UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Ilha Solteira - SP

ANDRÉIA BRASIL ALVES FERREIRA

PREVISÃO MÚLTIPLA EM SÉRIES TEMPORAIS DE ENERGIA ELÉTRICA COM *TEMPORAL FUSION TRANSFORMER*

Ilha Solteira
2024

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Ilha Solteira - SP

ANDRÉIA BRASIL ALVES FERREIRA

PREVISÃO MÚLTIPLA EM SÉRIES TEMPORAIS DE ENERGIA ELÉTRICA COM *TEMPORAL FUSION TRANSFORMER*

Tese apresentada à Universidade Estadual Paulista (UNESP), Faculdade de Engenharia, Ilha Solteira, para obtenção do título de Doutora em Engenharia Elétrica.

Área do Conhecimento: Automação

Orientador: Prof. Dr. Jonatas Boas Leite

Ilha Solteira
2024

FICHA CATALOGRÁFICA
Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

F383p

Ferreira, Andréia Brasil Alves.
Previsão múltipla em séries temporais de energia elétrica com temporal fusion transformer / Andréia Brasil Alves Ferreira. -- Ilha Solteira: [s.n.], 2024
188 f. : il.

Tese (doutorado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Automação, 2024

Orientador: Jonatas Boas Leite

Inclui bibliografia

1. Inteligência artificial. 2. Aprendizado profundo. 3. Mecanismos de atenção. 4. Análise de séries temporais. 5. Previsão de demanda. 6. Transformers.

Amanda Sertori dos Santos
Bibliotecária - CRB/8-9061
Seção Técnica de Referência, Atendimento ao Usuário e Documentação
Diretoria Técnica de Biblioteca e Documentação

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Ilha Solteira

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Previsão Múltipla de Energia Elétrica em Séries Temporais com Temporal Fusion Transformer

AUTORA: ANDRÉIA BRASIL ALVES FERREIRA
ORIENTADOR: JONATAS BOAS LEITE

Aprovada como parte das exigências para obtenção do Título de Doutora em Engenharia Elétrica, área: Automação pela Comissão Examinadora:

Prof. Dr. JONATAS BOAS LEITE (Participaçao Presencial)
Departamento de Engenharia Eletrica / Faculdade de Engenharia de Ilha Solteira - UNESP

Dr. MARIO ANDRES MEJIA ALZATE (Participaçao Virtual)
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira - UNESP

Prof. Dr. JULIO CESAR LOPEZ QUIZHPI (Participaçao Virtual)
Department of Electrical and Computing Engineering / Iowa State University - U.S.A

PROFESSOR DOUTOR MADSON CORTES DE ALMEIDA (Participaçao Virtual)
Departamento de Sistemas de Energia Elétrica / UNIVERSIDADE ESTADUAL DE CAMPINAS

Prof. Dr. RENZO AMILCAR VARGAS PERALTA (Participaçao Virtual)
Departamento de Engenharia de energia / Universidade Federal do ABC - UFABC

Ilha Solteira, 04 de setembro de 2024

DEDICATÓRIA

A Deus, pela luz e força que me acompanharam ao longo desta jornada.

A todos que acreditaram em mim, oferecendo apoio, amor e incentivo em cada etapa do caminho.

A todos aqueles que buscam a verdade através da ciência, mantendo acesa a chama da curiosidade e preservando a integridade do espírito científico, convictos de que o verdadeiro progresso não se encontra no caminho mais fácil, mas na perseverança de quem ousa trilhar o desconhecido.

A ciência é, acima de tudo, um exercício de coragem, dedicação e questionamento constante. Por isso, dedico este trabalho àqueles que, dentro das paredes da academia, permanecem firmes em suas convicções. Que esta obra seja uma prova de que o espírito científico pode sempre se renovar e florescer, mesmo em terrenos áridos.

AGRADECIMENTOS

A todos que, direta ou indiretamente, contribuíram para a concretização deste trabalho, manifesto minha mais profunda e sincera gratidão.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

"Meu cérebro é mais do que algo meramente mortal,
e o tempo mostrará isso."

Ada Lovelace

RESUMO

A previsão da demanda energética é essencial para a eficiência e estabilidade dos sistemas elétricos de potência. Tradicionalmente, as metodologias de previsão analisam as séries temporais de forma isolada, sem considerar o contexto global do sistema e as possíveis interações entre séries semelhantes. Este estudo examina a complexidade da demanda elétrica, sublinhando a importância das interdependências dinâmicas e das correlações presentes nas séries temporais do setor energético. Utilizando o *Temporal Fusion Transformer* (TFT), uma técnica avançada de aprendizado de máquina desenvolvida para lidar com dados periódicos complexos, a metodologia proposta investiga como a integração de diferentes sequências temporais de consumo influencia a precisão das previsões. Experimentos realizados com dados históricos de duas bases de dados distintas visaram obter projeções de demanda com um horizonte preditivo de 24 horas. Os resultados indicam que a abordagem múltipla resulta em uma melhoria significativa na precisão das previsões, evidenciando que, no contexto dos sistemas elétricos, a amplitude e a variedade das séries temporais utilizadas são fundamentais para alcançar previsões mais precisas e confiáveis.

Palavras-chave: inteligência artificial; aprendizado profundo; mecanismos de atenção; análise de séries temporais; previsão de demanda; *transformers*.

ABSTRACT

Energy demand forecasting is essential for the efficiency and stability of electrical power systems. Traditionally, forecasting methodologies analyze time series in isolation, without considering the overall context of the system and the possible interactions between similar series. This study examines the complexity of electricity demand, highlighting the importance of the dynamic interdependencies and correlations present in the energy sector's time series. Using the Temporal Fusion Transformer (TFT), an advanced machine learning technique developed to deal with complex periodic data, the proposed methodology investigates how the integration of different consumption time sequences influences forecast accuracy. Experiments carried out with historical data from two different databases aimed to obtain demand projections with a predictive horizon of 24 hours. The results indicate that the multiple approach results in a significant improvement in forecast accuracy, showing that, in the context of electricity systems, the breadth and variety of the time series used are fundamental to achieving more accurate and reliable forecasts.

Keywords: artificial intelligence; deep learning; attention mechanisms; time series analysis; demand forecasting; transformers.

LISTA DE FIGURAS

Figura 1 Exemplo de Série Temporal. 28
Figura 2 Exemplo de Série Temporal Decomposta: Série Original. 30
Figura 3 Exemplo de Série Temporal Decomposta: Componentes. 31
Figura 4 Classificação e Aplicações da Previsão de Carga. 39
Figura 5 Fatores Exógenos que Influenciam a Carga Elétrica. 41
Figura 6 Principais Tipos de Aprendizado de Máquina. 49
Figura 7 Tipos de Aprendizado Indutivo. 51
Figura 8 Neurônio Biológico. 55
Figura 9 Neurônio Artificial de McCulloch e Pitts. 55
Figura 10 Arquitetura de uma Rede Neural Artificial. 58
Figura 11 Conexão *Feedforward* e Conexão *Feedback*. 59
Figura 12 Exemplo de Descida do Gradiente. 60
Figura 13 Exemplo de Função de Ativação Linear. 62
Figura 14 Exemplo de Função de Ativação Limiar. 62
Figura 15 Exemplo de Função de Ativação Tangente Hiperbólica. 63
Figura 16 Exemplo de Função de Ativação Sigmoide. 64
Figura 17 Exemplo de Função de Ativação *Softmax*. 64
Figura 18 Exemplo de Função de Ativação ReLU. 65
Figura 19 Treinamento por *Backpropagation*. 67
Figura 20 Arquitetura de uma Rede Neural Profunda. 69
Figura 21 Mecanismo de Autoatenção. 73
Figura 22 Autoatenção Multifacetada. 74
Figura 23 Arquitetura do *Temporal Fusion Transformer*. 76
Figura 24 Rede de Seleção de Variáveis 77
Figura 25 *Gate Residual Network* (GRN). 80
Figura 26 Atenção Multicabeça Interpretável. 82
Figura 27 Sistema Previsor de Única Série Temporal (SPUST). 92
Figura 28 Sistema Previsor de Múltiplas Séries Temporais (SPMST). 92
Figura 29 Visualização da Série Temporal ST1A. 96
Figura 30 Visualização da Série Temporal ST2A. 96
Figura 31 Visualização da Série Temporal ST3A. 97
Figura 32 Visualização da Série Temporal ST4A. 97

Figura 33 Visualização da Série Temporal ST1B. 97
Figura 34 Visualização da Série Temporal ST2B. 98
Figura 35 Visualização da Série Temporal ST3B. 98
Figura 36 Visualização da Série Temporal ST4B. 98
Figura 37 Experimento 1 - Janela Retrospectiva e Previsão da ST1A. 108
Figura 38 Experimento 1 - Janela Retrospectiva e Previsão da ST2A. 108
Figura 39 Experimento 1 - Janela Retrospectiva e Previsão da ST3A. 109
Figura 40 Experimento 1: Curvas de Carga Real e Prevista da ST1A. 109
Figura 41 Experimento 1: Curvas de Carga Real e Prevista da ST2A. 110
Figura 42 Experimento 1: Curvas de Carga Real e Prevista da ST3A. 110
Figura 43 Gráfico de Dispersão do Experimento 1. 111
Figura 44 Experimento 2 - Janela Retrospectiva e Previsão da ST1B. 112
Figura 45 Experimento 2 - Janela Retrospectiva e Previsão da ST2B. 112
Figura 46 Experimento 2 - Janela Retrospectiva e Previsão da ST3B. 113
Figura 47 Experimento 2: Curvas de Carga Real e Prevista da ST1B. 113
Figura 48 Experimento 2: Curvas de Carga Real e Prevista da ST2B. 114
Figura 49 Experimento 2: Curvas de Carga Real e Prevista da ST3B 114
Figura 50 Gráfico de Dispersão do Experimento 2. 115
Figura 51 Experimento 3 - Janela Retrospectiva e Previsão da ST1A. 116
Figura 52 Experimento 3 - Janela Retrospectiva e Previsão da ST2A. 117
Figura 53 Experimento 3 - Janela Retrospectiva e Previsão da ST3A. 117
Figura 54 Experimento 3: Curvas de Carga Real e Prevista da ST1A. 118
Figura 55 Experimento 3: Curvas de Carga Real e Prevista da ST2A. 118
Figura 56 Experimento 3: Curvas de Carga Real e Prevista da ST3A. 119
Figura 57 Gráfico de Dispersão do Experimento 3. 119
Figura 58 Experimento 4 - Janela Retrospectiva e Previsão da ST1A. 120
Figura 59 Experimento 4 - Janela Retrospectiva e Previsão da ST2A. 121
Figura 60 Experimento 4 - Janela Retrospectiva e Previsão da ST3A. 121
Figura 61 Experimento 4 - Janela Retrospectiva e Previsão da ST4A. 122
Figura 62 Experimento 4: Curvas de Carga Real e Prevista da ST1A. 122
Figura 63 Experimento 4: Curvas de Carga Real e Prevista da ST2A. 123
Figura 64 Experimento 4: Curvas de Carga Real e Prevista da ST3A. 123
Figura 65 Experimento 4: Curvas de Carga Real e Prevista da ST4A. 124
Figura 66 Gráfico de Dispersão do Experimento 4. 125
Figura 67 Experimento 5 - Janela Retrospectiva e Previsão da ST1B. 126

Figura 68 Experimento 5 - Janela Retrospectiva e Previsão da ST2B. 126
Figura 69 Experimento 5 - Janela Retrospectiva e Previsão da ST3B. 127
Figura 70 Experimento 5: Curvas de Carga Real e Prevista da ST1B. 127
Figura 71 Experimento 5: Curvas de Carga Real e Prevista da ST2B. 128
Figura 72 Experimento 5: Curvas de Carga Real e Prevista da ST3B 128
Figura 73 Gráfico de Dispersão do Experimento 5. 129
Figura 74 Experimento 6 - Janela Retrospectiva e Previsão da ST1B. 130
Figura 75 Experimento 6 - Janela Retrospectiva e Previsão da ST2B. 130
Figura 76 Experimento 6 - Janela Retrospectiva e Previsão da ST3B. 131
Figura 77 Experimento 6 - Janela Retrospectiva e Previsão da ST4B. 131
Figura 78 Experimento 6: Curvas de Carga Real e Prevista da ST1B. 132
Figura 79 Experimento 6: Curvas de Carga Real e Prevista da ST2B. 132
Figura 80 Experimento 6: Curvas de Carga Real e Prevista da ST3B 133
Figura 81 Experimento 6: Curvas de Carga Real e Prevista da ST4B 133
Figura 82 Gráfico de Dispersão do Experimento 6. 134
Figura 83 Comparação do MAPE entre Experimentos com a Base de Dados A. . . 136
Figura 84 Percentuais de Melhoria Entre Experimentos com a Base de Dados A. . 137
Figura 85 Comparação do MAPE entre Experimentos com a Base de Dados B. . . 139
Figura 86 Percentuais de Melhoria Entre Experimentos com a Base de Dados B. . 141
Figura 87 Experimento 1: Importância dos Recursos - Codificador. 145
Figura 88 Experimento 1: Importância dos Recursos - Decodificador. 146
Figura 89 Experimento 3: Importância dos Recursos - Codificador. 146
Figura 90 Experimento 3: Importância dos Recursos - Decodificador. 147
Figura 91 Experimento 4: Importância dos Recursos - Codificador. 147
Figura 92 Experimento 4: Importância dos Recursos - Decodificador. 148
Figura 93 Experimento 2: Importância dos Recursos - Codificador. 149
Figura 94 Experimento 2: Importância dos Recursos - Decodificador. 149
Figura 95 Experimento 5: Importância dos Recursos - Codificador. 150
Figura 96 Experimento 5: Importância dos Recursos - Decodificador. 150
Figura 97 Experimento 6: Importância dos Recursos - Codificador. 151
Figura 98 Experimento 6: Importância dos Recursos - Decodificador. 151

LISTA DE TABELAS

Tabela 1 Modelos de RNAs Profundas para Previsão de Demanda Elétrica. . . . 19
Tabela 2 Modelos de Previsão de Carga com Mecanismos de Atenção. 23
Tabela 3 Análise Comparativa do Desempenho do TFT. 25
Tabela 4 Principais Características das Bases de Dados. 95
Tabela 5 Período da Análise Experimental, ... 95
Tabela 6 Nomenclatura das Séries Temporais. 96
Tabela 7 Distribuição dos Dados. 99
Tabela 8 Informações de Treinamento por Teste Experimental. 101
Tabela 9 Covariáveis Utilizadas na Base de Dados A. 102
Tabela 10 Covariáveis Utilizadas na Base de Dados B. 102
Tabela 11 Configuração dos Hiperparâmetros. 104
Tabela 12 Configurações Experimentais. 106
Tabela 13 Métricas de Desempenho do Experimento 1. 111
Tabela 14 Métricas de Desempenho do Experimento 2. 115
Tabela 15 Métricas de Desempenho do Experimento 3. 120
Tabela 16 Métricas de Desempenho do Experimento 4. 124
Tabela 17 Métricas de Desempenho do Experimento 5. 129
Tabela 18 Métricas de Desempenho do Experimento 6. 134
Tabela 19 Percentuais de Melhoria nos Experimentos da Base de Dados A. 136
Tabela 20 Percentuais de Melhoria nos Experimentos da Base de Dados B. 140
Tabela 21 Valores Observados da Base de Dados B. 171
Tabela 22 Valores Observados da Base de Dados A. 172
Tabela 23 Valores Previstos no Experimento 1. 173
Tabela 24 Valores Previstos no Experimento 2. 174
Tabela 25 Valores Previstos no Experimento 3. 175
Tabela 26 Valores Previstos no Experimento 4. 176
Tabela 27 Valores Previstos no Experimento 5. 177
Tabela 28 Valores Previstos no Experimento 6. 178

LISTA DE ABREVIAÇÕES E SIGLAS

ARIMA	*Autoregressive Integrated Moving Average*
CNN	*Convolutional Neural Network*
DL	*Deep Learning*
DNN	Redes Neurais Profundas
GLU	*Gated Linear Units*
GRN	*Gate Residual Network*
GRU	*Gated Recurrent Unit*
IA	Inteligência Artificial
LSTM	*Long Short-Term Memory*
LTLF	*Long-term Load Forecasting*
MAE	Erro Médio Absoluto
MAPE	Erro Absoluto Médio Percentual
ML	*machine learning*
MLP	*Multilayer Perceptron*
MTLF	*Medium-term Load Forecasting*
PE	Previsão Energética
PMST	Previsão de Múltiplas Séries Temporais
ReLU	*Rectified Linear Unit*
RFE	Eliminação Recursiva de Características
RNA	Rede Neurais Artificial
RNN	Rede Neurais Recorrente
RMSE	Raiz do Erro Quadrático Médio
SARIMA	Média Móvel Integrada Autorregressiva Sazonal
SEP	Sistemas Elétricos de Potência
Seq2Seq	Sequência a Sequência
sMAPE	Erro Percentual Médio Absoluto Simétrico
SPMST	Sistema Previsor de Múltiplas Séries Temporais
SPUST	Sistema Previsor de Única Série Temporal
ST	Série Temporal
STLF	*Short-term Load Forecasting*
TCN	*Temporal Convolutional Network*
TFT	*Temporal Fusion Transformer*
VMD	Decomposição Modal Variacional
XGBoost	*Extreme Gradient Boosting*

SUMÁRIO

1 INTRODUÇÃO 8
1.1 Justificativa da Pesquisa 9
1.2 Proposta do Trabalho 10
1.3 Contribuições 11
1.4 Estrutura do Documento 12

2 REVISÃO DE LITERATURA 14
2.1 Introdução 14
2.2 Redes Neurais Profundas 16
2.3 Mecanismo de Atenção 20
2.4 *Temporal Fusion Transformer* 24
2.5 Síntese do Capítulo 26

3 SÉRIES TEMPORAIS 27
3.1 Introdução 27
3.2 Elementos das Séries Temporais 29
3.3 Decomposição de Séries Temporais 32
3.4 Modelos de Previsão de Séries Temporais 32
3.4.1 Avaliação de Precisão 34
3.5 Propriedades dos Dados de Séries Temporais 35
3.6 Séries Temporais de Energia Elétrica 36
3.6.1 Previsão de Carga Elétrica 37
***3.6.1.1 Classificação da Previsão de Carga* 38**
***3.6.1.2 Fatores que Influenciam a Previsão de Carga* 40**
***3.6.1.3 Técnicas de Previsão de Carga* 42**
3.7 Síntese do Capítulo 44

4 REDES NEURAIS DE APRENDIZADO PROFUNDO 45
4.1 Introdução 46

4.2 Aprendizado de Máquina 47
4.2.1 Aprendizado Supervisionado 50
4.3 Redes Neurais Artificiais 53
4.3.1 O Neurônio Artificial 54
4.3.2 Arquitetura 57
4.3.3 Descida do Gradiente 60
4.3.4 Funções de Ativação 61
4.3.5 Treinamento 66
4.4 Aprendizado Profundo 68
4.5 Rede Neural Google *Transformer* 70
4.6 *Temporal Fusion Transformer* 75
4.6.1 Redes de Seleção de Variáveis 77
4.6.2 Mecanismos de Bloqueio 78
4.6.3 *Encoders* de Covariáveis Estáticas 80
4.6.4 Processamento Temporal 80
4.6.4.1 Atenção Multicabeça Interpretável 80
4.6.4.2 Camada de Sequência a Sequência 82
4.6.4.3 Camadas Adicionais 83
4.6.5 Intervalos de Previsão 84
4.7 Síntese do Capítulo 85

5 PREVISÃO DE DEMANDA ELÉTRICA COM TFT 86
5.1 Introdução 86
5.2 Descrição dos Dados 87
5.2.1 Base de Dados A 87
5.2.2 Base de Dados B 88
5.2.3 Pré-processamento dos Dados 89
5.3 Arquitetura Neural Utilizada 89
5.3.1 Configurações do Modelo 90
5.4 Desenvolvimento dos Modelos de Previsão Propostos 91
5.4.1 Modelagem Matemática da Previsão 93
5.5 Procedimentos e Definições do Arranjo Experimental 94

5.5.1 Divisão dos Dados 94
5.5.2 Variáveis Exógenas Utilizadas 101
5.5.3 Critérios de Avaliação 103
5.5.4 Definições de Treinamento do Modelo Neural 103
5.5.5 Ferramentas e Ambiente de Desenvolvimento 105
5.6 Síntese do Capítulo 105

6 RESULTADOS E DISCUSSÃO 106
6.1 Introdução 106
6.2 Análise dos Resultados 107
6.2.1 Resultados Experimentais do SPUST 107
6.2.1.1 Experimento 1: Base de Dados A 107
6.2.1.2 Experimento 2: Base de Dados B 111
6.2.2 Resultados Experimentais do SPMST 116
6.2.2.1 Experimento 3: Base de Dados A (3 STs) 116
6.2.2.2 Experimento 4: Base de Dados A (4 STs) 120
6.2.2.3 Experimento 5: Base de Dados B (3 STs) 125
6.2.2.4 Experimento 6: Base de Dados B (4 STs) 129
6.3 Discussão dos Resultados 135
6.3.1 Considerações sobre os Experimentos com a Base de Dados A 135
6.3.2 Considerações sobre os Experimentos com a Base de Dados B 138
6.3.3 Considerações Gerais 142
6.4 Análise de Interpretabilidade em Nível de Recursos 144
6.5 Síntese do Capítulo 153

7 CONCLUSÃO 154

REFERÊNCIAS 157

APÊNDICE A - VALORES OBSERVADOS E PREVISTOS 171

1 INTRODUÇÃO

A previsão da demanda elétrica constitui um componente essencial para a operação eficiente dos sistemas de potência. À medida que as redes de energia se tornam progressivamente mais complexas e integradas, a capacidade de prever com precisão o consumo futuro torna-se imprescindível para assegurar o equilíbrio entre oferta e demanda, minimizar custos operacionais e mitigar o impacto ambiental. No entanto, essa tarefa é complexa e desafiadora, uma vez que as séries temporais energéticas são influenciadas por uma multiplicidade de fatores, como condições climáticas, padrões de consumo, eventos sazonais e atividades econômicas.

