservatórios Sudeste/Centro-Oeste.

1.4 Resultados esperados

Espera-se que as técnicas aplicadas contribuam na geração de modelos preditivos no escopo do desenvolvimento de modelos de aprendizado de máquina para a análise e previsão da energia armazenada em reservatórios hidrelétricos. Por meio dessa aplicação, busca-se avançar na capacidade de previsão da energia armazenada, permitindo uma gestão mais eficiente dos recursos hídricos e energéticos. Assim, almeja-se que o trabalho produzido possa contribuir com a tomada de decisão que impacte diretamente na otimização da geração hidrelétrica e na minimização dos custos de produção de energia.

1.5 Estrutura do trabalho

No início do capítulo, é feita uma introdução ao assunto do estudo, destacando a relevância de se prever a energia armazenada em reservatórios hidrelétricos com o uso de métodos de aprendizado de máquina. Os desafios e a importância da utilização dessas técnicas são debatidos durante o processo. Adicionalmente, os propósitos do estudo e a organização do texto são mostrados para auxiliar o leitor na compreensão.

No segundo capítulo, é apresentada uma revisão da literatura, com destaque para os conceitos essenciais ligados à matriz energética e às estratégias de previsão. Nesta parte, são abordados os fundamentos teóricos que dão suporte à pesquisa, fornecendo uma compreensão minuciosa das técnicas empregadas e de como elas são aplicadas no contexto das usinas hidrelétricas.

No capítulo três, a metodologia utilizada para elaborar o trabalho é explicada em detalhes. Nesta parte, são descritos o método de coleta e análise de dados, a aplicação de modelos de aprendizado de máquina e as técnicas utilizadas para avaliar os resultados

alcançados. O quarto capítulo é reservado para a exibição dos resultados, ressaltando-se que os resultados da previsão da produção de energia armazenada nos reservatórios são comparados entre diferentes abordagens de algoritmos aplicados.

Por fim, no quinto capítulo, são expostas as considerações do estudo, enfatizando os resultados e contribuições principais alcançadas durante a pesquisa. Adicionalmente, são abordadas as restrições e propostas de possíveis caminhos para pesquisas futuras, com o objetivo de melhorar e ampliar a compreensão na área.

2 REFERENCIAL TEÓRICO

Neste capítulo, a revisão da literatura se une à fundamentação teórica, discutindo os conceitos essenciais para a compreensão do estudo. Depois, são mostrados os avanços mais recentes, enfatizando os projetos realizados na área de estudo.

2.1 Fundamentação teórica

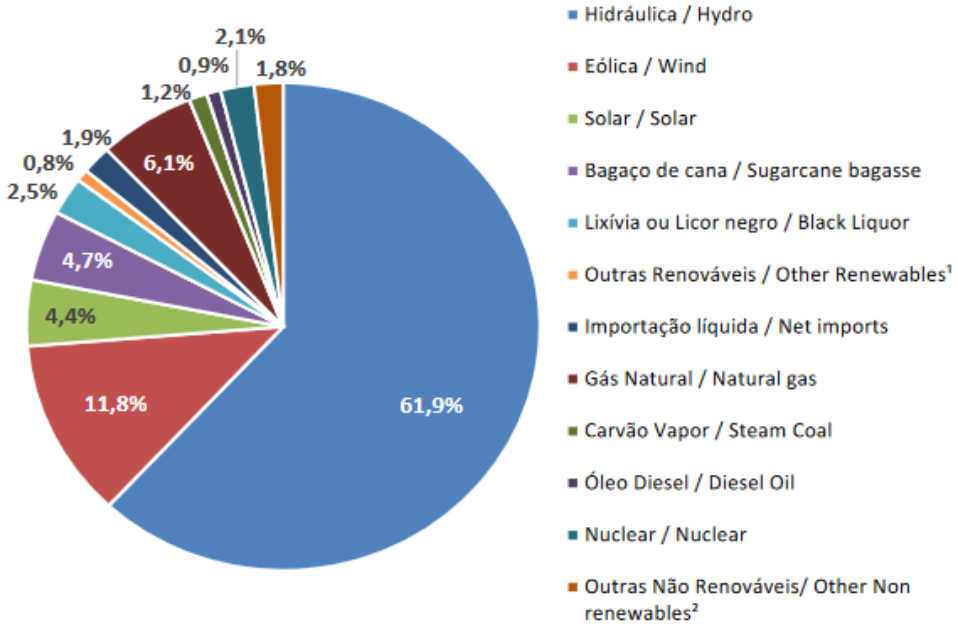
A fundamentação teórica desta pesquisa se debruça sobre três pilares: mineração de dados, matriz energética e técnicas de predição. Cada um desses temas será detalhadamente definido nas seções subsequentes.

2.1.1 Matriz Energética Brasileira

O uso de fontes de energia limpa vem crescendo nas matrizes energéticas do mundo todo, principalmente por causa da necessidade de proteger o meio ambiente e combater o aquecimento global. No entanto, essas fontes podem ter produção variável, o que pode ser um desafio (Martín *et al.*, 2011).

O cenário energético brasileiro é caracterizado por uma matriz vasta e diversificada. Predominantemente, a energia no Brasil é gerada a partir de recursos renováveis, com as hidrelétricas desempenhando um papel central como a principal fonte de produção de energia, conforme ilustrado na Figura 2.

Figura 2 – Oferta Interna de Energia Elétrica por Fonte



Fonte: (Empresa de Pesquisa Energética , 2023).

Entretanto, as fontes não renováveis também contribuem significativamente para a oferta interna de eletricidade. O gás natural, o carvão e o óleo diesel, entre outros, continuam a ser importantes componentes dessa matriz energética. Segundo dados da (Empresa de Pesquisa Energética , 2023), a participação da fonte hídrica na oferta interna de energia elétrica atingiu 64% em 2022, evidenciando a importância dessa fonte para o suprimento energético do país.

2.1.2 Reservatórios

Segundo (Lopes; Santos, 2002), os reservatórios hidrelétricos exercem um papel crucial na regulação das vazões naturais, armazenando água durante períodos pluviais para compensar a escassez durante períodos de estiagem. Esses reservatórios são, primordialmente, constituídos por barragens em cursos d'água, e suas características físicas, notadamente a capacidade de armazenamento, estão intrinsecamente associadas às particularidades topográficas do local de implantação. O nível e o volume de água armazenada emergem como os principais indicadores do estado do reservatório, fornecendo dados essenciais para sua operação e gestão.

De acordo com (Campagnoli; Diniz, 2012), um reservatório hidrelétrico é formado pela construção de um barramento artificial em um corpo d'água, direcionando a vazão do rio para as turbinas. Essa transformação de energia potencial em cinética e, posteriormente, em energia elétrica, comprova a Lei de Lavoisier e representa um compromisso da engenharia.

(Lopes; Borges, 2014) destacam que a quantidade de energia gerada por uma turbina está diretamente atrelada à vazão de água que a impulsiona. Em outras palavras, quanto maior a vazão, maior a produção de energia. Essa relação intrínseca conecta a geração de energia ao sistema hidrológico, onde a vazão depende do nível dos rios, que, por sua vez, é influenciado pela precipitação pluviométrica nas nascentes.

2.1.3 Mineração de Dados

A Mineração de Dados emerge da convergência de três áreas distintas: bancos de dados, estatística e inteligência artificial (Fayyad; Piatetsky-Shapiro; Smyth, 1996). Essa confluência deu origem a uma disciplina focada na extração de informações úteis e relevantes de grandes conjuntos de dados. Segundo (Tan; Steinbach; Kumar, 2009), a mineração de dados é uma tecnologia que utiliza métodos de análise para processar grandes conjuntos de dados, visando extrair informações relevantes, para auxiliar na tomada de decisões.

Segundo (Fayyad; Piatetsky-Shapiro; Smyth, 1996), os dois principais objetivos da mineração de dados, na prática, são a previsão e a descrição. A previsão envolve usar variáveis de um banco de dados para prever valores desconhecidos ou futuros de outras

variáveis de interesse. Já a descrição se concentra em encontrar padrões interpretáveis por humanos que descrevam os dados. Entre as diversas tarefas da Mineração de Dados, a predição se destaca como uma área de grande importância.

2.1.4 Aprendizado de máquina

Esta seção inicia-se definindo os conceitos básicos da área de inteligência artificial e aprendizado de máquina.

Segundo (Whitby, 2009), a inteligência artificial é caracterizada como o campo de estudo das ações inteligentes em diferentes entidades, como seres humanos, animais e máquinas, com o objetivo de replicar ou modelar esses comportamentos em instrumentos tecnológicos. O aprendizado de máquina é um subconjunto da área de inteligência artificial, cujo objetivo é desenvolver técnicas e algoritmos que permitem ao computador aprender e executar ações (Amorim; Barone; Mansur, 2008).

De acordo com (Samuel, 1959), AM é definido como a capacidade de o computador aprender com as experiências e se adaptar, aperfeiçoando-se gradativamente com o passar do tempo, sem que seja necessário programar todos os passos explicitamente. O AM é a extração de regras e padrões em conjuntos de dados, através de algoritmos dedutivos que se baseiam em estatística (Haykin, 1998).

A indução é o processo de se chegar a conclusões gerais a partir da observação de um conjunto de amostras, extraindo conhecimento por meio das observações dos exemplos apresentados, podendo ou não preservar a verdade através das hipóteses geradas pela indução (Monard; Baranauskas, 2003).

De acordo com (Monard; Baranauskas, 2003), o cérebro humano utiliza a indução como um dos recursos para prover conhecimentos novos, porém mantendo cautela em relação aos resultados obtidos, devido à possibilidade de falta de informação. Mesmo assim, a inferência indutiva é usada para descobrir possíveis ações futuras. O aprendizado indutivo se desdobra em duas partes: aprendizado supervisionado e aprendizado não supervisionado. No aprendizado supervisionado, são fornecidos exemplos para realizar o treinamento, sendo que as técnicas aplicadas neste trabalho utilizam aprendizado supervisionado.

Outras definições a serem exploradas incluem séries temporais e aprendizado profundo. Segundo (LeCun; Bengio; Hinton, 2015), o aprendizado profundo é uma abordagem em que modelos computacionais, formados por várias camadas de processamento, são capazes de extrair e entender informações dos dados em diferentes níveis de complexidade.

(Shumway; Stoffer; Stoffer, 2000) Descrevem uma série temporal como um conjunto de valores organizados em ordem cronológica, coletados ao longo do tempo.

