

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE ENGENHARIA
CÂMPUS DE ILHA SOLTEIRA

MATEUS BERTIPAGLIA TIRITAN

**ANÁLISE COMPARATIVA DE TÉCNICAS DE *MACHINE LEARNING* APLICADOS
À PREVISÃO DE PREÇOS DE ENERGIA ELÉTRICA NO MERCADO ELÉTRICO
PERUANO**

MATEUS BERTIPAGLIA TIRITAN

**ANÁLISE COMPARATIVA DE TÉCNICAS DE *MACHINE LEARNING* APLICADOS
À PREVISÃO DE PREÇOS DE ENERGIA ELÉTRICA NO MERCADO ELÉTRICO
PERUANO**

Trabalho de Graduação apresentado à
Faculdade de Engenharia do Câmpus de Ilha
Solteira – UNESP como parte dos requisitos
para obtenção do título de Bacharel em
Engenharia Elétrica.

Prof. Dr. Leonardo Brain Gracia Fernández
Orientador

FICHA CATALOGRÁFICA
Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

T597a Tiritan, Mateus Bertipaglia.
Análise comparativa de técnicas de machine learning aplicados à previsão de preços de energia elétrica no mercado elétrico peruano / Mateus Bertipaglia Tiritan. -- Ilha Solteira: [s.n.], 2025
72 f. : il.

Trabalho de conclusão de curso (Graduação em Engenharia Elétrica) - Universidade Estadual Paulista (UNESP), Faculdade de Engenharia, Ilha Solteira, 2025

Orientador: Leonardo Brain Gracia Fernández

Inclui bibliografia

1. Previsão de preços. 2. Energia elétrica. 3. Machine learning. 4. SVM. 5. LSTM. 6. ARTMAP Fuzzy.

ATA DE DEFESA DE TRABALHO DE GRADUAÇÃO

Aos dezessete dias do mês de junho do ano de dois mil e vinte e cinco, o discente **Mateus Bertipaglia Tiritan**, matriculado sob o nº 191051667, tendo como banca examinadora o seu orientador, o *Dr. Leonardo Brain García Fernández*, o *Prof. Dr. Carlos Roberto Minussi* e a *Doutoranda Mayla Awana dos Santos*, apresentou o Trabalho de Graduação intitulado **"ANÁLISE COMPARATIVA DE TÉCNICAS DE MACHINE LEARNING APLICADOS À PREVISÃO DE PREÇOS DE ENERGIA ELÉTRICA NO MERCADO ELÉTRICO PERUANO"**, obtendo a nota 9,5 (Nove vírgula cinco) e conceito Aprovado.

Documento assinado digitalmente
 **LEONARDO BRAIN GARCIA FERNANDEZ**
Data: 17/06/2025 11:21:14-0300
Verifique em <https://validar.itl.gov.br>

Dr. Leonardo Brain García Fernández
- Orientador -



Mateus Bertipaglia Tiritan
- Discente -



Prof. Dr. Carlos Roberto Minussi
- Membro da Banca -

Documento assinado digitalmente
 **MAYLA AWANA DOS SANTOS**
Data: 17/06/2025 11:28:46-0300
Verifique em <https://validar.itl.gov.br>

Doutoranda Mayla Awana dos Santos
- Membro da Banca -

Impacto potencial desta pesquisa

A estrutura preditiva apresentada neste estudo possui potencial para apoiar decisões estratégicas dos agentes do setor elétrico, otimizar a gestão operacional dos sistemas por parte do operador nacional, e contribuir para a formulação de políticas públicas direcionadas ao segmento energético. Esses avanços podem trazer ganhos econômicos relevantes, como maior eficiência no uso da energia, redução dos custos operacionais e fomento ao desenvolvimento sustentável no país.

Potential impact of this research

The predictive framework presented in this study has the potential to support strategic decision-making by electricity market agents, enhance the operational management of power systems by the national system operator, and contribute to the development of public policies aimed at the energy sector. These advancements may lead to significant economic benefits, such as greater energy efficiency, lower operational costs, and promotion of sustainable development in the country.

Impacto potencial de esta investigación

La estructura predictiva presentada en este estudio tiene el potencial de apoyar la toma de decisiones estratégicas por parte de los agentes del mercado eléctrico, mejorar la gestión operativa de los sistemas por parte del operador nacional, y contribuir al desarrollo de políticas públicas orientadas al sector energético. Estos avances pueden generar beneficios económicos significativos, como una mayor eficiencia en el uso de la energía, reducción de los costos operativos y el fomento del desarrollo sostenible en el país.

AGRADECIMENTOS

Agradeço, com sinceridade, a todos que contribuíram para a realização deste trabalho.

Ao Prof. Dr. Leonardo Brain Gracia Fernández, pela orientação atenciosa, pelas críticas construtivas, pela paciência e dedicação em todas as etapas deste trabalho.

Aos meus pais Marilene e João pelo amor incondicional, apoio e tudo que fizeram por mim durante a trajetória da minha vida. Sem eles eu não chegaria onde estou.

A minha república, que considero como minha segunda família, onde cresci e aprendi muito e fiz amizades inesquecíveis.

As pessoas incríveis que conheci durante a trajetória da graduação, que carrego no coração.

À Universidade Estadual Paulista – UNESP, campus de Ilha Solteira, pelo ensino de excelência e pela formação acadêmica que levarei por toda a vida.

Encerro com gratidão a todos os colegas que, direta ou indiretamente, contribuíram com sugestões, revisões, trocas de ideias ou mesmo pelo companheirismo durante essa jornada.

RESUMO

A previsão do comportamento dos preços de energia elétrica em mercados de curto prazo é um desafio enfrentado por operadores e agentes do setor energético. Com a crescente disponibilidade de dados e o avanço das tecnologias computacionais, o uso de técnicas de *Machine Learning* tem ganhado espaço como alternativa às abordagens tradicionais de modelagem. Neste trabalho, foram aplicados três modelos distintos de *ML*: *Support Vector Machine* (SVM), *Long Short-Term Memory* (LSTM) e *ARTMAP Fuzzy* Modificada, com o objetivo de prever os preços de energia do mercado elétrico peruano. A base de dados utilizada foi extraída do site do operador do sistema peruano: *Comité de Operación Económica del Sistema* (COES), contendo registros de preços, carga e geração hídrica com resolução de 30 minutos entre agosto de 2022 e setembro de 2024. Após o pré-processamento e segmentação da base, os modelos foram treinados e avaliados em dois horizontes: previsão de 1 dia e 1 semana. Os resultados mostraram que a SVM obteve melhor desempenho para horizontes curtos, enquanto a rede ARTMAP demonstrou maior robustez nas previsões semanais, refletindo corretamente padrões típicos como os baixos preços durante a madrugada. Já a LSTM apresentou desempenho inferior, com previsões excessivamente suavizadas e correlação negativa em horizontes longos.

Palavras-chave: previsão de preços; energia elétrica; *machine learning*; SVM; LSTM; ARTMAP Fuzzy.

ABSTRACT

The forecasting of electricity prices in short-term markets is a challenge faced by operators and stakeholders in the energy sector. With the increasing availability of data and the advancement of computational technologies, the use of Machine Learning techniques has emerged as a promising alternative to traditional modeling approaches. In this study, three distinct *ML* models were applied: Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Modified Fuzzy ARTMAP, with the goal of forecasting electricity prices in the Peruvian electricity market. The dataset was obtained from the *Comité de Operación Económica del Sistema* (COES), the operator of the Peruvian electric system, and included records of prices, load, and hydroelectric generation at 30-minute intervals from August 2022 to September 2024. After preprocessing and dataset segmentation, the models were trained and evaluated over two forecasting horizons: one day and one week. The results showed that SVM performed better for short-term forecasts, while the ARTMAP network demonstrated greater robustness in weekly forecasts, correctly capturing typical patterns such as low prices during nighttime hours. LSTM showed inferior performance, producing overly smoothed forecasts and negative correlation over longer horizons.

Keywords: price forecasting; electric energy; machine learning; SVM; LSTM; Fuzzy ARTMAP

LISTA DE FIGURAS

Figura 1 - Potência instalada no SEIN 2024.	14
Figura 2 – Classificação do SVM	24
Figura 3 - Ilustração do hiperplano ótimo para padrões linearmente separáveis.	26
Figura 4 - Hiperplano de margem suave.	28
Figura 5 - Rede LSTM "muitos para muitos".	32
Figura 6 - Célula LSTM.	33
Figura 7 - Arquitetura ARTMAP <i>Fuzzy</i>	36
Figura 8 – Fluxograma da RNA ARTMAP <i>Fuzzy</i> Modificada	39
Figura 9 – Variação de preços de energia da base de dados.	41
Figura 10 – Variação da carga da base de dados.	42
Figura 11 – Variação da geração hídrica da base de dados.	42
Figura 12 – Validação do treino SVM.	52
Figura 13 – Previsão para um dia SVM.	53
Figura 14 – Previsão para a semana SVM.	54
Figura 15 – Validação do treino LSTM para previsão de um dia.	55
Figura 16 - Validação do treino LSTM para previsão de uma semana.	56
Figura 17 - Previsão para um dia LSTM.	58
Figura 18 - Previsão para a semana LSTM.	59
Figura 19 – Validação do treino ARTMAP <i>Fuzzy</i>	61
Figura 20 - Previsão para um dia ARTMAP <i>Fuzzy</i>	62
Figura 21 – Previsão para a semana ARTMAP <i>Fuzzy</i>	63

LISTA DE TABELAS

Tabela 1 – Separação de dados.	40
Tabela 2 – Métricas de validação do treino SVM.	52
Tabela 3 – Métricas da previsão 1 dia SVM.	53
Tabela 4 - Métricas da previsão 1 semana SVM.	54
Tabela 5 - Métricas de validação do treino LSTM para um dia.	56
Tabela 6 – Métricas de validação do treino LSTM para uma semana.	56
Tabela 7 - Métricas da previsão 1 dia LSTM.	58
Tabela 8 - Métricas da previsão 1 semana LSTM.	59
Tabela 9 - Métricas de validação do treino ARTMAP <i>Fuzzy</i>	61
Tabela 10 - Métricas da previsão 1 dia ARTMAP <i>Fuzzy</i>	62
Tabela 11 - Métricas da previsão 1 semana ARTMAP <i>Fuzzy</i>	63

LISTA DE SIGLAS E ABREVIATURAS

ARTMAP	<i>Adaptive Resonance Theory Map</i>
COES	<i>Comité de Operación Económica del Sistema Interconectado Nacional (Peru)</i>
FAM	<i>Fuzzy ARTMAP Modified</i>
LSTM	<i>Long Short-Term Memory</i>
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MATLAB	<i>Matrix Laboratory</i>
ML	<i>Machine Learning</i>
PCC	<i>Pearson Correlation Coefficient</i>
RMSE	<i>Root Mean Square Error</i>
SEIN	<i>Sistema Eléctrico Interconectado Nacional (Peru)</i>
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>

Sumário

1.	INTRODUÇÃO	13
1.1	MOTIVAÇÕES.....	13
1.2	OBJETIVOS.....	15
1.2.1	Objetivo geral	15
1.2.2	Objetivos específicos	15
2	REVISÃO DA LITERATURA	16
3.	FUNDAMENTAÇÃO TEÓRICA	22
3.1	<i>MACHINE LEARNING</i>	22
3.1.1	Aprendizado supervisionado.....	23
3.1.2	Aprendizado não supervisionado	23
3.1.3	Aprendizado por reforço	23
3.2	<i>SUPPORT VECTOR MACHINE</i>	24
3.2.1	Hiperplano de separação ótimo.....	25
3.2.2	Hiperplano de separação ótimo pra padrões não separáveis	27
3.2.3	SVM visto como um <i>Kernel Machine</i>	30
3.3	<i>LONG SHORT TERM MEMORY</i>	31
3.4	<i>ARTMAP FUZZY</i>	34
3.4.1	Teoria da ressonância adaptativa e rede neural ART	34
3.4.2	Rede neural ARTMAP <i>Fuzzy</i>	34
3.4.3	Algoritmo de treinamento da RNA ARTMAP <i>Fuzzy</i> - Modificada ..	36
4.	METODOLOGIA.....	40
4.1	REDE SVM	44
4.1.1	Pré-processamento	44
4.1.2	Configuração e treinamento	45
4.2	REDE LSTM	46

4.2.1	Pré-processamento dos Dados	46
4.2.2	Configuração e treinamento	47
4.3	REDE ARTMAP <i>FUZZY</i> MODIFICADA	48
4.3.1	Pré-processamento dos Dados	48
4.3.2	Configuração e treinamento	49
4.4	CRITÉRIOS DE COMPARAÇÃO	50
5.	RESULTADOS	52
5.1	PREVISÃO SVM.....	52
5.1.1	Validação do treinamento	52
5.1.2	Previsão de um dia	53
5.1.3	Previsão de uma semana	54
5.2	PREVISÃO LSTM.....	55
5.2.1	Validação do treinamento	55
5.2.2	Previsão de um dia	57
5.2.3	Previsão de uma semana	59
5.3	ARTMAP <i>FUZZY</i> MODIFICADA.....	60
5.3.1	Validação do treinamento	60
5.3.2	Previsão de um dia	62
5.3.3	Previsão de uma semana	63
5.4	COMPARAÇÃO DOS MODELOS DE PREVISÃO E DISCUSSÃO.....	64
6	CONCLUSÃO E SUGESTÕES PARA FUTUROS TRABALHOS	66
6.1	CONCLUSÃO	66
6.2	SUGESTÕES PARA FUTUROS TRABALHOS	67

1. INTRODUÇÃO

1.1 MOTIVAÇÕES

O setor elétrico desempenha papel estratégico no desenvolvimento econômico e social das nações, sendo responsável por garantir o fornecimento contínuo de energia a centros urbanos, zonas industriais e regiões rurais, segundo estudos realizados por Maciejowska *et al.* (2022). Contudo, a operação dos sistemas interligados enfrenta desafios crescentes para equilibrar oferta e demanda em tempo real, especialmente devido à volatilidade dos preços de energia elétrica, que variam diariamente em função de múltiplos fatores, exemplificando, disponibilidade de fontes de geração, capacidade de geração térmica, manutenção ou falha técnica. Segundo estudos realizados, esse equilíbrio delicado depende não apenas da produção e do consumo, mas também de variáveis externas, como condições meteorológicas e atividade comercial.

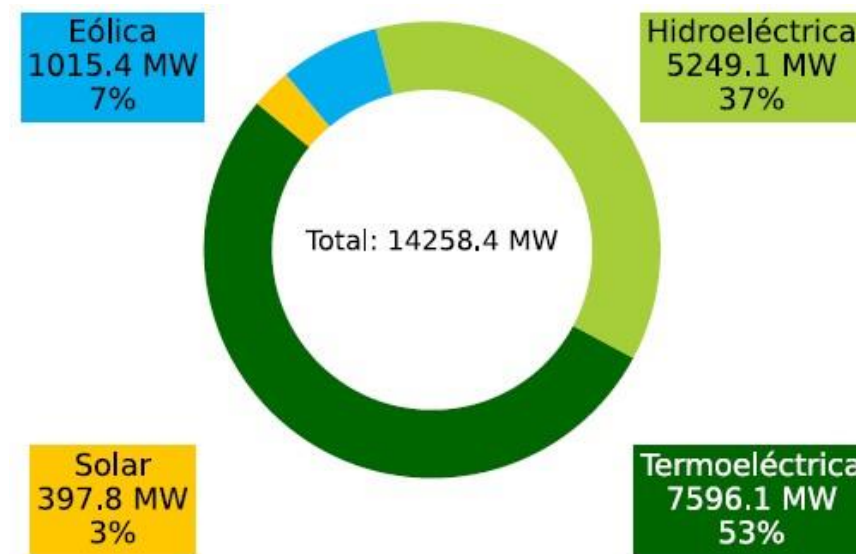
A acurácia da previsão dos preços desempenha um papel central, no âmbito do setor elétrico, uma vez que permite antecipar oscilações de mercado, tomar decisões operacionais mais eficientes e reduzir riscos financeiros para todos os agentes do setor, mais precisamente geradores que otimizam o despacho das usinas, comercializadores ajustam seus portfólios de contratos, consumidores industriais minimizam custos, e reguladores preservam a estabilidade sistêmica. Conforme estudos realizados por Granger (2022) previsões acuradas contribuem para alocações de recursos mais eficazes, decisões de investimento sustentáveis e mitigação de desequilíbrios operacionais.

De acordo com Jahangir *et al.* (2020). os preços da eletricidade apresentam comportamentos estocásticos, apresentando picos e vales abruptos, múltiplas sazonalidades e relações complexas entre variáveis. Fatores como feriados, variações de temperatura, sazonalidade e eventos não programados também afetam significativamente a formação dos preços, tornando os modelos tradicionais de previsão frequentemente insuficientes (Chang *et al.*, 2019).

A seleção de um modelo adequado requer uma avaliação criteriosa, uma vez que a metodologia apresenta limitações e pontos fortes que variam conforme a granularidade dos dados, o horizonte de previsão e as características do mercado em análise. Dada a diversidade da matriz peruana, essa análise é particularmente

desafiadora: sua matriz energética diversificada, a forte dependência da geração hidrelétrica, a crescente inserção de variáveis renováveis e uma estrutura regulatória em constante evolução geram condições altamente voláteis que dificultam a previsão precisa de preços. Segundo o COES (2024), órgão responsável pela coordenação, operação e planejamento da geração e transmissão elétrica no sistema interligado peruano (SEIN), essas características tornam o cenário ainda mais desafiador.

Figura 1 - Potência instalada no SEIN 2024.



Fonte: Coes (2024).

A ausência de estudos comparativos abrangentes aplicados especificamente ao mercado peruano representa uma lacuna crítica na literatura, limitando a capacidade dos agentes locais de identificar e adotar as técnicas de *Machine Learning* (ML) mais eficazes.

Diante desse panorama, este trabalho é motivado pela necessidade de preencher essa lacuna, avaliando de forma sistemática o desempenho de diferentes algoritmos de ML na previsão de preços de energia elétrica no Peru. Espera-se que os resultados obtidos subsidiem a formulação de estratégias operacionais mais robustas, contribuindo para a transparência, estabilidade e eficiência do setor elétrico peruano.

A falta de estudos comparativos abrangentes voltados especificamente para o mercado peruano constitui uma lacuna significativa na literatura, dificultando que os agentes locais identifiquem e adotem as técnicas de aprendizado de máquina mais eficazes.

1.2 OBJETIVOS

1.2.1 Objetivo geral

Realizar uma análise comparativa entre três técnicas de *ML* aplicado à previsão de preços de energia do mercado elétrico peruano visando à procura do melhor desempenho desde uma perspectiva metodológica das seguintes técnicas:

- *Support Vector Machine*
- *Long short-term Memory*
- *ARTMAP Fuzzy*

1.2.2 Objetivos específicos

- Estudar e aprender sobre a composição estrutural das técnicas de Machine Learning e seu impacto decorrente desde uma perspectiva metodológica na modelagem do problema;
- Estudar a situação atual do mercado elétrico peruano e compreender como são definidos os preços no mercado de curto prazo;
- Adquirir, analisar e organizar um banco de dados de preços de energia elétrica do mercado peruano;
- Implementar uma metodologia na modelagem dos preços da energia utilizando as técnicas de *ML* na procura do melhor desempenho;
- Avaliar o desempenho das técnicas e seus erros percentuais buscando uma previsão assertiva dos preços de energia elétrica do mercado.