Erros nas projeções podem acarretar uma série de consequências adversas, como sobrecarga na rede elétrica, falhas no fornecimento, desperdício de recursos e custos operacionais elevados. Por exemplo, subestimar a demanda pode ocasionar apagões e instabilidades, enquanto superestimar o consumo pode gerar custos excedentes com a geração e armazenamento de energia. Portanto, aprimorar a precisão das previsões é crucial para garantir a eficiência e a estabilidade do sistema elétrico.

A crescente disponibilidade de dados de alta frequência e alta resolução, oriundos de dispositivos de medição avançada, como os *smart meters*, proporciona oportunidades para o desenvolvimento de modelos preditivos mais robustos e precisos. Entretanto, a amplitude e o volume desses dados também impõem desafios significativos. Os métodos convencionais de previsão muitas vezes não conseguem lidar com a complexidade e a riqueza dos dados, resultando na incapacidade de capturar as nuances presentes nas informações.

Além disso, as abordagens tradicionais frequentemente analisam as séries temporais de forma isolada, ou seja, geram previsões para uma única sequência histórica em cada iteração, sem considerar o contexto global ou as possíveis confluências entre séries temporais correlacionadas. Essa abordagem restringe a capacidade do modelo de capturar padrões complexos e dependências dinâmicas entre as diferentes entidades que compõem o sistema, o que pode resultar em previsões imprecisas e, consequentemente, em desafios para a gestão eficiente da infraestrutura. Com o avanço das tecnologias de aprendizado de máquina e a crescente disponibilidade de dados históricos detalhados, surge a oportunidade de aprimorar as metodologias de previsão por meio do englobamento de interdependências e informações contextuais.

A utilização de técnicas avançadas de aprendizado de máquina, como o *Temporal Fusion Transformer* TFT, apresenta-se como uma solução promissora para aprimorar as previsões de demanda energética. O TFT é uma arquitetura de rede neural projetada para lidar com dados temporais e suas complexas inter-relações. Ao considerar simultaneamente diversos conjuntos de dados relacionados, o TFT é capaz de capturar os padrões temporais e as relações não lineares entre as variáveis, proporcionando uma análise abrangente das tendências de consumo.

Neste contexto, o presente estudo tem como objetivo investigar os padrões de associação

entre dados temporais do setor energético, utilizando a rede neural TFT. Assim, a metodologia proposta baseia-se na fusão de séries históricas correlacionadas, com o propósito de capturar as interdependências entre variáveis e aprimorar a capacidade do modelo de se adaptar a variações e tendências do sistema elétrico. Dessa forma, esta pesquisa visa explorar e avaliar, com um horizonte preditivo de 24 horas, a combinação de dados integrados de consumo e técnicas avançadas de aprendizado de máquina, com o intuito de refinar as previsões de demanda e contribuir para uma gestão mais eficiente dos sistemas elétricos de potência.

1.1 Justificativa da Pesquisa

Historicamente, a previsão de cargas elétricas tem se fundamentado predominantemente em modelos que tratam cada série temporal de forma isolada. Embora tais abordagens possam ser eficazes em contextos específicos, apresentam limitações substanciais em situações onde as séries temporais interagem e se influenciam mutuamente. Por exemplo, prever a demanda elétrica de uma região sem levar em consideração o consumo nas regiões vizinhas pode resultar em estimativas imprecisas. Eventos em uma região específica, como picos de consumo devido a condições climáticas extremas, feriados ou grandes eventos esportivos, podem impactar consideravelmente o consumo em múltiplas áreas simultaneamente. Modelos que desconsideram essas influências cruzadas podem falhar em capturar as complexas dinâmicas do sistema, resultando em subestimações ou superestimações significativas da demanda total, o que pode levar a erros na gestão da oferta e no planejamento do sistema elétrico.

Além disso, os padrões de consumo de energia apresentam variações significativas entre os consumidores residenciais, comerciais e industriais, e essas variações estão frequentemente correlacionadas. Por exemplo, um aumento na atividade industrial pode elevar a demanda por energia, impactando os preços e, consequentemente, influenciando o consumo nos setores residencial e comercial. Também, a variabilidade na geração de energia renovável, como solar e eólica, em uma determinada região pode exigir ajustes imediatos em outras regiões para garantir a estabilidade da rede elétrica. Modelos que não considerem essas interações entre diferentes tipos de consumidores e fontes de energia podem revelar-se inadequados para prever com precisão as necessidades reais do sistema em termos de redistribuição de carga e ajustes operacionais.

A previsão múltipla aborda essas interdependências, capturando não apenas a evolução temporal de cada série histórica, mas também o impacto que mudanças em uma série podem ter sobre as demais. Essa abordagem revela-se particularmente vantajosa no setor de energia elétrica, onde diversas séries temporais, como a demanda de energia, a temperatura, os níveis de atividade econômica e a produção de energia renovável, estão interligadas. Ao incorporar essas variáveis, modelos de múltiplas séries temporais oferecem uma visão mais abrangente, identificando padrões e tendências que podem não ser reconhecidos em abordagens convencionais.

Recentemente, a previsão de múltiplas séries temporais tem atraído considerável interesse no campo do aprendizado de máquina, devido à crescente necessidade de analisar dados complexos em diversas áreas, como finanças, saúde e meteorologia. O uso de redes neurais avançadas, como o modelo TFT, tem se mostrado particularmente eficaz para essa finalidade. Esse modelo permite capturar as complexas interdependências entre múltiplas séries temporais, facilitando uma análise detalhada e precisa da importância relativa de cada variável, além de proporcionar previsões mais robustas e interpretáveis.

Neste contexto, a pesquisa proposta justifica-se pela necessidade de aprimorar as previsões de cargas elétricas por meio de abordagens que considerem as interações complexas entre múltiplas séries temporais. Tal abordagem não apenas melhora a precisão das previsões, mas também contribui para uma gestão mais eficiente e sustentável da energia. Isso garante uma infraestrutura elétrica capaz de responder de maneira ágil e eficaz às demandas dinâmicas do mercado energético, promovendo uma alocação de recursos mais equilibrada e reduzindo desperdícios.

1.2 Proposta do Trabalho

A presente proposta visa a aplicação do modelo TFT para a previsão de múltiplas séries temporais (PMST) no setor de energia elétrica. O objetivo central é demonstrar que, dentro deste contexto específico, a abordagem PMST é mais vantajosa em relação às metodologias que consideram apenas uma única série temporal, devido à capacidade do PMST de capturar e explorar as interações entre diferentes séries temporais. Com esta abordagem, espera-se aprimorar a precisão dos resultados e contribuir para a criação de uma infraestrutura elétrica mais adaptável às variações dinâmicas do setor energético.

Para alcançar este objetivo, foram desenvolvidos dois sistemas de previsão baseados em redes neurais artificiais (RNAs) do tipo TFT. O primeiro sistema, denominado Sistema Previsor de Única Série Temporal (SPUST), concentra-se na previsão do comportamento futuro de uma única série temporal, utilizando informações exógenas e dados de cargas elétricas processados de forma isolada. Em contraste, o segundo sistema, denominado Sistema Previsor de Múltiplas Séries Temporais (SPMST), integra diversas séries temporais em um modelo unificado. Esta abordagem permite explorar as interações entre diferentes conjuntos de dados por meio de uma estratégia de fusão de informações, na qual cada sequência temporal é tratada como uma variável de entrada no modelo TFT.

A implementação e o treinamento desses modelos preditivos visam avaliar e comparar a eficácia das duas abordagens de previsão no contexto do setor elétrico. Para validar os resultados, foram realizados testes em diferentes níveis do sistema elétrico de potência, utilizando duas bases de dados distintas: uma referente ao nível de distribuição e outra ao nível de transmissão. Estas bases de dados incluem informações sobre o consumo de energia e variáveis exógenas

que refletem o comportamento das cargas, permitindo uma validação abrangente e robusta da aplicabilidade do método proposto.

1.3 Contribuições

As principais contribuições deste trabalho são:

1. Desenvolvimento e Implementação de Modelos *Transformers*: Esta pesquisa se dedica ao desenvolvimento e à implementação de dois modelos preditivos baseados na arquitetura *Transformers*: o SPUST e o SPMST. Uma comparação detalhada dessas abordagens é realizada, destacando suas respectivas vantagens e limitações. Esta contribuição é significativa para o avanço da previsão de cargas elétricas, uma vez que os *Transformers* oferecem novas perspectivas e melhorias substanciais na precisão preditiva;

2. Avanço na Previsão de Demanda Energética: Este trabalho apresenta e avalia a aplicação do modelo TFT para a previsão de múltiplas séries temporais no setor de energia elétrica. O TFT captura e explora interações complexas entre diversas séries temporais, oferecendo uma abordagem que potencialmente aprimora a precisão das previsões em comparação com métodos convencionais;

3. Avaliação Abrangente e Validada: A pesquisa conduz uma validação extensiva dos métodos propostos por meio de testes em diferentes níveis do sistema elétrico: distribuição e transmissão. Essa abordagem demonstra a aplicabilidade dos modelos de múltiplas séries temporais em contextos fundamentais da infraestrutura elétrica;

4. Exploração das Interdependências Temporais e Contextuais: A tese investiga a importância de capturar padrões temporais e interdependências complexas entre variáveis relacionadas ao consumo de energia, como condições climáticas, padrões de atividade econômica e produção de energia renovável. Essa análise proporciona percepções sobre os fatores que influenciam a demanda elétrica e como essas interações são essenciais para melhorar a precisão das previsões;

5. Análise de Interpretabilidade: A pesquisa realiza uma análise detalhada da interpretabilidade dos modelos, elucidando os mecanismos subjacentes às decisões preditivas. Essa análise é fundamental para a transparência e a compreensão das contribuições individuais de cada variável no processo preditivo, proporcionando uma melhor interpretação dos resultados e maior confiança nos modelos desenvolvidos;

6. Importância dos Recursos na Abordagem Múltipla: A tese apresenta uma análise comparativa da importância dos recursos (*features*) nos modelos de previsão de Única Série Temporal e Múltiplas Séries Temporais. Essa comparação revela como a consideração de

múltiplas séries temporais pode alterar significativamente a relevância das variáveis para o modelo, oferecendo *insights* sobre os fatores determinantes na demanda de energia;

7. Contribuição para o Avanço da Pesquisa em Aprendizado de Máquina Aplicado à Energia: Este trabalho avança o campo de aprendizado de máquina aplicado à previsão de demanda energética, demonstrando a aplicabilidade de técnicas avançadas, como o *Temporal Fusion Transformer*. A pesquisa abre novos caminhos para futuras investigações sobre a aplicação de modelos preditivos complexos em outros contextos relacionados à gestão de energia e áreas que envolvem séries temporais.

Em síntese, esta pesquisa apresenta avanços substanciais na metodologia de previsão de cargas elétricas, evidenciando a superioridade da abordagem de múltiplas séries temporais e a eficácia do modelo *Temporal Fusion Transformer* no setor energético. As contribuições desta pesquisa têm o potencial de influenciar práticas futuras e proporcionar melhorias significativas na gestão e operação dos sistemas elétricos.

1.4 Estrutura do Documento

Este documento está estruturado em sete capítulos. A seguir, apresenta-se uma breve descrição do conteúdo de cada capítulo, com o objetivo de orientar o leitor na compreensão do fluxo e da lógica da pesquisa.

- *Capítulo 1: Introdução* - Este capítulo proporciona uma visão geral do tema da pesquisa, destacando sua relevância e os objetivos principais. Apresenta a justificativa para a realização do estudo, a proposta específica e as principais contribuições esperadas. Dessa forma, estabelece a base necessária para contextualizar as questões que são abordadas ao longo do trabalho;

- *Capítulo 2: Revisão de Literatura* - Neste capítulo, realiza-se uma revisão detalhada da literatura existente sobre os tópicos pertinentes à pesquisa. São abordados os conceitos fundamentais e avanços recentes em redes neurais profundas, mecanismos de atenção e o modelo TFT. O objetivo é fornecer uma base teórica sólida, contextualizando o estudo no cenário acadêmico e tecnológico atual;

- *Capítulo 3: Séries Temporais* - O terceiro capítulo explora os elementos fundamentais das séries temporais, incluindo sua decomposição e modelos de previsão. São examinadas as propriedades dos dados de séries temporais no setor de energia elétrica, com uma ênfase na previsão da demanda de carga elétrica. Diversas técnicas de previsão de carga e os fatores que influenciam essas previsões são analisados para estabelecer o contexto da aplicação prática deste estudo;

- *Capítulo 4: Redes Neurais de Aprendizado Profundo* - Este capítulo aprofunda-se nas redes neurais de aprendizado profundo, abordando conceitos essenciais, como aprendizado de máquina supervisionado, estruturas de redes neurais artificiais e técnicas de treinamento. É dada atenção ao modelo *Google Transformer* e ao TFT, com ênfase em suas características e componentes específicos;

- *Capítulo 5: Previsão de Demanda Elétrica com TFT* - No quinto capítulo, descrevem-se os dados utilizados na pesquisa, incluindo as bases de dados empregadas e o pré-processamento realizado. A arquitetura neural aplicada aos modelos preditivos é detalhada, assim como o desenvolvimento dos modelos de previsão propostos, abrangendo a modelagem matemática e os procedimentos experimentais. Este capítulo também aborda as variáveis exógenas utilizadas e os critérios de avaliação adotados;

- *Capítulo 6: Análise dos Resultados* - O sexto capítulo apresenta e discute os resultados obtidos a partir dos experimentos realizados com os modelos desenvolvidos. São analisados os desempenhos dos sistemas preditivos, com uma comparação das duas abordagens analisadas na previsão de demanda elétrica.

- *Capítulo 7: Conclusão* - Neste capítulo, são apresentadas as conclusões da pesquisa, com destaque para as principais implicações dos resultados obtidos. Além disso, são sugeridas direções para pesquisas futuras.

Além dos capítulos mencionados, este trabalho inclui o Apêndice A, que apresenta uma compilação detalhada dos valores reais e previstos obtidos em cada experimento realizado durante a pesquisa, fornecendo uma visão abrangente dos resultados experimentais e facilitando a análise e a verificação dos dados.

2 REVISÃO DE LITERATURA

Este capítulo dedica-se ao estado da arte na área de redes neurais profundas, com foco na problemática abordada nesta pesquisa. Portanto, está estruturado de maneira a fornecer uma visão abrangente da literatura especializada no que se refere à utilização de técnicas de aprendizado profundo para a previsão de demandas elétricas. O capítulo está dividido em diferentes subseções. Na Subseção 2.1, são explorados os avanços das RNAs e dos modelos de aprendizado profundo, considerando as principais evoluções que essas abordagens impulsionaram. Em seguida, na Subseção 2.2, são apresentadas pesquisas sobre o uso de aprendizado profundo na previsão de demanda de energia elétrica, além de discutir as tendências nesse domínio.

A Subseção 2.3 explora a utilização do mecanismo de atenção no contexto deste estudo, abrangendo uma revisão dos métodos que obtiveram sucesso na melhoria da precisão das previsões e na otimização do uso de recursos energéticos. A Subseção 2.4 examina a rede neural TFT, trazendo os principais trabalhos que utilizam a referida arquitetura. Esta pesquisa se concentra na exploração do TFT, uma das técnicas mais avançadas para análise de séries temporais. Portanto, na referida subseção, são apresentadas as principais publicações correlatas à proposta apresentada neste trabalho e que serviram de base para o seu desenvolvimento. Finalmente, a Subseção 2.5 recapitula os pontos principais abordados no capítulo.

2.1 Introdução

A evolução das RNAs representa um dos avanços mais significativos no campo da inteligência artificial (IA) e da ciência da computação. Desde suas primeiras concepções teóricas até as implementações práticas mais recentes, as RNAs passaram por uma transformação expressiva, impactando uma ampla gama de áreas do conhecimento e setores da sociedade. Em particular, as RNAs de aprendizado profundo têm se destacado, oferecendo capacidades de aprendizado e generalização excepcionais em tarefas complexas, como reconhecimento de imagem, processamento de linguagem natural e até mesmo na condução autônoma. Esse avanço tem impulsionado a IA a novos patamares de desempenho e aplicabilidade, possibilitando soluções mais sofisticadas e eficientes em diversos domínios.

No campo da medicina, essas abordagens têm desempenhado um papel fundamental na análise de imagens médicas e no diagnóstico precoce de doenças. Estudos demonstram sua eficácia na detecção de câncer, como no caso do câncer de mama (Ragab *et al.*, 2019; Desai; Shah, 2021), e de tumores cerebrais (Khan *et al.*, 2024; Badža; Barjaktarović, 2020). Além disso, essas técnicas têm sido utilizadas no diagnóstico da COVID-19, apresentando resultados promissores na identificação rápida e precisa da doença (Marques; Agarwal; Díez, 2020; Khanday; Sofi, 2021). As RNAs têm contribuído significativamente para avanços no diagnóstico e no tratamento de diversas condições (Hireš *et al.*, 2022; Arnaout *et al.*, 2021; Miller; Blott;

Hames, 1992; Yang; Yu, 2021; Anand *et al.*, 2024), melhorando a precisão e a eficiência dos processos médicos.

No setor financeiro, as redes neurais têm revolucionado a previsão de tendências de mercado (Cheng *et al.*, 2022; Zulqarnain *et al.*, 2020; Oyewole *et al.*, 2024; Wang *et al.*, 2021), a análise de risco (Lai *et al.*, 2006; Feng; Chen; Jiang, 2021; Li; Wang; Yang, 2023), a detecção de fraudes (Motie; Raahemi, 2023; Chen; Lai, 2021; Innan *et al.*, 2024), e a personalização de serviços financeiros (Kaur; Kumar; Kaur, 2023; Yue, 2023; Zanke; Sontakke; Hassan, 2024). A aplicação de algoritmos de aprendizado profundo permite a análise de grandes quantidades de dados, aumentando a precisão das previsões e facilitando a tomada de decisões estratégicas e a identificação de oportunidades de investimento.