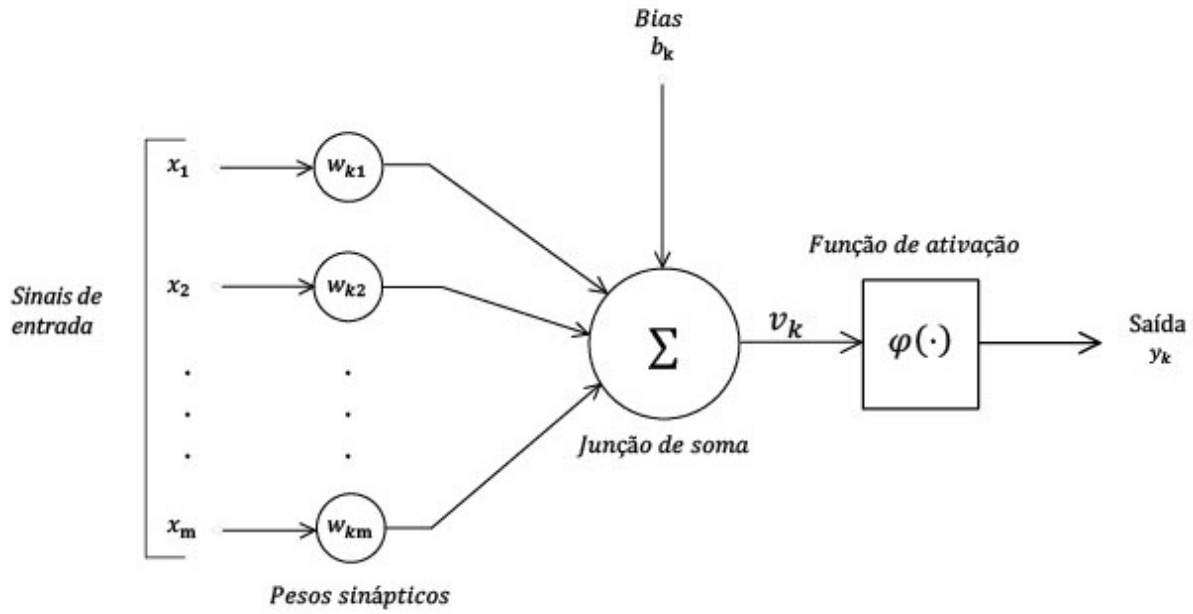
2.1.4.1 Redes Neurais Artificiais

Uma rede neural é um sistema computacional projetado para processar informações de forma paralela e distribuída. Esse sistema é composto por modelos de processamento simples, que têm a capacidade de armazenar informações experienciadas. Ela se assemelha ao cérebro, pois o conhecimento é adquirido pela rede, através da vivência e das experiências em seu ambiente. As conexões entre os neurônios artificiais, conhecidas como pesos sinápticos, são usadas para armazenar essas experiências conquistadas (Haykin, 1998).

2.1.4.2 Neurônio Artificial

Um neurônio é uma unidade capaz de realizar o processamento em uma rede neural. Os sinais de entrada são multiplicados pelos pesos sinápticos correspondentes, que indicam a força dos estímulos. Após a multiplicação, os sinais ponderados são aplicados ao somador, que soma todos os produtos ponderados, combinando-os em uma única entrada para o neurônio. Em seguida, é aplicada a função de ativação para produzir uma saída (Haykin, 1998).

Figura 3 – Neurônio não linear de um modelo



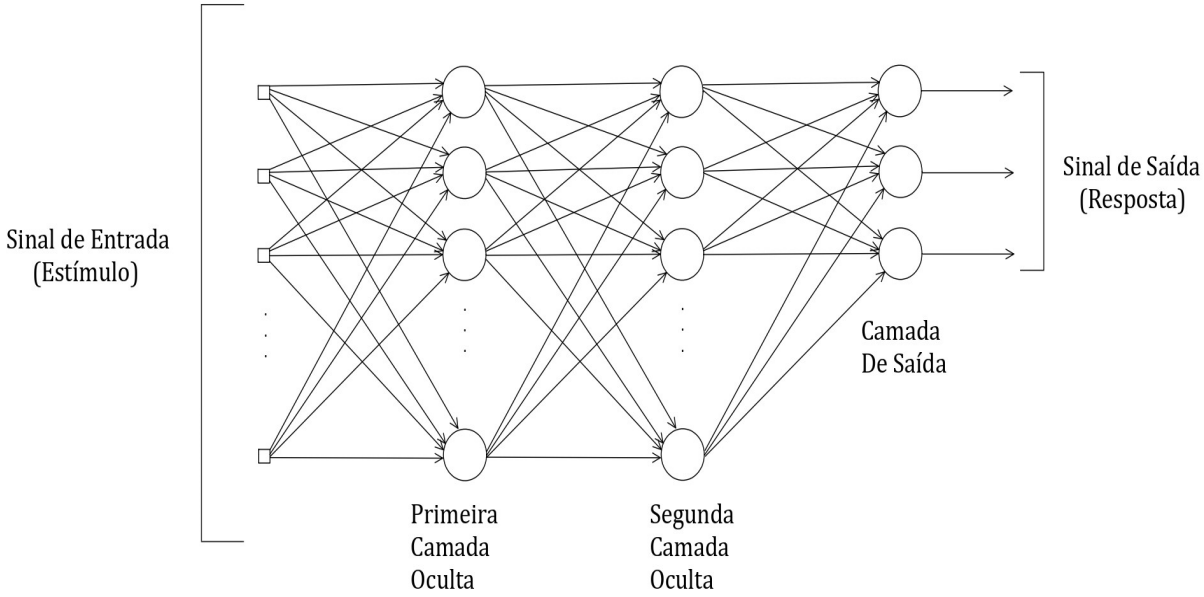
Fonte: Adaptado de (Haykin, 1998).

De acordo com (Haykin, 1998), um modelo de um neurônio aplicado à não linearidade limita o espaço de aplicação dessas redes. Uma rede neural *feedforward* distingue-se das demais pela composição de uma ou mais camadas ocultas, em que a

camada oculta refere-se ao pedaço da rede que não é visto na entrada ou na saída. Com a composição de camadas ocultas, a rede é habilitada a extrair estatísticas de ordem maior da entrada.

O *Multilayer Perceptrons* (MLP) consiste em três componentes principais. Cada neurônio, no modelo, utiliza uma função de ativação não linear, e há uma ou mais camadas intermediárias entre os neurônios de entrada e de saída. Essas camadas intermediárias são altamente interconectadas, com os sinais sendo transmitidos por meio dos pesos sinápticos de uma camada para a próxima até chegarem à camada de saída. Durante a fase de retropropagação, o erro calculado na saída é propagado de volta através da rede, ajustando os pesos sinápticos conforme necessário para minimizar o erro (Haykin, 1998).

Figura 4 – MLP



Fonte: Adaptado de (Haykin, 1998).

2.1.4.3 Rede Neural Recorrente

De acordo com (Gomes, 2005), uma RNN (*Recurrent Neural Network*) se assemelha à estrutura de uma rede *feedforward*, a qual apresenta uma modificação em sua estrutura, em que as camadas passam a ter retroalimentação. As retroalimentações podem pegar as saídas de neurônios de uma determinada camada e inseri-las novamente como entrada ao modelo nas camadas iniciais ou nele mesmo.

2.1.4.4 Long Short-Term Memory

Segundo (Graves; Schmidhuber, 2005), uma variação das redes neurais recorrentes é a arquitetura LSTM (*Long Short-Term Memory*), projetada para superar as

limitações do aprendizado de dependências temporais, devido ao desaparecimento ou explosão dos gradientes no processo de retropropagação.

De acordo com (Graves; Schmidhuber, 2005), uma camada LSTM é formada por blocos de memória interconectados, que funcionam como unidades para armazenar informações. Cada bloco pode conter uma ou mais células de memória, juntamente com três portões: de entrada, de saída e de esquecimento, cuja funcionalidade é realizar o controle sobre as informações armazenadas.

2.1.4.5 Decomposição STL

De acordo com (Pessanha; Almeida, 2021), uma série temporal pode ser representada por três componentes principais: tendência, sazonalidade e sua componente irregular. A separação dessas componentes pode ser realizada utilizando-se o método STL (*Seasonal-Trend decomposition using Loess*), baseado em regressão local, estimando as suas componentes.

2.1.4.6 Transformada de Fourier

Segundo (Cerqueira *et al.*, 2000), a filtragem baseada na Transformada de Fourier consiste em converter um sinal do domínio do tempo para o domínio da frequência. Para isso, aplica-se a Transformada de Fourier direta, obtendo-se o espectro de frequências do sinal analítico.

2.1.4.7 Métricas de Avaliação

De acordo com (Lago *et al.*, 2021), no campo da avaliação da precisão de previsões, são comumente utilizadas métricas de avaliação como *Mean Absolute Error* (MAE, do português, Erro Absoluto Médio), *Root Mean Squared Error* (RMSE, do português, Raiz do Erro Quadrático Médio) e *Mean Absolute Percentage Error* (MAPE, do português, Erro Percentual Absoluto Médio), onde:

- N_d é o número de dias,
- $p_{d,h}$ é o valor observado no dia d e hora h,
- $\hat{p}_{d,h}$ é o valor predito no dia d e hora h.

A fórmula para o *Mean Absolute Error* (MAE) é dada por:

$$\text{MAE} = \frac{1}{24N_d} \sum_{d=1}^{N_d} \sum_{h=1}^{24} |p_{d,h} - \hat{p}_{d,h}| \tag{1}$$

A fórmula para o *Root Mean Squared Error* (RMSE) é dada por:

$$\text{RMSE} = \sqrt{\frac{1}{24N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24}\left(p_{d,h} - \hat{p}_{d,h}\right)^2} \tag{2}$$

A fórmula para o *Mean Absolute Percentage Error* (MAPE) é dada por:

$$\text{MAPE} = \frac{1}{24N_d}\sum_{d=1}^{N_d}\sum_{h=1}^{24}\frac{|p_{d,h} - \hat{p}_{d,h}|}{|p_{d,h}|} \tag{3}$$

Conforme as equações descritas acima, a comparação entre bases de dados distintas é complicada quando se utilizam erros absolutos. Nesse sentido, o MAE (1) e o RMSE (2) podem não ser muito úteis devido à falta de informação adicional. A interpretação de métricas baseadas em erros quadráticos torna-se difícil e, para problemas de previsão, essas métricas podem não representar com precisão a qualidade dos modelos, por exemplo, o RMSE (Lago *et al.*, 2021).

Segundo (Lago *et al.*, 2021), o erro médio absoluto em escala (MASE) é uma forma eficaz de analisar se as previsões foram menores ou maiores, sendo considerado um ótimo fator analítico de desempenho do modelo. A fórmula do MASE é descrita abaixo.

A fórmula para o Erro Médio Absoluto em Escala (MASE) é dada por:

$$\text{MASE} = \frac{1}{N}\sum_{k=1}^{N}\frac{|p_k - \hat{p}_k|}{\frac{1}{n-1}\sum_{i=2}^{n}|p_i - p_{i-1}|} \tag{4}$$

onde:

- p_k é o valor real observado,
- $\hat{p}_k$ é o valor previsto,
- N é o número total de observações,
- p_i e p_{i-1} são os valores consecutivos na série temporal,
- n é o número total de observações na série temporal.

O MASE é dependente da amostra de dados utilizada. Métodos de previsão com janelas variáveis consideram diferentes agrupamentos de dados na amostra, resultando em fatores de escala distintos para o MASE de cada modelo. Esse conceito também se aplica a modelos com e sem janelas de rolagem: diferentes janelas de ajuste podem tornar o MASE indefinido devido à variabilidade da janela de ajuste. Portanto, comparar séries temporais pode apresentar diversos problemas (Lago *et al.*, 2021).