2 REVISÃO DA LITERATURA

A precisão na previsão de preços de energia é intrinsecamente ligada tanto à escolha do modelo preditivo quanto à metodologia de sua aplicação (Weron, 2014). Nesse contexto, a literatura tem explorado extensivamente o uso de diversas técnicas de *ML*. Esta revisão aborda estudos recentes que aplicam as técnicas *Support Vector Machine* (SVM), *Long Short-Term Memory* (LSTM) e *ARTMAP Fuzzy* para a previsão de preços, analisando suas propostas, conjuntos de dados, horizontes de previsão e principais contribuições.

Em Chang *et al.* (2019) aplicam um modelo híbrido denominado *WT-Adam-LSTM*, combinando transformada *wavelet* e redes neurais LSTM otimizadas com o algoritmo *Adam*, com o objetivo de melhorar a acurácia na previsão de preços de eletricidade. A aplicação da transformada *wavelet* permite decompor a série original em componentes com variância mais estável, favorecendo o aprendizado da rede neural. O modelo foi validado em quatro estudos de caso, utilizando bases de dados dos mercados de energia de *New South Wales* (Austrália) e da França. Foram realizados testes comparativos com diferentes otimizadores (SGD, RMSProp e *Adam*), bem como com modelos híbridos já existentes, como *AutoRegressive Moving Average* (ARMA), *Differential Evolution Long Short-Term Memory* (DE-LSTM) e *AutoRegressive Integrated Moving Average Artificial Neural Network* (ARIMA-ANN). Em todos os casos, o *WT-Adam-LSTM* apresentou os menores valores de erro, com destaque para o MAPE, superando consistentemente os modelos de referência. A performance superior foi atribuída tanto à estabilidade conferida pela decomposição *wavelet* quanto à eficiência do otimizador *Adam* em encontrar melhores mínimos locais no treinamento da LSTM. O artigo se destaca por apresentar uma abordagem robusta e eficaz para tratar a variabilidade de dados de preços em curto prazo, especialmente em mercados com comportamento irregular. No entanto, os autores reconhecem que o modelo ainda não incorpora explicitamente fatores fundamentais como demanda, clima ou variáveis econômicas externas.

Zahid *et al.* (2019) é feito uma abordagem de dois modelos distintos para previsão de curto prazo da carga e do preço da eletricidade, com foco em melhorar a acurácia e estabilidade das previsões por meio de técnicas avançadas de aprendizado profundo e mineração de dados. O modelo de previsão de carga utiliza *Enhanced Convolutional Neural Network* (ECNN), enquanto o modelo de previsão de preços

emprega *Enhanced Support Vector Regression* (ESVR), ambos combinados com estratégias robustas de seleção e extração de atributos. O estudo utilizou técnicas como *XGBoost* (XGB), *Random Forest* (RF), *Decision Trees* (DT) e *Recursive Feature Elimination* (RFE) para selecionar as variáveis mais relevantes, de modo a reduzir redundâncias e a dimensionalidade dos dados. Além disso, os parâmetros dos classificadores foram ajustados via *Grid Search* com validação cruzada, a fim de evitar *overfitting* e otimizar o desempenho. A avaliação dos modelos foi realizada com base em métricas como MSE, RMSE, MAE e MAPE, e os resultados demonstraram que o ECNN superou o CNN em 2% de acurácia na previsão de carga, enquanto o ESVR foi 1% mais preciso do que o SVR tradicional na previsão de preços. O artigo se destaca pela aplicação de um *pipeline* completo de pré-processamento, otimização e validação para aprimorar modelos preditivos em sistemas elétricos. Além disso, os autores realizaram análises de estabilidade dos modelos utilizando diferentes tamanhos de conjunto de dados, reforçando a robustez das soluções propostas.

Em Zhou *et al.* (2019) desenvolveram um modelo híbrido para previsão de preços de eletricidade no mercado *Pennsylvania - New Jersey – Maryland* (PJM) nos Estados Unidos. O modelo propõe a decomposição dos dados via *Ensemble Empirical Mode Decomposition* (EEMD), segregando os componentes de alta e baixa frequência. Para cada componente, foi projetada uma arquitetura LSTM diferente, mais profunda para séries altamente não lineares e mais simples para séries com baixa não linearidade. Além disso, a seleção de hiperparâmetros foi realizada por meio do algoritmo *Sequence Model-Based Optimization* (SMBO), evitando o ajuste empírico comum em outros estudos. A base de dados utilizada incluiu séries horárias do mercado PJM entre fevereiro e abril de 2018, com foco em previsões 1 hora e 24 horas à frente. O modelo foi comparado com *Support Vector Regression* (SVR), redes *Backpropagation Neural Networks Gradient Boosting Regressor* (BPNN), *Decision tree regressor* (DTR), *Gradient boosting regressor* (GTB) e variantes LSTM (*shallow*, *stacked* e EEMD-LSTM sem otimização). Os resultados indicaram que a abordagem proposta (EEMD-LSTM-SMBO) superou significativamente os demais modelos em termos de MAPE e RMSE, com redução média superior a 30% em relação às arquiteturas LSTM simples e melhoria de estabilidade mesmo em cenários de alta volatilidade.

Jahangir *et al.* (2020) propuseram uma abordagem híbrida para previsão de preços de eletricidade de curto prazo, combinando técnicas de mineração de dados,

redes neurais profundas e redução de dimensionalidade. O estudo se destaca por aplicar o método de *Grey Correlation Analysis* (GCA) para seleção de variáveis relevantes, seguido por *Stacked Denoising Auto-Encoders* (SDAEs) para remoção de ruído e redução de dimensão para detectar as principais características de entradas e separar características desnecessárias. Após esse pré-processamento, os autores implementaram uma rede neural artificial com estrutura *fuzzy* (*Rough ANN*), utilizando neurônios com pesos intervalares para lidar com a incerteza inerente aos dados do mercado. A base de dados utilizada refere-se ao mercado canadense de Ontário (IESO), com histórico horário entre 2016 e 2018, incluindo variáveis como geração por fonte (gás, eólica, solar, biocombustível), temperatura, umidade, demanda total, e preços previstos para uma hora a três horas. Os resultados indicam que a combinação da seleção de variáveis com GCA, remoção de ruído via SDAE e previsão com neurônios intervalares é eficaz na previsão de preços, especialmente em dias com picos acentuados. O modelo proposto apresentou redução de erro entre 6% e 12% em relação aos dados de referência do operador (IESO).

Zhang *et al.* (2020) foi desenvolvido uma estrutura híbrida baseada em *deep learning* para previsão do preço de energia elétrica com horizonte de 24 horas, composta por quatro módulos principais: pré-processamento de variáveis, previsão pontual com redes neurais profundas, compensação de erro e previsão probabilística. O módulo inicial aplica técnicas de detecção de *outliers*, como *Isolation Forest* e seleção de variáveis (*Lasso*) para refinar os dados de entrada. Em seguida, três modelos de *Deep Learning*, são estes *Deep Belief Network* (DBN), *Long Short-Term Memory* (LSTM) e *Convolutional Neural Network* (CNN) que são utilizados para extrair padrões não lineares complexos. Um módulo de compensação de erro é empregado para ajustar as previsões a partir dos resíduos gerados na validação. Por fim, a previsão probabilística é realizada via regressão quantílica, permitindo representar incertezas relevantes para estratégias de lance no mercado. O estudo foi aplicado ao mercado PJM dos EUA, com dados horários entre 2018 e 2019. Os modelos foram avaliados usando métricas como MAPE, MAE, RMSE, além de *Average Coverage Probability* (ACP) e *Interval Sharpness* (IS) para as previsões probabilísticas. Os resultados indicam que o modelo CNN obteve o melhor desempenho preditivo entre os modelos testados, superando técnicas tradicionais como SVR, BPNN e KNN, além de outras redes profundas como DBN e LSTM. Destacam-se que o uso combinado de extração de características, ajuste de resíduos e análise probabilística melhorou tanto

a acurácia quanto a robustez do modelo. A principal contribuição do artigo reside na integração das quatro etapas de processamento em uma única estrutura modular, o que permite lidar simultaneamente com ruídos nos dados, não linearidades, instabilidades na previsão e incertezas inerentes aos mercados de energia.

Li e Becker (2021) é abordado um conjunto de modelos híbridos baseados em redes neurais recorrentes do tipo LSTM para previsão de preços de energia no mercado do dia seguinte (*day-ahead*) no sistema *Nord Pool*, incorporando variáveis de acoplamento de mercado entre países europeus. O estudo utilizou dados de mercado do *Nord Pool* e de mercados acoplados como o alemão e o belga, abrangendo fatores como geração, carga, radiação solar, temperatura, fluxo de energia, além de atributos externos de países vizinhos. A previsão foi avaliada com as métricas RMSE e MAPE, e os autores compararam diferentes arquiteturas baseadas em LSTM, com seleção de atributos via *Recursive Feature Elimination* combinado com *Support Vector Regression* (RFE-SVR), codificador automático (*autoencoders*) e regressão Lasso, obtendo ganhos expressivos em acurácia. O modelo com dois estágios de seleção apresentou o melhor desempenho, com MAPE inferior a 8% na maioria dos casos. O artigo se destaca por integrar técnicas de aprendizado profundo com estratégias robustas de seleção de atributos e por avaliar explicitamente o impacto do acoplamento de mercado na previsão de preços. Essa abordagem é particularmente relevante em sistemas interligados como o europeu. Contudo, o modelo assume a existência de integração energética e dados de alta qualidade entre mercados vizinhos sendo uma limitação importante ao se considerar sistemas sul-americanos com menor grau de interconectividade, como o peruano.

Memarzadeh e Keynia (2021) propuseram um modelo híbrido para previsão de curto prazo da carga e do preço da eletricidade, combinando três técnicas principais: decomposição por transformada *wavelet* (WT), seleção de atributos via entropia e informação mútua *Feature selection* (FS), e redes neurais profundas do tipo LSTM. A abordagem visa contornar a natureza não linear e altamente volátil das séries temporais de preços e cargas, eliminando ruídos via WT, reduzindo redundâncias e destacando variáveis relevantes via FS, e utilizando a memória de longo prazo do LSTM para capturar padrões temporais persistentes. O modelo foi avaliado em três mercados: PJM (EUA), *operador de mercado eléctrico designado* (OMIE) na Espanha e a província de Hormozgan (Irã), demonstrando acurácia superior à de métodos anteriores. No caso do PJM em 2006, o modelo alcançou um MAPE médio de 0,4%

para carga e 0,93% para preço, com coeficiente de determinação (R^2) superior a 0,99 em todas as estações. Em 2018, mesmo com alta volatilidade, o modelo manteve boa performance resultando um MAPE médio de 4,37%. Na Espanha (2002), o MAPE médio foi de 2,20%, e no Irã, apesar de eventos extremos (chuvas intensas), o modelo mostrou resiliência, com erros aceitáveis, especialmente nas estações menos afetadas. O destaque do artigo está na integração eficiente de técnicas de pré-processamento, seleção de atributos e redes neurais profundas, o que resulta em previsões robustas e com baixa variância. Contudo, uma limitação importante é a ausência de variáveis exógenas como temperatura, eventos regulatórios ou indicadores econômicos.

Nogueira e Lima (2021) propuseram um modelo baseado em redes neurais recorrentes (RNN) para previsão de preços no mercado livre de energia brasileiro. O estudo utilizou uma base de dados histórica da CCEE entre 2015 e 2020, considerando variáveis como carga, geração hídrica e PLD anterior. A métrica de desempenho adotada foi o RMSE, e os autores compararam seu modelo com regressão linear e redes neurais tradicionais, obtendo uma redução média de erro de 18%. O artigo se destaca pela aplicação de uma abordagem moderna de aprendizado profundo em um contexto brasileiro, embora não tenha explorado variações sazonais ou a influência de fatores regulatórios. Além disso, os bons resultados do modelo LSTM indicam que técnicas baseadas em séries temporais com memória podem ser apropriadas para o escopo deste trabalho.

Tschora *et al.* (2022) investigaram a aplicação de técnicas de *Machine Learning* para prever preços de eletricidade no mercado europeu de curto prazo para o dia seguinte (*day-ahead*). O estudo introduz uma abordagem multivariada que incorpora variáveis anteriormente negligenciadas, como históricos de preços de países vizinhos, visando aprimorar a precisão das previsões. O modelo proposto foi aplicado a conjuntos de dados da França, Alemanha e Bélgica, abrangendo dois períodos distintos. As técnicas avaliadas incluíram *Random Forest* (RF), *Convolutional Neural Networks* (CNN), *Deep Neural Networks* (DNN) e *Support Vector Regression* (SVR). Os resultados indicaram que DNN e SVR foram os métodos mais eficazes, conseguindo identificar padrões significativos e se adaptar a mudanças abruptas no mercado, como a elevação dos preços no segundo semestre de 2021. Por outro lado, RF e CNN apresentaram desempenho inferior, falhando na extração de contribuições significativas. Um dos destaques do estudo é o uso da técnica *SHapley Additive*

exPlanations (SHAP) para análise da importância das variáveis, proporcionando maior transparência na interpretação dos modelos. O estudo demonstra que variáveis de países vizinhos, especialmente os preços suíços e do gás, desempenham um papel determinante na acurácia da previsão. No entanto, o estudo não incorporou variáveis relacionadas à capacidade de transferência entre países ou preços de *commodities* como carvão ou carbono, elementos apontados como promissores para pesquisas futuras.

Esta revisão salienta a crescente relevância das técnicas de *ML* para a previsão de preços de energia elétrica, especialmente em mercados voláteis como o peruano. Embora diversos modelos sofisticados tenham sido desenvolvidos e testados globalmente, as características únicas do setor elétrico do Peru, como sua matriz energética diversificada, a forte dependência da geração hidrelétrica e o cenário regulatório em constante evolução, exigem uma análise comparativa focada.

Os capítulos subsequentes detalharão a fundamentação teórica, que sustentam os conceitos das diferentes técnicas empregadas neste trabalho. Em seguida, a metodologia empregada para avaliar o desempenho das técnicas *Support Vector Machine* (SVM), *Long Short-Term Memory* (LSTM) e *ARTMAP Fuzzy* neste contexto específico, buscando preencher a lacuna de pesquisa existente e contribuir para estratégias operacionais mais robustas e para a formulação de políticas no mercado de energia peruano.

3. FUNDAMENTAÇÃO TEÓRICA

Nesta seção é apresentada toda fundamentação teórica necessária para a compreensão do estudo apresentado neste trabalho.

3.1 MACHINE LEARNING

Segundo Alpaydin (2020), nas últimas décadas, o *ML* se tornou um dos pilares da tecnologia da informação embora geralmente oculta das nossas vidas, mas se apresenta de diversas formas grande parte delas está em reduzir a variedade de problemas muito diferentes em conjuntos de protótipos restritos tendo alta precisão na previsão. Algumas de suas vastas aplicações incluem finanças, análise bancária de dados históricos para aplicações de crédito, detecção de fraude e previsão de estoque de mercado. Além disso, no ramo da medicina, programas de aprendizado para diagnóstico médico. No ramo da telecomunicação, o reconhecimento de padrões é utilizado para otimização da rede, a fim de maximizar a qualidade do serviço. Na ciência, grandes volumes de dados em física, astronomia e biologia podem ser analisados rapidamente por computadores. Adicionalmente, a rede mundial de computadores apresenta escala global e está em constante crescimento; desse modo, a busca por informações requer novos métodos, pois não podem ser feitos manualmente. Contudo, não se trata apenas de um problema de base de dados, mas é parte da inteligência artificial, que aprende e se modela com as mudanças do ambiente, não sendo necessário ao desenvolvedor de sistemas prever todos os cenários prováveis.

Em outras palavras, o *ML* pode ser descrito como um ramo da Inteligência artificial que propõe a implementação de aprendizado eficiente a partir de algoritmos de dados para modelos de previsão. Houve um aumento do uso desta ferramenta devido à abundância de dados disponíveis e ao crescente poder computacional para processá-los (Tschora *et al.*, 2022). Adicionalmente, o *ML* utiliza recursos computacionais para reconhecer automaticamente padrões por meio de processos inteligentes de tomada de decisões que se baseiam em um processo de treinamento. Portanto, permite automatizar tarefas por meio de um modelo capaz de analisar e executar em ordem cognitiva (Dua; Du, 2016).

A fim de alcançar o efetivo aprendizado de máquina efetivo, existem diferentes métodos que são comumente classificados como supervisionado, não supervisionado e por reforço.

3.1.1 Aprendizado supervisionado

De acordo com Haykin (2009), o aprendizado supervisionado, também conhecido como aprendizado com professor, pode ser imaginado da seguinte forma: o professor possui conhecimento sobre o ambiente em que a rede será aplicada. Esse conhecimento é representado como um conjunto de exemplos de entradas e saídas. Durante o processo de ajuste, o professor fornece a resposta esperada para o vetor de entrada, e os parâmetros são ajustados pela influência tanto do vetor de treinamento quanto do sinal de erro, que é a diferença entre a resposta desejada e a real da rede. Esse processo é feito de forma iterativa e repetida até que a rede neural emule o professor. Assim, o conhecimento do professor é transferido para a máquina, permitindo que ela aprenda a lidar com o ambiente de forma autônoma.

3.1.2 Aprendizado não supervisionado

Em relação ao aprendizado não supervisionado, os pesos e desvios são alterados em resposta apenas às entradas da rede. Não há saídas de destino disponíveis para comparação, como descrito no tópico anterior. A maioria desses algoritmos executa um tipo de operação de agrupamento. Eles aprendem a categorizar os padrões de entrada em um número finito de classes. Haykin (2009) afirma que, para este tipo de treinamento, apenas o ajuste de pesos é considerado em conformidade com os padrões de entrada.

3.1.3 Aprendizado por reforço

O aprendizado por reforço é utilizado em diferentes aplicações para alcançar a melhor solução e escolher o melhor caminho para ser adotado em cada caso. No aprendizado supervisionado, os dados de treinamento possuem a saída esperada, e o algoritmo utiliza essas “chaves de resposta” no processo de treinamento. Entretanto, o aprendizado por reforço não possui essas chaves de resposta. Dado isso, não há escolha de algoritmo para o aprendizado, exceto a experiência conforme descrito por Haykin (2009).

Portanto, no aprendizado por reforço, um problema é definido e o algoritmo tenta solucioná-lo por tentativa e erro.

3.2 SUPPORT VECTOR MACHINE

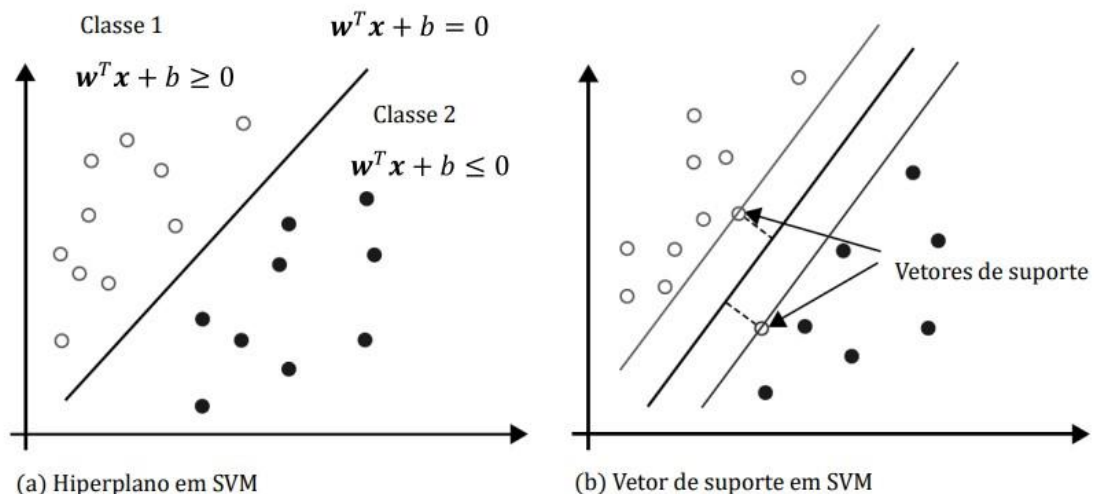
Em Haykin (2009) a principal ideia por trás desta máquina é:

Dada uma amostra de treinamento, a máquina de vetores de suporte constrói um hiperplano como superfície de decisão de tal forma que a margem de separação entre exemplos positivos e negativos seja maximizada (Haykin, 2009, p. 269)

É uma máquina de aprendizado para problemas de classificação de dois grupos, que conceitualmente implementa a ideia de vetores de entrada mapeados para um espaço de características de alta dimensão, onde é construída uma superfície de decisão linear garantindo uma alta capacidade de generalização. Essa concepção foi originalmente proposta por Cortes e Vapnik (1995), no contexto em que os dados de treinamento são linearmente separáveis, e posteriormente estendida para tratar casos não separáveis, considerando transformações de entrada polinomiais.