Na esfera tecnológica, as RNAs têm impulsionado o desenvolvimento de sistemas de reconhecimento de voz e imagem, ao mesmo tempo em que aprimoram a personalização de serviços e produtos. Tecnologias de reconhecimento de voz, como as implementadas por Lyashenko *et al.* (2021) e Dhanjal e Singh (2024), e sistemas de reconhecimento de imagem, como os desenvolvidos por Galić, Stojanović e Čajić (2024) e Tian (2020), têm transformado significativamente a interação humano-máquina, proporcionando experiências mais intuitivas e adaptativas para os usuários. Além disso, as redes neurais têm sido empregadas para melhorar a personalização de serviços e produtos, conforme evidenciado nos trabalhos de Vaskovsky, Chvanova e Rebezov (2020), Gao, Wei e Zhou (2020) e Li, Yang e Zeng (2021).

O setor automotivo também se beneficia significativamente dos sistemas neurais, com a criação de veículos autônomos (Chen *et al.*, 2021; Xiao *et al.*, 2022; Ravindran; Santora; Jamali, 2020) e sistemas avançados de assistência ao motorista, como evidenciado nos trabalhos de Ivanko *et al.* (2022) e Ayachi, Said e Abdelaali (2020). A capacidade das RNAs de processar informações em tempo real e tomar decisões rápidas e precisas tem sido crucial para o desenvolvimento de tecnologias de condução autônoma, melhorando a segurança e a eficiência dos transportes.

O setor de energia elétrica também tem sido impactado por essas abordagens, proporcionando avanços em diversas atividades, como a distribuição e o consumo de energia, a previsão de falhas em redes elétricas e a integração de fontes de energia renovável (Ghadami *et al.*, 2021; Huang; Wang, 2022; Kumar *et al.*, 2023; Liao *et al.*, 2021; Kang *et al.*, 2021; Luo *et al.*, 2022; Rácz *et al.*, 2018; Chandrasekaran *et al.*, 2021; Balakrishnan *et al.*, 2023; Kong *et al.*, 2021). Essas tecnologias promovem uma gestão mais eficiente e sustentável dos recursos energéticos, contribuindo para a redução de custos e a mitigação dos impactos ambientais.

A aplicação de redes neurais artificiais na previsão de cargas elétricas e em outros aspectos da previsão energética tem sido amplamente validada na literatura, evidenciando a precisão das RNAs em comparação com métodos clássicos. As redes neurais de aprendizado profundo, ou *deep learning*, constituem uma vertente avançada dessas técnicas, caracterizadas por múltiplas

camadas de unidades de processamento. Essa arquitetura oferece uma capacidade significativamente superior de aprender e representar informações complexas, tornando-se uma escolha robusta para projeções de consumo no âmbito energético. Portanto, no contexto da problemática abordada nesta pesquisa, os próximos segmentos deste capítulo serão dedicados à apresentação de uma seleção de estudos da literatura que utilizam técnicas de aprendizado profundo para a previsão de demandas elétricas.

2.2 Redes Neurais Profundas

Nos últimos anos, diferentes modelos de redes neurais artificiais profundas têm sido amplamente utilizados na previsão de séries temporais de energia elétrica. Essas abordagens destacam-se pela capacidade de capturar padrões complexos e dinâmicos presentes nos dados, superando muitas das limitações dos métodos tradicionais de previsão. Portanto, nesta subseção, são examinadas as arquiteturas profundas comumente empregadas na literatura, visando uma análise abrangente do estado atual da pesquisa neste campo e uma compreensão aprofundada das tendências e avanços mais recentes.

No trabalho apresentado por Yan *et al.* (2019), uma rede neural do tipo *Long Short-Term Memory* (LSTM) híbrida foi utilizada para realizar a previsão do consumo energético em residências. Nesta proposta, a transformada wavelet estacionária foi combinada com a referida rede, visando reduzir a volatilidade e ampliar as dimensões dos dados. Já na proposta apresentada por Bandara, Bergmeir e Hewamalage (2020), buscou-se investigar a superioridade das abordagens baseadas em aprendizado profundo, mais especificamente o LSTM, em comparação aos algoritmos tradicionais. Constatou-se que o LSTM supera técnicas clássicas como o modelo ARIMA (*Autoregressive Integrated Moving Average*). Além disso, os autores identificaram que o número de épocas no aprendizado profundo não afetou o desempenho desta rede, que exibiu um comportamento genuinamente aleatório.

Em Lago *et al.* (2018), foi proposto um modelo preditivo de energia solar a curto prazo utilizando uma rede neural *Perceptron* Multicamadas com mais de uma camada oculta (*Deep Multilayer Perceptron*). Devido à dispersão geográfica dos geradores de energia solar, a obtenção de medições diretas no solo para realizar previsões torna-se uma tarefa desafiadora. Para contornar essa limitação, a abordagem proposta utiliza dados geográficos obtidos por meio de satélites, que são inseridos na estrutura neural profunda. Esta, por sua vez, é treinada utilizando um pequeno subconjunto de dados terrestres. O estudo demonstrou que o modelo foi capaz de generalizar as informações inseridas para um número significativamente maior de locais sem dados geográficos disponíveis.

Na pesquisa conduzida por Fekri *et al.* (2021), foi concebida uma abordagem para prever cargas elétricas, utilizando dados provenientes de medidores inteligentes. Este modelo, baseado

em uma rede neural recorrente profunda do tipo LSTM, foi desenvolvido para operar em tempo real, adaptando-se de forma contínua a novos padrões de entrada. Diferentemente dos modelos convencionais que requerem treinamento *offline*, o modelo proposto realiza aprendizado contínuo por meio de um fluxo de dados fornecido por *smart meters*. Os resultados evidenciaram que essa abordagem alcançou previsões mais precisas em comparação com o modelo LSTM *offline* e superou outros cinco algoritmos *online* utilizados para a mesma finalidade.

O estudo de Chitalia *et al.* (2020), apresentou um modelo robusto baseado em redes neurais recorrentes profundas para prever a demanda energética em edifícios localizados em áreas distintas. Utilizando dados de cinco edifícios comerciais diferentes, o modelo foi capaz de fornecer soluções que resultaram em uma melhoria significativa no desempenho da previsão de carga. Mais especificamente, os resultados dos algoritmos de aprendizado profundo demonstraram uma melhoria de 20% a 45% em relação às previsões encontradas na literatura especializada.

Na investigação realizada por Bai *et al.* (2021), foram aplicados diferentes modelos de redes neurais recorrentes para prever o consumo diário de energia em empresas chinesas, variando em escala de consumo de eletricidade. Os resultados indicaram que os modelos LSTM e GRU (*Gated Recurrent Unit*) superam a metodologia padrão RNN (Redes Neurais Recorrentes), conforme avaliado pelas métricas de erro quadrático médio normalizado, erro percentual médio absoluto e estatística de limiar.

No trabalho conduzido por Wang *et al.* (2017), é apresentado um método híbrido para prever a produção de energia em sítios fotovoltaicos na Bélgica. Este método combina a transformada wavelet com uma rede neural convolucional profunda. A transformada wavelet é empregada para decompor as séries temporais em diversas frequências, enquanto a rede neural convolucional busca identificar padrões não lineares e estruturas invariantes em cada frequência dos dados de entrada. Os resultados demonstraram uma notável estabilidade e robustez, com um erro absoluto médio percentual (MAPE) de 3,85%.

No estudo de Koprinska, Wu e Wang (2018), a previsão da energia fotovoltaica foi realizada utilizando um modelo de redes neurais convolucionais para prever a produção 24 horas à frente. Utilizando dados de séries temporais da energia solar de três países, o desempenho da técnica proposta foi comparado com outras metodologias. Os resultados destacaram que a rede neural convolucional ofereceu previsões precisas e eficientes em comparação com abordagens tradicionais.

Em Marcos e Júnior (2021), os autores propuseram uma investigação com o objetivo de prever séries temporais de consumo energético na região nordeste do Brasil durante o período de 2004 a 2019. Para isso, foram empregados tanto modelos clássicos quanto técnicas de *machine learning*, incluindo os métodos autorregressivos ARIMA e SARIMA (Média Móvel Integrada Autorregressiva Sazonal), bem como modelos baseados em redes neurais, como *Multilayer Perceptron* (MLP), LSTM e *Convolutional Neural Network* (CNN). Os resultados revelaram

que a rede neural profunda CNN superou todas as outras técnicas de previsão examinadas, registrando um MAPE de 1,867%.

Na pesquisa realiza por Marcos e Júnior (2021), é apresentado um método que emprega uma Rede de Crença Profunda (DBN) para prever as cargas elétricas de subestações na África do Sul. Neste trabalho, foi investigado o impacto do fator de temperatura no erro da previsão, utilizando três medidas como base: o erro percentual médio absoluto simétrico (sMAPE), o erro médio absoluto (MAE) e o erro quadrático médio (RMSE). As soluções alcançaram um erro próximo de 4%, com os melhores resultados de MAPE, MAE e RMSE obtidos pela metodologia proposta. Os autores concluíram que o modelo neural de aprendizado profundo DBN oferece resultados promissores na previsão das cargas elétricas de subestações.

No trabalho de Laitsos *et al.* (2023), é introduzido um modelo híbrido avançado para a previsão de carga de curto prazo, empregando técnicas de aprendizado profundo. O núcleo do estudo é um modelo ConvLSTM do tipo *encoder-decoder*, que se destaca ao alcançar um MAPE de apenas 2,15%. Além disso, foram desenvolvidos e otimizados diversos modelos de aprendizado profundo, como LSTM, *Multilayer Perceptron* e CNN, visando aprimorar a previsão. Notavelmente, os modelos recorrentes, incluindo LSTM e os híbridos, demonstraram uma precisão superior, evidenciando sua eficácia na tarefa em questão.

O estudo realizado por Hu *et al.* (2023) apresentou um modelo que utiliza uma combinação de mineração de dados e aprendizado profundo para melhorar a precisão na previsão de cargas elétricas. Assim, foi proposta uma técnica denominada LVMD-DBFCM, que combina *clustering* com Decomposição Modal de Variação Local utilizando duplo *Clustering Fuzzy* e C-*means*. Além disso, os autores desenvolveram outro modelo híbrido, o CNN-IVIA-BLSTM, que integra uma CNN com o Algoritmo de Otimização da Imunidade ao Vírus Influenza (IVIA), juntamente com uma rede LSTM Bidirecional. Os resultados obtidos mostraram uma taxa de erro de apenas 2%, evidenciando a eficácia dos modelos profundos híbridos propostos.

A publicação de Yang, Sparrow e Wallom (2023) apresenta um modelo generalizado de aprendizado profundo baseado em GRU para prever a carga elétrica em regiões propensas a incêndios florestais. Trata-se de uma abordagem multifatorial, que considera a estrutura dos dados de entrada, os efeitos do calendário e as condições de temperatura. Os resultados indicaram que o modelo supera significativamente o LSTM, com uma redução de 30,73% no MAPE, alcançando uma precisão média de cerca de 3%, representando uma potencial economia de energia para a região em análise.

Na investigação realizada por Wang, Xue e Li (2023), é proposto um método de previsão de carga de curto prazo que combina Decomposição Modal Variacional (VMD) com um modelo profundo híbrido que combina CNN e Bi-LSTM (LSTM Bidirecionais). Os resultados indicaram que a abordagem proposta melhora a precisão da previsão ao acompanhar efetivamente as mudanças na carga elétrica, superando outros métodos existentes. A integração da VMD com

redes neurais convolucionais e redes LSTM Bidirecionais permite capturar tanto padrões locais quanto dependências temporais de longo prazo, resultando em previsões mais precisas.

O trabalho apresentado por Mohsenimanesh, Entchev e Bosnjak (2022) propôs um modelo híbrido para a previsão de carga de curto prazo de uma frota de veículos elétricos (EV). O modelo, denominado SD-CEEMDAN-BiLSTM, combina a seleção do dia semelhante (SD), a decomposição completa em modo empírico com adaptação de ruído (CEEMDAN) e redes neurais profundas (DNN), incluindo Unidades Recorrentes com *Gate* (GRUs), LSTM e BiLSTM. Os resultados evidenciaram que a abordagem proposta superou outros modelos com um MAPE de 2,63%, sendo uma abordagem híbrida viável e eficaz para prever a carga de EV.

Com base nas informações apresentadas nesta subseção, verifica-se que diversos métodos preditivos para sistemas de energia foram desenvolvidos utilizando técnicas variadas de aprendizagem profunda. A Tabela 1 resume os métodos destacados nesta área, conforme a revisão bibliográfica realizada.

Tabela 1 - Modelos de RNAs Profundas para Previsão de Demanda Elétrica.

Referência	**Método**
Yan *et al.* (2019)	LSTM com transformada wavelet
Bandara, Bergmeir e Hewamalage (2020)	LSTM
Lago *et al.* (2018)	Deep MLP
Fekri *et al.* (2021)	LSTM
Chitalia *et al.* (2020)	Deep RNN
Bai *et al.* (2021)	LSTM, GRU
Wang *et al.* (2017)	CNN com transformada wavelet
Koprinska, Wu e Wang (2018)	CNN
Marcos e Júnior (2021)	CNN
Laitsos *et al.* (2023)	ConvLSTM, LSTM, CNN-1D
Hu *et al.* (2023)	LVMD-DBFCM, CNN-IVIA-BLSTM
Yang, Sparrow e Wallom (2023)	GRU
Wang, Xue e Li (2023)	VMD com CNN e Bi-LSTM
Mohsenimanesh, Entchev e Bosnjak (2022)	SD-CEEMDAN-BiLSTM

Fonte: Autoria própria.

Observa-se o amplo uso dos modelos LSTM e CNN, que têm sido adotados com sucesso, proporcionando avanços significativos em termos de precisão preditiva. Além disso, algumas pesquisas, exemplificadas pelos trabalhos de Yan *et al.* (2019) e Wang *et al.* (2017), concentram-se na exploração e aprimoramento desses métodos já estabelecidos. Destaca-se também a ascen-

são de abordagens híbridas, que combinam diferentes arquiteturas de redes neurais e técnicas de pré-processamento de dados. Esses modelos, como os propostos por Laitsos *et al.* (2023), Hu *et al.* (2023), Wang, Xue e Li (2023) e Mohsenimanesh, Entchev e Bosnjak (2022), representam um avanço significativo na busca por previsões mais precisas e robustas.

2.3 Mecanismo de Atenção

Recentemente, os mecanismos de atenção emergiram como uma ferramenta promissora no campo da previsão, trazendo novas perspectivas para a análise de séries temporais. Dotados da capacidade de focar em partes específicas dos dados e de considerar as interações entre diferentes variáveis, esses mecanismos destacam-se como uma abordagem eficaz para capturar nuances e padrões sutis nos dados de cargas elétricas. Nesta análise, diversas arquiteturas e implementações de mecanismos de atenção serão exploradas, examinando seu papel no contexto do consumo energético.

O estudo realizado por Wu, Yang e Wang (2023) apresenta um modelo de previsão de carga elétrica que utiliza uma rede neural do tipo LSTM com um mecanismo de atenção dupla, combinado com uma seleção de características. Assim, o método proposto incorpora a eliminação recursiva de características (RFE) para remover fatores redundantes e utiliza mecanismos de atenção dupla para fundir e explorar conexões entre as características e a carga elétrica. Os resultados experimentais indicam uma melhoria significativa na precisão das previsões em comparação com modelos tradicionais.

No trabalho de Gao *et al.* (2023), é desenvolvido o Adaptive-TgDLF, um *framework* adaptativo que incorpora uma RNA *Transformer* juntamente com o conhecimento de domínio para a realização da previsão de carga. Esta abordagem leva em consideração as tendências e flutuações na demanda elétrica, visando uma previsão mais precisa. Os resultados demonstraram que os métodos de aprendizado adaptativo são capazes de capturar as dependências de longo prazo nas séries de energia, resultando em uma melhora significativa nas projeções. O Adaptive-TgDLF mostrou-se 16% mais preciso que modelos anteriores, mantendo sua robustez mesmo diante de 50% de ruído meteorológico, e ainda contribuiu para uma redução significativa no tempo de treinamento.

Na pesquisa apresentado por Xiong e Zhang (2023), uma estrutura unificada de aprendizado profundo baseada em atenção é proposta para realizar a previsão de cargas elétricas. O modelo é composto por três módulos principais: ponderação de características, um modelo STLF (*Seasonal and Trend Decomposition using Loess Forecasting*) com atenção temporal hierárquica e correção de erros por meio de aprendizado de transferência. Os resultados indicaram uma alta capacidade de generalização e precisão, destacando-se na previsão de cargas em redes elétricas. Além disso, o *design* modular da estrutura facilitou a personalização e redução dos custos de

treinamento, superando métodos existentes em dois conjuntos de dados públicos.

O estudo realizado por Zhou e Wu (2022) propõe um modelo de previsão de carga que combina uma rede neural LSTM com mecanismos de atenção. O objetivo foi aprimorar o desempenho em comparação com os algoritmos de previsão já existentes. Os resultados obtidos destacaram a eficácia da abordagem proposta em extrair características de dependência local, o que resulta em previsões mais precisas. Experimentos realizados com conjuntos de dados reais corroboraram que o modelo proposto supera de forma significativa os algoritmos tradicionais de previsão de carga.

Na investigação realizada por Yang *et al.* (2023), é apresentado o MFAMNet, um modelo neural que integra mecanismos de atenção em múltiplas escalas para aprimorar a previsão de carga de curto prazo. Assim, essa abordagem emprega módulos de atenção em duas vertentes para realçar informações relevantes e suprimir dados irrelevantes. Os experimentos realizados demonstraram que o MFAMNet supera os métodos existentes em termos de precisão e robustez, evidenciando sua eficácia em diferentes conjuntos de dados de teste.

O trabalho de Zhang *et al.* (2023) propõe uma abordagem para previsão de carga de curto prazo utilizando redes GRU com mecanismo de atenção. O modelo combina a capacidade do GRU de capturar dependências temporais de longo prazo com a eficácia do mecanismo de atenção em quantificar correlações temporais. Os resultados indicaram que a aplicação dessas técnicas resulta em uma melhoria significativa na precisão da previsão de cargas elétricas, demonstrando um desempenho superior em comparação com métodos tradicionais.

No estudo realizado por Wang (2023), é proposto um algoritmo de previsão de demanda elétrica que combina o WOA (*Whale Optimization Algorithm*) com BiLSTM (*Bidirectional Long Short-Term Memory*) e mecanismos de atenção. O WOA é empregado para otimizar a seleção de hiperparâmetros, enquanto o mecanismo de atenção atribui diferentes pesos aos estados implícitos do BiLSTM. Os resultados evidenciaram que a abordagem proposta resulta em melhorias na precisão, eficiência e confiabilidade da previsão de cargas.

A pesquisa conduzida por Liu *et al.* (2022) descreve um modelo de previsão de carga de curto prazo que combina um algoritmo de otimização aprimorado, o IWOA (*Improved Whale Optimization Algorithm*), com BiLSTM (*Bidirectional Long Short-Term Memory*) e mecanismos de atenção, formando assim o IWOA-*Attention*-BiLSTM. Os resultados indicaram que o modelo proposto supera métodos como LSTM e Seq2Seq, alcançando um MAPE (*Mean Absolute Percentage Error*) de 1,44%. Além disso, os autores constataram que o uso de fatores meteorológicos, tipos de data e hiperparâmetros otimizados contribuem para a alta precisão e desempenho do previsor de demanda energética.

Na publicação de Zhang *et al.* (2022), é proposto um modelo híbrido que combina uma CNN, um LSTM Bidirecional (BiLSTM) e mecanismos de atenção para realizar a previsão

de carga de curto prazo em sistemas de energia. Os resultados permitiram identificar que a abordagem proposta melhora a precisão das previsões, superando outros modelos comumente adotados, uma vez que foi capaz de reduzir a perda de memória.