As métricas de avaliação que calculam os erros do modelo permitem analisar a qualidade das previsões. Dessa forma, quanto menores forem os valores, melhor será o desempenho do modelo. Para as métricas que avaliam o ajuste do modelo, como o coeficiente de determinação e o POCID, os resultados são analisados de forma que, quanto maiores forem os valores obtidos, melhores serão a precisão e a aderência às variações dos dados.

2.1.5 Metodologia ágil

A metodologia ágil adotada neste trabalho foi o Scrum, que inclui o *Product Backlog* e a *Sprint*. (Pereira; Torreão; Marçal, 2007) explicam que o *Product Backlog* é como uma lista de compras em que cada item tem seu valor e importância. As *sprints*, por outro lado, são como etapas de uma receita, dividindo o trabalho em partes menores e mais fáceis de completar. Com o Scrum, podemos ser mais ágeis e flexíveis, respondendo às mudanças com rapidez e criatividade. É a maneira ideal de trabalhar em projetos complexos e inovadores, onde a colaboração e a adaptabili-dade são fundamentais.

2.2 Estado da arte

O primeiro trabalho citado é sobre Previsão de geração de energia fotovoltaica no Brasil por meio de modelos de aprendizado de máquina, desenvolvido por (Knupp, 2023). Esse trabalho consistiu em realizar a extração e o tratamento dos dados, sendo escolhidas duas técnicas de aprendizado de máquina: árvores de decisão e florestas aleatórias. A avaliação dos modelos gerados para a previsão de geração de energia fotovoltaica utilizou as métricas de avaliação como acurácia, erro médio absoluto (MAE) e erro médio quadrático (RMSE). Os resultados obtidos demonstraram que os modelos de árvores de decisão e florestas aleatórias foram promissores na previsão de geração de energia solar, demonstrando serem capazes de fornecer previsões precisas, com coeficiente de determinação acima de 60%.

O segundo trabalho citado é sobre Previsão da eficiência dos módulos fotovoltaicos utilizando técnicas de aprendizado de máquina, desenvolvido por (Santos, 2023). Esse trabalho consistiu na coleta dos dados, processamento da base de dados e aplicação de técnicas de aprendizado de máquina, selecionando-se três técnicas: Redes Neurais Artificiais (RNA), *Support Vector Machine* (SVM) e Regressão Linear (RL). Os modelos foram treinados com base nos dados históricos. Para a avaliação dos modelos para determinar a eficácia, foram utilizadas as métricas como Desvio Quadrático Médio (*Root Mean Squared Error*- RMSE), Erro Médio Absoluto (*Mean Absolute Error* - MAE), Erro Quadrático Médio (*Mean Squared Error* - MSE) e o Coeficiente de Determinação (R^2). O modelo RNA teve o maior desempenho.

O terceiro trabalho citado é Previsão de Geração de Energia Fotovoltaica Utilizando Método de Aprendizado de Máquina, desenvolvido por (Brolese, 2019). Este estudo envolveu uma análise detalhada e o processamento de uma base de dados. O autor utilizou a regressão *stepwise* para selecionar as variáveis mais relevantes, identificando aquelas que têm a maior correlação com a variável-resposta. O treinamento da rede neural foi realizado utilizando-se o método de aprendizado supervisionado, aplicando a técnica de retropropagação do erro e a função de treinamento conhecida como *bayesian regularization*. Esta abordagem foi escolhida com o objetivo de reduzir o erro nas previsões.

Os resultados mostraram um desempenho satisfatório do modelo, com um MAPE (erro percentual absoluto médio) de 12,97%. O modelo foi capaz de prever a geração de energia fotovoltaica para um período de seis meses, o que demonstra a eficácia das redes neurais na previsão desse tipo de geração de energia.

O quarto trabalho mencionado é Previsão de Geração de Energia Hidráulica no Brasil: um Estudo de Caso usando Redes Neurais Artificiais e Regressão Linear, desenvolvido por (Pimentel *et al.*, 2023). Neste estudo, foi utilizada uma base de dados de geração hidráulica fornecida pelo Instituto de Pesquisa Econômica Aplicada. Modelos de previsão, baseados em algoritmos de Regressão Linear Múltipla e Redes Neurais Artificiais, foram implementados no *software* WEKA. Os resultados obtidos dos dois modelos foram comparados por meio das métricas RMSE (*Root Mean Squared Error*), MAPE (*Mean Absolute Percent Error*) e MAE (*Mean Absolute Error*). Para um horizonte de curto prazo (seis meses), verificou-se que o modelo MLP apresentou melhor desempenho, com um Erro Percentual Médio Absoluto menor que o do modelo MLR (MAPE-MLP = 3,59% e MAPE-MLR = 5,7%).

O quinto trabalho mencionado é intitulado Modelos de Aprendizado de Máquinas Híbridos aplicados à previsão de curto prazo da vazão do Rio Zambeze afluente à barragem hidroelétrica de Cahora-Bassa, em Moçambique, desenvolvido por (Martinho, 2023). Concentra-se na criação de modelos híbridos para previsão de vazões naturais dos corpos de água do rio Zambeze, especificamente na barragem hidrelétrica de Cahora-Bassa. Para isso, foram utilizados valores de evaporação, vazões afluentes e umidade como variáveis de entrada, considerando cinco modelos para análise: *Extreme Gradient Boosting* (XGB), *Extreme Learning Machine* (ELM), *Support Vector Regression* (SVR), *Elastic Net linear* e (EN) *Multivariate Adaptive Regression Splines* (MARS). Foram empregados algoritmos de otimização evolutivos/bioinspirados, como *Grey Wolf Optimization* (GWO), Algoritmo Genético, *Differential Evolution* e *Particle Swarm Optimization*, para selecionar os parâmetros internos dos cinco modelos. Os resultados indicaram que todos os modelos obtiveram desempenhos significativos na previsão das vazões do rio, mostrando que a integração de algoritmos evolutivos e bioinspirados é uma alternativa eficaz para produzir previsões precisas. Entre os modelos híbridos XGB com GWO e suas variações com seleção de variáveis de entrada, XGB-LASSO e XGB-PMI superam os demais modelos hibridos descritos, evidenciando a melhor performace para previsões de 1,3,5 e 7 passos à frente.

O sexto trabalho citado é Aplicação de Técnicas de Aprendizagem de Máquinas na Previsão de Vertimento em Usinas Hidrelétricas, desenvolvido por (Nascimento, 2021). Este estudo se concentrou em aplicar métodos de aprendizado de máquina supervisionado para prever, com cinco horas de antecedência, como seria a operação de vertimento em uma usina hidrelétrica. Foram testados dois algoritmos principais: Floresta Aleatória (*Random Forest*) e Perceptron Multicamada (*Multilayer Perceptron*), além de combinar esses algoritmos para ver como trabalhariam juntos. A pesquisa envolveu processar os

dados necessários para treinar esses modelos. No final das contas, a técnica de *Random Forest* se destacou, mostrando um desempenho melhor que o do Perceptron Multicamada, independentemente de como os dados de treinamento foram tratados.

Com base nas pesquisas mencionadas anteriormente, o presente estudo está alinhado com o estado da arte ao empregar técnicas de previsão da energia armazenada em reservatórios hidrelétricos.

3 METODOLOGIA

O objetivo deste capítulo é descrever o desenvolvimento da metodologia, abordando as seguintes seções. Na Seção 3.1, apresenta-se a classificação da pesquisa. Na Seção 3.2, discute-se a solução do problema descrito. A Seção 3.3 detalha as tecnologias e materiais utilizados no desenvolvimento. Por fim, a Seção 3.4 descreve os métodos e procedimentos aplicados na solução do problema.

3.1 Classificação da pesquisa

O presente trabalho se caracteriza como uma pesquisa aplicada, pois busca gerar conhecimento prático e soluções para problemas específicos na geração de energia elétrica em reservatórios, alinhando-se à definição de (Moresi *et al.*, 2003). A pesquisa se alinha à quantitativa pelo fato de que emprega dados numéricos e técnicas estatísticas para analisar a geração de energia, incluindo métodos de Mineração de Dados para extrair conhecimento de grandes conjuntos de dados, conforme (Moresi *et al.*, 2003). O trabalho também se alinha a uma pesquisa qualitativa devido à discussão dos resultados, que tem por objetivo esclarecer as causas que fundamentam os padrões identificados nas análises estatísticas.

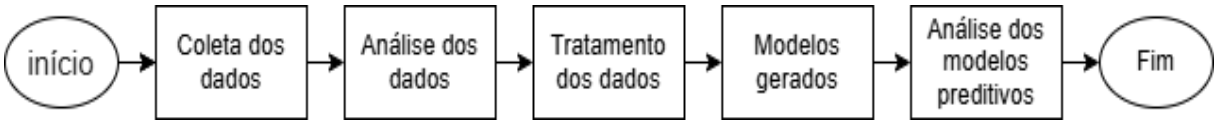
Já em relação aos objetivos da pesquisa, pode ser entendido como descritivo, pois visa descrever como a base de dados de previsão da produção de energia pelo volume de água presente nos reservatórios contribui para a tomada de decisões e identifica correlações entre as variáveis, caracterizando-se como pesquisa descritiva. Os meios de investigação da pesquisa são similares à pesquisa experimental pelo fato de que aplica técnicas de Mineração de Dados para analisar os modelos gerados.

3.2 Solução da Pesquisa

O projeto consiste no desenvolvimento de modelo preditivo utilizando técnica de aprendizado de máquina, com o uso do algoritmo de rede neural para a previsão de energia armazenada nos reservatórios da Região Sudeste/Centro-Oeste do país.

O desenvolvimento do modelo envolve a utilização da base de dados disponibilizada pelo ONS, em que são realizadas a análise e o tratamento desses dados, os quais serão empregados como *inputs* para alimentar o modelo preditivo. Após a etapa de treinamento, o modelo gerado será avaliado utilizando-se as métricas de avaliação. Esse processo será descrito no fluxo do processo de desenvolvimento do projeto, descrito na Figura 5, a seguir.

Figura 5 – Fluxo de desenvolvimento do projeto



Fonte: Elaborado pelo Autor, 2024.

O fluxo do desenvolvimento do projeto começa com a etapa inicial, dada pelo processo de obtenção dos dados, que é crucial para fornecer informações aos modelos de aprendizado de máquina (AM) e permitir a previsão da energia armazenada no reservatório. Após a obtenção dos dados, passa-se para a próxima etapa, que visa manter a integridade e a qualidade dos dados para o treinamento dos modelos.

O próximo ciclo, descrito pelo fluxograma da Figura 5, refere-se ao treinamento dos modelos, buscando encontrar as melhores configurações, por meio das métricas de avaliação e observando quais delas obtiveram os melhores resultados. Finalmente, o processo avança para a etapa de comparação entre os modelos, na qual se analisa qual modelo se destacou na previsão.

3.3 Materiais e Tecnologias

Na seção atual, apresentam-se as tecnologias e os materiais utilizados no desenvolvimento do trabalho. A Seção 3.3.1 aborda as tecnologias empregadas, enquanto a Seção 3.3.2 descreve o ambiente de experimentos e os equipamentos empregados.