Dados pontos de X em um espaço dimensional n , o SVM separa estes pontos com um plano dimensional $n - 1$. O principal objetivo é classificar esses dados com um hiperplano que tenha a máxima distância ao ponto mais próximo do outro lado, sendo consequentemente também chamado de classificador linear de máxima margem (Dua; Du, 2016).

Figura 2 – Classificação do SVM



Fonte: Adaptado de Dua e Du (2016)

Qualquer hiperplano pode ser escrito com um conjunto de pontos X que satisfazem:

$$w^T x + b = 0 \quad (1)$$

Onde w é um vetor normal perpendicular ao hiperplano e b é a constante de deslocamento do ponto de origem do hiperplano ao longo do vetor normal, conforme ilustrado na Figura 2a. Rotulando os pontos dos dados de X em duas classes, dado o par de (w^T, b) pode-se classificar os dados de X como o sinal da função (positivos e negativos). Adicionalmente, a distância dos pontos de separação é dada pela razão entre o hiperplano definido pela equação (1) e o vetor $\hat{w}$. Os pontos mais próximos ao hiperplano são chamados de vetores de suporte e a distância entre os vetores de suporte é chamada de margem de separação conforme mostrado na Figura 2b (Dua; Du, 2016).

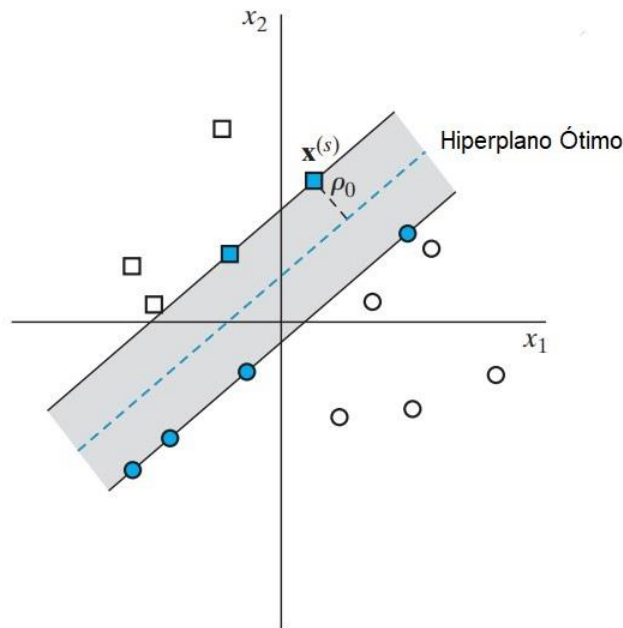
3.2.1 Hiperplano de separação ótimo

Considera-se a amostra de treinamento sendo o par de entradas e saídas $\{(x_i, d_i)\}_{i=1}^N$, onde x_i são as entradas e d_i a correspondente desejada resposta partindo que se pode assumir que os padrões são representados por $d_i = +1$ e $d_i = -1$ são linearmente separáveis a equação do plano de decisão na forma de hiperplano representado pela Equação 1 em que w é um vetor de peso ajustável, podemos escrever as seguintes equações:

$$\begin{cases} w^T x_i + b \geq 0 & \text{para } d_i = +1 \\ w^T x_i + b < 0 & \text{para } d_i = -1 \end{cases} \quad (2)$$

O objetivo do SVM é encontrar o hiperplano particular para o qual a margem de separação é maximizada, sobre essa condição o plano de decisão é tido como hiperplano ótimo (Haykin, 2009).

Figura 3 - Ilustração do hiperplano ótimo para padrões linearmente separáveis.



Fonte: Adaptado de Haykin (2009).

Reescrevendo a Equação 1 tem-se a função discriminante:

$$g(x) = w_0^T + b_0 \quad (3)$$

A partir da equação acima é possível obter uma medida algébrica da distância de x ao hiperplano ótimo, sendo expressa como:

$$x = x_p + r \frac{w_0}{\|w_0\|} \quad (4)$$

Em que x_p é a projeção normal de x para o hiperplano ótimo e r é a distância algébrica desejada (Haykin, 2009). O valor r é positivo se x está no lado positivo do hiperplano ótimo se negativo se estiver do lado negativo. Uma vez que, por definição, $g(x_p) = 0$, obtêm-se as seguintes expressões:

$$\begin{cases} g(x) = w_0^T + b_0 = r\|w_0\| \\ \text{Equivalentemente } r = \frac{g(x)}{\|w_0\|} \end{cases} \quad (5)$$

Em termos conceituais, os vetores de suporte são os pontos de dados que ficam mais próximos do hiperplano ótimo e são, portanto, os mais difíceis de classificar. Como tal, eles têm uma relação direta com a localização ótima da superfície de decisão (Haykin, 2009). A partir da Equação 5 a distância algébrica do vetor de suporte $x^{(s)}$ ao hiperplano ótimo é:

$$r = \frac{g(x^{(s)})}{\|w_0\|} = \begin{cases} \frac{1}{\|w_0\|} & \text{se } d^{(s)} = +1 \\ -\frac{1}{\|w_0\|} & \text{se } d^{(s)} = -1 \end{cases} \quad (6)$$

Denotando ρ o valor ótimo da margem de separação entre duas classes da amostra de treinamento.

$$\rho = 2r = \frac{2}{\|w_0\|} \quad (7)$$

Então maximizar a margem de separação entre classes binárias é equivalente a minimizar a norma euclidiana do vetor de peso w (Haykin, 2009).

3.2.2 Hiperplano de separação ótimo pra padrões não separáveis

Considerando que há casos em que não é possível construir um hiperplano de separação sem ocorrer erros de classificação, ou seja, onde os dados não podem ser linearmente separáveis, então devemos minimizar a probabilidade de erros de classificação (Alpaydin, 2020; Haykin, 2009).

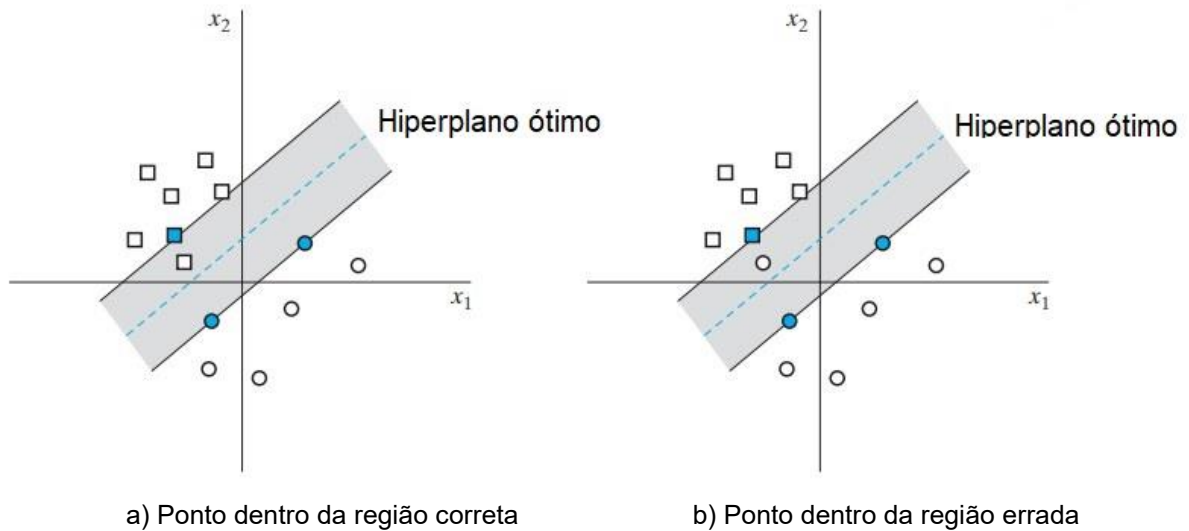
A margem de separação entre as classes é considerada suave se um ponto de dados infringe a seguinte condição que parte da Equação 2 para o exemplo de treinamento $T = \{(x_i, d_i)\}_{i=1}^N$:

$$d_i(w^t x_i + b) \geq 1 \text{ para } i = 1, 2, \dots, N \quad (8)$$

Esta violação pode ocorrer de duas maneiras: o primeiro caso é quando o ponto de dado (x_i, d_i) está dentro da região de separação, mas no lado correto da área de

decisão. O segundo caso ocorre quando o ponto de dado (x_i, d_i) está do lado errado da área de decisão (Haykin, 2009). Conforme Figura 4:

Figura 4 - Hiperplano de margem suave.



Fonte: Adaptado de Haykin (2009).

Introduzindo um novo conjunto de variáveis escalares não negativas $\{\xi_i\}_{i=1}^N$, a fim de preparar um tratamento formal aos pontos não separáveis, sendo assim complementando a Equação 8 da seguinte forma (Alpaydin, 2020):

$$d_i(w^t x_i + b) \geq 1 - \xi_i, i = 1, 2, \dots, N \quad (9)$$

Definindo ξ_i como variável de folga, qual mede o desvio do ponto em relação à condição ideal de separação: $0 < \xi_i \leq 1$ significa que o ponto está no lado correto da região de separação, conforme a Figura 4a, entretanto, quando $\xi_i > 1$, o ponto está dado contrário da região de separação, de acordo com a Figura 4b). Note que, se um exemplo com $\xi_i > 0$ não for considerado nos exemplos de treinamento, a superfície de decisão irá mudar. Sendo assim, os vetores de suporte são definidos exatamente da mesma maneira para ambos os casos: linearmente e não linearmente separáveis. Logo o objetivo é encontrar um hiperplano de separação para que o erro de classificação incorreta seja minimizado em relação à amostra de treinamento, desse modo, é minimizado a função (Haykin, 2009).

$$\underset{w, b, \xi}{\text{Minimizar}} \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^N \xi_i \right) \quad (10)$$

A constante C impõe um peso à minimização dos erros, pois é um termo de regularização no conjunto de treinamento em relação à minimização da complexidade do modelo. O termo $\sum_{i=1}^N \xi_i$ também pode ser visto como uma minimização dos erros marginais no problema de otimização, pois um valor de ξ_i maior que zero (até um) indica um dado representado entre as margens (Lorena; De Carvalho, 2007). O problema de otimização gerado é quadrático, e sua solução envolve passos matemáticos semelhantes aos apresentados anteriormente. Contudo, com a adição de uma função Lagrangiana, suas derivadas parciais tornam-se nulas, resultando no seguinte problema dual:

$$\underset{\alpha}{\text{maximizar}} Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j d_i d_j x_i^T \cdot x_j \quad (11)$$

Com as seguintes restrições:

$$\begin{cases} 0 \leq \alpha_i \leq C, & \forall i = 1, \dots, N \\ \sum_{i=1}^N \alpha_i d_i = 0 \end{cases} \quad (12)$$

Observa-se que nem as variáveis de folga ξ_i nem os próprios multiplicadores de Lagrange aparecem no problema dual representado acima. Sendo assim, o problema dual para casos não separáveis é semelhante ao caso de padrões linearmente separáveis, porém com uma pequena, mas importante, diferença: a restrição $\alpha_i \geq 0$ é substituída por uma restrição mais rigorosa $0 \leq \alpha_i \leq C$. A otimização restrita para o caso não separável e os cálculos dos valores ótimos do vetor de pesos w e da constante b procedem da mesma forma que no caso linearmente separável. Note que os vetores de suporte são definidos da mesma maneira que anteriormente (Haykin, 2009).

3.2.3 SVM visto como um *Kernel Machine*

Inicialmente, para um *kernel* de produto interno, x será denotado como um vetor extraído do espaço de entrada de dimensão m_0 . Além disso, tem-se $\{\varphi(x)\}_{j=1}^{\infty}$ definido como um conjunto de funções não lineares que transformam o espaço de entrada de dimensão m_0 em um espaço de características de dimensionalidade infinita. Sendo assim, pode-se definir um hiperplano que age como uma superfície de decisão, conforme visto a seguir:

$$\sum_{j=1}^{\infty} \omega_j \varphi_j(x) = 0 \quad (13)$$

Onde $\{\omega_j\}_{j=1}^{\infty}$ representa um conjunto infinitamente grande de pesos que transforma o espaço de características em espaço de saída. A decisão sobre se o vetor de entrada pertence ao lado negativo ou positivo das classes é feita nesse espaço (Haykin, 2009).

A fim de simplificar a representação, define-se o *bias* igual a zero. Considerando $\phi(x)$ o vetor de características e N_s o número de vetores de suporte, a área de decisão no espaço de saídas como:

$$\sum_{i=1}^N \alpha_i d_i \phi^T(x_i) \phi(x) = 0 \quad (14)$$

Formalmente definido por Shawe-Taylor (2004):

“O *kernel* $k(x, x_i)$ é uma função que comuta o produto interno de imagens feitas no espaço de características sobre o ϕ anexado a dois pontos de dados no espaço de entradas (Shawe-Taylor, 2004).”

Sendo assim, o multiplicador escalar de $\phi^T(x) \phi(x)$ como *kernel* de produto interno, e a Equação 14 pode ser rearranjada da seguinte forma:

$$\sum_{i=1}^N \alpha_i d_i k(x, x_i) = 0 \quad (15)$$

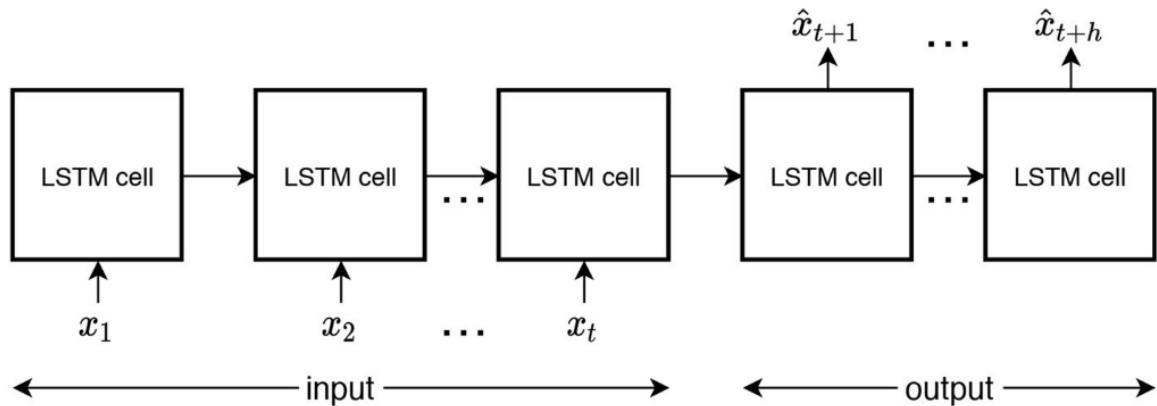
Continuamente, estabelece-se que o *kernel* $k(x, x_i)$ é uma função que tem duas propriedades básicas: a primeira é que a função *kernel* é simétrica sobre o centro de pontos x_i , ou seja, $k(x, x_i) = k(x_i, x)$ para todo x_i e atinge o ponto máximo em $x = x_i$; a segunda propriedade é que o volume sob a superfície da função $k(x, x_i)$ é uma contante. Com relação à segunda propriedade, se for apropriadamente dimensionado para ser igual à unidade, então terá propriedades similares às de uma função de densidade probabilística de uma variável aleatória (Haykin, 2009).

3.3 LONG SHORT TERM MEMORY

O *Long Short Term Memory* (LSTM) é uma rede neural recorrente cujo principal objetivo é solucionar o problema de desaparecimento de gradiente, que pode causar atrasos no treinamento de uma rede neural. Este método busca criar uma memória de curto prazo para redes neurais com capacidade de dura por um tempo indefinido. A estrutura da rede neural recorrente imita o padrão de pensamento de que a compreensão humana é baseada na experiência e na memória. É principalmente usada para processar dados sequenciais relevantes em contexto; além das informações de entrada atuais, também são consideradas as informações de entrada anteriores, que são mantidas através da função de “memória”. Adicionalmente, a estrutura de rede conectada em série é apropriada para processar dados com informações contextuais (Yu *et al.*, 2019).

Uma rede LSTM, dependendo do número de resultados a serem obtidos, pode ser estruturada de diferentes maneiras. Pode ser estruturado com uma entrada e uma saída (um para um), muitas entradas e uma saída (muitas para uma), uma entrada e muitas saídas (uma para muitas) ou muitas entradas e muitas saídas (muitos para muitos). Considerando que o objetivo fundamental deste trabalho é prever os próximos valores de h com base em um conjunto de dados históricos, deparamo-nos com um problema “muitos para muitos”, conforme Figura 5.

Figura 5 - Rede LSTM "muitos para muitos".

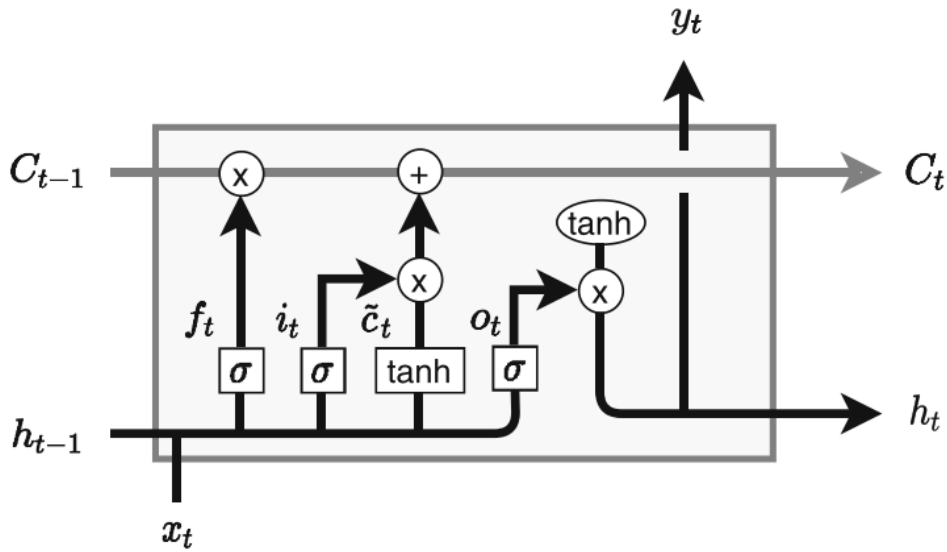


Fonte: (Torres; Martínez-Álvarez; Troncoso, 2022).