No estudo de Wang, Jia e Ren (2021), é proposto um modelo de previsão de carga de curto prazo baseado em BiLSTM com mecanismo de atenção. Foi possível evidenciar que a abordagem melhora significativamente a precisão da previsão ao reforçar o impacto de características-chave. Os resultados de simulação demonstram que a abordagem proposta supera o BiLSTM tradicional em várias métricas de avaliação, destacando a importância do mecanismo de atenção na estrutura LSTM.

A investigação conduzida por Li, Sun e Liao (2022) apresenta um modelo de previsão de carga de curto prazo que combina uma rede TCN (*Temporal Convolutional Network*) com LSTM e mecanismos de atenção. O modelo foi concebido para capturar eficientemente dependências de longo prazo nos dados de carga, enquanto reduz as distrações relacionadas à incerteza. Estudos de caso realizados em conjuntos de dados públicos demonstram a viabilidade e os benefícios do modelo proposto, evidenciando sua capacidade de aprimorar significativamente a precisão da previsão de carga.

O estudo conduzido por Qin *et al.* (2022) propôs uma estrutura de previsão de carga de curto prazo que integra uma CNN, GRU e um mecanismo de atenção, formando um modelo denominado CNN-GRU-*Attention*. Essa proposta visa aprimorar o desempenho da previsão, capacitando-o a extrair características dinâmicas e atender às exigências temporais. Os resultados demonstraram que o modelo superou outros métodos de ponta em métricas comuns, destacando sua eficácia em um conjunto de dados reais de redes de distribuição.

No trabalho de Xiong *et al.* (2021), um modelo de previsão de carga é proposto utilizando uma rede neural do tipo *encoder-decoder* com mecanismo de atenção. Tal abordagem foi projetada para selecionar adaptativamente características relevantes e informações temporais, superando esquemas de previsão existentes na competição global de energia de 2014. Os resultados evidenciaram melhorias significativas na precisão da previsão de carga de curto prazo, destacando a capacidade do modelo de selecionar e adaptar informações temporais e características de maneira eficiente.

Observa-se que a literatura acadêmica dedicada à previsão de demandas elétricas abrange uma ampla gama de modelos avançados que integram mecanismos de atenção. Os estudos revisados demonstram progressos significativos na previsão de cargas elétricas, destacando a eficácia de técnicas baseadas em redes neurais, mecanismos de atenção e otimização de hiperparâmetros. Em particular, a inclusão de mecanismos de atenção mostra-se particularmente importante, permitindo uma seleção e ponderação mais precisas das características relevantes, o que melhora significativamente o desempenho dos modelos de previsão.

A Tabela 2 apresenta uma visão sintetizada dos diferentes métodos discutidos ao longo desta revisão acadêmica.

Tabela 2 - Modelos de Previsão de Carga com Mecanismos de Atenção.

Referência	**Método**	**Características**
Wu *et al.* (2023)	LSTM com Atenção	Integração de atenção dupla e RFE.
Gao *et al.* (2023)	*Adaptive*-TgDLF	Uso de *Transformer* e conhecimento de domínio com resiliência a ruído.
Xiong *et al.* (2023)	STLF com Atenção	Módulos de ponderação de características, STLF com atenção temporal e correção de erros.
Zhou e Wu (2022)	LSTM com Atenção	Extração de características de dependência local para previsões mais precisas.
Yang *et al.* (2023)	MFAMNet	Mecanismos de atenção em múltiplas escalas para previsão de curto prazo.
Zhang *et al.* (2023)	GRU com Atenção	Combinação de GRU e atenção para capturar dependências temporais de longo prazo.
Wang (2023)	Bi-LSTM e WOA com Atenção	Otimização de hiperparâmetros com WOA e uso de atenção para pesos dos estados.
Liu *et al.* (2022)	IWOA com Atenção e Bi-LSTM	Combinação de IWOA, BiLSTM e atenção com foco em fatores meteorológicos.
Zhang *et al.* (2022)	CNN com BiLSTM e Atenção	Combinação de CNN, Bi-LSTM e atenção para melhorar a precisão e reduzir perda de memória.
Wang *et al.* (2021)	BiLSTM com Atenção	Aprimoramento da precisão através do reforço de características-chave.
Li *et al.* (2022)	TCN e LSTM com Atenção	Captura de dependências de longo prazo e redução de incertezas.
Qin *et al.* (2022)	CNN-GRU-Attention	Integração de CNN, GRU e atenção para extração de características dinâmicas.
Xiong *et al.* (2021)	*Encoder-Decoder* com Atenção	Seleção adaptativa de características relevantes e informações temporais.

Fonte: Autoria própria.

2.4 *Temporal Fusion Transformer*

O TFT é um modelo avançado de aprendizado profundo desenvolvido pela Google em 2021 (Lim *et al.*, 2021), especificamente projetado para lidar com séries temporais. Em aplicações práticas, como a previsão de demandas elétricas, o *Temporal Fusion Transformer* tem se mostrado uma ferramenta poderosa, capaz de realizar projeções futuras com alta precisão. Até o momento, há um número limitado de trabalhos publicados na literatura especializada que abordam a previsão de séries temporais de energia elétrica utilizando o TFT. Portanto, nesta seção, são apresentadas as principais publicações relacionadas à proposta deste estudo, que serviram de base para o desenvolvimento desta pesquisa.

No estudo conduzido por Huy *et al.* (2022), foi analisada a efetividade do TFT em comparação com métodos estatísticos tradicionais e arquiteturas neurais, incluindo LSTM, Seq2Seq e *Prophet*. A pesquisa incorporou dados históricos de carga, temperatura, umidade e variáveis temporais como meses, dias da semana e feriados. Os resultados indicaram que o TFT supera modelos tradicionais e outras RNNs, apresentando erros de previsão significativamente menores (TFT: 3,38%, LSTM: 4,64%, Seq2Seq: 6,02%, Prophet: 13,14%, ARIMA: 13,42%).

No trabalho realizado por Liao e Radhakrishnan (2022), foi desenvolvida a previsão de carga de curto prazo no contexto da distribuição de energia, avaliando diferentes cenários de carga: leve, média e pesada. Para aferir o desempenho das abordagens TFT, *Baseline*, *Prophet* e *Neural-Prophet*, o método foi testado com dados anonimizados de um campus universitário, onde anomalias e dados faltantes foram imputados com o método *K-nearest Neighbor* (KNN). O TFT demonstrou maior precisão, apresentando um MAPE inferior a 5% em todos os cenários de carga testados. No cenário de carga média, o TFT mostrou uma melhoria de 9,21% no MAPE em relação ao modelo *baseline*.

De forma similar, o trabalho apresentado por Valencia e Sanchez-Galan (2022) abordou a previsão de carga em curto prazo utilizando o modelo estatístico SARIMA e o TFT. Ambas as técnicas foram comparadas, e os resultados indicaram que o modelo SARIMA é apropriado para previsões de curto prazo. No entanto, a precisão das previsões melhora significativamente com o uso do modelo TFT, que alcançou um MAPE de 0,044 e um RMSE de 69,2 MWh.

O estudo realizado por Jittanon, Mensin e Termritthikun (2023) teve como objetivo analisar o desempenho do TFT na antecipação da demanda de energia elétrica, comparando-o com dois métodos amplamente adotados em problemas de previsão de séries temporais, o N-BEATS e o N-HiTS. A análise do MAPE revelou que o TFT alcançou o menor MAPE (4,5980%), superando os modelos N-BEATS (5,0266%) e N-HiTS (7,9865%).

Na pesquisa desenvolvida por Nazir *et al.* (2023), o *Temporal Fusion Transformer* foi empregado para estimar as demandas diárias, semanais e mensais de clientes de redes inteligentes. O desempenho do TFT foi comparado com as metodologias neurais LSTM, LSTM interpretável

e TCN. Os resultados revelaram que a arquitetura TFT superou os demais modelos, registrando valores de o sMAPE significativamente menores (TFT: 26,46%, LSTM: 29,78%, LSTM interpretável: 31,10%, TCN: 36,42%).

O estudo de Santos *et al.* (2022) utilizou o TFT para prever a carga horária de energia solar fotovoltaica em seis instalações localizadas na Austrália e na Alemanha. A performance do TFT foi comparada a outros modelos, como ARIMA, LSTM, MLP e XGBoost. Os resultados mostraram que o TFT apresentou uma redução de 46%, 48% e 48% nas métricas RMSE, MAE e MASE na Austrália, e de 20%, 26% e 54% na Alemanha, em relação aos outros modelos.

A Tabela 3 oferece um resumo das informações mais relevantes dos estudos revisados.

Tabela 3 - Análise Comparativa do Desempenho do TFT.

Estudo	**Métodos**	**Métricas**	**Desempenho**
Huy *et al.* (2022)	LSTM Seq2Seq *Prophet* ARIMA	MAPE	TFT: 3,38% LSTM: 4,64% Seq2Seq: 6,02% *Prophet*: 13,14% ARIMA: 13,42%
Liao *et al.* (2022)	*Baseline* *Prophet* *Neural-Prophet*	MAPE	MAPE <5% em todos os cenários de carga (Melhoria de 9,21% em carga média).
Valencia *et al.* (2022)	SARIMA	MAPE RMSE	MAPE: 0,044 RMSE: 69,2 MWh
Jittanon *et al.* (2023)	N-BEATS N-HiTS	MAPE	TFT: 4,5980% N-BEATS: 5,0266% N-HiTS: 7,9865%
Nazir *et al.* (2023)	LSTM LSTM I[1] TCN	sMAPE	TFT: 26,46% LSTM: 29,78% LSTM I[1]: 31,10% TCN: 36,42%
Santos *et al.* (2022)	ARIMA LSTM MLP XGBoost	RMSE MAE MASE	Inferior em: RMSE: 46% e 20% MAE: 48% e 26% MASE: 48% e 54%

[1]LSTM Interpretável (LSTM I).

Fonte: Autoria própria.

Com base na revisão dos estudos apresentados, é possível constatar que o TFT oferece desempenho superior na previsão de carga em comparação com métodos tradicionais e outras arquiteturas neurais, como LSTM, *Multi-Layer Perceptron*, Seq2Seq, *Prophet*, *Extreme Gradient Boosting* e ARIMA, evidenciando erros consideravelmente menores. Além disso, o TFT demonstrou elevada acurácia em diversos cenários e contextos de carga.

Os estudos discutidos nas Seções 2.3 e 2.4 enfatizam a importância crescente dos mecanismos de atenção em modelos de previsão de cargas elétricas. Esses mecanismos permitem que o algoritmo foque em aspectos específicos dos dados, facilitando a identificação de padrões relevantes e, assim, aprimorando a precisão das previsões. Tal avanço é fundamental para o refinamento dos sistemas de gestão de energia elétrica, sendo essencial para a sustentabilidade e segurança do fornecimento de energia elétrica em uma perspectiva futura.

2.5 Síntese do Capítulo

Neste capítulo são apresentados os principais avanços e técnicas no campo das redes neurais profundas, com ênfase na previsão de demandas elétricas. Primeiramente, discute-se a evolução das RNAs e dos modelos de aprendizado profundo, destacando suas contribuições mais significativas. Em seguida, explora-se a aplicação dessas abordagens especificamente na previsão de demanda de energia elétrica, analisando tendências e pesquisas atuais. O uso do mecanismo de atenção é abordado, evidenciando as melhorias na precisão das previsões. Por fim, é examinada a rede neural TFT, com uma revisão dos principais trabalhos que utilizam essa arquitetura para análise de séries temporais de demanda elétrica, ressaltando as publicações que fundamentaram a proposta deste estudo.

3 SÉRIES TEMPORAIS

Este capítulo dedica-se à investigação dos conceitos fundamentais inerentes às séries temporais (ST), destacando os alicerces teóricos indispensáveis para a compreensão da problemática abordada nesta pesquisa. Portanto, está dividido em diferentes subseções. Inicialmente, a Subseção 3.1 oferece uma visão ampla das STs, estabelecendo os fundamentos para as análises subsequentes. Em seguida, a Subseção 3.2 detalha os principais componentes das sequências temporais, a fim de promover uma compreensão aprofundada de sua estrutura e comportamento.

A Subseção 3.3 aborda as técnicas aplicadas na decomposição das séries temporais em seus elementos constituintes, facilitando tanto a análise quanto a modelagem. Posteriormente, na Subseção 3.4, são explorados os modelos utilizados para prever valores futuros com base em dados históricos, bem como as métricas mais usadas para mensurar a acurácia das projeções.

A Subseção 3.5 examina as características intrínsecas dos dados temporais que influenciam a seleção e o desempenho dos modelos preditivos. Em seguida, a Subseção 3.6 concentra-se nas séries temporais associadas ao setor de energia elétrica. Dentro deste contexto, é abordada a previsão de carga elétrica, incluindo sua categorização com base em diferentes tipos de previsões. Além disso, é realizada uma análise dos diversos elementos que podem influenciar a precisão das previsões neste âmbito específico. Ainda na Subseção 3.6, são discutidas as diferentes metodologias aplicadas na previsão de carga elétrica. Por fim, o capítulo é concluído com uma síntese que recapitula os principais pontos abordados, fortalecendo a compreensão dos conceitos apresentados.

3.1 Introdução

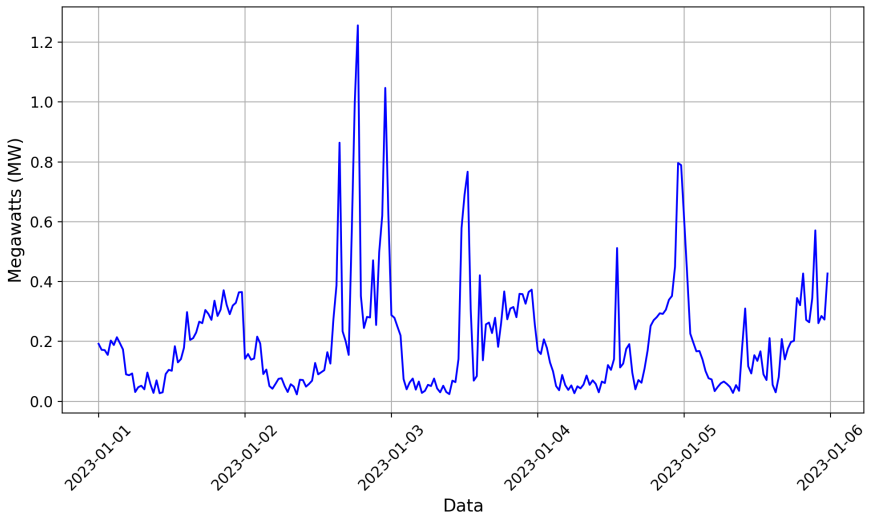
Uma série temporal constitui uma sequência de observações de uma variável ao longo do tempo, geralmente coletadas em intervalos regulares. Essa estrutura de dados proporciona uma visão dinâmica da evolução de fenômenos e processos ao longo de períodos cronológicos específicos. A análise e a antecipação de padrões temporais são atividades essenciais para embasar decisões estratégicas em uma ampla gama de áreas, como economia, finanças, meteorologia, engenharia e saúde. Dessa forma, a interpretação e a previsão de dados temporais não apenas fornecem percepções sobre o passado e o presente, mas também possibilitam uma preparação mais eficaz para eventos futuros e uma adaptação mais ágil a mudanças (Armstrong, 2001; Morettin; Toloi, 2022).

As séries são constituídas por dados obtidos pela medição e registro sequencial de uma variável ao longo do tempo, dentro de um período fixo denominado intervalo amostral. Dessa forma, podem ser definidas como um conjunto de observações ordenadas cronologicamente, dispostas em sequência temporal. Essas observações possuem características de natureza con-

tínua ou discreta, dependendo do método de amostragem e do intervalo de tempo entre os elementos (Morettin; Toloi, 2022).

Séries temporais contínuas referem-se a dados registrados de maneira ininterrupta ao longo do tempo, sem intervalos entre as medições. Nesse caso, os dados são capturados continuamente, proporcionando um registro regular e detalhado das variáveis observadas. Em contraste, séries temporais discretas são aquelas em que as observações são registradas em intervalos de tempo fixos e regulares. Para essa abordagem, um intervalo de tempo constante é previamente definido para a coleta de dados, assegurando que as observações sejam espaçadas uniformemente no tempo t (Morettin; Toloi, 2022). A Figura 1 ilustra um exemplo de série temporal.

Figura 1 - Exemplo de Série Temporal.



Fonte: Autoria própria.

Os dados de séries temporais observados no mundo real apresentam uma grande variabilidade, tanto em termos de estrutura quanto de conteúdo. Em uma análise abrangente, esses dados podem ser classificados em duas categorias principais: de processos naturais e de atividades humanas. No que se refere aos processos naturais, incluem-se variáveis climáticas como radiação solar, velocidade do vento e padrões sazonais. Em contraste, as atividades humanas abrangem dados como consumo de energia, produção industrial e comportamentos de mercado (Armstrong, 2001; Lim; Zohren, 2021).

As observações coletadas em intervalos amostrais anteriores, conhecidas como séries temporais históricas, são utilizadas para a análise e compreensão de eventos passados, bem como para a previsão de eventos futuros. Neste contexto, o estudo de séries temporais pode ser abordado de duas maneiras principais: análise e previsão. A análise de séries temporais tem como objetivo compreender a estrutura subjacente que gerou a série de dados. Isso envolve a identificação de componentes essenciais, como tendências, ciclos e sazonalidades. Para realizar essa análise, os dados históricos da variável em questão são examinados detalhadamente, permitindo

detectar padrões que presumivelmente continuarão a se manifestar no futuro (Armstrong, 2001; Chatfield, 2000).

A previsão de séries temporais, por outro lado, é uma área de estudo que se dedica à construção de modelos matemáticos capazes de projetar valores futuros da série com base nos dados históricos. Esses modelos podem variar desde métodos simples de extrapolação, como médias móveis e suavização exponencial, até técnicas mais avançadas, como modelos autorregressivos (AR). Além disso, com o avanço da inteligência artificial e do aprendizado de máquina, métodos mais sofisticados, como redes neurais artificiais e algoritmos de séries temporais baseados em aprendizado profundo, têm sido explorados para melhorar a precisão das previsões (Armstrong, 2001; Lim; Zohren, 2021).

Essa abordagem envolve a análise e modelagem dos dados históricos para compreender seus padrões e antecipar condições futuras, com o objetivo de mitigar a incerteza e fundamentar a tomada de decisão. Com os avanços na tecnologia computacional, tal prática tem sido amplamente adotada em diversos domínios, incluindo previsão de vendas, gestão de recursos hídricos, monitoramento da qualidade do ar e análise do mercado de ações, entre outras aplicações.

O objetivo central da previsão é extrair padrões e tendências dos dados históricos para realizar projeções sobre o comportamento futuro da série temporal. Essas antecipações são fundamentais em uma ampla gama de aplicações práticas, uma vez que a capacidade de antecipar com precisão eventos futuros torna-se uma atividade essencial na tomada de decisões e na formulação de estratégias eficazes sobre os eventos futuros. Neste contexto, o presente estudo se propõe a investigar a previsão de séries temporais relacionadas à demanda de energia elétrica, utilizando redes neurais artificiais profundas como método de previsão. Portanto, as próximas seções deste capítulo serão dedicadas à explanação dos conceitos fundamentais das séries temporais e da previsão da demanda de energia elétrica, em consonância com os objetivos específicos definidos para esta pesquisa.