3.3.1 Tecnologias

O *Visual Studio Code* é um editor de código utilizado para escrever e editar códigos. É multiplataforma e está disponível para os sistemas operacionais Linux, *macOS* e *Windows*. Oferece suporte a diversas linguagens de programação, auxiliando no desenvolvimento de códigos por meio de extensões (Microsoft, 2024). A instalação do *Visual Studio Code* pode ser realizada no *site* oficial[1].

Python é uma linguagem de programação de alto nível, interpretada e multiparadigma, que abrange paradigmas como funcional, imperativo e orientado a objetos, entre outros. Pode ser utilizada em diversas aplicações, incluindo a área de estudo do projeto, que contempla análise de dados e aprendizado de máquina (Python Software Foundation, 2023). A instalação da linguagem Python pode ser realizada a partir do *site* oficial[2].

Pandas é uma biblioteca do Python, de código aberto, que dispõe de estruturas de dados e ferramentas de análise de dados, tornando a manipulação destes mais interativa e fácil, auxiliando no processo de análise. As principais estruturas são o *DataFrame*

[1] https://code.visualstudio.com/
[2] https://www.python.org/

e a Series (The Pandas Development Team, 2024). Para mais informações, consulte a documentação no *site* oficial[3].

TensorFlow é uma biblioteca desenvolvida pela Google para facilitar o desenvolvimento de modelos de aprendizado de máquina. Ela oferece uma ampla gama de algoritmos e ferramentas que auxiliam no processo de treinamento e na produção de modelos (Martín Abadi *et al.*, 2015). Para mais informações, consulte a documentação no *site* oficial[4].

3.3.2 Ambiente de experimento e desenvolvimento

Todo o desenvolvimento deste trabalho, incluindo o processamento da base de dados, o treinamento dos modelos de aprendizado de máquina, a realização dos experimentos e a escrita do documento, foi realizado em um *notebook* pessoal com as seguintes configurações:

- **Modelo:** HP 246 G6 Notebook PC
- **Tipo de Sistema:** Sistema operacional de 64 bits, processador baseado em x64
- **Sistema Operacional:** Windows 10
- **Processador:** *Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz 2.71 GHz*
- **Memória RAM:** 8,00 GB
- **Unidade de Estado Sólido:** *WD Green M.2 2280 480GB*

Para a aplicação da pesquisa, foram utilizadas as seguintes ferramentas e suas respectivas versões:

- **Python:** 3.12.3
- **Visual Studio Code:** 1.92.0
- **Pandas:** 2.2.2
- **TensorFlow:** 2.16.1

3.4 Método de Trabalho e Diretrizes

Na seção atual, são apresentadas as etapas do processo de pesquisa. A Seção 3.4.1 descreve o processo de desenvolvimento da pesquisa, seguindo a metodologia ágil Scrum. A Seção 3.4.2 aborda a obtenção do conjunto de dados utilizado no treinamento

[3] https://pandas.pydata.org/
[4] https://www.tensorflow.org/

dos modelos. A Seção 3.4.5 detalha o processo de tratamento dos dados, visando melhorar os resultados do treinamento dos modelos. A Seção 3.4.8 explica a configuração do treinamento e arquitetura do modelo, com os parâmetros dos modelos. Por fim, a Seção 4 descreve os resultados obtidos no processo de treinamento e avaliação de cada técnica aplicada nos experimentos.

3.4.1 Planejamento da pesquisa

O desenvolvimento da pesquisa, conforme definido na seção de Referencial Teórico, segue a metodologia ágil Scrum, com a divisão das *sprints* em 12 meses. O *Product Backlog* deste trabalho é apresentado no Quadro 1.

Quadro 1 – Product Backlog

ID	**Tarefa**	**Tempo estimado (meses)**	**Peso (0 a 5)**
1	Revisão bibliográfica	12	5
2	Reuniões periódicas	12	5
3	Pré-processamento da base	1	5
4	Aplicação das Técnicas LSTM	1	5
5	Análise dos Resultados Obtidos	1	5
6	Escrita da Monografia	12	5
7	Defesa da Monografia	1	5

Fonte: Elaborado pelo Autor, 2024.

Conforme descrito no Quadro 1, ela é composta por um ID de identificação, tarefas a serem executadas até a conclusão da pesquisa, o tempo estimado para a realização de cada tarefa e o peso que indica o grau de importância delas. Todas as tarefas possuem peso igual a cinco, de modo que a produção deste documento depende da conclusão de todas as etapas anteriores.

A tarefa 1 descreve a revisão bibliográfica, com um prazo de duração de 12 meses, considerando a consulta a obras literárias e acadêmicas desde o início até o fim do desenvolvimento do documento. Na tarefa 2, realizaram-se reuniões periódicas semanais com o professor orientador para auxiliar na escrita do documento e no desenvolvimento da pesquisa, conciliando as melhores abordagens a serem aplicadas.

Para as tarefas 3 a 7, foi destinado 1 mês ou mais para a realização de cada etapa. Nesse período, efetuou-se o pré-processamento da base de dados disponível no site do ONS[5], a fim de verificar a consistência dos dados e realizar o tratamento necessário na tentativa de melhorar os resultados dos modelos e evitar qualquer influência negativa que pudesse ocorrer devido a dados nulos ou inconsistentes.

Na tarefa 4, realizou-se a implementação do modelo RNA, utilizando as bibliotecas descritas na Seção 3.3.1, para auxiliar no desenvolvimento, abstraindo a implementação

[5] https://dados.ons.org.br/dataset/ena-diario-por-subsistema

manual e focando no treinamento do modelo. Na tarefa 5, executou-se a análise dos resultados preditivos, observando-se as métricas alcançadas. Nas tarefas 6 e 7, finalizou-se a escrita do documento, sendo que, após sua conclusão e produção, efetuou-se a defesa do trabalho de conclusão de curso.

3.4.2 Coleta dos Dados

Nesta seção, apresenta-se o processo da coleta dos dados, ressaltando-se que o conjunto de dados está presente no *site* do ONS, no qual os dados estão disponíveis nos formatos CSV e XLS, referentes aos anos de 2000 a 2024, sendo atualizados diariamente.

O conjunto é composto por observações diárias, contendo informações relativas aos subsistemas Norte, Nordeste, Sul e Sudeste/Centro-Oeste. Coletados os dados referentes a cada ano de 2000 a 2024, foram agrupados em um único arquivo CSV, contendo apenas as informações pertinentes ao subsistema Sudeste/Centro-Oeste.

3.4.3 Categorização dos dados

Nesta seção, são descritos os dados presentes no conjunto de dados, o qual contém 6 atributos e um atributo-classe, descritos na Tabela 1, abaixo.

Tabela 1 – Descrição das colunas da base de dados

Índice	**Nome da Coluna**	**Tipo de Dado**
0	id_subsistema	object
1	nom_subsistema	object
2	ena_data	object
3	ena_bruta_regiao_mwmed	float64
4	ena_bruta_regiao_percentualmlt	float64
5	ena_armazenavel_regiao_mwmed	float64
6	ena_armazenavel_regiao_percentualmlt	float64

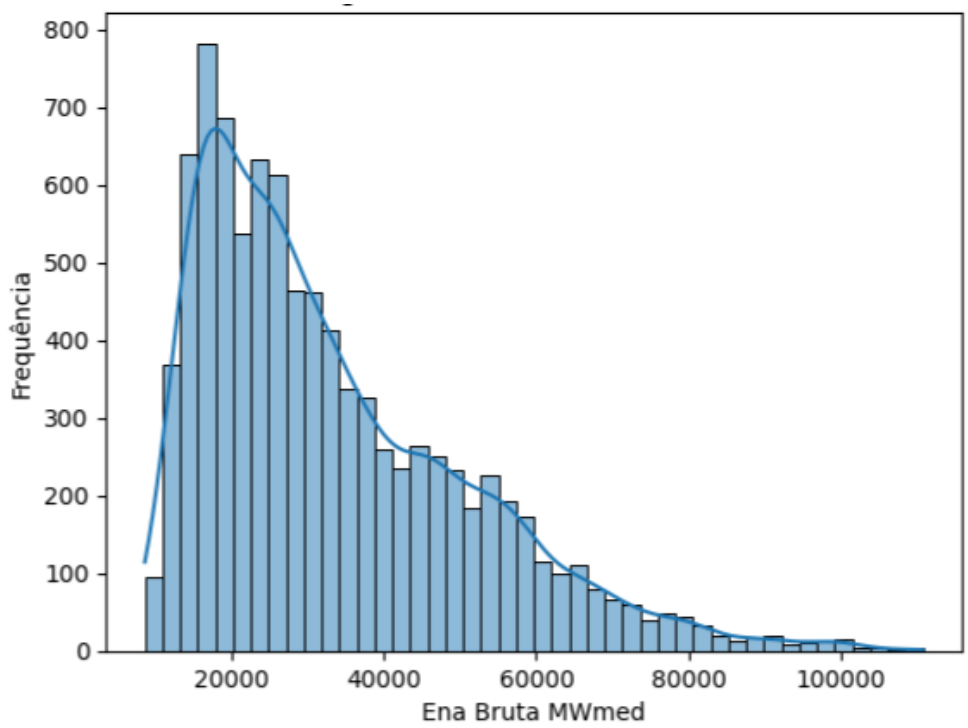
Fonte: Elaborado pelo autor, adaptado de (ONS. . . , 2024).

A Tabela 1 contém informações pertinentes ao conjunto de dados, no qual o índice 0 é representado pelo identificador de cada subsistema, e o mesmo vale para o índice 1, que representa, por extenso, o nome do sistema ao qual se refere. No índice 2, temos a coluna "data", que representa o dia da observação da coleta dos dados, sendo que os índices 3 e 4 representam os valores registrados da ENA bruta, que representa a energia que pode ser gerada a partir da vazão afluente, desconsiderando perdas e limitações operativas, variável-alvo que desejamos prever; no índice 4, é a representação dessa variável em percentual. Os índices 5 e 6 contêm a ENA Armazenada, representando a parte da ENA bruta que pode ser armazenada para geração de energia, descontando perdas e limitações operativas. No índice 6, está a representação dessa variável em percentual.

3.4.4 Análise Exploratória dos dados

Nesta seção, é realizada uma análise para se compreender as características e como os dados se distribuem, buscando entender se seguem uma distribuição normal. De acordo com (Miot, 2017), os dados que seguem uma distribuição normal têm como características sua forma de sino e o fato de possuírem uma única moda na qual se coincidem a média e a mediana, sendo que, é através do histograma que os dados são observados visualmente. Na Figura 6, pode-se notar uma semelhança a um sino, porém, é possível visualizar uma assimetria positiva, devido à presença de uma cauda que se estende para a direita no histograma. Além de aparentar uma única moda, não coincide com os valores da média e mediana.

Figura 6 – Histograma do Conjunto de Dados ENA Bruta MWMED



Fonte: Elaborado pelo Autor, 2025.

Para se ter maior certeza, foi utilizado método estatístico para determinar qual tipo de distribuição esse conjunto de dados segue, sendo escolhido o teste de Kolmogorov-Smirnov.

Tabela 2 – Resumo estatístico da série de dados de ENA bruta

Estatística	**Valor**
Média	33.038
Mediana	28.365
Desvio padrão	16.968
Assimetria	0,89
Teste de Kolmogorov-Smirnov (D)	0,11 ($p < 0,0001$)

Fonte: Elaborado pelo autor, 2025.