Cada uma das células LSTM recebe a modificação de informações modeladas das células anteriores, sendo estas C_{t-1} e h_{t-1} , bem como dados no instante de tempo atual x_t . Dependendo de um conjunto de portas lógicas, é determinado o grau de influência que os dados em instantes de tempo anteriores têm sobre os dados no instante de tempo a ser previsto, modelando, desse modo, o comportamento da rede (Hochreiter; Schmidhuber, 1997).

O desenho esquemático de uma célula LSTM pode ser visto a seguir (Figura 6), em que f_t é a porta de esquecimento, i_t é a porta de entrada e o_t é a porta de saída. Em seguida, f_t que decide quais informações devem ser descartadas ou salvas. Um valor próximo de zero significa que a informação passada será esquecida, enquanto um valor próximo de um significa que ela permanece. i_t decide qual nova informação o_t usar para atualizar o estado da memória c_t . Portanto, c_t é atualizado usando ambos f_t e i_t . Por fim, o_t decide qual é o valor de saída que será utilizado como a entrada da próxima camada oculta.

Figura 6 - Célula LSTM.



Fonte: (Torres; Martínez-Álvarez; Troncoso, 2022).

As informações da camada oculta anterior h_{t-1} e a informação da entrada atual x_t são combinadas e processadas por meio da função de ativação *sigmóide* (σ , com saída entre 0 e 1), a fim de gerar os valores das portas de entrada, esquecimento e saída. Já a função de ativação tangente hiperbólica ($\tanh$, com saída entre -1 e 1), é aplicada para calcular o vetor de nova informação o_t , atualizando h_t . A saída final da célula LSTM resulta da combinação dessas operações, como mostrado nas seguintes equações:

$$\tilde{c}_t = \tanh(W_c[a_{t-1}, x_t] + b_c) \quad (16)$$

$$i_t = \sigma(W_u[a_{t-1}, x_t] + b_u) \quad (17)$$

$$f_t = \sigma(W_f[a_{t-1}, x_t] + b_f) \quad (18)$$

$$o_t = \sigma(W_o[a_{t-1}, x_t] + b_o) \quad (19)$$

$$c_t = i_t * \tilde{c}_t + f_t * c_{t-1} \quad (20)$$

$$a_t = o_t * \tanh(c_t) \quad (21)$$

Onde W_u , W_f e W_o são os pesos e b_u , b_f e b_o são os *bias* que, respectivamente, governam o comportamento das portas i_t , f_t e o_t . Além disso, W_c e b_c são, respectivamente, os pesos e *bias* da célula de memória o_t . Para uma análise mais detalhada sobre a rede LSTM, suas portas lógicas e operação detalhada pode ser encontrada em (Hochreiter; Schmidhuber, 1997).

3.4 ARTMAP FUZZY

A fim de introduzir a rede neural ARTMAP *Fuzzy*, é necessário explicar um pouco sobre a Teoria da Ressonância Adaptativa (*Adaptive Resonance Theory* - ART), a qual sustenta toda a fundamentação teórica proposta e desenvolvida.

3.4.1 Teoria da ressonância adaptativa e rede neural ART

A Teoria da Ressonância Adaptativa (*Adaptive Resonance Theory* – ART) foi desenvolvida por Carpenter e Grossberg (1988) para solucionar o dilema entre plasticidade e estabilidade nos sistemas de aprendizado. A plasticidade significa a capacidade da rede neural de aprender novos padrões, agrupando entradas em categorias específicas sem comprometer o conhecimento adquirido anteriormente. Já a estabilidade assegura que os novos padrões aprendidos sejam corretamente diferenciados dos já existentes, garantindo, desse modo, um estado consistente na rede à medida que os pesos são ajustados.

As redes ART foram projetadas para realizar reconhecimento estável em tempo real, pois utiliza-se de mecanismos de auto-organização para responder a padrões arbitrários. Entre os princípios adotados para lidar com o dilema da plasticidade e estabilidade, evidenciam-se o aprendizado por correspondência, o aprendizado baseado em erro e o aprendizado rápido e estável. Além disso, na formulação matemática das redes ART, a notação vetorial é tradicionalmente representada por linhas em vez de colunas, pois proporciona maior clareza nas operações matemáticas envolvidas, de acordo com os autores Carpenter e Grossberg (1988).

3.4.2 Rede neural ARTMAP *Fuzzy*

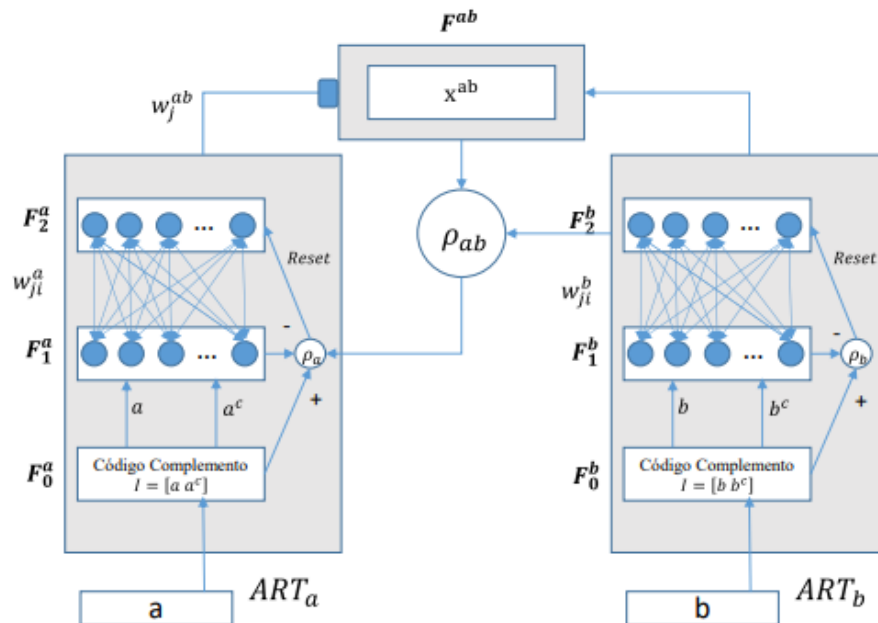
O modelo modificado *Fuzzy* ARTMAP foi projetado para otimizar o processo de treinamento de forma racional e eficiente pois é uma extensão das redes neurais baseadas na Teoria da Ressonância Adaptativa (ART). O modelo é inspirado no

funcionamento do cérebro, visto que ativa neurônios inativos à medida que novos conhecimentos são adquiridos, consequentemente mantendo a propriedade fundamental das redes ART, que é a solução do problema da plasticidade e estabilidade. Isso é feito por meio de três módulos internos supervisionados: os módulos *Fuzzy ART_a* e *ART_b* processam os vetores de entrada e saída esperados, respectivamente, criando categorias de reconhecimento estáveis. O terceiro módulo, Inter-ART (*F_{ab}*), atua como um regulador, garantindo a correspondência entre *ART_a* e *ART_b* por meio de um mecanismo de rastreamento adaptativo (*match-tracking*), ajustando dinamicamente os parâmetros para manter a coerência do aprendizado (Carpenter; Grossberg; Rosen, 1991).

Inicialmente, o treinamento começa com um único cluster, que se expande dinamicamente para acomodar novos dados, otimizando assim o processo de aprendizado com base em um critério de vigilância predefinido. Trata-se de um mecanismo que reduz a quantidade de cálculos necessários para a função de ativação. A estrutura do ARTMAP *Fuzzy* inclui diversos parâmetros que influenciam diretamente a velocidade, eficiência e eficácia do aprendizado. São eles: o parâmetro α , que controla a busca entre os nós da camada F_2 e deve ser positivo ($\alpha > 0$); a taxa de aprendizado β , no intervalo maior que zero a um, que tem a função de regular a adaptação dos pesos - valores menores resultam em aprendizado mais lento; por fim, o parâmetro de vigilância ρ , também no intervalo maior que zero a um, que define a granularidade das categorias criadas: valores mais altos aumentam a especificidade, enquanto valores menores favorecem a generalização da rede. Esse critério de vigilância é aplicado aos módulos *ART_a* (ρ_a), *ART_b* (ρ_b) e Inter-ART (ρ_{ab}).

O treinamento do *Fuzzy ARTMAP* envolve dois módulos *Fuzzy ART* interconectados por um módulo regulador Inter-ART (*F_{ab}*), responsável por garantir a correspondência entre as categorias formadas. A dinâmica desse processo pode ser representada na Figura 7 a seguir (Carpenter *et al.*, 1992).

Figura 7 - Arquitetura ARTMAP Fuzzy



Fonte: Adaptado de Carpenter et al (1991).

3.4.3 Algoritmo de treinamento da RNA ARTMAP Fuzzy - Modificada

O algoritmo da ARTMAP Fuzzy Modificada propõe um comportamento mais racional inspirado na ativação progressiva de neurônios biológicos, otimizando o treinamento da rede (Carpenter *et al.*, 1992; Moreno, 2010). Inicialmente, apenas um cluster é ativado com os dados de entrada, e novos padrões são incorporados conforme necessário, reduzindo o número de cálculos de ativação e testes de vigilância (Moreno, 2010).

A seguir, é descrito o algoritmo de treinamento da Rede Neural Artificial ARTMAP Fuzzy Modificada, conforme proposto por (Moreno, 2010). Além disso, a Figura 8 ilustra o fluxograma correspondente ao processo de treinamento, composto por dois módulos ART Fuzzy interligados por um módulo Inter-ART.

Os sistemas Fuzzy são compostos por três campos: F_0 , que representa o vetor de entrada; F_1 , que recebe sinais *bottom-up* de F_0 e sinais *top-down* de F_2 ; F_2 , qual é responsável por representar a categoria ativa. A camada de entrada (F_0) é responsável por receber os vetores de entrada, realizando a normalização e a codificação de complemento dos dados.

$$I = \frac{a}{|a|} \quad (22)$$

Em que a é o vetor de entrada e $|a|$ é a função norma para o vetor de entrada, definidos por:

$$a = [a_1 \ a_2 \ \dots \ a_M] \quad (23)$$

$$|a| = \sum_{i=1}^M a_i \quad (24)$$

A fim de preservar a amplitude da informação recebida, é feita a codificação do complemento a_i^c . Desse modo, o vetor de entrada de dimensão M terá em F_1 uma dimensão $2M$, devido às equações a seguir:

$$a_i^c = 1 - a_i \quad (25)$$

$$I = [a \ a^c] \quad (26)$$

O vetor de ativação da camada F_2 é representado da seguinte forma:

$$y = [y_1 \ y_2 \ \dots \ y_N] \quad (27)$$

Sendo N o número de categorias criadas em F_2 , ou seja, indicando o número de *clusters* ativos, obtém-se a seguinte relação:

$$y_j = \begin{cases} 1, & \text{se o nó } j \text{ de } F_2 \text{ é ativo} \\ 0, & \text{se não é ativo} \end{cases}; \quad (28)$$

Logo após, é feita a inicialização dos pesos, onde apenas um *cluster* possui dados do primeiro vetor de entrada. Sendo assim, apenas um *cluster* fica ativo com os dados do primeiro padrão:

$$w_{11}(0) = I_{11}; w_{12}(0) = I_{12}; \dots w_{jM}(0) = I_{1,2M}; \quad (29)$$

Visto que a função escolha é definida tendo um vetor I em F_1 para cada nó em F_2 , obtém-se a seguinte equação:

$$T_j = \frac{|I \wedge w_j|}{\alpha + |w_j|} \quad (30)$$

Em que o operador $\wedge$ atua similar a porta lógica *AND*.

Para fins de escolha do *cluster*, seleciona-se a que possuir menor índice e a categoria escolhida sendo o nó J ativo. É possível encontrar mais de uma categoria ativa, sendo assim, a equação a seguir é utilizada:

$$J = \arg_{\max}\{T_j\} \quad (31)$$

Para $j = 1, 2, \dots N$.

Na sequência, tem-se o critério da vigilância, que, se for satisfeito, verifica-se a atividade em F_2 .

$$\frac{|I \wedge w_j|}{I} \geq \rho \quad (32)$$

Caso contrário, é verificado se todos os *clusters* foram testados para assim criar um novo *cluster*, pois significa que nenhum dos clusters existentes comporta a entrada atual.

Caso o critério de vigilância seja satisfeito (ressonância), ocorre a atualização dos pesos de acordo com as equações a seguir:

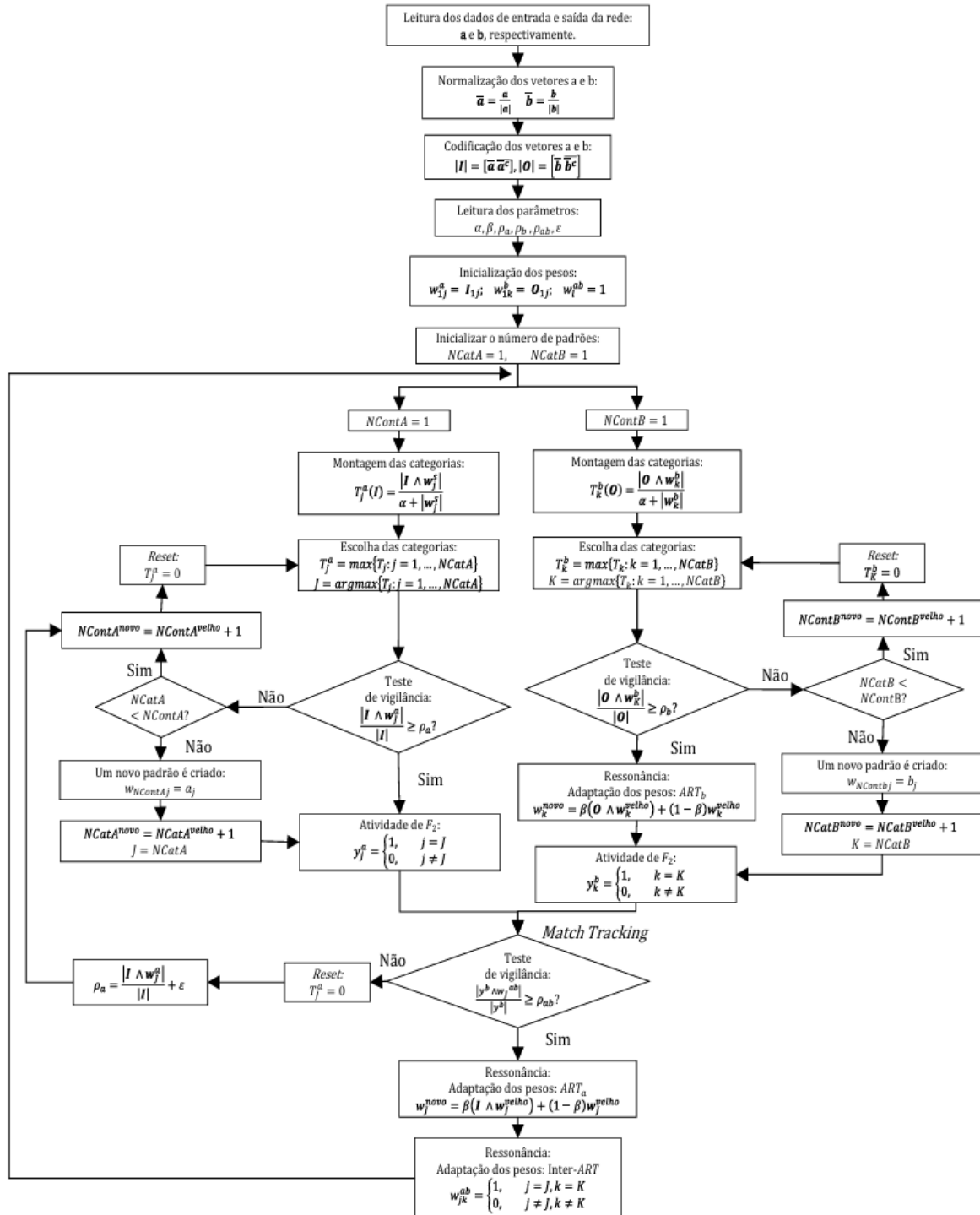
$$w_j^{novo} = \beta(I \wedge w_j^{velho}) + (1 - \beta)w_j^{velho} \quad (33)$$

A seguir, o vetor de atividade no nó F_2 é feito, sendo assim:

$$y = [y_1 \ y_2 \ \dots \ y_N], \text{ para } y_j = \begin{cases} 1, & \text{se } j = J \\ 0, & \text{se } j \neq J \end{cases} \quad (34)$$

Onde N é o número de *clusters* criadas em F_2 . A Figura 8 apresenta o fluxograma do algoritmo de treinamento da ARTMAP Fuzzy Modificada, adaptado de Moreno (2010), o qual sistematiza essas etapas de forma visual e sequencial.

Figura 8 – Fluxograma da RNA ARTMAP Fuzzy Modificada



Fonte: Adaptação de Moreno (2010).

4. METODOLOGIA

Neste capítulo, são apresentadas as metodologias utilizadas para cada um dos sistemas previsores empregados neste trabalho: SVM, LSTM e ARTMAP *Fuzzy*. Para tal, foi utilizada uma base de dados extraída do site do COES, entidade que coordena a operação do Sistema Elétrico Interligado Nacional (SEIN) e atua como operador e administrador do mercado elétrico do Peru. O COES disponibiliza dados diários e públicos de toda a operação do sistema elétrico peruano, desde geração, demanda, preços, taxas, comissões, além de análises diárias e relatórios mensais e anuais sobre toda a operação do país. A partir disto, foram extraídos dados diários de carga, geração hídrica e preços de meia em meia hora, do período de 01/08/2022 00:30 até 01/10/2024 00:00, do barramento Santa Rosa 220kV, para serem utilizados como entradas para treinamento dos modelos previsores. Além disso, os dados para treinamento e previsão foram separados, conforme a Tabela 1.

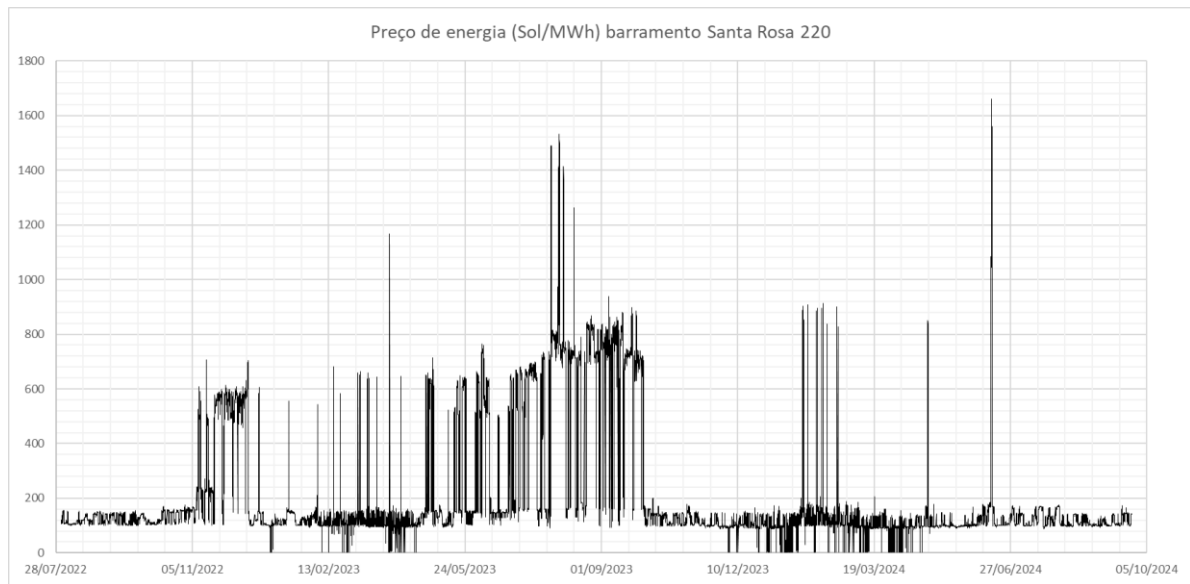
Tabela 1 – Separação de dados.