3.2 Elementos das Séries Temporais

As séries temporais exibem uma variedade de padrões, conhecidos como componentes, e podem ser categorizados em três tipos principais (Morettin; Toloi, 2022; Dagum, 2010):

- Tendência: Tendência refere-se à direção geral e persistente de uma série temporal ao longo de um período prolongado. Esse padrão pode manifestar-se com um aumento ou diminuição sistemática da variável observada ao longo do tempo. Por exemplo, o crescimento gradual na demanda por energia elétrica ao longo de várias décadas devido ao aumento da urbanização e do desenvolvimento industrial, ou a redução na produção de energia solar durante os meses de inverno devido à menor incidência de luz solar;

- Sazonalidade: Caracteriza-se por padrões recorrentes e previsíveis que ocorrem em intervalos regulares de tempo, geralmente associados a fatores sazonais como estações do ano, feriados ou eventos cíclicos. Por exemplo, o crescimento da demanda por eletricidade durante os meses de verão devido ao uso generalizado de sistemas de ar condicionado ou o aumento na geração de energia hidrelétrica durante os períodos de chuva;

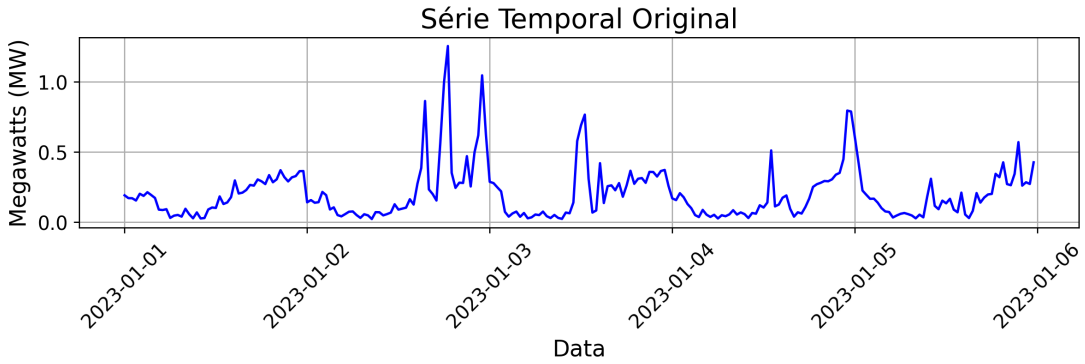
- Ciclos: Representa flutuações recorrentes, porém não necessariamente periódicas, na série temporal que não podem ser atribuídas a fatores sazonais. Estas variações podem ser influenciadas por fatores econômicos, políticos ou sociais mais amplos e podem ocorrer ao longo de vários anos. Por exemplo, as flutuações nos preços do petróleo que impactam os custos de produção da energia, ou as oscilações na produção de energia eólica devido a mudanças nos padrões climáticos globais.

Além desses componentes, as séries temporais também podem incluir um quarto elemento:

- Erro Aleatório: Refere-se às flutuações irregulares e imprevisíveis na série temporal que não podem ser explicadas pelos padrões identificados nas três categorias anteriores. Essas variações são atribuídas a causas não identificadas ou aleatórias e são consideradas parte natural da variabilidade inerente aos dados observados.

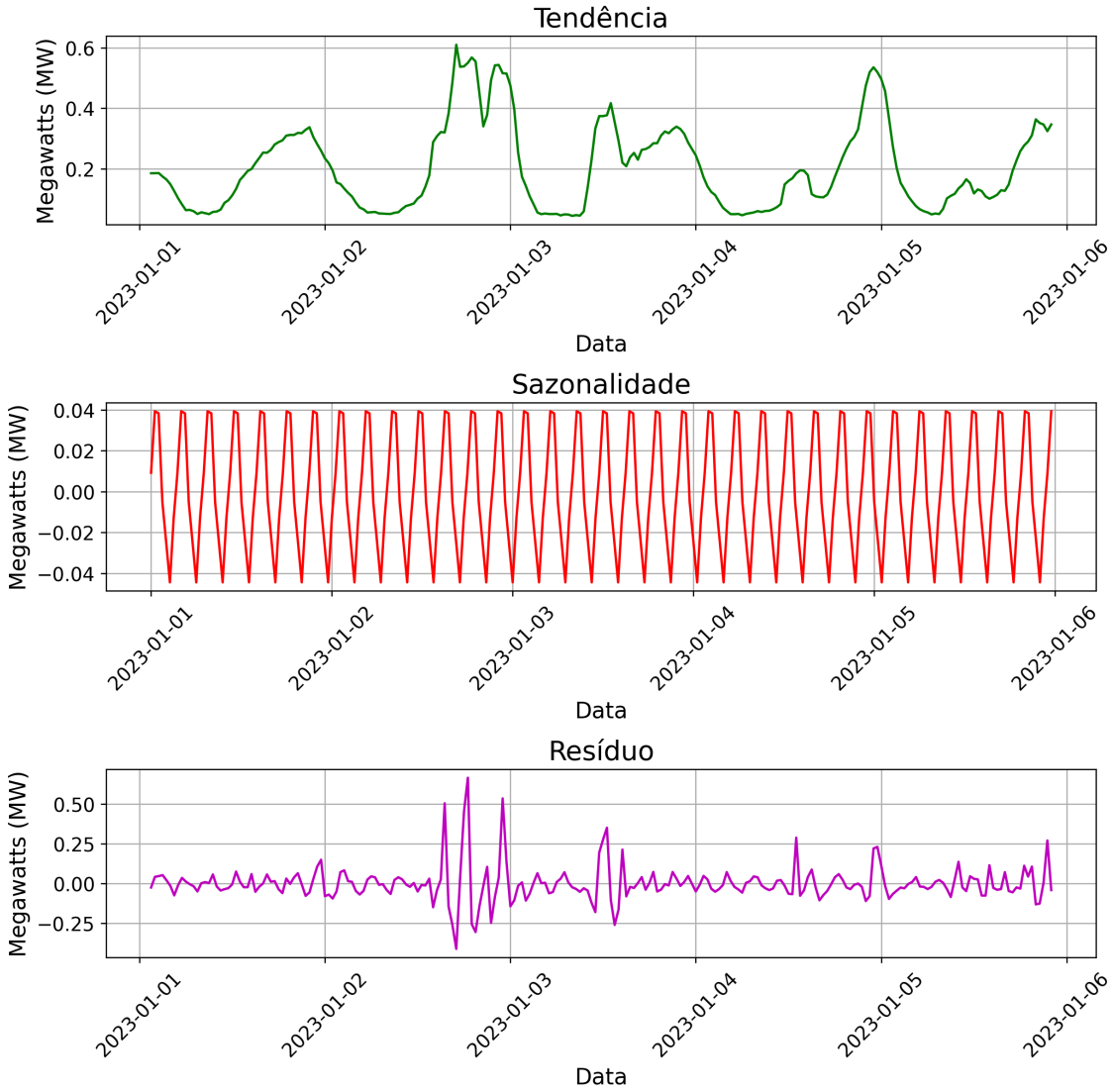
As Figuras 2 e 3 exibem um exemplo de série temporal decomposta em suas componentes de tendência, sazonalidade e resíduo.

Figura 2 - Exemplo de Série Temporal Decomposta: Série Original.



Fonte: Autoria própria.

Figura 3 - Exemplo de Série Temporal Decomposta: Componentes.



Fonte: Autoria própria.

A análise detalhada desses componentes proporciona uma visão mais clara das dinâmicas subjacentes aos dados. Assim, para compreender e prever os comportamentos das variáveis ao longo do tempo, é importante identificar e diferenciar os componentes de tendência, sazonalidade, ciclos e erro aleatório, permitindo desenvolver modelos mais precisos e confiáveis. Essa abordagem não só facilita a interpretação dos dados históricos, mas também aprimora a capacidade de projeção futura.

3.3 Decomposição de Séries Temporais

A decomposição de séries temporais é uma técnica que permite visualizar os diferentes componentes que contribuem para o comportamento observado ao longo do tempo em uma sequência cronológica. Esse interesse surge da necessidade de compreender e modelar as séries de maneira mais precisa. Assim, a decomposição divide a série temporal em seus distintos componentes, facilitando a análise individual de cada uma dessas partes e proporcionando uma compreensão mais profunda do fenômeno em estudo (Zarnowitz; Ozyildirim, 2006).

A decomposição clássica é uma abordagem amplamente utilizada para decompor séries temporais em suas componentes principais. Esta técnica pode ser aplicada de duas maneiras: aditiva ou multiplicativa. Na decomposição aditiva, a série é dividida em três componentes principais: tendência, sazonalidade e erro. Por outro lado, na decomposição multiplicativa, a série é expressa como o produto dessas três componentes (Morettin; Toloi, 2022; Dagum, 2010).

No caso da decomposição aditiva, determinada série temporal $\{x_1, x_2, \ldots, x_n\}$, denotada por x_t, é decomposta conforme expresso na equação (1):

$$x_t = S_t + T_t + E_t \qquad t = 1, 2, \ldots, n \tag{1}$$

Em que:

S_t: representa o componente de sazonalidade no intervalo t;

T_t: simboliza o componente de tendência no período t;

E_t: representa o componente de erro ou resíduo no tempo t.

Por outro lado, na decomposição multiplicativa, a série temporal é expressa como o produto desses três componentes, como representado na equação (2):

$$x_t = S_t \times T_t \times E_t \qquad t = 1, 2, \ldots, n \tag{2}$$

A abordagem multiplicativa é preferível em contextos nos quais a amplitude da sazonalidade varia temporalmente.

3.4 Modelos de Previsão de Séries Temporais

Os modelos de previsão de séries temporais têm como objetivo principal estimar valores futuros de uma variável alvo $y_{i,t}$ para uma determinada entidade i em um instante temporal t. Cada entidade i representa um agrupamento lógico de dados temporais, que pode variar

conforme o contexto da aplicação. Conforme delineado por Lim e Zohren (2021), no caso mais simples, os modelos de previsão de um passo à frente podem ser representados pela equação (3):

$$\hat{y}_{i,t+1} = f(y_{i,t-k:t}, x_{i,t-k:t}, s_i) \tag{3}$$

Em que:

$\hat{y}_{i,t+1}$: corresponde a previsão obtida pelo modelo;

$y_{i,t-k:t} = \{y_{i,t-k}, \ldots, y_{i,t}\}$: representam as observações passadas da variável alvo;

$x_{i,t-k:t} = \{x_{i,t-k}, \ldots, x_{i,t}\}$: denota as observações passadas das variáveis exógenas;

k: refere-se ao tamanho da janela retrospectiva;

s_i: simboliza os metadados estáticos associados à entidade i;

$f(\cdot)$: corresponde a função de previsão aprendida pelo modelo.

As redes neurais profundas são eficazes na captura e análise de relações preditivas complexas, devido à sua estrutura composta por várias camadas não lineares. Tais camadas são capazes de processar e transformar as características observadas, gerando representações intermediárias das mesmas. Em contextos que envolvem séries temporais, essa capacidade pode ser interpretada como a habilidade da rede em codificar informações históricas em uma variável latente, denotada por z_t, a qual encapsula os padrões e tendências presentes nos dados ao longo do tempo. Essa variável latente, z_t, é então utilizada como entrada única para a geração da previsão final. Matematicamente, esse processo ocorre em conformidade com as equações (4) e (5):

$$f(y_{t-k:t}, x_{t-k:t}, s) = g_{\text{dec}}(z_t), \tag{4}$$

$$z_t = g_{\text{enc}}(y_{t-k:t}, x_{t-k:t}, s). \tag{5}$$

Em que:

$f(\cdot)$: representa a função de previsão;

$y_{t-k:t}$: corresponde ao histórico das variáveis alvo;

$x_{t-k:t}$: denota o histórico das variáveis de entrada;

s: simboliza outras informações relevantes;

$g_{\text{enc}}(\cdot)$: corresponde a função codificadora;

$g_{\text{dec}}(\cdot)$: representa a função decodificadora;

z_t: simboliza a variável latente.

No contexto das equações (4) e (5), a notação foi simplificada removendo o subscrito *i*. Por exemplo, $y_{i,t}$ é substituído por y_t, conforme Lim e Zohren (2021).

Os codificadores e decodificadores são componentes básicos nas estruturas das redes neurais profundas, constituindo os elementos responsáveis pela representação e manipulação dos dados. Ao codificar, esses componentes convertem os dados de entrada em representações latentes de alta dimensionalidade, enquanto, na fase de decodificação, reconstroem ou geram as saídas desejadas a partir dessas representações. Portanto, a escolha da arquitetura adequada da rede neural é um fator determinante para o modelo (Lim; Zohren, 2021).

3.4.1 Avaliação de Precisão

A etapa de avaliação de desempenho dos modelos preditivos permite não apenas aferir a capacidade do sistema em gerar projeções precisas, mas também compreender a qualidade das previsões, identificando a relevância das mesmas para o problema em questão. Nesta subseção, são apresentadas as principais métricas de desempenho, embasadas na obra de Hyndman *et al.* (2006), e em conformidade com o escopo desta pesquisa. Tais métricas proporcionam uma visão abrangente de quão bem os modelos capturam os padrões subjacentes nos dados temporais, proporcionando, assim, uma base sólida para a avaliação e o aprimoramento contínuo dos modelos de previsão.

Supondo que y_i é uma observação real de uma série temporal no período *i*, e $\hat{y}_i$ a previsão correspondente de y_i. O erro de previsão no ponto *i* é definido como expresso na equação (6):

$$e_i = y_i - \hat{y}_i \tag{6}$$

Neste contexto, para avaliar o desempenho geral de um modelo preditivo de séries temporais, são utilizadas diferentes métricas de erro, como o erro médio absoluto (MAE), o raiz do erro quadrático médio (RMSE), o MAPE, entre outras. Cada um desses indicadores oferece uma perspectiva distinta sobre a precisão e a adequação do modelo em relação aos dados observados. O MAE, por exemplo, mede a média das magnitudes dos erros sem considerar sua direção, fornecendo uma visão clara dos erros médios absolutos. O RMSE, por sua vez, apresenta os erros na mesma unidade dos dados originais, facilitando a interpretação. Já o MAPE expressa o erro médio absoluto como uma porcentagem dos valores reais, sendo particularmente útil para comparações entre diferentes séries temporais ou modelos.

Portanto, pode-se dividir os critérios de precisão em dois grupos principais: métricas base-

adas em escala e métricas baseadas em percentuais. No primeiro grupo, o erro de previsão (e_i) possui a mesma unidade da variável resposta (y_i). As duas principais métricas de erro baseadas em escala são o MAE e a RMSE, obtidas conforme as equações (7) e (8):

$$MAE = \text{média}(|e_i|) \tag{7}$$

$$RMSE = \sqrt{\text{média}(e_i^2)} \tag{8}$$

Ao conduzir uma análise que visa avaliar a qualidade do ajuste e a precisão de diferentes modelos aplicados a uma mesma série temporal, o uso de métricas baseadas em escala é adequado e não apresenta problemas significativos. No entanto, ao expandir a análise para incluir a performance de um ou mais modelos em diversas séries temporais com escalas distintas, torna-se necessário adotar métricas baseadas em percentuais. Isso ocorre porque métricas percentuais, como o MAPE, permitem uma comparação direta entre séries que possuem diferentes magnitudes. Em contraste, métricas absolutas, como o MAE, podem ser distorcidas pelas diferenças nas escalas das séries temporais, dificultando uma avaliação comparativa precisa da performance dos modelos. Portanto, para garantir a consistência e a relevância dos resultados em contextos de múltiplas séries temporais, é fundamental a utilização de métricas que normalizem os erros em relação aos valores das próprias séries. Neste contexto, sendo p_i definido como expresso na equação (9), a métrica MAPE pode ser obtida conforme a equação (10):

$$p_i = \frac{e_i}{y_i} \times 100 \tag{9}$$

$$MAPE = \text{média}(|p_i|). \tag{10}$$

3.5 Propriedades dos Dados de Séries Temporais

Os dados de séries temporais são pontos amostrais extraídos de um processo contínuo ao longo do tempo. Conforme discutido por Längkvist, Karlsson e Loutfi (2014), essas informações possuem características únicas que os diferenciam de outros tipos de dados:

- Ruído e Alta Dimensionalidade: Dados de séries temporais frequentemente apresentam alta dimensionalidade e são suscetíveis a ruído. Abordagens como redução de dimensionalidade e filtragem podem ser empregadas para mitigar o impacto do ruído e simplificar a complexidade. Contudo, tais técnicas podem resultar na perda de informações essenciais e requerem conhecimento especializado para uma aplicação adequada;

- Incerteza nas Informações: A incerteza é uma constante na análise de dados temporais, muitas vezes decorrente da falta de informações abrangentes para uma compreensão completa do processo subjacente;
- Dependência Temporal: A dependência temporal é uma característica intrínseca das séries temporais, uma vez que seus valores são explicitamente vinculados ao tempo. Portanto, ao realizar previsões com base em dados passados, é fundamental considerar as dependências temporais. Isso implica que os algoritmos devem incorporar uma quantidade significativa de dados históricos ou ter a capacidade de reter informações sobre as entradas anteriores, o que pode aumentar consideravelmente a complexidade do modelo;
- Não Estacionariedade: A não estacionariedade é uma característica comum em séries temporais, onde propriedades como média, variância e frequência podem oscilar ao longo do tempo. Em alguns casos, trabalhar no domínio da frequência pode ser mais vantajoso do que no domínio do tempo, permitindo uma análise mais eficaz das flutuações e padrões temporais dos dados;
- Invariância: Em séries temporais, as características precisam permanecer consistentes independentemente de translações no tempo. Isso contrasta com domínios, como o da visão computacional, onde a invariância a translações, rotações e escalas é fundamental.

Em síntese, a análise e previsão de séries temporais enfrentam desafios complexos devido à variedade de características singulares que definem os dados que compõem essas sequências amostrais. Portanto, é fundamental adotar estratégias eficazes e possuir um entendimento profundo das técnicas de modelagem, visando assegurar resultados precisos e relevantes.

3.6 Séries Temporais de Energia Elétrica

As séries temporais de energia são conjuntos de dados que delineiam a progressão temporal das características energéticas reais, abrangendo uma ampla gama de fatores exógenos. Estes incluem, mas não se limitam a, mudanças sazonais, oscilações climáticas, dinâmicas econômicas e avanços tecnológicos. Tais conjuntos de dados desempenham um papel fundamental na análise e previsão da demanda e oferta de energia, fornecendo *insights* cruciais para a gestão eficiente e estratégica dos recursos energéticos (Motlagh; Berry; O'Neil, 2019; Li *et al.*, 2020).

Neste contexto, as séries temporais fornecem uma representação detalhada da dinâmica da infraestrutura elétrica ao longo de uma sequência cronológica, englobando uma variedade de componentes do sistema, dos quais a carga elétrica se destaca como um indicador fundamental do consumo de energia em diferentes momentos. Estes dados são coletados em intervalos temporais específicos e incorporam uma gama diversificada de influências externas que moldam os registros históricos de demanda (Li *et al.*, 2020).

Nas últimas décadas, houve uma evolução significativa na tecnologia de coleta de dados, passando dos métodos tradicionais de pesquisa e medição de campo para a era atual do *big data*. Inicialmente, as informações sobre o consumo da eletricidade eram obtidas principalmente por meio de pesquisas pontuais e medições periódicas realizadas *in loco*. Contudo, com os avanços tecnológicos, especialmente com a introdução dos medidores inteligentes, a capacidade de coleta de dados aumentou exponencialmente. Esses dispositivos permitem a captura contínua e detalhada da demanda de energia dos consumidores, fornecendo séries temporais de alta resolução que registram o consumo em intervalos frequentes, como por hora ou até mesmo por minuto (Motlagh; Berry; O'Neil, 2019).

Além disso, as *smart grids*, ou redes inteligentes, que representam um avanço significativo na entrega de energia elétrica, geram uma grande quantidade de dados provenientes de infraestruturas avançadas de monitoramento, com destaque para as séries temporais de demanda. Esse alto volume de dados é essencial para análises detalhadas e precisas, particularmente na previsão de carga elétrica, uma tarefa fundamental para otimizar o desempenho da rede elétrica em diferentes níveis, desde residências individuais até a gestão global da demanda e armazenamento de energia (Gasparin; Lukovic; Alippi, 2022).

Neste contexto, considerando a relevância da previsão da demanda de energia para os sistemas elétricos de potência, esta pesquisa busca contribuir para o progresso do entendimento e das metodologias relacionadas à previsão energética. Dessa forma, os próximos tópicos desta subseção explorarão este tema de forma mais abrangente, detalhando aspectos fundamentais no âmbito do escopo desta pesquisa.