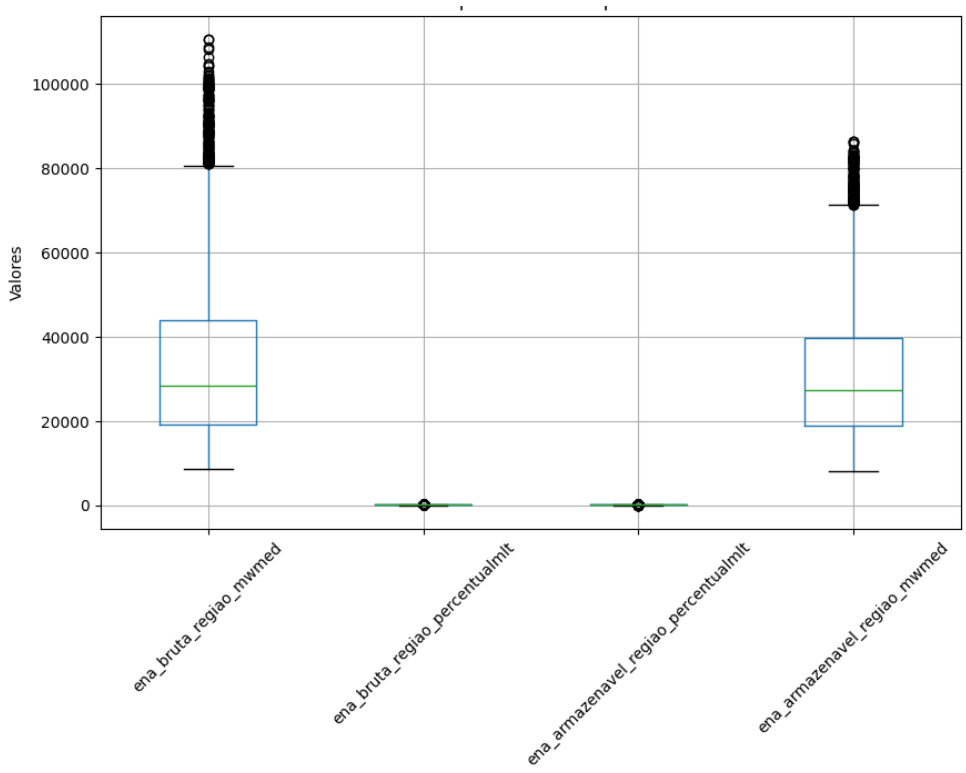
Com base nos dados apresentados, observa-se que a média difere da mediana, o que já indica uma possível assimetria na distribuição. A assimetria com valor maior que zero indica que a distribuição é positivamente inclinada, ou seja, há uma cauda mais longa à direita. Além disso, o teste de Kolmogorov-Smirnov demonstra que os dados não seguem uma distribuição normal, pois o p-valor obtido é menor que 0,05. Em resumo, os resultados indicam que os dados apresentam uma concentração maior de valores mais baixos, com desvios significativos em relação à normalidade.

3.4.5 Tratamento dos dados

Nesta seção, será abordado o processo de tratamento do conjunto de dados, após a finalização da coleta dos dados descritos na Seção 3.4.2, na qual será realizada a remoção de informações que não foram utilizadas para o treinamento do modelo, sendo removidas as colunas referentes ao Identificado e nome do subsistema.

De acordo com o processo de análise exploratória para identificar a característica desse conjunto de dados, percebe-se a necessidade de avaliar melhor e tratar esses valores extremos presentes na base. De acordo com (Neto *et al.*, 2017), o boxplot é uma técnica gráfica que pode ser adotada na análise dos dados, na qual ele permite identificar se o conjunto de dados contém valores extremos. Aplicando o gráfico boxplot no conjunto de dados, constatou-se a presença desses valores, como é possível visualizar na Figura 7, a seguir.

Figura 7 – Identificação de valores extremos no conjunto de dados por meio do gráfico box plot



Fonte: Elaborado pelo Autor, 2025.

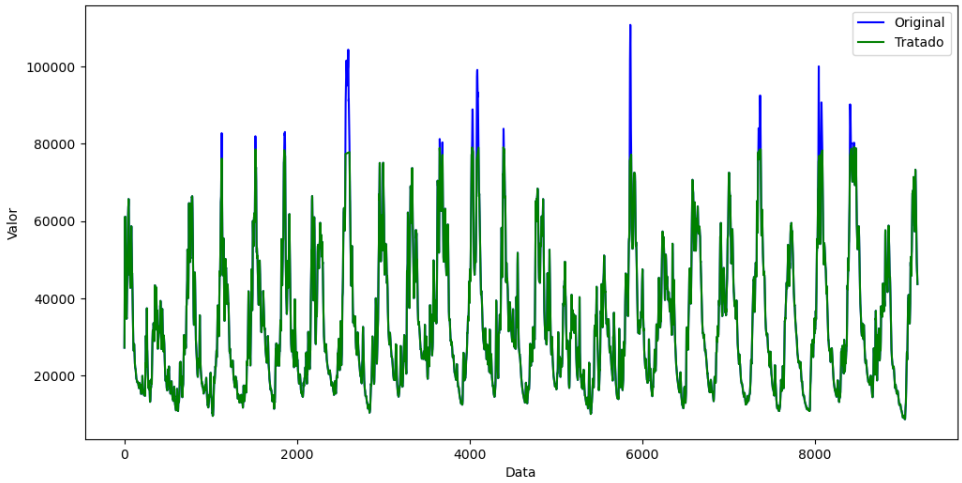
Após a identificação dos *outliers*, realizou-se a remoção desses valores, a fim de não prejudicar a previsão do modelo com esses valores. Com a remoção destes, quebramos a representação sequencial dessas observações diárias do conjunto de dados. Dessa forma, podemos utilizar métodos para estimar esses valores e preencher essas lacunas, sendo que o método proposto para este trabalho foi a interpolação linear.

$$P(x) = y_0 + \frac{(y_1 - y_0)}{x_1 - x_0}(x - x_0) \tag{5}$$

A interpolação linear se baseia em uma reta, assumindo que há uma variação constante entre esses dois pontos que a interligam. Essa reta conecta os pontos x0 e x1, cujos valores correspondentes y0 e y1 são conhecidos. Entre esses dois pontos, considera-se um valor intermediário x, para o qual desejamos estimar o valor de y, representado por P(x).

Após a aplicação da interpolação linear para estimar os novos valores removidos ao serem identificados como *outliers*, o conjunto de dados foi submetido novamente ao gráfico box plot, no qual é possível ver uma redução dos valores extremos que foram removidos e estimados por interpolação, como demonstrado na Figura 8.

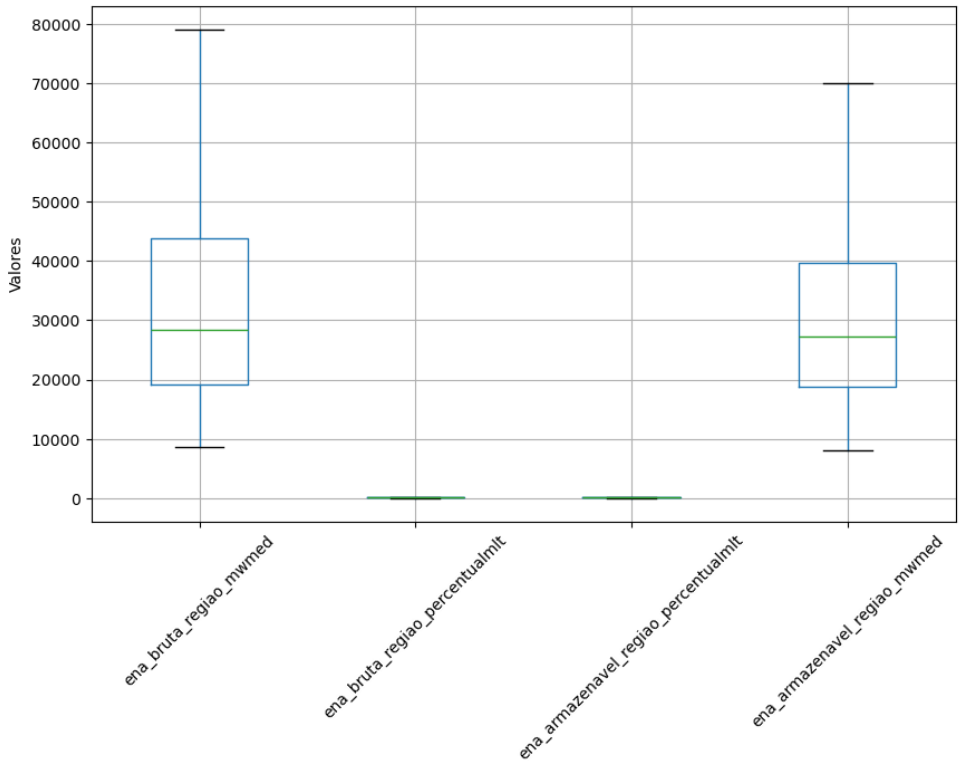
Figura 8 – Remoção de valores extremos no conjunto de dados



Fonte: Elaborado pelo Autor, 2025.

O novo conjunto de dados com o devido tratamento foi submetido novamente ao gráfico box plot, para realizar a verificação e confirmação da remoção dos valores extremos, que pode ser visualizado na Figura 9.

Figura 9 – Verificação da remoção de valores extremos no conjunto de dados



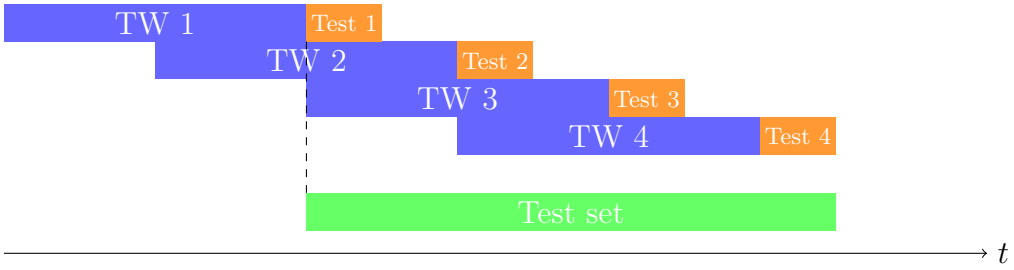
Fonte: Elaborado pelo Autor, 2025.

3.4.6 Janelamento

Nesta seção, vamos falar a respeito do janelamento dos dados, que foi aplicado neste trabalho.

No contexto de séries temporais, torna-se uma prática comum a divisão de dados em partes menores, fornecendo mais informação sequencial dos dados para o modelo, com intuito de capturar maior relação e padrões no decorrer do tempo dado pela sequência amostral. Nesse caso, este trabalho fez uso da técnica de janelas deslizantes, na qual se define uma quantidade de observações sequenciais para prever os próximos valores a serem seguidos, conforme o exemplo na Figura 10.

Figura 10 – Representação da técnica de janelas deslizantes



Fonte: Adaptado de (Markudova *et al.*, 2021).