Período	Tamanho	Conjunto de dados
01/08/2022 até 30/09/2024	38016	Total
01/08/2022 até 31/01/2024	26384	Treino
01/02/2024 até 23/09/2024	11296	Validação
24/09/2024 até 30/09/2024	336	Previsão

Fonte: Elaborado pelo autor.

O gráfico apresentado na Figura 9 representa a evolução do preço da energia elétrica (Sol/MWh) no barramento Santa Rosa 220kV entre julho de 2022 e outubro de 2024. Observa-se uma série de flutuações acentuadas nos preços, com picos significativos que ultrapassam os 1400 Sol/MWh, especialmente entre o final de 2022 e meados de 2023. Posteriormente, verifica-se uma redução na volatilidade, ainda que com ocorrências pontuais de aumentos abruptos.

Figura 9 – Variação de preços de energia da base de dados.

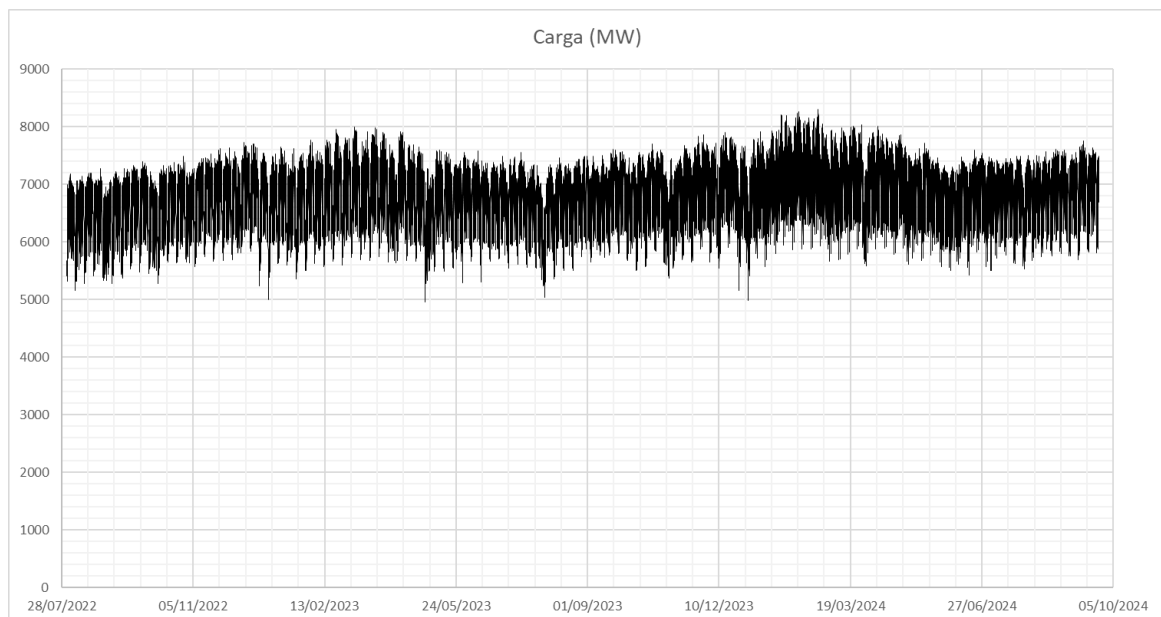


Fonte: Elaborado pelo autor.

Esse comportamento pode estar associado a fatores climáticos, operacionais e de mercado, como apontado por Bello *et al.* (2021), que destacam a influência do regime hidrológico e da oferta térmica no comportamento dos preços horários na América Latina. Além disso, segundo Weron (2014), sistemas isolados ou regiões com restrições de transmissão podem apresentar variações mais acentuadas nos preços devido à limitação de alternativas de geração.

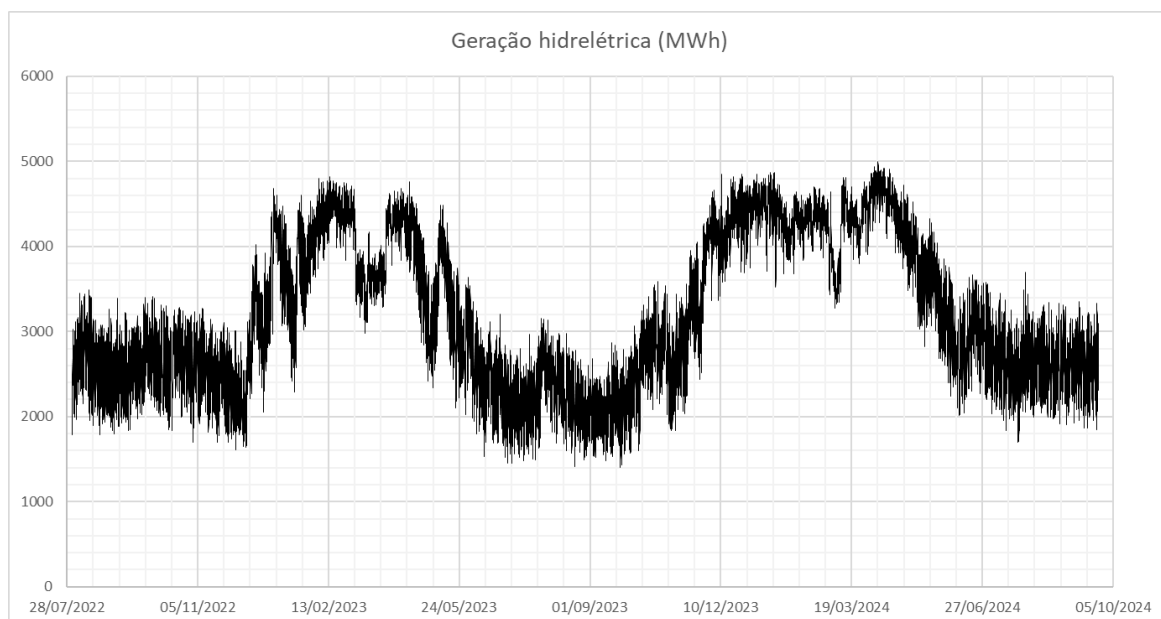
Verifica-se uma correspondência nas oscilações sazonais do mercado peruano, atribuídas principalmente à disponibilidade hídrica e ao despacho térmico. No entanto, o barramento Santa Rosa 220 kV parece ter sido mais sensível a eventos extremos de mercado, como períodos de seca ou picos de demanda não previstos, o que justifica os valores acima da média.

Figura 10 – Variação da carga da base de dados.



Fonte: Elaborado pelo autor.

Figura 11 – Variação da geração hídrica da base de dados.



Fonte: Elaborado pelo autor.

Conforme apresentado na Figura 9, é possível conferir a evolução do preço da energia elétrica no barramento Santa Rosa 220kV (em Sol/MWh), entre julho de 2022 e outubro de 2024. Nota-se um comportamento altamente instável, especialmente entre o início de 2023 e setembro do mesmo ano, período em que se verificam

sucessivos picos acima de 1000 Sol/MWh. Após esse intervalo, o preço médio tende à estabilidade, embora ainda haja episódios pontuais de elevação.

A volatilidade dos preços pode estar associada à combinação de fatores estruturais e conjunturais do setor elétrico. De acordo com Owolabi *et al.* (2023), o aumento da utilização de fontes renováveis influencia a estrutura de preços, pois há maior variabilidade na oferta. Entretanto, a volatilidade tende a ser mitigada em mercados com reservas regulatórias bem definidas ou maior interligação.

Logo, o comportamento dos preços observado no barramento Santa Rosa 220kV é um reflexo da transição energética marcada por incertezas, flutuações sazonais, variações abruptas e ajustes de mercado, fatores que reforçam a importância de políticas públicas e mecanismos de mercado de energia capazes de mitigar os riscos de volatilidade excessiva.

Utilizou-se a última semana da base de dados, a qual foi reservada exclusivamente para comparação entre os valores previstos e os dados reais observados. A previsão referente a esse período foi realizada com base nas variáveis data (período do ano), geração hídrica e carga. Para realizar este estudo e simulações, foram utilizados a versão 24.2.0.2923080 do MATLAB (2024b) contendo as bibliotecas *Deep Learning Toolbox*, *Statistics and Machine Learning Toolbox*, além disso, um computador com processador Intel® Core™ i5-10400F CPU @3,98 GHz 16 GB de RAM e placa de vídeo Nvidia RTX 3060 12GB.

Para avaliar a eficácia dos modelos desenvolvidos adotaram-se as métricas clássicas de erro absoluto médio (MAE), raiz quadrada do erro quadrático médio (RMSE) e erro percentual absoluto médio (MAPE), amplamente utilizadas em estudos sobre previsão de séries temporais e preços de energia.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (35)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (36)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - x_i|}{x_i} \quad (37)$$

$$PCC = \frac{cov(x_i, y_i)}{\sigma_{x_i} \sigma_{y_i}} \quad (38)$$

Onde:

$$\begin{cases} y_i \rightarrow \text{previsão} \\ x_i \rightarrow \text{real} \\ n \rightarrow n^\circ \text{ de pontos} \end{cases}$$

4.1 REDE SVM

4.1.1 Pré-processamento

Para a preparação dos dados para a rede SVM, realizou-se a importação dos dados brutos do arquivo Excel extraído do site COES utilizado como base de dados, contendo variáveis de carga, geração hidrelétrica e preço horário de energia do barramento Santa Rosa 220kV. A limpeza dos dados consistiu na remoção de todas as linhas que apresentavam valores faltantes (*NaN*) após a criação das variáveis atrasadas (*lags* semanais: preço, carga e geração hidrelétrica, para uma a quatro semanas anteriores), assegurando a integridade total dos dados de entrada. Em seguida, o formato textual dos preços, que utilizavam vírgulas como separador decimal, foi convertido para o formato numérico do MATLAB utilizando *directImportOptions*. As datas foram, então, convertidas ao formato *datetime* para viabilizar a extração de características de data, como hora, dia da semana e mês, no MATLAB, garantindo a integridade total dos dados utilizados (Fan *et al.*, 2016; Tschora *et al.*, 2022).

Posteriormente, aplicou-se a técnica de *winsorização*, com substituição dos valores extremos pelos percentis 1 e 99, a fim de evitar distorções no aprendizado causadas por *outliers*. Adicionalmente, foi realizada a normalização pelo método *Z-score*, utilizando média e desvio padrão do treino calculados exclusivamente com dados do conjunto de treino. Isso visou a aplicação à validação para evitar vazamento de dados futuros, ou *data leakage*, como é conhecido mundialmente (Alpaydin, 2020). Em seguida, dividiu-se sequencialmente os dados em conjuntos de treino e validação, com 70% e 30%, respectivamente (Zhang *et al.*, 2023). Sendo assim, garantiu-se o rigor metodológico e uma simulação realista de previsão futura, a fim de que o modelo seja aplicável a outros dados e possua previsões reais (Liu *et al.*, 2022).

4.1.2 Configuração e treinamento

A fim de obter uma boa configuração do modelo SVM, ou seja, encontrar uma função de regressão que minimiza erros dentro de uma margem de tolerância (*Epsilon*) através da maximização da margem entre vetores chamados vetores de suporte (o que dá origem ao nome do modelo) para garantir uma boa capacidade de generalização do modelo, optou-se por permitir que o tipo de *kernel* fosse escolhido automaticamente por otimização *bayesiana*. Desse modo, o algoritmo testou diferentes funções como linear, polinomial, gaussiano, selecionando a que melhor se adapta à relação entre as variáveis de entrada, pois há melhor flexibilidade em capturar relações complexas presentes nos dados de preços para os horários de energia elétrica (Alpaydin, 2020; Haykin, 2009).

Os principais hiperparâmetros configurados foram a Regularização (C), que controla o equilíbrio entre minimizar erros de treinamento e evitar sobreajuste (*overfitting*). Adicionalmente, foram configurados o *kernel* escolhido pela função *KernelFunction*, o parâmetro do próprio *kernel* selecionado e o *Epsilon*. Estes hiperparâmetros foram otimizados automaticamente pela busca *bayesiana*. Isso permitiu ao algoritmo testar diferentes combinações e tipos de *kernel* (linear, polinomial, gaussiano), possibilitando selecionar aqueles que oferecem melhor desempenho para o conjunto de dados analisado.

A escolha e o ajuste de cada parâmetro são dinâmicos e dependem do *kernel* selecionado durante o processo de otimização, tornando o modelo mais flexível e adequado à complexidade dos dados energéticos analisados. No MATLAB, a utilização da “*Statistics and Machine Learning Toolbox*” permitiu uma busca eficiente dos hiperparâmetros citados anteriormente, melhorando significativamente o desempenho preditivo final. Contudo, isso resultou em um custo de processamento computacional mais elevado e, conseqüentemente, em maior tempo de treinamento (Alpaydin, 2020; Cortes; Vapnik, 1995).

Na sequência, foram feitos o treinamento e a validação. Foi utilizada validação cruzada (*k-fold* = 5) do tipo *walk-forward*, a fim de validar os modelos e garantir a generalização. Os dados de treino foram divididos em cinco subconjuntos iguais que se alternam entre treinamento e validação, com a finalidade de assegurar a robustez do modelo (Alpaydin, 2020). Durante o treinamento, adotou-se como critério de parada

a convergência do processo de otimização bayesiana, com limite máximo de 30 iterações, a fim de reduzir o tempo de processamento da rede, prevenindo treinamentos extensos e ineficientes (Snoek *et al.* 2012).

Com o modelo SVM treinado, a previsão foi realizada para um horizonte de sete dias utilizando uma abordagem *multi-step* direta, na qual cada ponto futuro foi estimado de forma independente com base em variáveis sazonais (hora, dia da semana e mês) e *lags* semanais históricos de carga, geração hidráulica e preços.

4.2 REDE LSTM

4.2.1 Pré-processamento dos Dados

Com o modelo SVM treinado, a previsão foi realizada para um horizonte de sete dias utilizando uma abordagem *multi-step* direta, na qual cada ponto futuro foi estimado de forma independente com base em variáveis sazonais (hora, dia da semana e mês) e *lags* semanais históricos de carga, geração hidráulica e preços. Foi realizada a conversão de formatos: a coluna de data foi transformada para *datetime* e as demais colunas para *double*, ajustadas para casos de números em formato de texto no Excel para melhor leitura no MATLAB. Em seguida, foi feita a remoção dos registros com valores ausentes (*missing*), especialmente aqueles não aplicáveis após a criação de defasagens temporais (*lags*), por meio de *rmmmissing*. Adicionalmente, foi realizada a ordenação cronológica a fim de garantir a sequência temporal correta (Memarzadeh; Keynia, 2021).

Aplicou-se, previamente, a técnica de *winsorização*, com substituição dos valores extremos pelos percentis 1 e 99, para evitar distorções no aprendizado causadas por outliers. Também foi aplicada Engenharia de Variáveis Exógenas, que geraram atributos sazonais como hora do dia, dia da semana e indicação de fim de semana, além de *lags* de preço anteriores (30min, 1h, 2h, 24h, 48h e 168h) para capturar dependências temporais de curto e médio prazo (Zhou *et al.*, 2019). Além disso, foi feita a Normalização Z-score, na qual se calculou a média (μ) e o desvio-padrão (σ) exclusivamente sobre o conjunto de treino e normalizou-se todo o conjunto de dados (treino, validação e teste), desse modo evitando vazamento de informação (*data leakage*) (Goodfellow *et al.*, 2016). Por fim, foi feito o particionamento temporal para a preparação dos dados, no qual se dividiu a base de dados em treino (70%),

validação (30%) e teste para uma semana após a base de dados, simulando um cenário real de previsão futura.

4.2.2 Configuração e treinamento

Para transformar os dados tabulares em sequências temporais compatíveis com a rede LSTM, adotou-se a técnica de janelas deslizantes (*sliding window*). Sendo assim, definiu-se um parâmetro de *lookback* (olhar para trás) e o horizonte de previsão, correspondentes ao número de passos a serem previstos, pois a rede LSTM utiliza o método *direct forecast*, realizando a previsão direta para o número de passos desejados (Zhang *et al.*, 2020).

No trabalho, são apresentados dois horizontes de previsão: tanto para um dia (ou seja, 24h ou 48 pontos), quanto para uma semana (168h ou 336 pontos). Por exemplo, foram adotados *Lookback* e *Horizon* igual a 48 para previsão de um dia e igual a 336 para a previsão de uma semana. Para cada instante t , construiu-se um vetor de entrada contendo as características de carga, geração hídrica, hora do dia, dia da semana e defasagens de preços, assim como mencionado anteriormente, representando o comportamento histórico até t . A saída correspondente à entrada é uma sequência futura de preços de energia de comprimento igual ao horizonte definido (Memarzadeh; Keynia, 2021; Zhou *et al.*, 2019).

Logo após, foi feita a etapa de configuração e treinamento do modelo LSTM, que adotou uma arquitetura composta por uma camada de entrada capaz de receber sequências de múltiplas variáveis ao longo do tempo, seguida por duas camadas de memória sequencial que aprendem padrões temporais em diferentes níveis de abstração, ou seja, contém 200 camadas ocultas (*hidden layers*). Essas camadas são intercaladas por mecanismos de “*dropout*” com taxa de 20%, a fim de evitar que o modelo se ajuste em excesso aos dados de treinamento. Ao final da arquitetura LSTM, há uma camada densa que converte as representações extraídas em previsões pontuais, e uma camada de regressão compara essas saídas aos valores reais para guiar o aprendizado (Zhou *et al.*, 2019).

Para otimizar os pesos, utilizou-se um algoritmo de adaptação de taxa de aprendizado, chamado otimizador *Adam*, que ajusta automaticamente o ritmo das atualizações conforme a convergência, complementado por uma estratégia de redução gradual dessa taxa a intervalos predefinidos para refinar o ajuste. A taxa

inicial foi de 0,005, dividida pela metade a cada 50 épocas (Li; Becker, 2021). Note que o treinamento foi conduzido em ciclos suficientes para capturar a dinâmica dos dados (*MaxEpochs*=250, *MiniBatchSize*=128 e *Shuffle*="never"), mantendo a ordem cronológica das amostras, requisito essencial para redes temporais. Parte dos dados foi reservada para validação periódica durante o treinamento, o que permite monitorar a performance e prevenir o sobreajuste (Zhang *et al.*, 2020). Fez-se uso de processamento gráfico para acelerar o fluxo de cálculos, pois a rede LSTM tem um grande custo computacional, assim garantindo resultados mais rápidos sem comprometer a robustez do modelo e um processamento eficiente (Torres *et al.*, 2022).

Após o treinamento, inicia-se a etapa de previsão propriamente dita, na qual é utilizada a última sequência de observações conhecidas do conjunto de treino — correspondente aos pontos mais recentes normalizados — como entrada da rede LSTM. Diferentemente de abordagens autoregressivas (*open-loop*), em que cada saída prevista realimenta a entrada do próximo passo, aqui adotou-se a estratégia *direct multi-step*, na qual a rede é treinada para prever simultaneamente todos os 336 passos futuros (7 dias) a partir de uma única sequência de entrada, sem utilizar como entrada suas próprias previsões. Essa abordagem evita o acúmulo de erro ao longo da série prevista, proporcionando maior estabilidade e consistência para horizontes longos (Weron, 2014). Simulando, assim, o avanço do tempo dentro da própria rede e permitindo que ela projete o comportamento futuro do preço de energia a partir dos padrões que aprendeu. A previsão gerada pela rede é inicialmente obtida em escala normalizada e, portanto, é necessário realizar a desnormalização para retornar à escala original dos preços. Essa operação é feita utilizando a média (μ) e o desvio padrão (σ) calculados a partir do conjunto de treino.