3.6.1 Previsão de Carga Elétrica

A previsão é uma atividade essencial no processo decisório da indústria, sobretudo no contexto do setor de energia elétrica. Dentro desse segmento, o planejamento e a operação do sistema elétrico dependem da capacidade de ajustar em tempo real a geração de energia para atender à demanda, garantindo, simultaneamente, a segurança do sistema. Portanto, a previsão é uma atividade essencial em diversos aspectos dos sistemas elétricos de potência (SEP), como a expansão das redes de transmissão, a realocação de carga, a programação de manutenção, entre outras aplicações (Hong *et al.*, 2020; Kamaev *et al.*, 2012).

A importância da previsão energética também se estende à formulação de diretrizes regulatórias referentes à produção e ao consumo da energia elétrica, frequentemente delineadas por políticas estatais. Portanto, a previsão de variáveis relacionadas à eletricidade emerge como um campo de estudo dentro da teoria econômica. Isso ocorre devido às características regionais específicas que moldam o mercado de energia de cada país, tais como clima, geografia e fatores sociodemográficos (Mason, 1955).

Devido à natureza estratégica, as matrizes energéticas têm sido objeto de investimentos consideráveis ao longo do tempo. Como resultado, a previsão de variáveis associadas à energia elétrica progrediu para um domínio de estudo específico denominado previsão energética (PE). A PE não apenas se apresenta como uma ferramenta fundamental, mas também como um campo diversificado e complexo (Hong, 2014).

No contexto da previsão energética, a previsão de carga, em particular, é de grande importância para o setor industrial, desempenhando um papel essencial em diversas atividades como o planejamento e a operação de sistemas energéticos, iniciativas de resposta à demanda, transações comerciais de energia e formulação de políticas energéticas. Uma ampla gama de entidades corporativas, abrangendo concessionárias de energia, órgãos reguladores, grandes empresas industriais, complexos comerciais e instâncias governamentais, fazem uso da projeção da demanda elétrica (Hong *et al.*, 2016).

3.6.1.1 Classificação da Previsão de Carga

A operação segura e eficiente dos SEP depende da capacidade de prever com precisão a demanda de energia em diferentes horizontes temporais. As atividades em um SEP são executadas em diversos períodos de tempo, com a tomada de decisão podendo variar desde respostas imediatas a mudanças na demanda até o planejamento de longo prazo para a expansão e modernização da infraestrutura. A tomada de decisão referente às operações diárias pode ocorrer de forma imediata ou dentro de algumas horas, enquanto o planejamento de expansão, que envolve estudos de retorno de capital e análise de receitas, pode se estender por meses ou até anos (Parlos *et al.*, 1996). Nesse contexto, a categorização da previsão de carga conforme o intervalo de tempo considerado desempenha um papel fundamental na gestão dos SEP, abrangendo períodos de minutos, horas, dias, semanas, entre outros.

A análise das diversas abordagens propostas na literatura revela uma ampla gama de escalas temporais adotadas para a previsão de carga elétrica. Entretanto, a classificação mais comumente adotada está relacionada ao tamanho do horizonte de previsão (Mishra *et al.*, 2020). Alguns autores, como Al-Kandari, Soliman e El-Hawary (2004), Murto (1998), e Hong e Wang (2014), delinearam uma classificação abrangente, destacando diferentes categorias. De acordo com Hong *et al.* (2016), a previsão de demanda elétrica é dividida em quatro grupos distintos, também conhecidos como horizontes de previsão:

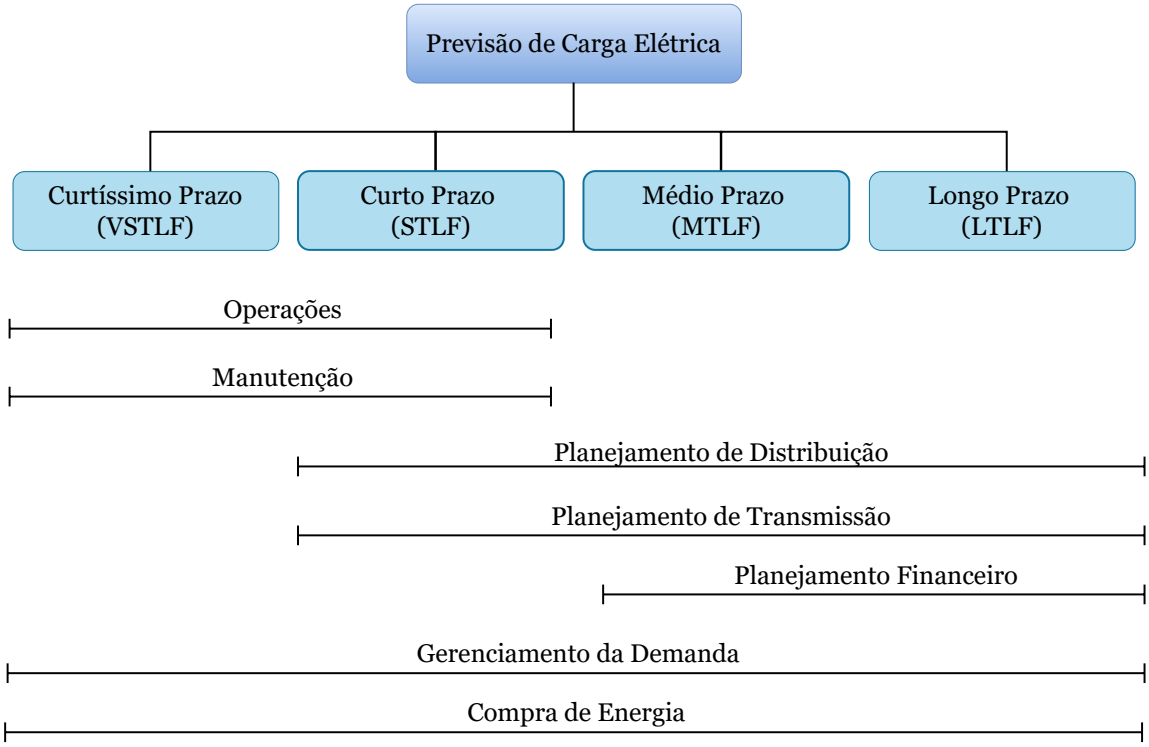
- *Very Short-term Load Forecasting* (VSTLF): A previsão de carga a curtíssimo prazo abrange intervalos de tempo extremamente curtos, que variam de alguns segundos até um dia à frente;
- *Short-term Load Forecasting* (STLF): Na previsão de carga a curto prazo, as antecipações ocorrem em intervalos que variam de um dia até duas semanas no futuro;

- *Medium-term Load Forecasting* (MTLF): A previsão de carga a médio prazo estende-se de duas semanas até três anos;
- *Long-term Load Forecasting* (LTLF): Já a previsão de carga a longo prazo trata-se de uma categoria que abrange horizontes temporais mais amplos, geralmente superiores a três anos, levando em consideração as tendências de consumo de energia a longo prazo e as necessidades de desenvolvimento futuro.

Os autores Hong *et al.* (2016) ainda propuseram uma categorização simplificada, dividindo-a em duas principais categorias temporais: curto prazo, que abrange desde segundos até 15 dias no futuro, e longo prazo, que compreende o intervalo de 15 dias até décadas subsequentes.

As previsões realizadas em cada um desses horizontes temporais atendem a diferentes propósitos, refletindo a complexidade e a diversidade das demandas enfrentadas pelos SEP. Assim, a escolha da escala temporal apropriada para a previsão de carga é crucial para assegurar o funcionamento eficiente e confiável dos SEP, além de orientar decisões estratégicas de planejamento e investimento. A Figura 4 ilustra a diversidade de propósitos e aplicações associadas a cada categoria de previsão de carga.

Figura 4 - Classificação e Aplicações da Previsão de Carga.



Fonte: Adaptação (Zor; Timur; Teke, 2017).

Neste contexto, esta pesquisa concentra-se especificamente na STLF. A escolha deste ho-

rizonte temporal é motivada pela importância crítica que a STLF possui para o controle operacional e a programação dos SEP. Conforme mencionado por Zheng, Yuan e Chen (2017), essa atividade é essencial para assegurar a estabilidade e a eficiência dos SEP, permitindo a adaptação imediata das operações às variações na demanda de energia.

3.6.1.2 Fatores que Influenciam a Previsão de Carga

A acurácia das previsões é um fator crítico na otimização operacional e na gestão financeira dos sistemas de energia elétrica. A imprecisão nesse domínio pode resultar em aumentos substanciais nos custos operacionais, acarretando um risco financeiro significativo para as empresas do setor energético. Portanto, é essencial que as projeções sejam não apenas precisas, mas também rápidas e confiáveis. No entanto, a obtenção de previsões pontuais é desafiada pela presença de fatores exógenos complexos, os quais introduzem não linearidades nas séries temporais de demanda, dificultando a modelagem (Bunn; Farmer, 1985).

A demanda de energia elétrica, conforme definida pelo Operador Nacional do Sistema Elétrico (ONS), representa a média da quantidade de eletricidade consumida em um período específico, geralmente expressa em quilowatts (kW) ou megawatts (MW). Por outro lado, o conceito de carga refere-se à quantificação da demanda de energia em intervalos horários específicos (ONS, 2024). Diferentes elementos externos variáveis ao longo do tempo, como condições climáticas, atividades humanas, processos industriais e características temporais e sazonais, exercem uma influência significativa sobre as cargas elétricas. Como resultado, as séries temporais que registram os padrões de consumo energético ao longo do tempo consistem em uma combinação de fatores lineares e não lineares (Bunn; Farmer, 1985; Gasparin; Lukovic; Alippi, 2022).

Os diversos segmentos de consumidores, incluindo residenciais, comerciais e industriais, estão integrados à rede elétrica e apresentam diferentes padrões de consumo, influenciados por uma variedade de elementos externos. A segmentação da previsão por classe de consumo revela uma diminuição na previsibilidade à medida que se aproxima do nível individual do consumidor. A determinação dos fatores que influenciam os consumidores residenciais é particularmente desafiadora devido à heterogeneidade dos perfis de consumo. Por outro lado, embora a demanda industrial seja mais previsível, ela ainda está sujeita a perturbações devido a eventos imprevistos, como interrupções na produção (Becken; Frampton; Simmons, 2001; Chou; Tran, 2018).

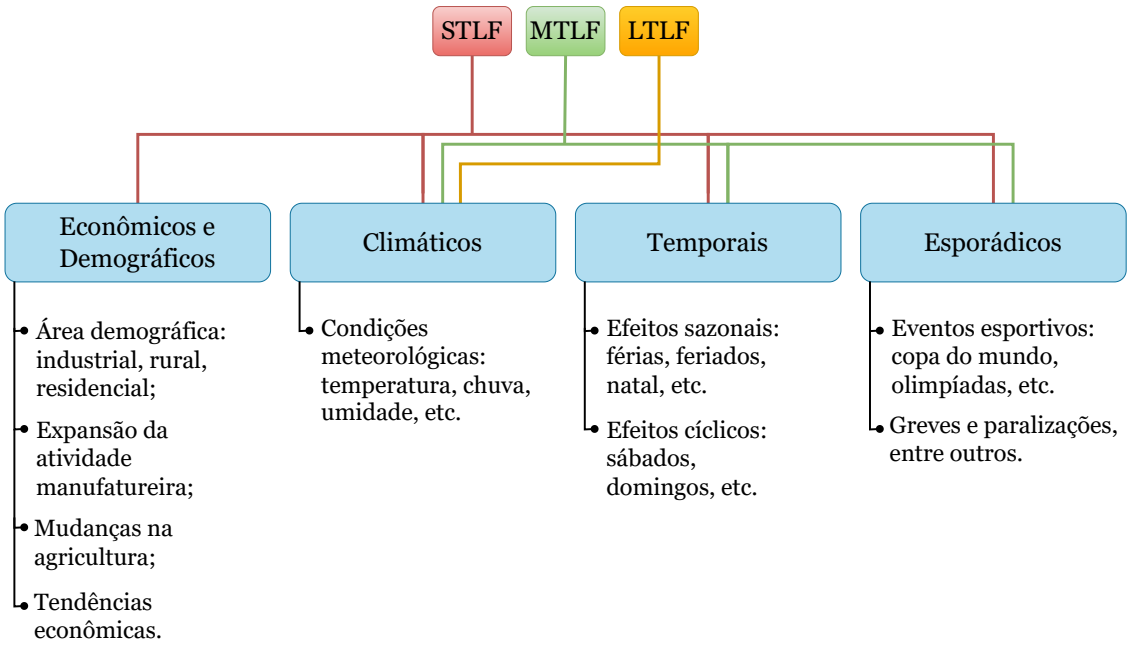
A influência dos fatores climáticos na demanda de energia pode variar significativamente em função do clima, da classe de consumidores e dos fatores socioeconômicos específicos de cada localidade. Em áreas caracterizadas por climas extremamente frios ou quentes, o consumo energético mostra-se altamente sensível a essas variáveis, devido à necessidade de manter o con-

forto térmico (Khatoon; Singh *et al.*, 2014). Nas áreas onde a predominância de consumidores pertence ao setor industrial, a demanda por energia é menos influenciada pelas variações meteorológicas. Em contrapartida, regiões cuja maioria dos clientes são residenciais e comerciais apresentam maior suscetibilidade a essas variações (Sadi *et al.*, 2010).

Em climas temperados, onde se manifesta uma considerável variação sazonal de temperatura, a relação entre a carga e a temperatura não segue um padrão linear (Moral-Carcedo; Vicéns-Otero, 2005). Nesse contexto, há uma temperatura crítica na qual a demanda de energia atinge seu mínimo, frequentemente observada entre 18ºC e 21ºC, variando de acordo com a região. Acima desse limiar térmico, tanto a intensificação quanto a diminuição da temperatura resultam em um aumento no consumo, seja para resfriamento, quando a temperatura aumenta, ou para aquecimento, quando a temperatura diminui (Andrade; Sant'Anna, 2013).

Com referência à pesquisa de Gross e Galiana (1987), a Figura 5 ilustra uma análise dos fatores exógenos que impactam o comportamento das cargas elétricas, evidenciando suas ramificações nos diversos tipos de previsões.

Figura 5 - Fatores Exógenos que Influenciam a Carga Elétrica.



Fonte: Autoria própria.

A análise da Figura 5 revela que as STLF são sensíveis a uma gama diversificada de influências, abrangendo aspectos econômicos e demográficos, além de variáveis climáticas, temporais e esporádicas. Por outro lado, a MTLF concentra-se principalmente na consideração dos impactos de fatores climáticos, temporais e esporádicos. Em contraste, a LTLF é predominantemente influenciada por fatores climáticos. É evidente que os fatores climáticos exercem a

maior influência sobre o comportamento das cargas elétricas, afetando diretamente os três tipos de previsão mencionados. Além disso, a STLF é a mais impactada pelos elementos exógenos.

3.6.1.3 Técnicas de Previsão de Carga

Nos últimos anos, dada a complexidade inerente às previsões de cargas elétricas, várias metodologias foram desenvolvidas e documentadas na literatura visando obter resultados significativos. Estas abordagens podem ser agrupadas em duas categorias principais (Gross; Galiana, 1987):

- Técnicas Estáticas: A previsão é tratada como um problema estático no tempo. Portanto, utiliza-se um conjunto fixo de dados históricos para construir o modelo, que prevê a carga elétrica com base nessas informações. Após o treinamento, o algoritmo permanece inalterado e usa apenas os dados disponíveis até aquele ponto específico para realizar previsões futuras. Essas técnicas empregam séries temporais discretas, onde os dados são coletados em intervalos de tempo fixos. Exemplos incluem a média móvel, a suavização exponencial e os modelos de regressão linear;
- Técnicas Dinâmicas: Refere-se às abordagens que introduzem uma dimensão adicional ao modelo, incorporando não apenas os dados históricos, mas também as informações emergentes ao longo do tempo. Esses algoritmos têm a capacidade de ajustar dinamicamente sua estrutura com base em novos dados, permitindo uma atualização contínua das previsões. Exemplos dessas técnicas incluem as RNAs, modelos de espaço de estado e filtros de Kalman.

No decorrer da evolução das técnicas de previsão, observou-se uma inclinação histórica para a aplicação de métodos estatísticos na antecipação de tendências de consumo. No entanto, com os avanços tecnológicos e a disponibilidade de *softwares* especializados em estatística, juntamente com o advento dos algoritmos de IA, novas estratégias preditivas surgiram, enriquecendo significativamente o campo da previsão de demanda elétrica. Neste contexto, conforme delineado por Hong *et al.* (2016), essas abordagens podem ser agrupadas em duas categorias principais: métodos convencionais e aqueles baseados em inteligência artificial.

Os métodos estatísticos tradicionais constituem um conjunto de técnicas matemáticas empregadas na previsão de cenários futuros com base em dados históricos. Essas abordagens se fundamentam na análise das relações entre diversas variáveis, visando identificar padrões passados que possam sugerir tendências futuras. Entre os métodos utilizados, destacam-se a regressão linear ou múltipla (Murto, 1998; Sfetsos, 2003), o filtro de Kalman (Moghram; Rahman, 1989; Murto, 1998), o alisamento exponencial (Moghram; Rahman, 1989) e os modelos de espaço de estado e séries temporais de Box-Jenkins (O'Donovan, 1983).

Métodos baseados em IA, por outro lado, empregam algoritmos capazes de aprender padrões complexos a partir de extensos conjuntos de dados, dispensando a necessidade de suposições prévias sobre a natureza das relações entre as variáveis. Em contraste com as abordagens estatísticas convencionais, que frequentemente demandam a necessidade de especificar explicitamente tais relações, os métodos baseados em IA possuem a habilidade de descobrir padrões de maneira autônoma, sendo particularmente vantajosos em cenários onde as relações entre variáveis são altamente não lineares e complexas (Damborg *et al.*, 1990; Carbonell; Michalski; Mitchell, 1983). Entre esses métodos, as redes neurais artificiais estão entre as abordagens mais promissoras (Haykin, 1998).

Ao longo das últimas décadas, as RNAs têm despertado um significativo interesse na comunidade científica devido à sua eficácia na abordagem de problemas complexos, superando diversas técnicas convencionais (Chae *et al.*, 2016; Jovanović; Sretenović; Živković, 2015). Esses algoritmos são concebidos como modelos computacionais compostos por unidades básicas de processamento, que emulam a estrutura do cérebro humano e a capacidade dos neurônios biológicos em adquirir e reter conhecimento (Haykin, 1998).

As RNAs têm se destacado como uma ferramenta eficaz na previsão de séries temporais. Em particular, esses avanços têm sido notáveis no campo da previsão do consumo energético, conforme evidenciado no Capítulo 2 deste trabalho. Entretanto, apesar das inúmeras vantagens que as RNAs oferecem, com o surgimento do *big data* e o aumento massivo no volume de dados de treinamento, observou-se certa instabilidade nas arquiteturas tradicionais. Para contornar essa dificuldade, surgiu uma nova abordagem baseada em redes neurais: o *deep learning* (Khodayar; Wang; Manthouri, 2018; Viscondi; Alves-Souza, 2019).

A abordagem descrita representa uma sofisticada extensão das redes neurais artificiais, que se destaca por sua profundidade e capacidade computacional. Essas características capacitam as redes neurais de aprendizado profundo a processar grandes conjuntos de dados e identificar padrões que seriam desafiadores de discernir utilizando métodos convencionais (Khodayar; Wang; Manthouri, 2018; Viscondi; Alves-Souza, 2019; Helbing; Ritter, 2018).