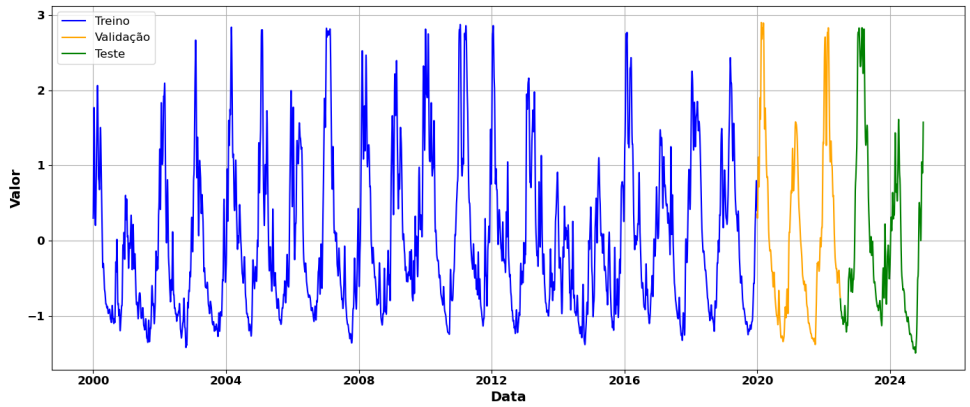
Neste trabalho, o intervalo definido de observações foi de 30 valores amostrais para prever os próximos 7 dias subsequentes, visando fornecer mais informação histórica ao modelo para que possa capturar os padrões e comportamentos na sequência temporal.

3.4.7 Particionamento dos Dados

Nesta seção, abordaremos o processo de divisão do conjunto de dados para o treinamento do modelo, o qual foi divido em três etapas: treinamento, validação e teste. Os conjuntos de dados foram utilizados para treinar o modelo com o conjunto de treinamento configurado com 80% dos dados, contendo informação referente a 1º de janeiro de 2000 até 30 de dezembro de 2019, contemplando 7304 observações diárias.

Para o conjunto de validação com 10% dos dados referentes ao intervalo de 1º de janeiro de 2020 até 30 de junho de 2022, houve 913 observações. No conjunto de teste, foi utilizado para ver a previsão do modelo e comparar a capacidade preditiva, representada por 10% dos dados, sendo composta pelos dados restantes após a última observação do conjunto validação até o último dia do ano de 2024, como é possível ver na Figura 11.

Figura 11 – Divisão do Conjunto de Dados: Treinamento, Validação e Teste



Fonte: Elaborado pelo Autor, 2024.

A normalização dos dados é uma etapa essencial no processo de treinamento do modelo no qual há presença de dados com multivariáveis. Cada informação pode ser representada em escalas muito diferentes, podendo atrapalhar no desempenho do modelo de aprendizado. Dada essa necessidade, foi utilizada a padronização que consiste em subtrair a média dos valores e dividi-los pelo desvio padrão, como demonstrado na Equação 6, a seguir.

$$z = \frac{x - \mu}{\sigma} \tag{6}$$

Esse método se mostra particularmente vantajoso, já que é menos suscetível aos efeitos de *outliers*. É importante ressaltar que o *StandardScaler* deve ser aplicado apenas ao conjunto de treinamento, garantindo, assim, que as informações dos demais dados não influenciem o processo de preparação (Géron, 2022).

A divisão dos dados foi realizada antes da aplicação da normalização. Para isso, utilizou-se o método *StandardScaler*, aplicado de forma independente ao conjunto de treinamento, e com o *StandScaler*, que foi aplicado ao conjunto de treinamento; somente assim, foi aplicado aos demais conjuntos de validação e teste.

3.4.8 Arquitetura

Nesta seção, abordaremos a respeito da arquitetura e configuração do modelo, falando sobre a arquitetura e as configurações utilizadas.

Neste trabalho, optamos por um modelo de redes neurais recorrentes com células LSTM. Para encontrar a configuração mais adequada, experimentamos diversas arquiteturas, adicionando ou removendo camadas até atingir o melhor desempenho.

Foi demonstrado melhor desempenho com poucas camadas LSTM e Densas para o modelo no qual a adição de muitas camadas não melhorava o desempenho da rede. Os pesos dos neurônios foram inicializados com uma semente a fim de facilitar os testes, para que os pesos iniciassem com os mesmos valores em todos eles.

A função de perda utilizada nos experimentos foi a Huber, que é a combinação das métricas MAE e MSE, utilizando a sensibilidade da métrica MSE para erros pequenos e MAE para erros grandes e *outliers*. A função é descrita a seguir, em que **a** representa o erro da predição ($a = y - \hat{y}$) e δ define o ponto de transição entre as métricas, sendo adotado $\delta = 0{,}5$.

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2, & \text{se } |a| \leq \delta \\ \delta\left(|a| - \frac{1}{2}\delta\right), & \text{se } |a| > \delta \end{cases} \tag{7}$$

A estrutura do modelo é apresentada na Tabela 3.

Tabela 3 – Estrutura das camadas do modelo

Camada	**Tipo**	**Neurônios**	**Função de Ativação**
1	LSTM	128	Tangente Hiperbólica (tanh)
2	LSTM	64	Tangente Hiperbólica (tanh)
3	Densa	128	$LeakyReLU(alpha{=}0.01)$
4	Densa	64	$LeakyReLU(alpha{=}0.01)$
5	Densa	7	Linear

Fonte: Elaborado pelo Autor, 2025.

A Função de ativação $LeakyReLU$ é uma variação da função $ReLU$ tradicional, onde, em vez de desativar completamente o neurônio, devido a entradas negativas ou iguais a zero, adiciona-se um pequeno valor *alpha*, permitindo que o neurônio permaneça ativo e contribuindo, mesmo que pouco.

$$\text{LeakyReLU}(x) = \begin{cases} x, & \text{se } x > 0 \\ 0{,}01x, & \text{se } x \leq 0 \end{cases} \tag{8}$$

A configuração do modelo conta com o otimizador *Adam*, definido com uma taxa de aprendizado inicial. Para tornar o processo de treinamento mais ajustável e adaptativo, foi utilizada uma função do módulo *callbacks*, e, da biblioteca *Keras*, a função *ReduceLROnPlateau*, que ajusta de forma dinâmica o valor da taxa de aprendizado pela metade caso a métrica monitorada não apresente melhoria num intervalo de 10 épocas consecutivas.

Tabela 4 – Configurações do modelo

Parâmetro	**Valor**
Tamanho da janela	30
Dias a prever	7
Taxa de aprendizado	0,001
Número de épocas	300
Tamanho do batch	32
Função de perda	Huber

Fonte: Elaborado pelo autor, 2025.

As configurações do modelo foram testadas para encontrar os melhores parâmetros, o que será descrito na seção de Resultados.

4 RESULTADOS

Nesta parte do trabalho, apresentaremos os resultados dos experimentos; a Seção 4.1 aponta quais configurações foram utilizadas no modelo e os resultados obtidos, além de explicar técnicas que auxiliaram nos referentes resultados. Ademais, na Seção 4.2, será abordada a discussão dos resultados, que descrevem os possíveis vieses e motivos que levaram à obtenção desses.

4.1 Resultados Experimentais

Os resultados obtidos nos experimentos foram baseados na arquitetura do modelo descrita na Seção 3.4.8. Os experimentos iniciais demonstraram que, com a utilização de uma camada LSTM e Densa, o modelo não conseguiu capturar os padrões da série, obtendo, dessa forma, resultados não significativos.

Com o acréscimo de mais camadas LSTM e Densa, possibilitou-se ao modelo performar melhor em relação à arquitetura minimalista. Foram realizados mais testes acrescentando camadas LSTM e Densa, o que não demonstrou melhorias significativas para os resultados. Dessa forma, adotou-se a arquitetura com 2 camadas LSTM e 2 camadas densas e uma densa de saída, a fim de reduzir o custo computacional.

4.1.1 Experimento 1

Nesse experimento, será descrita a execução do treinamento para configurar os hiperparâmetros, a fim de determinar a quantidade de valores de entrada e de *batch size* repassada ao modelo antes de atualizar os pesos para cada época. Os experimentos contavam com 3 tamanhos para a janela, sendo eles 15, 30 e 60 amostras para prever 7 valores futuros.

Para o parâmetro *batch size*, foram considerados 3 valores possíveis: 16, 32 e 64. O *batch size* igual a 8 também seria interessante para realizar os testes, pois efetua a atualização dos pesos a cada 8 janelas fornecidas para o modelo, porém a desvantagem é o aumento do custo computacional.

Os testes foram aplicados com uma quantidade máxima de 100 épocas e uma taxa de aprendizado igual a 0,0010 inicial, contendo os dados originais da base descrita na Seção 3.4.5. A coluna-alvo foi removida do conjunto de entrada para o modelo, e as métricas dos erros foram calculadas considerando-se as previsões acumuladas ao longo das janelas sobrepostas, devido ao janelamento de 1 posição por vez.

Tabela 5 – Resultados dos experimentos com *batch size* 16 e diferentes tamanhos de janela

Janela	MAE	MSE	RMSE	R^2	MAPE (%)	sMAPE (%)	MASE	POCID (%)
15	**2.7017**	**15.7990**	**3.9748**	**0.9815**	**3.5329**	**3.5181**	**2.6145**	75.6837
30	2.8125	16.5289	4.0656	0.9807	3.7016	3.7330	2.7006	75.1121
60	3.1391	16.9301	4.1146	0.9803	4.3174	4.2669	2.9757	**76.2232**

Fonte: Elaborado pelo autor, 2025.

Tabela 6 – Resultados dos experimentos com *batch size* 32 e diferentes tamanhos de janela

Janela	MAE	MSE	RMSE	R^2	MAPE (%)	sMAPE (%)	MASE	POCID (%)
15	3.0305	**15.4530**	**3.9310**	**0.9819**	4.3510	4.2923	2.9327	75.7576
30	**2.9319**	20.4760	4.5250	0.9761	**3.6229**	**3.7103**	**2.8153**	76.0837
60	3.6590	20.2350	4.4983	0.9764	5.4076	5.2724	3.4686	**76.9113**

Fonte: Elaborado pelo autor, 2025.

Tabela 7 – Resultados dos experimentos com *batch size* 64 e diferentes tamanhos de janela

Janela	MAE	MSE	RMSE	R^2	MAPE (%)	sMAPE (%)	MASE	POCID (%)
15	2.8172	**14.5611**	**3.8159**	**0.9830**	3.8675	3.8447	2.7263	**75.7576**
30	**2.7619**	15.5754	3.9466	0.9819	**3.6561**	**3.6547**	**2.6520**	75.0374
60	4.4764	28.7790	5.3646	0.9665	6.7139	6.4878	4.2434	73.9297

Fonte: Elaborado pelo autor, 2025.

Os resultados mostraram que o modelo não estava capturando bem os padrões da série, sendo ineficaz com os dados aparentes até o momento, necessitando de mais informações que complementassem o modelo e permitissem que ele capturasse melhor o comportamento da série.