4.3 REDE ARTMAP FUZZY MODIFICADA

4.3.1 Pré-processamento dos Dados

Inicialmente, os dados brutos foram submetidos a um processo de pré-processamento que incluiu a verificação de valores faltantes, correção de formatos e normalização. Os valores de entrada e saída foram normalizados no intervalo de 0 a 1, a fim de ter compatibilidade com o modelo ARTMAP-Fuzzy, utilizando a técnica de reescala *min-max*. Adicionalmente, aplicou-se previamente a técnica de *winsorização*,

com substituição dos valores extremos pelos percentis 1 e 99, a fim de evitar distorções no aprendizado causadas por outliers, conforme descrito por Müller *et al.* (2020), visando evitar a proliferação de categorias na rede devido a variações abruptas de escalas.

O conjunto de dados foi então dividido em dois subconjuntos: dados para treinamento do modelo e dados para validação e teste. A separação seguiu critérios temporais, garantindo que os dados de validação correspondessem a períodos posteriores aos dados de treinamento, evitando assim *data leakage*. A definição dos vetores de entrada e saída da rede seguiu uma abordagem baseada em defasagens temporais (*lags*), conforme aplicado por (Fernández *et al.*, 2023). Foram adotadas como entradas variáveis de atraso do tipo semanais (de uma a quatro semanas), representando o preço da energia, carga e geração hídrica, com o objetivo de capturar a influência sazonal de curto prazo. Ou seja, a saída desejada para cada padrão de entrada corresponde ao preço sazonal da energia na hora correspondente ao instante considerado. Tal configuração caracteriza um modelo de previsão direta supervisionada (*direct mapping*), no qual o valor da variável-alvo no instante atual é estimado a partir de múltiplas defasagens temporais. Essa abordagem é típica em tarefas de previsão de curto prazo (*short-term load/price forecasting*) e permite capturar padrões sazonais e comportamentais complexos na série temporal.

Além disso, conforme proposto por (Fernández *et al.*, 2023), todos os vetores foram codificados utilizando o *complement coding*, dobrando a dimensionalidade de entrada e garantindo a preservação da granularidade informacional dos dados após normalização.

4.3.2 Configuração e treinamento

Para a previsão do preço de energia, foi utilizada uma rede neural baseada na arquitetura ARTMAP-Fuzzy Modificada (FAM). A configuração do modelo considerou uma abordagem supervisionada direta (*direct mapping*), na qual o valor atual do preço de energia foi previsto a partir de variáveis históricas defasadas.

O vetor de entrada da rede foi composto por múltiplas variáveis defasadas temporalmente. Foram utilizados os valores de quatro semanas anteriores (*lags* de 336, 672, 1008 e 1344 intervalos de 30 minutos) das variáveis de preços, carga e geração hídrica. Além disso, o instante atual foi representado por uma codificação binária composta pelas informações de data e hora (dia do ano, mês, ano, hora e

minuto), convertidas em vetores binários para representar características sazonais e horários de forma categórica. O vetor de saída foi definido como o valor normalizado do preço de energia correspondente ao mesmo instante dos dados de entrada, caracterizando uma previsão direta no tempo atual.

A separação dos dados foi realizada com base temporal: os primeiros 70% da base foram utilizados para treinamento, e os 30% finais para validação, garantindo que os dados de teste pertencessem a períodos posteriores aos dados de treino, evitando assim o data leakage. Posteriormente, a rede foi treinada com os seguintes hiperparâmetros: parâmetro de escolha (α) igual a 0,1; e critérios de vigilância ($\rho_a, \rho_b, \rho_{ab}$) iguais a 0,98; 1 e 1, respectivamente. Esses valores foram definidos com base em experimentação empírica e referências da literatura que aplicaram configurações semelhantes ao contexto de previsão energética (Abreu *et al.*, 2019; Fernández *et al.*, 2023). Após isto foi feita previsão aplicando a rede ARTMAP-Fuzzy, previamente treinada, sobre novos dados de entrada organizados com a mesma estrutura utilizada no treinamento.

4.4 CRITÉRIOS DE COMPARAÇÃO

Com o objetivo de obter uma comparação metodologicamente coerente e justa entre as técnicas de *ML* utilizadas neste estudo, foram adotadas metodologias uniformes tanto de pré-processamento quanto de seleção de variáveis e avaliação de desempenho. Isso visou assegurar que as diferenças de desempenho observadas nos diferentes modelos estivessem relacionadas exclusivamente às características intrínsecas dos algoritmos. Além disso, a divisão do conjunto em treino e validação é mantida para os três modelos, garantindo equivalência temporal e, assim, evitando qualquer forma de vazamento de informação (*data leakage*).

Além disso, aplicou-se a mesma técnica de *winsorização* (percentis 1% e 99%) para atenuar o efeito de *outliers* extremos, seguido de normalização *Z-score* ou *Min-Max*, conforme exigência da arquitetura algorítmica de cada modelo. O processo de geração de variáveis de entrada (*features*) seguiu um padrão equivalente, com o uso de *lags* temporais múltiplos para as variáveis de preço, carga e geração hídrica. Informações sazonais como hora, dia da semana e mês, e codificações especiais, como *complement coding* para ARTMAP e vetores de sequência para LSTM, foram ajustadas para cada modelo, porém com equivalência informacional.

Com relação ao horizonte de previsão, todos os modelos foram avaliados para dois cenários distintos:

- Curto prazo (1 dia = 48 passos de 30 minutos)
- Médio prazo (1 semana = 336 passos de 30 minutos)

A abordagem de previsão foi mantida como *direct forecast* para os três modelos, com os modelos treinados para prever todos os pontos futuros simultaneamente, evitando a propagação de erros acumulativos que ocorrem em estratégias autorregressivas.

Por fim, a avaliação do desempenho preditivo foi realizada utilizando as mesmas quatro métricas mencionadas anteriormente (MAE, MAPE, RMSE e PCC), pois essa padronização metodológica garante que os resultados obtidos e analisados nos capítulos seguintes representem uma comparação robusta, transparente e reproduzível, conforme sugerido por (Lago *et al.*, 2018; Tschora *et al.*, 2022) em estudos de *benchmarking* entre modelos preditivos no setor energético.

5. RESULTADOS

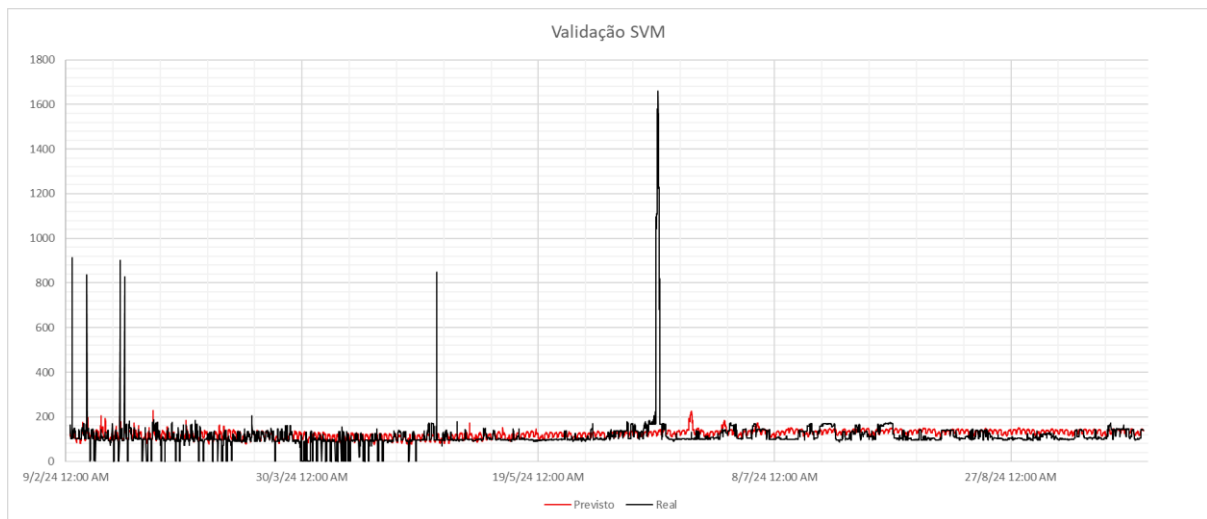
Nesta seção, são apresentados e discutidos os principais resultados obtidos a partir da aplicação das técnicas de *ML* selecionadas para a previsão de preços de energia elétrica no mercado peruano. Os modelos foram treinados e testados com dados históricos do mercado, considerando variáveis relevantes que influenciam o comportamento dos preços, conforme definido na metodologia (Zhang *et al.*, 2020).

5.1 PREVISÃO SVM

Após o treinamento do modelo SVM é feita a validação utilizando as métricas sugeridas na metodologia. Os dados reais são representados pela linha preta e os dados previstos pelo modelo são representados pela linha vermelha, conforme a Figura 12.

5.1.1 Validação do treinamento

Figura 12 – Validação do treino SVM.



Fonte: Elaborado pelo autor.

Tabela 2 – Métricas de validação do treino SVM.

MAE	MAPE	RMSE	PCC
28,20	137,28	82,20	0,215

Fonte: Elaborado pelo autor.

Percebe-se que o modelo conseguiu seguir relativamente o comportamento geral dos dados reais, exceto em alguns pontos específicos onde ocorreram grandes

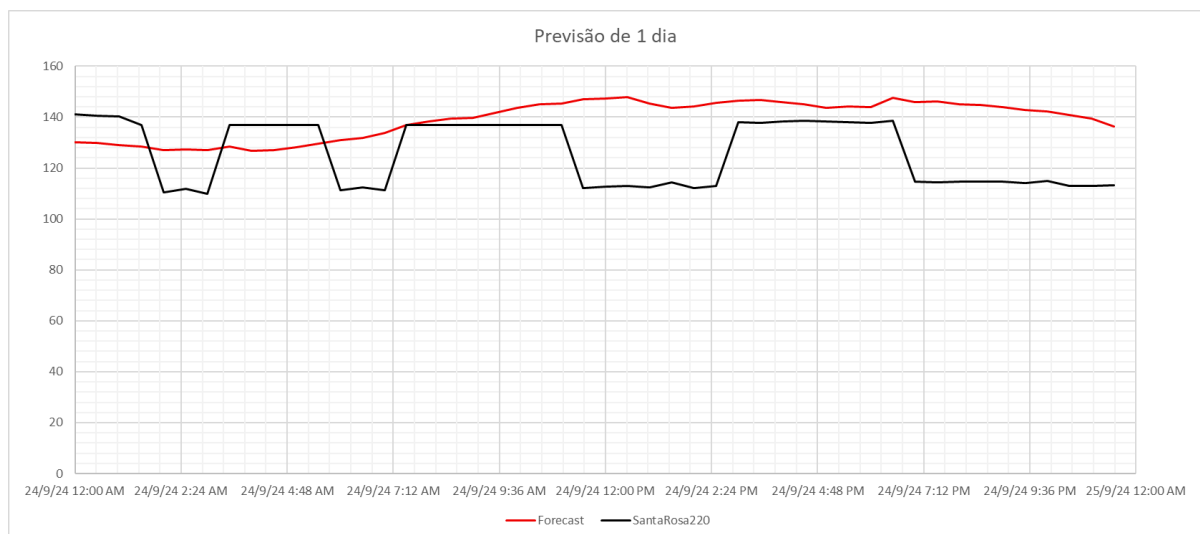
picos os quais o modelo apresentou dificuldades significativas, sugerindo sensibilidade a valores extremos ou eventos atípicos que são comuns para preços de energia (Müller, 2018).

As métricas quantitativas mostram um MAE moderado, porém um MAPE alto indicando dificuldades do modelo em lidar com valores extremos pontuais sendo reforçado pelo RMSE representando as discrepâncias pontuais entre valores reais e previstos. Além disso, o PCC é baixo, sugerindo que o modelo não é eficaz em capturar tendências gerais dos preços, devido a altas variações dos preços (Amorim, 2019).

5.1.2 Previsão de um dia

Assim feito a simulação de previsão de preços de energia para um dia após na Figura 13, isto significa que são 24 horas a frente, logo 48 pontos em passos de 30 minutos.

Figura 13 – Previsão para um dia SVM.



Fonte: Elaborado pelo autor.

Tabela 3 – Métricas da previsão 1 dia SVM.

MAE	MAPE	RMSE	PCC
21,22	18,03	24,66	0.001

Fonte: Elaborado pelo autor.

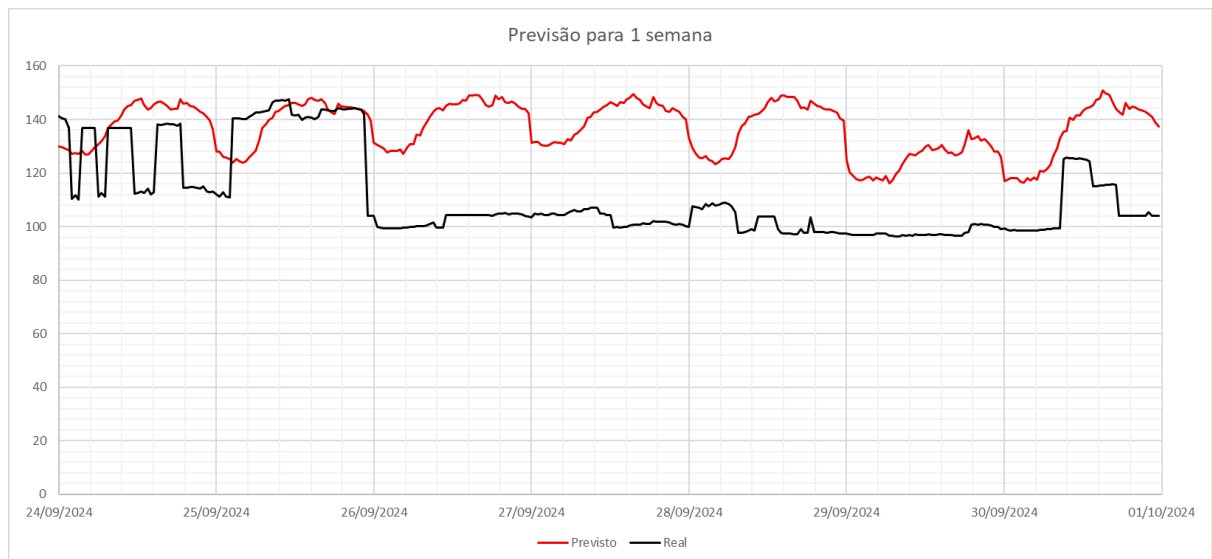
Observa-se que a previsão é consideravelmente mais suave e menos responsiva em relação às oscilações reais dos dados, indicando uma dificuldade do modelo em captar a variabilidade da série de preços de energia.

Essa limitação é reforçada pelas métricas quantitativas apresentadas na Tabela 3 que apesar de o erro absoluto médio e o erro percentual médio absoluto estarem numericamente aceitáveis, o erro quadrático médio ainda evidencia a existência de discrepâncias relevantes em pontos específicos da série o que fica evidente quanto ao indicador mais crítico que é o coeficiente de correlação de Pearson (PCC).

Este comportamento é típico de *underfitting* (suavização), que ocorre quando o modelo é demasiadamente simples para representar os padrões subjacentes aos dados (Tschora *et al.*, 2022; Zahid *et al.*, 2019).

5.1.3 Previsão de uma semana

Figura 14 – Previsão para a semana SVM.



Fonte: Elaborado pelo autor.

Tabela 4 - Métricas da previsão 1 semana SVM

MAE	MAPE	RMSE	PCC
27,36	26,35	30,72	0,240

Fonte: Elaborado pelo autor.

Nota-se que a curva de previsão apresenta um padrão durante a previsão devido a função de regressão utilizada durante o treinamento que busca minimizar os erros dentro de uma margem de tolerância ao redor da série real favorecendo

comportamento suavizado e contínuo. Como a entrada do modelo é composta por defasagens semanais, ele é guiado por padrões médios recorrentes semanais. Isso gera uma previsão com ritmo próprio, semelhante a um ciclo, resultando em uma onda contínua com periodicidade aproximada ao histórico observado (Tschora *et al.*, 2022; Zhou *et al.*, 2019).

O modelo SVM utiliza previsão *multistep direct*, desse modo, cada valor futuro ao longo do horizonte de uma semana foi estimado de forma independente, com base exclusivamente em variáveis exógenas (hora, dia, mês) e em *lags* históricos reais de carga, geração hidráulica e preços defasados em múltiplas semanas. especialmente em horizontes mais longos assim como de uma semana, logo, como não há reutilização das saídas previstas é evitado a propagação de erros, entretanto a ausência de realimentação também impede o modelo de ajustar dinamicamente sua trajetória de previsão em função de desvios sequenciais, o que contribui para um padrão suavizado nas estimativas. Embora tenha sido aplicada normalização dos dados e otimização bayesiana dos hiperparâmetros com validação cruzada do tipo *walk-forward*, o modelo ainda apresentou indícios de *underfitting*, especialmente visíveis na baixa correlação entre as séries real e prevista e na incapacidade de capturar picos e quedas abruptas. Essa limitação pode ser associada à própria estrutura do SVM, que, sem mecanismos de memória ou estrutura dinâmica, pode não ser o ideal para capturar padrões temporais de longo prazo, como também apontado por (Tschora *et al.*, 2022; Zhou *et al.*, 2019)

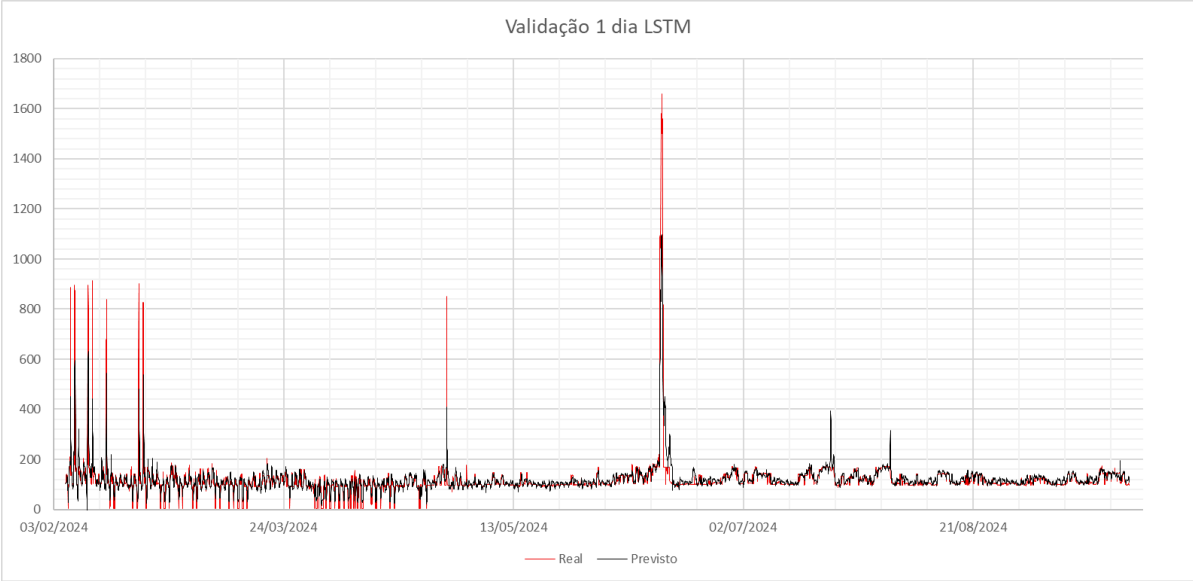
Esse comportamento também é apontado em diversos estudos que destacam as limitações de modelos de aprendizado de máquina ao lidar com a volatilidade do mercado de energia elétrica, principalmente de preços.