Recentemente, o campo do aprendizado profundo (*deep learning*) tem experimentado avanços significativos, impulsionados pela adoção de arquiteturas que integram mecanismos de atenção. Nesse contexto, um marco crucial foi estabelecido pelo trabalho *"Attention Is All You Need"*, publicado pela Google em 2017 (Vaswani *et al.*, 2017), o qual introduziu o *Transformer*, uma arquitetura de rede neural que se fundamenta exclusivamente em tais mecanismos para a análise de dados sequenciais. Embora originalmente concebidos para aplicações no processamento de linguagem natural (NLP), esses avanços têm o potencial de oferecer melhorias significativas na capacidade de previsão em séries temporais, incluindo a demanda energética.

Neste contexto, a implementação dos mecanismos de atenção em problemas de previsão revela uma promissora potencialidade para aprimorar a precisão das projeções, especialmente

em ambientes onde os padrões de consumo são influenciados por uma variedade de fatores interconectados e dinâmicos, como comportamentos dos consumidores, condições climáticas e sazonalidades. Dentro dessa perspectiva e com o objetivo de alcançar resultados promissores na previsão da demanda energética, esta pesquisa propõe uma metodologia que utiliza redes neurais profundas com mecanismos de atenção, visando uma modelagem adaptável aos padrões complexos e variáveis do consumo de energia.

3.7 Síntese do Capítulo

Este capítulo oferece uma explanação dos conceitos fundamentais das séries temporais, destacando sua relevância para a investigação conduzida neste estudo. As seções delineiam a importância das séries temporais na análise de dados e contextualizam a modelagem temporal, com particular ênfase nas técnicas de decomposição e modelagem preditiva. Além disso, o texto explora os modelos de previsão baseados em histórico temporal, acompanhados de métricas para a avaliação de precisão. Também são examinadas as propriedades inerentes aos dados temporais e seu impacto no desempenho dos modelos preditivos.

Especificamente direcionado ao domínio da energia elétrica, o capítulo investiga a previsão de carga elétrica, categorizando diferentes modalidades de previsões e analisando os determinantes de sua acurácia. Também, são discutidas diversas metodologias empregadas na previsão de carga elétrica, proporcionando uma visão das abordagens disponíveis.

4 REDES NEURAIS DE APRENDIZADO PROFUNDO

Este capítulo tem como propósito explorar os conceitos fundamentais relacionados às redes neurais de aprendizado profundo, delineando aspectos essenciais para a compreensão do escopo deste estudo. Além disso, visa fornecer os alicerces conceituais necessários para a compreensão dos métodos empregados nesta pesquisa. A estrutura do capítulo compreende várias subseções, cada uma dedicada a um tema específico e fundamental para o desenvolvimento da argumentação proposta.

A primeira Subseção, 4.1, estabelece o contexto e a relevância das redes neurais no âmbito da inteligência artificial contemporânea. A Subseção 4.2 examina distintas abordagens e técnicas de aprendizado de máquina, com ênfase no aprendizado supervisionado. Na Subseção 4.3, são inicialmente abordados os conceitos de neurônio artificial que, neste contexto, constituem as unidades básicas de computação. A discussão prossegue com a análise da arquitetura das redes neurais, descrevendo a organização das diferentes camadas e conexões para a execução de tarefas específicas. Também, a técnica de descida do gradiente é explorada, abordando seu papel na otimização dos parâmetros da rede. Além disso, são discutidas as funções de ativação e o processo de treinamento, que envolve a contínua otimização da rede.

A Subseção 4.4 aborda o aprendizado profundo, um subcampo do aprendizado de máquina que emprega redes neurais com múltiplas camadas para modelar as complexidades presentes nos dados. Na Subseção 4.5, são apresentadas as redes neurais *Transformer*, uma inovação significativa no processamento de dados sequenciais, destacando os princípios fundamentais dos mecanismos baseados em atenção.

A Subseção, 4.6, dedica-se ao TFT, cuja arquitetura é o foco desta pesquisa, detalhando componentes específicos como redes de seleção de variáveis, mecanismos de bloqueio, *encoders* de covariáveis estáticas e diversos aspectos do processamento temporal. São abordados a atenção multifacetada interpretável, que melhora a capacidade da rede de se concentrar em partes relevantes da sequência de entrada, a camada de sequência a sequência, que facilita a previsão de séries temporais, e outras camadas adicionais que melhoram o desempenho do modelo.

O capítulo é concluído na Subseção 4.7, que sintetiza os principais pontos discutidos, reforçando a compreensão dos conceitos abordados. Desta forma, este capítulo oferece uma base teórica abrangente para a compreensão das técnicas e conceitos subjacentes ao desenvolvimento e aplicação de redes neurais de aprendizado profundo, além dos métodos específicos empregados nesta pesquisa.

4.1 Introdução

Atualmente, observa-se uma crescente integração da IA em diferentes esferas do conhecimento, desempenhando um papel expressivo na evolução de diversos setores da sociedade. A capacidade da IA de sistematizar e automatizar uma ampla gama de tarefas, a posiciona como uma ferramenta aplicável em múltiplos domínios da atividade intelectual humana. Dessa forma, a IA emerge como um campo de vasto estudo e de alcance universal, com potencial para impactar significativamente o entendimento e a prática de áreas diversas como saúde, educação, economia, tecnologia, entre outras (Russell; Norvig, 2013).

Nos primórdios da discussão acerca da capacidade das máquinas em demonstrar inteligência, Ada Lovelace iniciou as primeiras reflexões sobre o tema em 1842. Lovelace expressou ceticismo quanto à perspectiva de que máquinas pudessem exibir qualquer forma de inteligência, argumentando que as mesmas eram limitadas a executar apenas as tarefas para as quais foram explicitamente programadas (Füegi; Francis, 2015).

Nos anos subsequentes, as hipóteses inicialmente levantadas por Ada Lovelace encontraram uma nova perspectiva com as investigações de Alan Turing. De forma pioneira, Turing propôs uma metodologia para avaliar a inteligência das máquinas, o que ficou consagrado como o Teste de Turing (Turing, 1950). No trabalho intitulado *"Computing Machinery and Intelligence"*, Turing vislumbrou um futuro no qual os computadores não só seriam capazes de aprender, mas também de exibir criatividade. Este marco significativo na história da inteligência artificial sinalizou uma mudança de paradigma, ampliando as possibilidades e as fronteiras do que até então era considerado possível para as máquinas.

Contudo, somente em 1956, os pesquisadores John McCarthy (Dartmouth College), Marvin Minsky (MIT), Nathaniel Rochester (IBM) e Claude Shannon (Bell Laboratories) cunharam o termo "Inteligência Artificial", marcando o início formal do estudo e da pesquisa sistemática sobre a IA (Nilsson, 1982). Nos dias atuais, apesar de a expressão "Inteligência Artificial" ser amplamente difundida tanto na literatura especializada quanto no senso comum, a definição precisa desta área de estudo ainda é um desafio.

Stuart Russell e Peter Norvig, em sua obra *"Artificial Intelligence: a modern approach"* (Russell; Norvig, 2016), amplamente reconhecida como uma referência desta disciplina científica, o termo "IA" é usado para descrever sistemas com a capacidade de simular o pensamento ou comportamento humano, ou até mesmo raciocinar de maneira análoga à humana. Contudo, a inteligência artificial pode ser definida como uma área de estudo que se dedica à criação de sistemas computacionais capazes de replicar e automatizar uma ampla gama de tarefas perceptivas, cognitivas e manipulativas que são executadas por organismos vivos, enfatizando principalmente o ser humano (Russell; Norvig, 2013). Portanto, trata-se de um campo de pesquisa que visa não apenas reproduzir as capacidades humanas em termos de percepção e raciocínio,

mas também desenvolver mecanismos que possam eventualmente superar habilidades naturais.

A campo da IA é bastante abrangente e engloba uma vasta gama de subáreas, entre as quais se destaca o aprendizado de máquina, do inglês, *machine learning* (ML). Este, por sua vez, é amplamente reconhecido como um dos pilares fundamentais da IA contemporânea. O aprendizado de máquina dedica-se ao desenvolvimento de algoritmos e técnicas que capacitam os sistemas computacionais a adquirir conhecimento a partir de dados brutos, habilitando-os a desempenhar tarefas específicas. Dentro do ML, as RNAs emergem como uma classe particularmente relevante de algoritmos. Inspiradas na arquitetura neural do cérebro humano, as RNAs são empregadas na modelagem de sistemas complexos e na resolução de uma extensa diversidade de problemas, consolidando-se como uma ferramenta promissora no domínio da IA (Murphy, 2012; Haykin, 2009).

4.2 Aprendizado de Máquina

A IA teve seus primeiros sucessos fundamentados no paradigma da programação clássica, também conhecida como IA simbólica. Nessa abordagem, especialistas desenvolviam conjuntos de regras para resolver problemas específicos. Essas instruções eram então traduzidas em modelos que representavam o problema em questão, permitindo assim que estes gerassem respostas quando confrontados com dados pertinentes. Embora essa metodologia possibilitasse a resolução de problemas para os quais havia sido programada, se mostrou inadequada ou inviável quando aplicada a desafios perceptivos. Estes problemas são caracterizados por uma maior complexidade, nos quais as regras formais não são claramente definidas ou envolvem elementos subjetivos (Goodfellow; Bengio; Courville, 2016).

A resolução de problemas perceptivos é uma tarefa que, embora seja natural para os seres humanos, apresenta uma complexidade considerável para os sistemas computacionais. A necessidade de representações detalhadas em linguagem de máquina torna esses desafios particularmente difíceis de serem abordados. Muitas vezes, não há instruções formais claras, o que os torna intrinsecamente subjetivos. Além disso, mesmo quando diretrizes formais estão presentes, a complexidade pode dificultar significativamente a tradução desses problemas para modelos matemáticos ou lógicos, exigindo um investimento considerável para se obter uma modelagem adequada (LeCun; Bengio; Hinton, 2015; Chollet, 2021).

A demanda por lidar com problemas perceptivos deu origem a um ramo alternativo dentro do campo da inteligência artificial, conhecido como aprendizado de máquina. Nesse paradigma, em vez de ser programado com regras fixas, o modelo adquire conhecimento através da exposição a um conjunto prévio de perguntas e as respectivas respostas. Esse enfoque permite que o algoritmo aprenda diretamente com os dados, adaptando-se a uma variedade de problemas. Portanto, este ramo da IA concentra-se na resolução de problemas perceptivos e subjetivos,

oferecendo uma abordagem mais eficaz para lidar com tarefas complexas e ambíguas (Chollet, 2021).

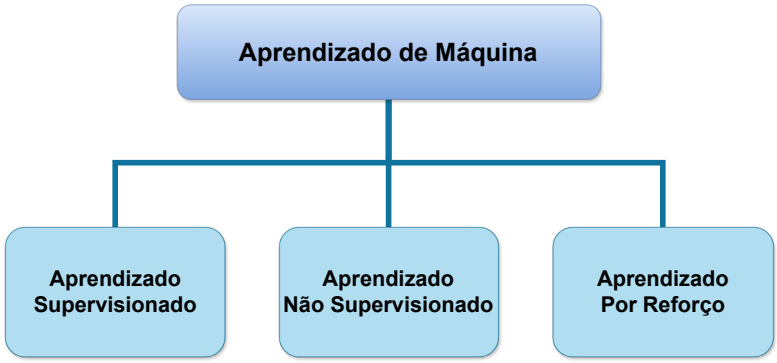
A aprendizagem de máquina tem raízes na teoria estatística, buscando construir modelos matemáticos para realizar inferências a partir de conjuntos de dados. Nesse cenário, o papel da ciência da computação é essencial em duas frentes: inicialmente, no treinamento, onde são necessários algoritmos para resolver o problema de otimização, juntamente com a habilidade de armazenar e processar grandes volumes de dados. Em seguida, após o processo de aprendizado, é crucial garantir a eficiência tanto na representação do modelo quanto nos algoritmos utilizados para realizar as inferências. Em certos contextos, a eficácia dos algoritmos de aprendizado ou de inferência, expressa pela complexidade temporal e espacial, pode ser tão relevante quanto a precisão (Alpaydin, 2020).

Desta forma, o ML se baseia em técnicas derivadas de modelos matemáticos embasados na teoria estatística. Essas técnicas permitem que os computadores automatizem tarefas ao identificar padrões nos conjuntos de dados ou em experiências prévias (Bhavsar *et al.*, 2017). Em essência, um sistema de aprendizado de máquina pode ser conceituado como um programa de computador que, utilizando-se de experiências derivadas da resolução bem-sucedida de problemas anteriores, adquire a capacidade de tomar decisões (Carbonell; Michalski; Mitchell, 1983). Esse processo permite que o sistema generalize o conhecimento adquirido para situações novas e complexas, ampliando a aplicabilidade para uma variedade de domínios.

Os algoritmos de ML visam melhorar o desempenho em tarefas específicas, através da otimização de critérios de desempenho mediante a análise de dados de exemplo. Esse processo envolve a criação de um modelo com parâmetros específicos, cuja aprendizagem ocorre quando o algoritmo é executado para ajustar esses critérios utilizando conjuntos de dados de treinamento. Tais abordagens podem ter uma natureza preditiva, permitindo fazer previsões sobre eventos futuros, ou descritiva, que extrai conhecimento dos dados. Além disso, é possível integrar ambas as funcionalidades em um único modelo (Alpaydin, 2020).

Existem diversas implementações de algoritmos de aprendizado de máquina, abrangendo uma ampla gama de tipos e aplicações. De acordo com as características de aprendizagem, esses modelos podem ser basicamente divididos em três subcategorias principais: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço (Fausett, 2006; Krose; Smagt, 1996; Sutton; Barto, 2018). Essas categorias representam diferentes abordagens que variam em como os modelos interagem com os dados e aprendem a partir deles. A Figura 6 esquematiza essa divisão, destacando as principais subcategorias no contexto do aprendizado de máquina.

Figura 6 - Principais Tipos de Aprendizado de Máquina.



Fonte: Autoria própria.

O aprendizado supervisionado destaca-se como a abordagem mais comum em *machine learning* (LeCun; Bengio; Hinton, 2015). Nesse método, o modelo é submetido ao processo de treinamento utilizando um conjunto de dados composto por entradas e as respectivas saídas. O objetivo desse processo é obter uma função capaz de mapear tais entradas para as saídas correspondentes. Isso é alcançado por meio do uso de conjuntos de dados previamente rotulados, o que permite o aprendizado e a compreensão das características inerentes às informações fornecidas. Consequentemente, à medida que mais registros são incorporados, e em consonância com a qualidade dos dados utilizados, espera-se uma maior precisão do modelo treinado (Russell; Norvig, 2016).

Já nos algoritmos de aprendizagem não-supervisionada, não há atribuição de rótulos aos dados de saída, neste caso, busca-se identificar padrões e similaridades entre os dados fornecidos. Esses algoritmos podem ser categorizados em algoritmos de transformação, que buscam criar uma nova representação dos dados para melhor interpretação ou desempenho, e algoritmos de agrupamento, que segmentam os dados em grupos com características similares. Os métodos de agrupamento podem variar de acordo com critérios pré-estabelecidos, como distância geométrica, distribuições estatísticas ou densidade de pontos. Em determinadas situações, visando a otimização do desempenho dos algoritmos não-supervisionados, é possível realizar a rotulação parcial dos dados. Tal abordagem, denominada aprendizado semi-supervisionado, constitui uma aplicação híbrida que combina elementos de ambas as modalidades de aprendizado, não-supervisionado e supervisionado (Müller; Guido, 2016).

O aprendizado por reforço, por sua vez, é um tipo de aprendizado de máquina que se concentra na interação entre um agente e um ambiente em constante mudança. O objetivo principal é que o agente aprenda a tomar ações que levem a maximizar uma recompensa numérica específica. Esse método tem três características principais: ciclo contínuo de interações (*loop*

fechado), a falta de instruções diretas sobre quais ações tomar e a consideração de consequências das ações ao longo do tempo. Assim, em vez de receber instruções explícitas sobre quais ações tomar, o agente aprende através da tentativa e erro e, em um processo iterativo e contínuo, ajusta o seu comportamento com base nas informações recebidas do ambiente. Ao contrário do aprendizado supervisionado, onde o modelo é ensinado por exemplos rotulados, e do aprendizado não supervisionado, que busca encontrar padrões em dados não rotulados, o aprendizado por reforço enfrenta o desafio de equilibrar a exploração, ou seja, experimentar novas ações para descobrir o impacto ou aproveitar as ações conhecidas para maximizar a recompensa (Sutton; Barto, 2018).

Outro aspecto importante em relação aos algoritmos de aprendizado de máquina diz respeito à medida em que os sistemas de aprendizado proporcionam compreensibilidade ao ser humano. Neste contexto, os sistemas de aprendizado podem ser divididos em duas amplas abordagens (Michalski, 1983; Kubat; Bratko; Michalski, 1998):

- Sistemas "caixa-preta": Esta categoria de sistemas caracteriza-se pela sua representação interna, que muitas vezes é de difícil compreensão para os seres humanos. Em outras palavras, tais sistemas não oferecem uma interpretabilidade clara do processo de reconhecimento ou tomada de decisão;
- Sistemas orientados ao conhecimento: Tratam-se de sistemas que são projetados com o objetivo de criar estruturas simbólicas compreensíveis para os seres humanos. Tais estruturas são concebidas para fornecer uma visão clara e interpretável do processo de aprendizado e decisão, facilitando assim a interação e compreensão por parte dos usuários.

Esta pesquisa se baseia em um tipo de aprendizado de máquina que envolve supervisão, portanto os algoritmos são treinados com exemplos rotulados para realizar previsões ou classificações. Assim, no próximo tópico, a discussão e a conceitualização desse método serão aprofundados, explorando suas implicações teóricas em maior detalhe.

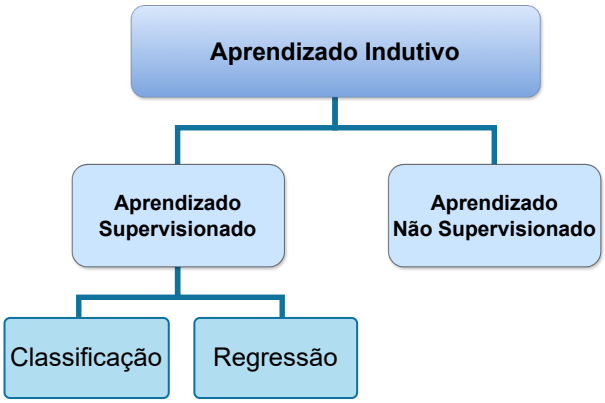
4.2.1 Aprendizado Supervisionado

A indução é um tipo de raciocínio lógico que permite fazer generalizações a partir de exemplos específicos. Portanto, consiste em aprender conceitos ao fazer inferências indutivas sobre os exemplos apresentados. No entanto, as hipóteses geradas por meio da indução podem ou não ser verdadeiras. Apesar disso, a indução é amplamente utilizada para derivar novos conhecimentos e prever eventos futuros. Grandes descobertas científicas, como as leis da hidrostática por Arquimedes, as leis do movimento planetário por Kepler e as leis da seleção natural por Darwin, foram feitas através da indução (Monard; Baranauskas, 2003).

O aprendizado indutivo ocorre quando se raciocina sobre exemplos fornecidos por um processo externo ao sistema de aprendizado. Assim, ele pode ser supervisionado ou não supervisionado. No aprendizado não supervisionado, o algoritmo examina os exemplos disponíveis e busca identificar se é possível agrupá-los de alguma forma, criando conjuntos ou *clusters*. Após a formação dos *clusters*, geralmente é necessário realizar uma análise adicional para determinar o significado de cada agrupamento dentro do contexto do problema em questão.

Já no aprendizado supervisionado, o algoritmo recebe um conjunto de exemplos de treinamento com rótulos de classe previamente conhecidos. Em termos gerais, cada instância é representada por um conjunto de características, ou atributos, acompanhada pelo rótulo que indica a classe correspondente. Dessa forma, o objetivo é desenvolver um classificador que possa identificar com precisão a classe de novas instâncias não rotuladas. Essa tarefa é conhecida como classificação para rótulos de classe discretos e regressão para valores contínuos (Monard; Baranauskas, 2003). Conforme ilustra a Figura 7.

Figura 7 - Tipos de Aprendizado Indutivo.