Analisando as Tabelas 5, 6 e 7, podemos ver que os melhores resultados para cada tamanho de *batch size* foram com janela de tamanho 15, para *batch size* 16, e, para janelas de tamanho 30, para *batch size* 32 e 64, que obtiveram os melhores resultados na métrica MASE. Observando os resultados, um bom ponto de partida é a utilização das configurações com 16 ou 32 para *batch size* e tamanho de janela de 15 e 30. Os resultados obtidos até o momento não são discrepantes, o que pode dar mais liberdade de se optar por uma configuração mais ajustável. Os valores escolhidos para o treinamento do modelo foram com tamanho de janela 30, que permite passar maior quantidade de informações de entrada ao modelo, e um *batch size* de 32, que é um meio-termo para reduzir o custo computacional, mesmo que os ganhos sejam poucos.

4.1.2 Experimento 2

Os experimentos descritos na Seção 4.1.1 demonstraram que o modelo não estava capturando bem os padrões da série. Nesse cenário, o ideal é buscar novas formas de incrementar a entrada para o modelo com novas informações.

Uma abordagem inicial foi incrementar com variáveis defasadas da nossa variável-alvo com valores do passado, com defasagem de 7, 15 e 30 dias do passado. Essa abordagem visa fornecer ao modelo uma noção do comportamento passado da série a fim de permitir que ele conheça padrões sazonais, tendência e dependências temporais que não seriam capturados com a variável-alvo apenas.

Tabela 8 – Métricas de avaliação no conjunto de validação com variáveis defasadas

MAE	MSE	RMSE	R^2	MAPE (%)	sMAPE (%)	MASE	POCID (%)
1946.64	8274443.61	2876.53	0.9741	5.68	5.69	3.04	73.81

Fonte: Elaborado pelo autor, 2025.

Os resultados obtidos na Tabela 8 evidenciaram que as novas entradas não surtiram efeitos no processo de aprendizado, aumentando os valores das métricas em relação ao experimento 4.1.1.

4.1.3 Experimento 3

Os resultados obtidos nos experimentos 4.1.1 e 4.1.2 demonstraram dificuldade do modelo em extrair informações da série. Diante disso, foi aplicada a técnica de transformada de Fourier, que permite representar a série no domínio da frequência.

Ao aplicar a transformada e plotar o espectro, observou-se uma concentração significativa de energia em frequências baixas, próximas de zero. Essas informações evidenciam as componentes de baixa frequência, concentradas em torno de zero, as quais indicam a presença de tendências de longo prazo e sazonalidades no sinal. As frequências altas estão associadas a mudanças rápidas, como variações bruscas e anomalias de curto prazo.

Dessa forma, as frequências altas, em torno de 0,08 ciclos por dia, descrevem variações de curto prazo, anomalias ou mudanças bruscas.

A relação entre frequência f e período T é:

$$T = \frac{1}{f} \tag{9}$$

$$T = \frac{1}{0{,}08} = 12{,}5 \text{ dias} \tag{10}$$

Dessa forma, a fim de remover ruídos e anomalias de curto prazo, foram utilizadas médias móveis no sinal-alvo, atualizando o sinal da variável-alvo a ser predita, suavizando anomalias e variações bruscas.

Foram utilizados 2 parâmetros de janela para a suavização do sinal, empregando médias móveis de 7 e de 15 dias, a fim de comparar os efeitos e ganhos no modelo. Os experimentos seguiram as características do experimento 4.1.1. A utilização da média móvel deve ser casual, restringindo acesso à informação do futuro, ou seja, considerar apenas valores do passado da série temporal. Esse processo é fundamental para garantir a integridade do treinamento, evitando vazamento de informação futura para o modelo durante o processo de treinamento.

Os resultados obtidos são descritos na Tabela 9, a seguir.

Tabela 9 – Métricas de avaliação no conjunto de validação Médias Móveis

Dias	MAE	MSE	RMSE	R^2	MAPE (%)	sMAPE (%)	MASE	POCID (%)
7	1.1555	3.1343	1.7704	0.9963	1.4832	1.4817	1.3647	89.4974
15	0.7481	1.5580	1.2482	0.9982	1.0472	1.0446	1.1009	93.5094

Fonte: Elaborado pelo autor, 2025.

Os resultados alcançados evidenciaram ganho significativo na capacidade de predição do modelo, porém a utilização da suavização do sinal remove componentes importantes de curto prazo, o que pode ser prejudicial à previsão.

4.1.4 Experimento 4

Diante dos experimentos anteriores, o experimento 4.1.3 demonstrou ser um bom caminho para se prever a variável-alvo; porém, aplicando-se a suavização por médias móveis, acaba removendo informações importantes do sinal.

Neste experimento, é proposta a aplicação da decomposição STL no sinal, a fim de separar o sinal em tendência, sazonalidade e resíduo, capturando as componentes que descrevem o sinal em um determinado período. Duas configurações de periodicidade foram testadas:

- **Período de 365 dias**: para capturar os padrões anuais;
- **Período de 7 dias**: para capturar os padrões semanais.

Dessa forma, o algoritmo de decomposição realiza a decomposição capturando ciclos anuais para período de 365 dias e semanais para período de 7 dias, evidenciando variações que se repetem a cada ciclo.

O treinamento iniciou aplicando-se a divisão do conjunto em treino, validação e teste. Após a divisão, foram criadas novas variáveis em cada conjunto e aplicada a decomposição do sinal. Dessa forma, foram incrementadas ao conjunto de dados as variáveis tendência, sazonalidade e ruído, aplicando de forma isolada em cada conjunto.

Após a aplicação da técnica, foi realizado o treinamento do modelo, no qual os resultados obtidos são demonstrados na Tabela 10 .

Tabela 10 – Métricas de avaliação no conjunto Validação, com decomposição STL

Período	MAE	MSE	RMSE	R^2	MAPE (%)	sMAPE (%)	MASE	POCID (%)
7	173.18	73293.88	270.72	0.99	0.51	0.51	0.27	92.08
365	366.18	275188.97	524.58	0.98	1.17	1.17	2.39	76.17

Fonte: Elaborado pelo Autor, 2025.

Os resultados obtidos decompondo o sinal no período de 7 dias foram melhores. Utilizando a decomposição de 7 dias, permite-se à série capturar as variações de curto prazo, o que auxilia o modelo a descrever as variações de curto prazo, fornecendo essas informações, sem as quais a técnica aplicada não conseguiria capturar sozinha. Porém, vale evidenciar que a utilização dessa técnica pode ser considerada vazamento de informação, mesmo que aplicada nos conjuntos separados. Utilizando-se a técnica diretamente sobre o sinal, permite-se ao modelo ter acesso a informações do futuro, pois, para decompor o sinal, ele tem acesso a todo instante da série, olhando tanto para o passado quanto para o futuro, para separar os padrões.

Vale ressaltar que, no conjunto de teste, o modelo conseguiu bons resultados, evidenciando que, mesmo com a possibilidade de vazamento de informação, o modelo conseguiu generalizar bem em relação aos conjuntos de validação e teste. Os resultados são descritos na tabela abaixo para o período de 7 dias.

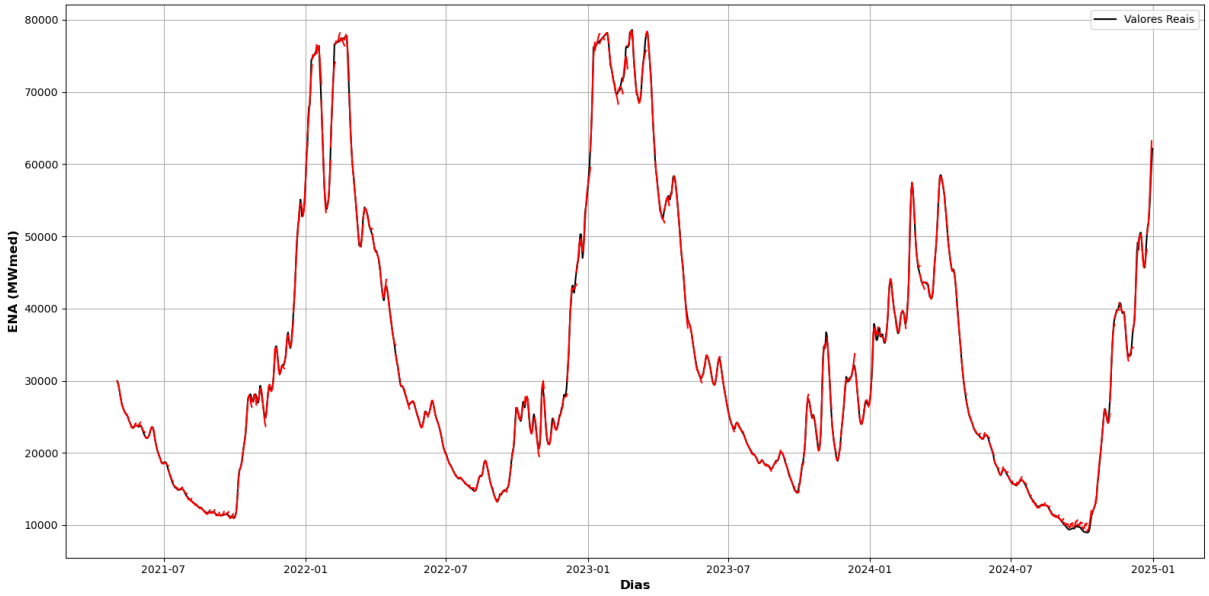
Tabela 11 – Métricas de avaliação no conjunto Teste, com decomposição STL

Período	MAE	MSE	RMSE	R^2	MAPE (%)	sMAPE (%)	MASE	POCID (%)
7	199.35	95231.80	308.59	0.99	0.71	0.70	0.32	89.84

Fonte: Elaborado pelo autor, 2025.

A Figura 12 demonstra a previsão do modelo em relação ao conjunto teste, no qual os dados reais estão na cor preta, e os valores previstos pelo modelo estão em vermelho.

Figura 12 – Valores reais e previstos no conjunto teste.



Fonte: Elaborado pelo Autor, 2025.

4.2 Discussão dos Resultados

Os resultados obtidos no experimento 4.1.1 demonstram que os dados utilizados para treinar o modelo, por si só, não conseguem extrair as informações necessárias para aprender conforme o comportamento desejado. Como expresso pelas métricas, observou-se que o modelo apresentou dificuldades, uma vez que a métrica MASE registrou valores superiores a 2 e o POCID obteve acurácia inferior a 80% na identificação da tendência. A dificuldade do modelo em capturar os padrões da série pode ser devido à presença de anomalias ou variações bruscas no sinal, o que torna o processo de aprendizado incapaz.

A adição de variáveis defasadas demonstrou ser ineficaz no experimento 4.1.2, uma vez que a presença dessas variações de curto prazo dificulta o modelo de capturar bem a tendência e padrões sazonais. Esse processo acaba por introduzir maior complexidade ao modelo, elevando os erros das métricas e resultando em um desempenho inferior ao observado no experimento 4.1.1, o que indica que o acréscimo de novas variáveis não contribuiu positivamente para a melhoria dos resultados.

No experimento 4.1.3, os resultados obtidos utilizando médias móveis, para remoção de variações de curto prazo, foram benéficos para a previsão, ao remover variações de curto prazo de 7 e 15 dias, suavizando o sinal, deixando mais evidentes as componentes de médio a longo prazo, devido à redução dos erros das métricas MAE, MSE e MASE com valor próximo de 1, em comparação ao modelo ingênuo, e ao aumento das métricas R^2 e POCID, que atingiu valor superior a 90%.