5.2 PREVISÃO LSTM

Assim como é mencionado anteriormente, é feito a validação do treino tanto para o horizonte de previsão de 24 horas quanto para uma semana (168 horas).

5.2.1 Validação do treinamento

Figura 15 – Validação do treino LSTM para previsão de um dia.



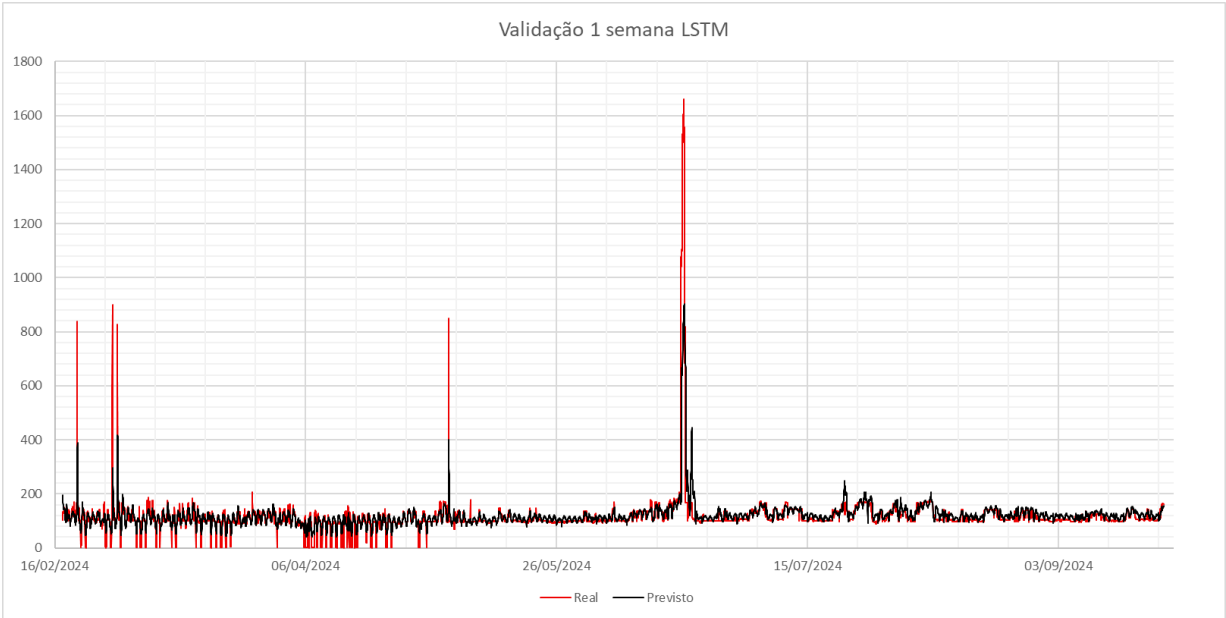
Fonte: Elaborado pelo autor.

Tabela 5 - Métricas de validação do treino LSTM para um dia.

MAE	MAPE	RMSE	PCC
28,68	105,77	84,42	0,386

Fonte: Elaborado pelo autor.

Figura 16 - Validação do treino LSTM para previsão de uma semana.



Fonte: Elaborado pelo autor.

Tabela 6 – Métricas de validação do treino LSTM para uma semana.

MAE	MAPE	RMSE	PCC
-----	------	------	-----

36,55	148,35	95,73	0,110
-------	--------	-------	-------

Fonte: Elaborado pelo autor.

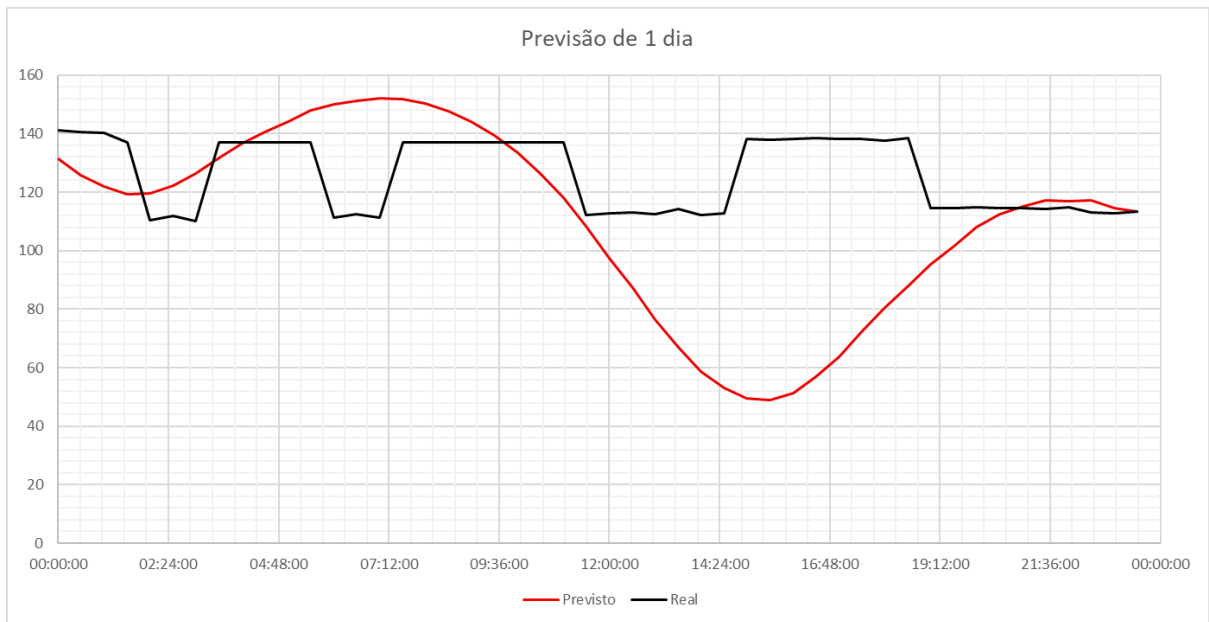
Observa-se que o modelo conseguiu acompanhar parcialmente o comportamento geral da série temporal de preços, especialmente em períodos de relativa estabilidade. Porém nos momentos em que ocorreram picos acentuados de preços, o modelo apresentou grande dificuldade em capturar a amplitude e o instante exato dessas variações, o que é consistente com a literatura sobre o comportamento suavizado das redes neurais recorrentes como a LSTM em séries altamente voláteis (Lago; *et al.*, 2018; Weron, 2014).

As métricas de validação indicam que, apesar do MAE moderado, o modelo apresentou alto MAPE e RMSE elevado, evidenciando dificuldade em prever variações extremas de preço. Além disso, o baixo PCC revela fraca correlação entre os valores previstos e reais, sugerindo que a LSTM não conseguiu capturar bem a tendência dos dados em cenários com forte volatilidade.

Tais resultados são esperados em aplicações que utilizam apenas variáveis históricas diretas e não incorporam variáveis exógenas mais explicativas, como condições hidrológicas, meteorologia ou políticas operacionais (Zhang *et al.*, 2020). Portanto, ainda que o modelo LSTM represente uma abordagem robusta para séries temporais, sua efetividade depende diretamente da qualidade e riqueza informacional das variáveis de entrada.

5.2.2 Previsão de um dia

Figura 17 - Previsão para um dia LSTM.



Fonte: Elaborado pelo autor.

Tabela 7 - Métricas da previsão 1 dia LSTM.

MAE	MAPE	RMSE	PCC
25,23	19,81	36,66	0,050

Fonte: Elaborado pelo autor.

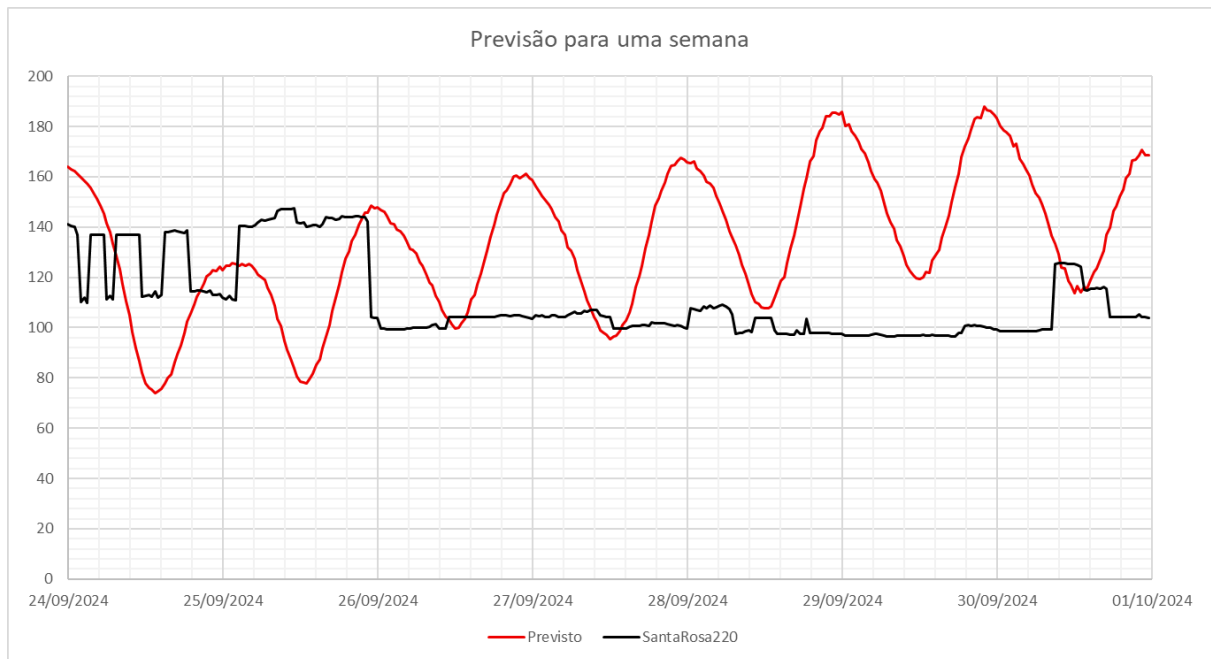
Observa-se que a previsão gerada pela LSTM para um horizonte de 24 horas apresenta um comportamento suavizado, com tendência senoidal e baixa responsividade às oscilações reais dos preços de energia. Embora o modelo consiga seguir parcialmente a tendência geral da série, ele demonstra limitação em capturar as variações abruptas e os patamares reais dos valores, comportamento este típico de modelos que sofrem de *underfitting*, nos quais a rede tende a suavizar excessivamente a previsão (Tschora *et al.*, 2022; Zahid *et al.*, 2019).

Esse padrão é confirmado pelas métricas apresentadas na Tabela 7. As métricas MAE e MAPE são relativamente baixos, sugerindo que o modelo é capaz de fornecer estimativas razoáveis em termos de magnitude média. No entanto, o RMSE ainda indica discrepâncias pontuais e, principalmente, o coeficiente de correlação mostra baixa aderência entre as formas da curva prevista e da curva real, reforçando a dificuldade do modelo em reproduzir o comportamento dinâmico dos preços mesmo em horizontes de curtíssimo prazo.

Essa limitação pode estar relacionada ao fato de a LSTM ser sensível à ausência de variáveis exógenas explicativas e à elevada suavidade do modelo em tarefas com janelas curtas de previsão, especialmente quando o comportamento da série é altamente volátil e não linear — características comuns no contexto dos preços de energia elétrica (Tschora *et al.*, 2022).

5.2.3 Previsão de uma semana

Figura 18 - Previsão para a semana LSTM.



Fonte: Elaborado pelo autor.

Tabela 8 - Métricas da previsão 1 semana LSTM.

MAE	MAPE	RMSE	PCC
37,13	37,85	44,33	-0,456

Fonte: Elaborado pelo autor.

Nota-se que a curva de previsão exibe um padrão fortemente cíclico e suavizado, com forma senoidal e frequência quase constante. Esse comportamento indica que o modelo não foi capaz de capturar as variações reais dos preços de energia durante esse período, adotando uma estratégia de média temporal em que o sinal previsto oscila de maneira regular, ignorando eventos abruptos, patamares reais e discontinuidades da série histórica.

As métricas quantitativas reforçam essa observação. O MAE e MAPE mostram que o modelo possui erros significativos em termos médios, mesmo em um horizonte mais amplo. O RMSE confirma a presença de discrepâncias entre os valores reais e previstos, principalmente nos períodos em que ocorrem mudanças bruscas de regime. O indicador mais preocupante, porém, é o PCC, que foi negativo, sinalizando que há inversão de tendência entre os dados reais e as previsões em alguns momentos — o que compromete fortemente a confiabilidade da previsão gerada (Weron, 2014).

Esse tipo de comportamento pode estar associado à ausência de variáveis exógenas relevantes no modelo e à dificuldade da LSTM em generalizar para longos horizontes de tempo. Além disso, redes LSTM podem induzir suavizações excessivas quando o horizonte de previsão extrapola a capacidade de memória útil da rede (Lago *et al.*, 2018; Zhang *et al.*, 2020).

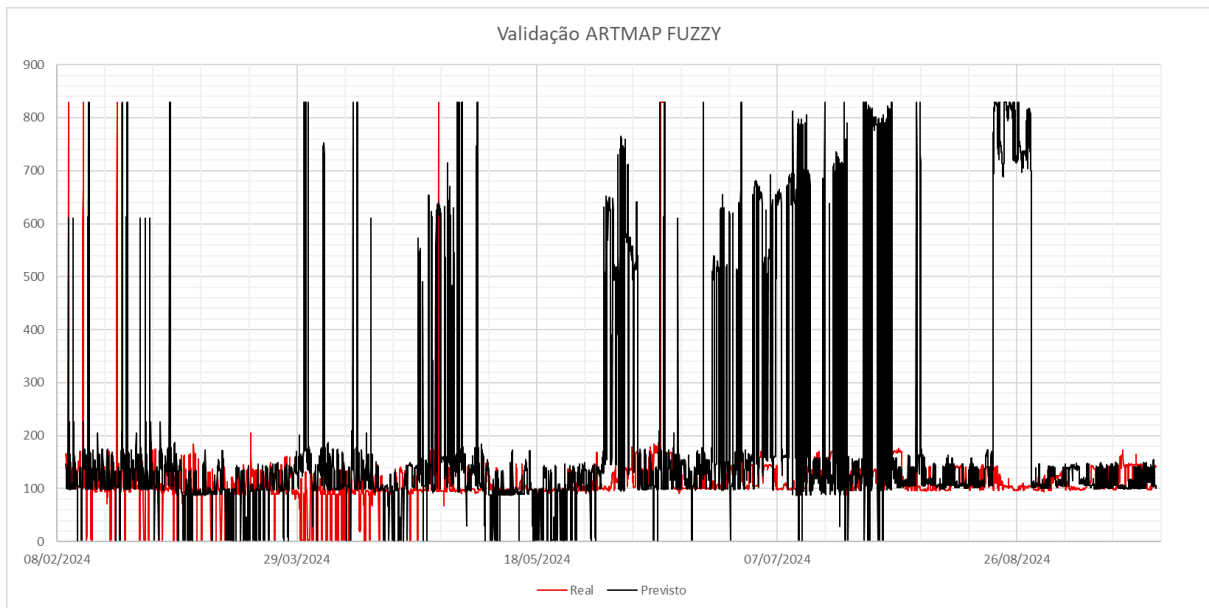
A baixa qualidade das previsões da LSTM para horizontes longos está relacionada à sua limitação em reter padrões temporais por muitos passos à frente, especialmente quando não há variáveis exógenas futuras disponíveis no conjunto de entrada. Esse efeito é agravado pelo acúmulo de incertezas ao longo da janela de previsão e pela tendência de suavização do modelo (*underfitting*), resultando em estimativas genéricas e cíclicas que se afastam dos padrões reais observados nos dados (Goodfellow *et al.*, 2016; Zhang *et al.*, 2020).

5.3 ARTMAP FUZZY MODIFICADA

Assim, como mencionado na metodologia, é feita a validação do treino para o modelo ARTMAP FUZZY.

5.3.1 Validação do treinamento

Figura 19 – Validação do treino ARTMAP *Fuzzy*.



Fonte: Elaborado pelo autor.

Tabela 9 - Métricas de validação do treino ARTMAP *Fuzzy*.

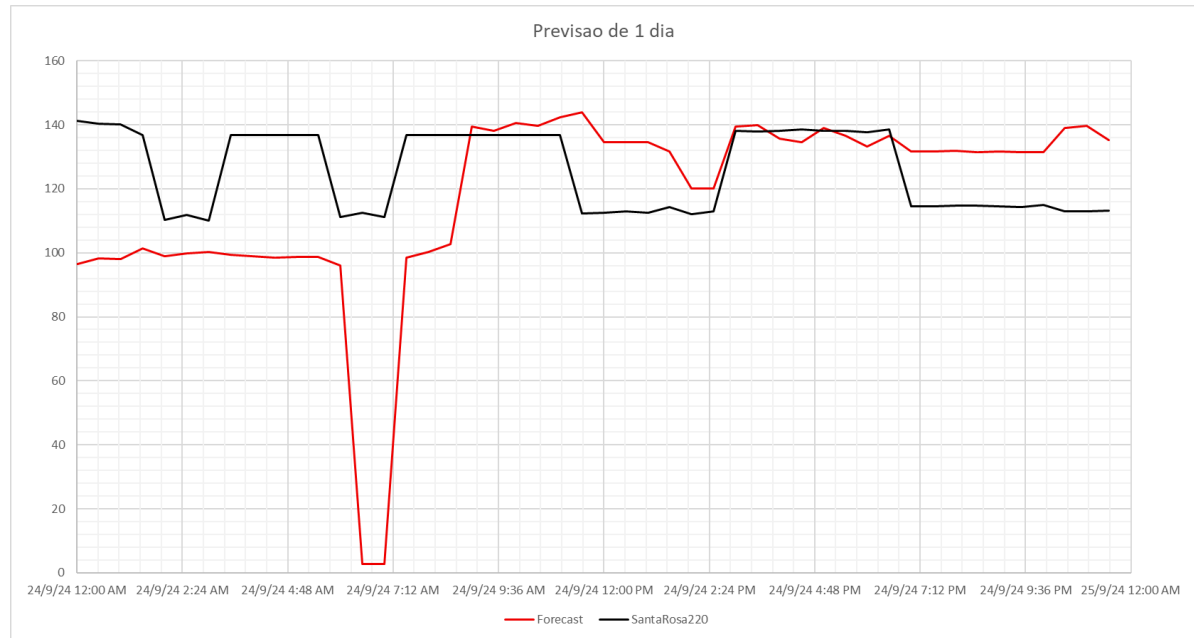
MAE	MAPE	RMSE	PCC
112,91	229,43	228,48	0,071

Fonte: Elaborado pelo autor.