O experimento 4.1.4 demonstrou-se benéfico ao modelo, utilizando a decomposição STL com período de 7 dias, permitindo dar acesso ao modelo sobre informações de

curto prazo. Isso, de fato, auxiliou o processo de aprendizado, evidenciando as variações bruscas e anomalias presentes no sinal. Dessa forma, a utilização da técnica de decomposição é uma abordagem vantajosa no processo de previsão; porém, uma sugestão para evitar o vazamento de informação é a utilização da decomposição em janelas, permitindo ao modelo ter acesso somente ao passado até o instante t, o que pode demandar um grande custo computacional.

5 CONCLUSÃO

A proposta deste trabalho visou à previsão de energia natural afluente em reservatórios hidrelétricos da Região Sudeste/Centro-Oeste, utilizando-se modelo de aprendizado de máquina para prever sete dias à frente.

Este trabalho teve como objetivos a coleta da base de dados, realizada no *site* do ONS, como descrito na Seção 3.4.2, coletando-se os dados referentes aos anos de 2000 a 2024. Além disso, realizar a análise sobre a base de dados, verificando a possibilidade de valores ausentes e *outliers*, efetuando-se o tratamento, como descrito na Seção 3.4.5.

Após a realização das etapas descritas acima, foram realizados o desenvolvimento do modelo e o treinamento, cujos resultados foram abordados na Seção de resultados 4. Os resultados obtidos demonstraram dificuldades do modelo em capturar os padrões e alcançar bons resultados; dessa forma, destaca-se o experimento 4.1.3, que utilizou as técnicas de médias móveis para suavizar o sinal a ser previsto, obtendo um bom resultado em relação aos experimentos 4.1.1 e 4.1.2. O experimento 4.1.3 alcançou bons resultados nas métricas MASE (1.1009) e POCID (93.5094), para médias móveis, suavizando o sinal em relação a 15 dias, e, para médias móveis, suavizando o sinal em relação a 7 dias, foram obtidos os resultados para as métricas MASE (1.3647) e POCID (89.4974).

Os achados obtidos no experimento 4.1.4 demonstraram bons resultados nas métricas MASE e POCID utilizando-se a decomposição STL. Fornecendo as componentes do sinal, foi possível repassar informações ao modelo sobre as variações de curto prazo, o que o auxiliou a compreender informações que, somente com os dados originais, o modelo não conseguiria. Vale evidenciar que o modelo conseguiu ter bons resultados em todos os experimentos, capturando bem a sua tendência.

Dessa forma, ressalta-se que a necessidade de informações extras para descrever comportamentos do sinal tornou-se útil, como descrito no experimento 4.1.4. Porém, vale destacar a possibilidade de vazamento de informação. Ao aplicar a decomposição STL sobre o sinal, capturando informações tanto do passado quanto do presente, a técnica utiliza a série inteira para realizar a decomposição. Dessa forma, dá acesso a informações futuras sobre o conjunto de treinamento.

5.0.1 Trabalhos futuros

Para trabalhos futuros, destaca-se a utilização de decomposição STL, utilizando janelas a fim de evitar a possibilidade de vazamento de informação ao modelo, utilizando informações do passado, o que pode auxiliar no aprendizado do modelo, evitar vazamento de informação e ter possíveis ganhos nos resultados. Outra abordagem é experimentar novos horizontes de previsão, como 3 dias, havendo a possibilidade de ter alguma melhora nos resultados. Avaliar outra arquitetura ou modelos como Redes Neurais Convolucionais, Prophet, são modelos que podem ter resultados relevantes na previsão do sinal.

Acrescentar novas variáveis ao modelo, como variáveis climáticas, pode auxiliar o modelo na capacidade de previsão, sendo uma abordagem que pode explicar os componentes de ruído presente no sinal.

Uma abordagem sugerida é a utilização das técnicas utilizadas no artigo (Yousefi *et al.*, 2022), que demonstrou bons resultados para dados de vazões em hidrelétricas, aplicando sobre os dados da ENA, de modo a avaliar os possíveis ganhos nos resultados.

REFERÊNCIAS

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL). **Energia Assegurada**. Brasília: Agência Nacional de Energia Elétrica (ANEEL), 2005. p. 18. (Cadernos Temáticos ANEEL, 3). Inclui bibliografia. Disponível em: https://biblioteca.aneel.gov.br/Busca/Download?codigoArquivo=177338&tipoMidia=0. Acesso em: 29 jul. 2024.

AMORIM, M. J.; BARONE, D.; MANSUR, A. U. Técnicas de aprendizado de máquina aplicadas na previsao de evasao acadêmica. *In:* 1. BRAZILIAN Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2008. v. 1, p. 666–674.

BARROS, D. P. **A (des) construção dos modelos regulatórios no setor de energia elétrica do Brasil: instabilidades, incertezas e a reforma institucional de 2004**. 2005. Tese (Doutorado).

BROLESE, R. R. Previsão de geração de energia fotovoltaica utilizando método de aprendizado de máquina, 2019.

CAMPAGNOLI, F.; DINIZ, N. C. **Gestão de reservatórios de hidrelétricas**. 1. ed. São Paulo: Oficina de Textos, 2012. E-book. Disponível em: https://plataforma.bvirtual.com.br. Acesso em: 22 jul. 2024.

CERQUEIRA, E. O. *et al.* Utilização de filtro de transformada de Fourier para a minimização de ruídos em sinais analíticos. **Química Nova**, SciELO Brasil, v. 23, p. 690–698, 2000.

EMPRESA DE PESQUISA ENERGÉTICA. **Balanço Energético Nacional 2023: Ano base 2022 / Brazilian Energy Balance 2023 Year 2022**. 274 p. : 182 il. ; 23 cm. 2023. Disponível em: https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-748/topico-687/BEN2023.pdf. Acesso em: 17 jul. 2024.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37–37, 1996.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems**. "O'Reilly Media, Inc.", 2022.

GOMES, D. T. Redes neurais recorrentes para previsão de séries temporais de memórias curta e longa. **Master's thesis, Department of Statistics, Campinas State University, Campinas, Brazil**, p. 153, 2005.

GRAVES, A.; SCHMIDHUBER, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. **Neural Networks**, v. 18, n. 5, p. 602–610, 2005. IJCNN 2005. ISSN 0893-6080. DOI: https://doi.org/10.1016/j.neunet.2005.06.042. Disponível em: https://www.sciencedirect.com/science/article/pii/S0893608005001206.

HAYKIN, S. **Neural networks: a comprehensive foundation**. Prentice Hall PTR, 1998.

KNUPP, H. K. Previsão de geração energia fotovoltaica no Brasil por meio de modelos de aprendizado de máquina. **Trabalho de Conclusão de Curso (Bacharelado em Engenharia Mecânica)-Instituto Politécnico, Universidade Federal do Rio de Janeiro, Macaé**, 2023.

LAGO, J. *et al.* Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. **Applied Energy**, v. 293, p. 116983, 2021. ISSN 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2021.116983. Disponível em: https://www.sciencedirect.com/science/article/pii/S0306261921004529.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015.

LOPES, J. E. G.; SANTOS, R. C. P. Capacidade de reservatórios. **São Paulo: Escola Politécnica da Universidade de São Paulo**, 2002.

LOPES, V. S.; BORGES, C. L. Impact of the combined integration of wind generation and small hydropower plants on the system reliability. **IEEE Transactions on Sustainable Energy**, IEEE, v. 6, n. 3, p. 1169–1177, 2014.

MARKUDOVA, D. *et al.* Preventive maintenance for heterogeneous industrial vehicles with incomplete usage data. **Computers in Industry**, v. 130, p. 103468, 2021. ISSN 0166-3615. DOI: https://doi.org/10.1016/j.compind.2021.103468. Disponível em: https://www.sciencedirect.com/science/article/pii/S0166361521000750.

MARTÍN, J. *et al.* Energy storage technologies for electric applications. **Renewable Energy and Power Quality Journal**, p. 593–598, maio 2011. DOI: 10.24084/repqj09.398.

MARTÍN ABADI *et al.* **TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**. 2015. Software available from tensorflow.org. Disponível em: https://www.tensorflow.org/.

MARTINHO, A. D. Modelos de aprendizado de máquinas híbridos aplicados à previsão de curto prazo da vazão do Rio Zambeze afluente à barragem hidroelétrica de Cahora-Bassa, em Moçambique, 2023.

MICROSOFT. **Visual Studio Code**. 2024. Acessado em: 13 ago. 2024. Disponível em: https://code.visualstudio.com/.

MIOT, H. A. **Avaliação da normalidade dos dados em estudos clínicos e experimentais**. v. 16. SciELO Brasil, 2017. p. 88–91.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.

MORESI, E. *et al.* Metodologia da pesquisa. **Brasília: Universidade Católica de Brasília**, v. 108, n. 24, p. 5, 2003.

NASCIMENTO, P. H. M. Aplicação de técnicas de aprendizagem de máquinas na previsão de vertimento em usinas hidrelétricas, 2021.

NETO, J. V. *et al.* Boxplot: um recurso gráfico para a análise e interpretação de dados quantitativos. **Revista Odontológica do Brasil Central**, v. 26, n. 76, 2017.

ONS. Energia Natural Afluente (ENA) Diária por Subsistema. Base de dados. 2024. Disponível em: https://dados.ons.org.br/dataset/ena-diario-por-subsistema. Acesso em: 17 mar. 2024.

PEREIRA, P.; TORREÃO, P.; MARÇAL, A. S. Entendendo Scrum para gerenciar projetos de forma ágil. **Mundo PM**, v. 1, n. 14, p. 64–71, 2007.

PESSANHA, J. F.; ALMEIDA, V. A. de. AJUSTE SAZONAL DIÁRIO NO CÁLCULO DE SÉRIES TEMPORAIS DE CARGA ELÉTRICA LIVRES DOS EFEITOS CALENDÁRIO E TEMPERATURA. Galoá, 2021.

PIMENTEL, C. C. *et al.* Previsão de geração de energia hidráulica no Brasil: um estudo de caso usando redes neurais artificiais e regressão linear. **Caderno Pedagógico**, v. 20, n. 10, p. 4568–4582, 2023.

PYTHON SOFTWARE FOUNDATION. **Python Documentation**. 2023. Accessed: 2024-07-22. Disponível em: https://docs.python.org/3/.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, IBM, v. 3, n. 3, p. 210–229, 1959.

SANTOS, R. G. d. Previsão da eficiência dos módulos fotovoltaicos utilizando técnicas de aprendizado de máquina, 2023.

SHUMWAY, R. H.; STOFFER, D. S.; STOFFER, D. S. **Time series analysis and its applications**. Springer, 2000. v. 3.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao datamining: mineração de dados**. Ciência Moderna, 2009.

THE PANDAS DEVELOPMENT TEAM. **Pandas**. 2024. Accessed: 2024-08-13. Disponível em: https://pandas.pydata.org/.

WHITBY, B. **Artificial intelligence**. The Rosen Publishing Group, Inc, 2009.

YOUSEFI, M. *et al.* Day-ahead inflow forecasting using causal empirical decomposition. **Journal of Hydrology**, Elsevier, v. 613, p. 128265, 2022.