Durante a etapa de validação é possível notar que a rede seguir relativamente o comportamento nos períodos com padrões similares aos do treinamento, sendo assim confirma a capacidade de reconhecimento, entretanto em momentos em que os dados de entrada são bem diferentes dos aprendidos, por exemplo variações abruptas devido à alta volatilidade dos preços, nota-se valores próximo de zero, devido ao aprendizado de preços conter valores próximos de zero. As métricas de desempenho evidenciam e reforça limitações do modelo com altos valores para MAE, MAPE e RMSE, além disso baixíssima correlação.

5.3.2 Previsão de um dia

Figura 20 - Previsão para um dia ARTMAP Fuzzy



Fonte: Elaborado pelo autor.

Tabela 10 - Métricas da previsão 1 dia ARTMAP Fuzzy.

MAE	MAPE	RMSE	PCC
23,17	18,45	31,49	0,071

Fonte: Elaborado pelo autor.

O modelo apresentou uma significativa melhora para a previsão de um dia em relação a validação geral obtendo métricas mais consistentes para os erros, obtendo um MAE e MAPE que indicam boa aderência aos valores reais na maior parte do dia, entretanto o RMSE indica que discrepâncias pontuais foram acentuadas o que é visível na Figura 20 onde a rede falhou em reconhecer o padrão de entrada. O coeficiente de correlação (PCC) é baixo, mas reflete uma correlação linear limitada pela rigidez do critério de vigilância, ainda assim, o resultado mostra que a rede consegue capturar padrões locais com maior precisão quando submetida a horizontes curtos de previsão, estando de acordo com a literatura (FERNÁNDEZ *et al.*, 2023; MORENO, 2010; MÜLLER, 2018)

5.3.3 Previsão de uma semana

Figura 21 – Previsão para a semana ARTMAP Fuzzy.



Fonte: Elaborado pelo autor.

Tabela 11 - Métricas da previsão 1 semana ARTMAP Fuzzy.

MAE	MAPE	RMSE	PCC
43,71	41,34	55,19	0,161

Fonte: Elaborado pelo autor.

Nota-se, pela Figura 21, que o modelo apresentou desempenho mais limitado frente ao aumento do horizonte temporal, agora para uma semana, embora o MAE e o MAPE indiquem um resultado intermediário, a métrica RMSE confirma que houve discrepâncias pontuais relevantes que pode ser notada pelos vales próximos de zero, ainda sim obteve um coeficiente de correlação superior ao observado na previsão de um dia, sugerindo que houve melhora ao captar tendências sazonais.

Também é possível notar um padrão de valores em zero durante as madrugadas, isto ocorre devido ao aprendizado do modelo com parâmetros da alta vigilância que aprende valores próximos de zero nestes períodos, isto ocorre devido ao contexto do sistema elétrico peruano, em que é comum o desligamento das usinas termelétricas flexíveis por despacho econômico, isto ocorre devido a redução da demanda de energia e à priorização de fontes mais econômicas, como a hidrelétrica ou renováveis, portanto resulta em preços de energia mais baixos, em momentos próximo de zero para o barramento Santa Rosa 220 kV que é comum o desligamento

das usinas termelétricas no período da madrugada, principalmente em períodos chuvosos que as hidrelétricas conseguem entregar mais geração hidrelétrica (Vicente, 2024).

5.4 COMPARAÇÃO DOS MODELOS DE PREVISÃO E DISCUSSÃO

Após a implementação dos modelos propostos é possível detectar importantes particularidades em cada um deles, a tabela a seguir resume as métricas de avaliação para cada modelo em diferentes horizontes.

Tabela 12 – Comparação dos modelos de previsão

Modelo	Horizonte	MAE	MAPE	RMSE	PCC
SVM	1 dia	21,22	18,03	24,66	0.001
	1 semana	27,36	26,35	30,72	0,240
LSTM	1 dia	25,23	19,81	36,66	0,050
	1 semana	37,13	37,85	44,33	-0,456
ARTMAP <i>Fuzzy</i>	1 dia	23,17	18,45	31,49	0,071
	1 semana	43,71	41,34	55,19	0,161

Fonte: Elaborado pelo autor.

O modelo SVM obteve melhor desempenho em termos numéricos para previsão de curto prazo, entretanto houve baixíssima correlação indicando a incapacidade significativa em acompanhar corretamente os padrões voláteis dos preços de energia elétrica, já para o horizonte semanal obteve um bom equilíbrio entre as métricas de erros, mas ainda apresenta limitações importantes na captura das variações pois há suavização excessiva da previsão. Isto ocorre devido à ausência temporal, mesmo que utilizando *lags*, pois trabalha exclusivamente com hiperplanos

lineares ou não-lineares que limita a capacidade em lidar com a natureza sequencial e altamente variável dos dados da série temporal.

Em seguida, para a rede LSTM é possível notar dificuldades importantes em ambos os horizontes, especialmente no semanal que apresenta correlação negativa, isto ocorre devido a estrutura ineficiente e entradas limitadas que geram uma suavização extrema dos padrões temporais complexos. Estes resultados insatisfatórios sugerem a necessidade do modelo para inclusão de variáveis explicativas como precipitação, temperatura, despacho térmico entre outros.

Enfim, para o modelo ARTMAP *Fuzzy*, apresenta um comportamento misto, apesar de não alcançar os menores erros absolutos, obteve uma correlação ligeiramente superior para o horizonte de curto prazo, embora apresente graves limitações pela ocorrência frequente de valores da previsão próximos a zero, principalmente na previsão semanal. Isto ocorre devido a sua estrutura altamente vigilante que aprende valores próximos a zero durante a madrugada devido ao despacho econômico onde as termelétricas são desligadas devido à baixa demanda neste período do dia durante épocas chuvosas que se tem maior geração hídrica.

Apesar do rigor metodológico empregado, os resultados obtidos não são plenamente eficazes na previsão de preços de energia elétrica do mercado peruano, isto ocorre devido a complexidade ao compor os preços de energia, sendo as principais limitações a alta variabilidade devido a volatilidade dos preços horários que são fortemente influenciadas por fatores não lineares e exógenos, como chuvas, temperatura e irradiação solar (Fernández *et al.*, 2023), consequentemente a ausência dessas variáveis dificultam a modelagem precisa de oscilações abruptas, além disso, fatores operacionais e estruturais como intercâmbios energéticos entre regiões e barramentos do sistema interligado, capacidade de transmissão, bem como eventos sazonais (Mendling; Mouratidis, 2018; Tschora *et al.*, 2022). Portanto, embora os modelos capturem certos padrões de comportamento típicos do mercado a limitação na incorporação de variáveis exógenas críticas devido ao alto custo computacional foi um fator determinante para a qualidade limitada das previsões, assim é reforçado a recomendação da literatura para uso de abordagens multivariadas e integradas a fim de obter maior riqueza contextual e operacional do sistema energético (Li; Becker, 2021; Memarzadeh; Keynia, 2021; Wood *et al.*, 2013).

6 CONCLUSÃO E SUGESTÕES PARA FUTUROS TRABALHOS

6.1 CONCLUSÃO

Este trabalho de graduação tem como objetivo realizar uma análise comparativa entre três técnicas distintas de *ML* aplicadas à previsão de preços de energia elétrica no mercado peruano, sendo estas *Support Vector Machine* (SVM), *Long Short-term Memory* (LSTM) e *ARTMAP Fuzzy* Modificada. Utilizando dados reais do mercado peruano que foram extraídos pelo site do COES aplicados as estruturas metodológicas de cada modelo, os resultados obtidos indicam que cada modelo possui características únicas que influenciam diretamente o desempenho da previsão, tanto para um dia quanto para uma semana.

O SVM apresenta resultados razoáveis para previsões de curtíssimo prazo, ou seja, até 24 horas, porém a baixa correlação com os dados reais evidencia limitação significativa em acompanhar oscilações bruscas e capturar tendências complexas. Já para o modelo LSTM, apesar de possuir arquitetura especializada em capturar padrões temporais, apresentou limitações relevantes relacionadas à suavização excessiva das previsões e dificuldades em lidar com a volatilidade elevada, especialmente para horizonte de uma semana, onde gera correlação negativa, desse modo pode-se destacar a importância crítica da seleção apropriada das variáveis explicativas e seleção de dados contendo padrões anuais de variações além da arquitetura do modelo para esse tipo específico de aplicação. Por fim, o modelo *ARTMAP Fuzzy* Modificado demonstrou capacidade intermediária, mantendo correlação positiva tanto em curto quanto em médio prazo e erros moderados, porém há ocorrência recorrente de previsões com valores próximos a zero, especialmente durante as madrugadas, note que isto não representa uma falha do modelo, mas sim um reflexo do seu aprendizado adequado em relação ao padrão real observado no mercado em que os preços se aproximam de zero nestes horários devido ao despacho econômico e ao desligamento das usinas termelétricas, entretanto repete este comportamento mesmo em períodos anuais em que não ocorre o desligamento e não há quedas no preço, indicando dificuldades em interpretar padrões anuais ou necessidade de treinamento com períodos maiores de tempo.

Portanto pode se concluir que nenhum dos modelos utilizados apresentou bons desempenhos para previsão de preços de energia do mercado peruano ou confiável para aplicações de previsão operacional em mercados com volatilidade devido à

complexidade para composição dos preços que envolvem a necessidade de mais variáveis, como condições climáticas, hidrológicas, alterações regulatórias e variáveis sociais e sazonais.

6.2 SUGESTÕES PARA FUTUROS TRABALHOS

Diante das limitações identificadas neste estudo, recomenda-se a exploração das seguintes abordagens para extensão metodológica em trabalhos futuros voltados à previsão de preços de energia elétrica. Algumas dessas melhorias são:

- Inclusão de variáveis exógenas no modelo: climáticas e ambientais como precipitação, temperatura, radiação solar e velocidade do vento que afetam tanto a geração quanto o consumo de energia elétrica. Hidrológicas e operacionais como níveis dos reservatórios, vazões naturais e despacho térmico. Econômicas e regulatórias como custo de combustíveis (gás natural, carvão, óleo), taxa de câmbio, inflação e alterações regulatórias. Variáveis relacionadas ao mercado interligado como dados de mercados vizinhos, intercâmbios energéticos e capacidade de transmissão. Sociais e sazonais como feriados nacionais, eventos esportivos ou culturais.
- Desenvolvimento de modelos híbridos ou multivariados: estudar abordagens que combinem diferentes arquiteturas de *ML*, como LSTM com ARTMAP ou SVM com redes profundas, por exemplo EEMD-LSTM descrito em Zhou *et al.* (2019). Métodos multimodais permitindo capturar tanto dependências temporais quanto relações espaciais e categóricas.
- Análise de interpretabilidade: estudar e avaliar a contribuição de cada variável exógena a fim de obter melhor previsão dos preços de energia com menor custo computacional.
- Expansão da base de dados e estudo de granulidade: utilizar janelas históricas de maior magnitude, por exemplo 5 ou 10 anos e abranger múltiplos barramentos para verificar a generalização do modelo, adicionalmente avaliar o desempenho para diferentes frequências temporais dos dados, como 15 minutos, uma hora ou um dia observando como impacta no resultado da previsão.

REFERÊNCIAS

- ABREU, T.; MOREIRA, J. R.; MINUSSI, C. R.; LOTUFO, A. D. P.; LOPES, M. L. M. Short-term multinodal load forecasting using a fuzzy-ARTMAP neural network. *In: IEEE PES INNOVATIVE SMART GRID TECHNOLOGIES CONFERENCE - LATIN AMERICA (ISGT LATIN AMERICA)*, 2019, Gramado. **Proceedings** [...]. Gramado: IEEE, 2019, p. 1-6.
- ALPAYDIN, E. **Introduction to machine learning**. 4. ed. Cambridge: MIT press, 2020.
- AMORIM, A. J. **Previsão de carga multinodal formulada via rede neural baseada na teoria da ressonância adaptativa com treinamento direto e reverso**. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, Ilha Solteira, 2019.
- BELLO, A.; RENESES, J.; MUÑOZ, A.; DELGADILLO, ANDRES, Probabilistic forecasting of hourly electricity prices in the medium-term using spatial interpolation techniques. **International Journal of Forecasting**, Amsterdam, v.32,n.3,p. 966-980, 2016.
- CARPENTER, G. A.; GROSSBERG, S. The ART of adaptive pattern recognition by a self-organizing neural network. **Computer**, Piscataway, v. 21, n. 3, p. 77–88, 1988.
- CARPENTER, G. A.; GROSSBERG, S.; MARKUZON, N.; REYNOLDS, J. H.; ROSEN, D. B. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. **IEEE Transactions on Neural Networks**, New York, v. 3, n. 5, p. 698–713, 1992.
- CARPENTER, G. A.; GROSSBERG, S.; ROSEN, D. B. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. **Neural Networks**, New York, v. 4, n. 6, p. 759–771, 1991.
- CHANG, Z.; ZHANG, Y.; CHEN, W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. **Energy**, Lanzhou, v. 187, p. 1-12, 2019.
- COES – COMITÉ DE OPERACIÓN ECONÓMICA DEL SISTEMA INTERCONECTADO NACIONAL. **Portal institucional**, 2024. Disponível em: <https://www.coes.org.pe/>. Acesso em 6 mar. 2025.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, Norwell, v. 20, p. 273–297, 1995.
- DUA, S.; DU, X. **Data mining and machine learning in cybersecurity**. New York : CRC press, 2016.
- FAN, G.-F.; PENG, L.-L.; HONG, W.-C.; SUN, F. Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. **Neurocomputing**, Henan, v. 173, p. 958–970, 2016. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S0925231215012230>. Acesso em: 21 mai. 2025.

FERNÁNDEZ, L. B. G.; LOTUFO, A. D. P.; MINUSSI, C. R. Development of a Short-Term Electrical Load Forecasting in Disaggregated Levels Using a Hybrid Modified Fuzzy-ARTMAP Strategy. **Energies**, Ilha Solteira, v. 16, n. 10, 2023.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. **Deep learning**. Cambridge: MIT press Cambridge, 2016.

GRANGER, M. Uncertainty in energy and other forecasts. Em: GRANGER, M. **Interdisciplinary Research on Climate and Energy Decision Making**. 1. ed. London: Routledge, 2022. p. 197–207.

HAYKIN, S. **Neural networks and learning machines**. 3. ed. New Jersey: Prentice-Hall, 2009.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, Cambridge, v. 9, n. 8, p. 1735–1780, 1997.

JAHANGIR, H.; TAYARANI, H.; BAGHALI, S.; AHMADIAN, A.; ELKAMEL, A.; GOLKAR, M. A.; CASTILLA, M. A Novel Electricity Price Forecasting Approach Based on Dimension Reduction Strategy and Rough Artificial Neural Networks. **IEEE Transactions on Industrial Informatics**, New York, v. 16, n. 4, p. 2369–2381, 2020.

LAGO, J.; RIDDER, F. D.; SCHUTTER, B. D. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. **Applied Energy**, London, v. 221, p. 386–405, 2018.

LI, W.; BECKER, D. M. Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. **Energy**, Trondheim, v. 237, 2021.

LIU, J.; ZHANG, S.; WANG, J. Development and Comparison of Two Computational Intelligence Algorithms for Electrical Load Forecasts with Multiple Time Scales. *In*: 2022 POWER SYSTEM AND GREEN ENERGY CONFERENCE (PSGEC), 2022, Shanghai, **Proceedings** [...]. Shanghai: IEEE, 2022. p. 637-643

LORENA, A. C.; DE CARVALHO, A. C. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, Porto Alegre, v. 14, n. 2, p. 43–67, 2007.

MACIEJOWSKA, K.; UNIEJEWSKI, B.; WERON, R. Forecasting electricity prices. **arXiv preprint**. Ithaca: Cornell University, v. 1, 2022. *In press*.

MEMARZADEH, G.; KEYNIA, F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. **Electric Power Systems Research**, Amsterdam, v. 192, p. 106995, 2021.

MENDLING, J.; MOURATIDIS, H. **Information Systems in the Big Data Era**. 1. Ed. Cham: Springer Cham, 2018.

MORENO, A. L. **Análise de estabilidade transitória via rede neural ART-ARTmap Fuzzy Euclidiana modificada com treinamento continuado**. 2010. Tese (Doutorado em Engenharia) – Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, Ilha Solteira, 2010.

MÜLLER, M. R. **Análise de desempenho da rede neural artificial ARTMAP Fuzzy aplicada para previsão multi-step de cargas elétricas em diferentes níveis de agregação**. 2018. Tese (Doutorado em Engenharia) – Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, Ilha Solteira, 2018.

NOGUEIRA, G. O.; LIMA, M. Previsão dos preços de abertura, mínima e máxima de índices de mercados financeiros usando a associação de redes neurais LSTM. **arXiv e-prints**, Ithaca: Cornell University, v. 1, 2021.

OWOLABI, O. *et al.* A. Role of Variable Renewable Energy Penetration on Electricity Price and its Volatility Across Independent System Operators in the United States. **Data Science in Science**, London, v. 2, n. 1, p. 2158145, 2023.

SHAWE-TAYLOR, J. **Kernel Methods for Pattern Analysis**. 1 ed. Cambridge: Cambridge University, 2004.

SNOEK, J.; LAROCHELLE, H.; ADAMS, R. Practical Bayesian Optimization of Machine Learning Algorithms. *In*: Advances in Neural Information Processing Systems (NIPS), **Proceedings** [...]. Lake Tahoe: NeurIPS, v. 25, 2012.

TORRES, J. F.; MARTÍNEZ-ÁLVAREZ, F.; TRONCOSO, A. A deep LSTM network for the Spanish electricity consumption forecasting. **Neural Computing and Applications**, London, v. 34, n. 13, p. 10533–10545, 2022.

TSCHORA, L.; PIERRE, E.; PLANTEVIT, M.; ROBARDET, C. Electricity price forecasting on the day-ahead market using machine learning. **Applied Energy**, London, v. 313, p. 118752, 2022.

VICENTE, J. R. F. **Crónica de los precios de energía en el Perú**, 2024. LinkedIn: @Jaime Raúl F. Disponível em: <https://www.linkedin.com/pulse/crónica-de-los-precios-energía-en-el-perú-jaime-raúl-flor-vicente-gpwie>. Acesso em 22 abr. 2025.

WERON, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. **International journal of forecasting**, Amsterdam v. 30, n. 4, p. 1030-1081, 2014.

WOOD, A. J.; WOLLENBERG, B. F.; SHEBLÉ, G. B. **Power Generation, Operation, and Control**. Minnessota: Jonh Wiley & Sons, 2013.

YU, Y.; SI, X.; HU, C.; ZHANG, J. A review of recurrent neural networks: LSTM cells and network architectures. **Neural computation**, Massachusetts, v. 31, n. 7, p. 1235–1270, 2019.

ZAHID, M.; AHMED, F.; JAVAID, N.; ABBASI, R. A.; KAZMI, H. S. Z.; JAVAID, A.; BILAL, M.; AKBAR, M.; ILAHI, M. Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids. **Electronics**, Basel, v. 8, n. 2, p. 101–109, 2019.

ZHANG, R.; LI, G.; MA, Z. A Deep Learning Based Hybrid Framework for Day-Ahead Electricity Price Forecasting. **IEEE Access**, Piscataway, v. 8, p. 143423–143436, 2020.

ZHANG, S.; LIU, J.; WANG, J. High-Resolution Load Forecasting on Multiple Time Scales Using Long Short-Term Memory and Support Vector Machine. **Energies**, Conventry, v. 16, n. 4, 2023.

ZHOU, S.; ZHOU, L.; MAO, M.; TAI, H. M.; WAN, Y. An Optimized Heterogeneous Structure LSTM Network for Electricity Price Forecasting. **IEEE Access**, Piscataway, v. 7, p. 108161–108173, 2019.