Fonte: Adaptação (Monard; Baranauskas, 2003).

Após o treinamento com os dados históricos, o algoritmo de aprendizado de máquina desenvolve uma função de classificação que define fronteiras no espaço de características, discriminando exemplos pertencentes a diferentes classes. Essa função, uma vez estabelecida, pode ser aplicada para fins de predição: ao identificar padrões nos dados históricos, pode-se generalizar para novas instâncias, presumindo que o futuro seguirá padrões similares ao passado. Portanto, o processo de aprendizado possibilita a extração de conhecimento, cuja função de classificação resultante é um modelo simplificado que oferece uma interpretação do processo subjacente aos dados, facilitando a identificação de redundâncias e padrões relevantes (Alpaydin, 2020).

Na classificação, o modelo recebe uma entrada e atribui uma categoria ou classe como resposta, representada numericamente como 0 ou 1, $C(x) \in \{0,1\}$. Entretanto, quando a saída desejada é um valor numérico, busca-se aprender não apenas uma classe, mas sim uma função que mapeia as entradas para saídas correspondentes. Isso é conhecido como regressão, nessa

abordagem o objetivo é encontrar uma função $f(x)$ que estime os valores de saída com base nos dados de entrada. Geralmente, essa função é desconhecida previamente, em vez disso, tem-se um conjunto de exemplos de treinamento representado por pares de entradas-saídas (Alpaydin, 2020).

Considerando um conjunto de dados de treinamento X, onde cada elemento é representado como uma tupla (x^t, r^t), com t variando de 1 a N, onde N é o número total de exemplos de treinamento. Nesse contexto, x^t é uma entrada específica e r^t é o valor numérico associado a essa entrada, interpolado no conjunto dos números reais ($\mathbb{R}$). O objetivo é encontrar uma função $f(x)$ que possa mapear cada entrada para o valor correspondente, tal que $r^t = f(x^t)$. Em outras palavras, busca-se uma função que seja capaz de prever ou representar os valores associados às entradas do conjunto de dados de treinamento.

Na interpolação polinomial, ao trabalhar com N pontos de dados, é possível encontrar um polinômio de grau $N-1$ que se ajusta adequadamente a esses pontos, permitindo prever a saída para qualquer valor de x dentro do intervalo dos dados. No entanto, quando o valor de x está fora do intervalo dos pontos x^t no conjunto de treinamento, surge uma situação conhecida como extrapolação. Um exemplo prático dessa situação é observado na previsão de séries temporais, onde os dados geralmente estão limitados até o presente. Ao realizar uma regressão para prever valores futuros, o objetivo é extrapolar além dos limites temporais dos dados de treinamento, antecipando assim o comportamento da série além do horizonte temporal observado. Essa extrapolação é crucial para projetar tendências ou padrões futuros com base nos dados históricos disponíveis.

No contexto da regressão, é uma prática comum introduzir um ruído na saída da função desconhecida, representada como $r^t = f(x^t) + \varepsilon$, onde $f(x) \in \mathbb{R}$ é a função desconhecida e ε é o termo de erro aleatório adicionado. Este ruído é atribuído à presença de variáveis latentes no mundo real que podem não ser completamente observáveis ou previsíveis. Dessa forma, a função real desconhecida pode ser modelada como $r^t = f^*(x^t, z^t)$, onde z^t representa essas variáveis latentes. Essa formulação reconhece a complexidade do mundo real, incorporando variáveis não diretamente observáveis na relação entre as entradas e as saídas do modelo de regressão.

Portanto, busca-se aproximar a saída usando um modelo $g(x)$. O erro empírico no conjunto de treinamento X é calculado como a média do quadrado das diferenças entre as saídas reais e as saídas previstas, matematicamente expresso na equação (11).

$$E(g \mid X) = \frac{1}{N} \sum_{t=1}^{N} [r^t - g(x^t)]^2 \tag{11}$$

Dado que tanto r quanto $g(x)$ são valores numéricos, é possível estabelecer uma ordem

definida entre esses valores e calcular a distância entre eles. Uma medida comum de distância é o quadrado da diferença, que oferece mais informações do que apenas determinar se eles são iguais ou diferentes, como ocorre na classificação. Outra função de erro comum é o valor absoluto da diferença. Assim, esse cálculo proporciona uma medida de quão bem o modelo está se ajustando aos dados disponíveis no conjunto de treinamento, e o objetivo é minimizar o erro, ajustando adequadamente os parâmetros do modelo $g(x)$ (Alpaydin, 2020).

4.3 Redes Neurais Artificiais

O cérebro humano pode ser conceituado como um sistema interconectado de processamento de informações, caracterizado por uma alta complexidade, não linearidade e capacidade de operação paralela. Diferentemente dos computadores digitais, o sistema nervoso central humano possui a habilidade de reconhecer padrões de forma rápida e eficiente, devido à estruturação de neurônios. Essa capacidade cognitiva é atribuída à habilidade do cérebro de criar e adaptar regras próprias com base em um vasto conjunto de experiências vivenciadas (Haykin, 2009).

As RNAs representam uma vertente computacional inspirada no complexo funcionamento do cérebro humano. Além de buscar replicar as interações e o processamento das informações entre os neurônios, esses modelos visam incorporar capacidades lógicas aos algoritmos, permitindo que realizem tarefas cognitivas de forma semelhante à mente humana. A notoriedade dessa abordagem é atribuída, principalmente, à capacidade de aprendizado e generalização do conhecimento adquirido (Haykin, 2009; Nielsen, 2015).

Uma das competências fundamentais no processo de aquisição de conhecimento humano é a regra de aprendizagem empírica, que envolve a experimentação e a correção de erros como método principal de aprendizado. Paralelamente, as RNAs representam uma forma de tecnologia computacional que demonstra capacidade de aprendizado contínuo, semelhante ao humano. De maneira análoga ao funcionamento cerebral, essas redes adquirem conhecimento por meio da interação com o ambiente. Tal abordagem é dotada de conexões sinápticas, cujos pesos são ajustados ao longo do processo de aprendizagem para reter o conhecimento assimilado (Chang; Tseng, 2017; Haykin, 2009).

Neste contexto, é importante ressaltar que o propósito das redes neurais artificiais não reside na reprodução da complexidade operacional dos sistemas biológicos, mas sim na utilização dos princípios e conhecimentos adquiridos acerca da funcionalidade das redes biológicas para enfrentar desafios de natureza complexa. Em outras palavras, as RNAs não buscam uma simulação exata do funcionamento do cérebro, mas sim uma abstração funcional que se inspire nos mecanismos biológicos para resolver problemas em diversos domínios. Assim, este paradigma explora os conceitos fundamentais subjacentes ao funcionamento cerebral para desenvolver soluções computacionais eficientes e adaptáveis, capazes de lidar com problemas que

frequentemente carecem de soluções algorítmicas avançadas (Basheer; Hajmeer, 2000).

O neurônio artificial, proposto por McCulloch e Walter Pitts em 1943 (McCulloch; Pitts, 1943), representa um marco originário no desenvolvimento das redes neurais artificiais. Os referidos autores introduziram o conceito de neurônio artificial, fundamentando-se tanto na compreensão da fisiologia e funcionamento cerebrais quanto nos princípios da lógica proposicional e da teoria da computação de Alan Turing (Turing, 1950). Este conceito pioneiro desempenhou um papel catalisador para uma série de avanços subsequentes no domínio das RNAs (Haykin, 1998; Russell; Norvig, 2016).

Desde então, as arquiteturas neurais têm sido alvo de constante desenvolvimento e refinamento, abrangendo uma ampla gama de aplicações em diversos domínios, como reconhecimento de padrões, processamento de linguagem natural, previsão de séries temporais, diagnóstico médico, sistemas de recomendação, otimização de processos industriais, entre outros. Esse crescimento exponencial demonstra a versatilidade e eficácia das RNAs como poderosas ferramentas para lidar com problemas complexos em diversas áreas do conhecimento (Haykin, 1998; Nielsen, 2015).

Este estudo tem como foco principal a aplicação de RNAs. Assim, os próximos segmentos desta subseção são dedicados ao aprofundamento dos conceitos associados a esta abordagem, explorando aspectos que contribuam para uma compreensão mais abrangente dos fundamentos pertinentes ao escopo desta pesquisa.

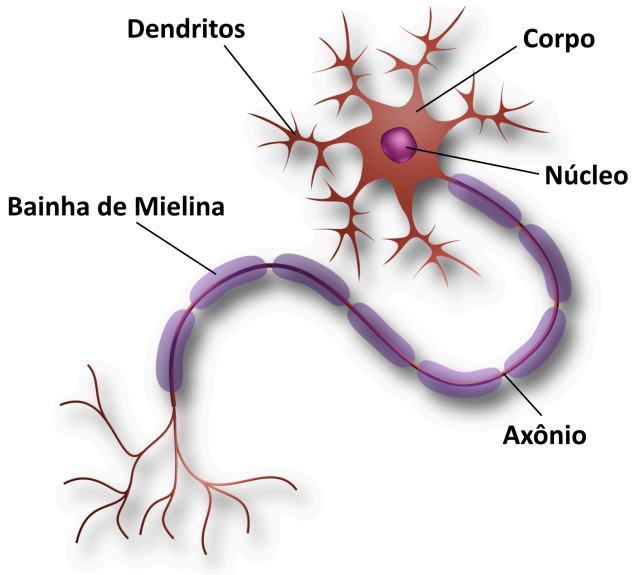
4.3.1 O Neurônio Artificial

Os elementos fundamentais do neurônio biológico, compreendem o corpo celular (ou soma), os dendritos e o axônio, cada um com papéis específicos na transmissão de sinais nervosos. Os dendritos atuam como receptores de estímulos provenientes de neurônios vizinhos, encaminhando tais sinais ao corpo celular. Este último, converte os estímulos em potenciais elétricos, que são conduzidos ao longo do axônio em direção a outras células nervosas. A interconexão entre células nervosas ocorre por meio de sinapses, cujo axônio de um neurônio estabelece contato com os dendritos de outro. Assim, o encéfalo humano constitui-se de um sistema neural intrincado, caracterizado por uma complexa rede de células nervosas interligadas, capazes de processar e transmitir informações através da geração e propagação de impulsos elétricos (Shepherd, 2003).

O modelo de neurônio artificial proposto por McCulloch e Pitts (McCulloch; Pitts, 1943) representa uma abstração computacional inspirada na estrutura e operação dos neurônios biológicos. Essa formulação, concebida para o processamento de informações, exibe um paralelismo com o comportamento observado nos neurônios biológicos, os quais recebem estímulos de outros neurônios e emitem impulsos nervosos em resposta (Wasserman, 1989; Krose; Smagt,

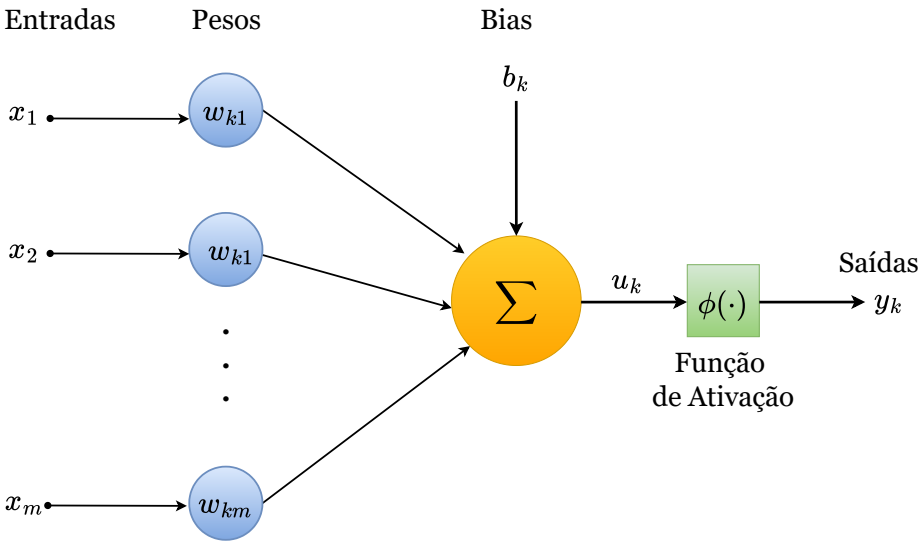
1996). A Figura 8 ilustra a estrutura do neurônio biológico, enquanto a Figura 9 apresenta a unidade de computação artificial desenvolvida por McCulloch e Pitts.

Figura 8 - Neurônio Biológico.



Fonte: Autoria própria.

Figura 9 - Neurônio Artificial de McCulloch e Pitts.



Fonte: Autoria própria.

Ao simplificar as complexas interações neuronais, o modelo de McCulloch e Pitts fundamentou diversos algoritmos que visam emular a capacidade adaptativa do cérebro humano. O ponto central desta estrutura reside na ponderação das entradas, uma etapa fundamental para determinar a relevância de cada estímulo para o neurônio em consideração. Esse processo envolve a atribuição de pesos sinápticos às diferentes entradas, refletindo a contribuição relativa

de cada estímulo nas operações do neurônio artificial. Em termos práticos, as entradas são multiplicadas pelos pesos sinápticos correspondentes, resultando numa combinação ponderada dos estímulos recebidos (Schmidhuber, 2015; Russell; Norvig, 2016).

Após esse estágio, a saída resultante é submetida a uma função de ativação, que desempenha o papel de processador interno do neurônio artificial. Tal função, frequentemente não linear, regula a saída final do neurônio, determinando sua ativação e intensidade, em resposta à combinação ponderada das entradas. Estes procedimentos, em conjunto, constituem os fundamentos do modelo de neurônio artificial proposto por McCulloch e Pitts (Russell; Norvig, 2016; Krose; Smagt, 1996).

O comportamento de um neurônio pode ser compreendido como uma transformação matemática dos sinais de entrada que recebe. Neste contexto, cada estímulo recebido (x_m), é multiplicado por um peso sináptico correspondente (w_{km}). Estes, por sua vez, são ajustáveis e definem a importância relativa de cada entrada para a saída do neurônio. Conforme Haykin (2009), pode-se expressar matematicamente o funcionamento básico de um neurônio artificial k por meio das equações (12) e (13).

$$u_k = \sum_{j=0}^{m} w_{kj} x_j \tag{12}$$

Em que:

u_k: representa a saída do combinador linear do neurônio k;

$x_1, x_2, ..., x_m$: correspondem aos sinais de entrada fornecidos ao neurônio;

$w_1, w_2, ..., w_{km}$: reflete os pesos associados aos estímulos recebidos pelo neurônio k.

Portanto, a soma ponderada dos sinais de entrada x_j multiplicados pelos seus respectivos pesos w_{kj} é realizada, com j variando de 0 a m, sendo m o número total de sinais de entrada.

A saída y_k do neurônio k, expressa na equação (13), é calculada após a execução de uma função de ativação $\phi(\cdot)$. Tal função é aplicada ao resultado do combinador linear $(u_k + b_k)$, onde b_k é o bias (ou desvio) do neurônio k. O bias permite ao neurônio ajustar sua resposta para além da soma ponderada dos sinais de entrada, incrementando uma componente adicional.

$$y_k = \phi(u_k + b_k) \tag{13}$$

No contexto da equação (13), o neurônio incorpora um componente polarizador denominado viés (b_k), que é adicionado ao resultado da soma ponderada. Esse elemento modula a influência geral das entradas, ajustando o limiar de ativação. O valor resultante, $u_k + b_k$ é então submetido a uma função de ativação, encarregada de determinar a saída do neurônio (represen-

tada por y_k). Tal função introduz não-linearidades na resposta neural, capacitando-o a discernir padrões complexos nos dados de entrada e a gerar respostas não-lineares à estimulação.

Em síntese, o neurônio opera como uma unidade de processamento que recebe múltiplos sinais de entrada (x_m), os quais podem originar-se de outros neurônios na rede ou de estímulos externos. Cada sinal é associado a um peso específico (w_{km}), que denota a importância relativa do sinal na computação do neurônio. Assim, quando um sinal de entrada é recebido, este é multiplicado pelo peso correspondente (ou seja, x_m é multiplicado por w_{km}), refletindo a influência desse sinal na atividade neural. Em seguida, o neurônio realiza a soma dos produtos de todos os sinais de entrada e pesos, resultando em um valor u_k. Este valor, por sua vez, representa a combinação linear dos sinais de entrada, ponderados pelos seus respectivos pesos, e reflete a influência total das entradas sobre o neurônio. Por fim, o resultado obtido u_k é processado juntamente com um componente de polarização b_k, através de uma função de ativação $\phi(\cdot)$, que determina a saída do neurônio y_k.

Uma RNA é um sistema complexo composto por diversas unidades de neurônios artificiais. Existe uma variedade de modelos disponíveis dentro dessa abordagem, cuja diferenciação ocorre principalmente com base em três elementos fundamentais: a arquitetura, que estabelece o arranjo de conexões entre os neurônios; o algoritmo de treinamento, que determinam como a rede ajusta suas conexões para aprender e realizar tarefas específicas; e as funções de ativação, que controlam a saída de cada neurônio e são essenciais para a capacidade da rede de realizar operações não lineares (Fausett, 2006). Neste contexto, os próximos segmentos desta subseção visam fornecer uma compreensão mais aprofundada do funcionamento das redes neurais artificiais, apresentando de forma mais detalhada cada um desses elementos.

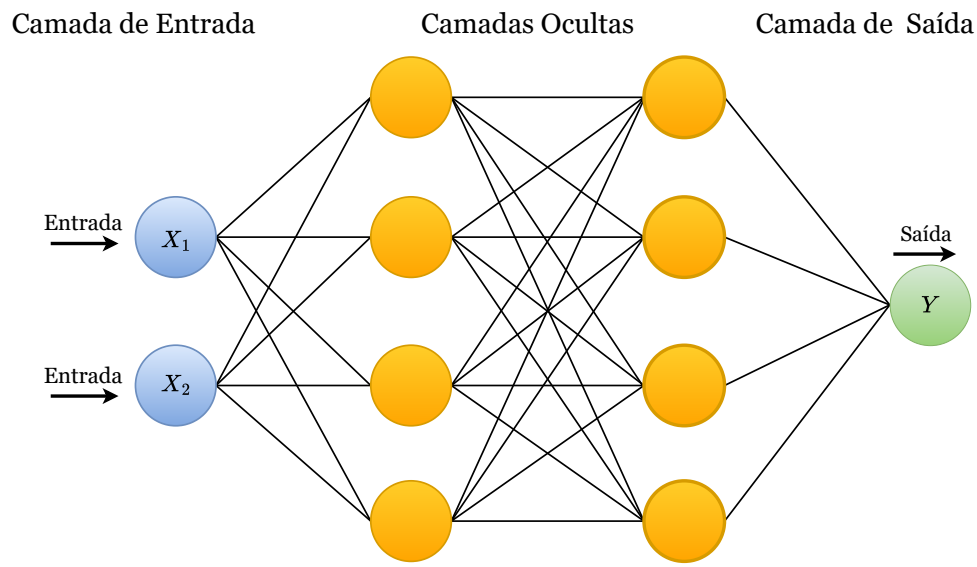
4.3.2 Arquitetura

A configuração das conexões entre neurônios, conhecida como arquitetura, é um aspecto fundamental na estrutura de uma rede neural artificial. Essa organização determina não apenas como os elementos interagem entre si, mas também como a informação é processada e transmitida através da rede. Desde modelos simples até redes profundas, o arranjo das conexões entre os neurônios desempenha um papel essencial na capacidade das RNAs em aprender e generalizar padrões complexos (Gurney, 2018; Krose; Smagt, 1996).

A estrutura de uma RNA é composta por neurônios interconectados, cujo funcionamento reflete o processo de propagação de sinais no cérebro humano. Quando um neurônio recebe um estímulo de entrada, este processa tal informação e, em seguida, transmite um sinal modificado como saída para outros neurônios. Desta forma, a dinâmica de processamento e transmissão de sinais em uma RNA ocorre de acordo com a arquitetura específica do modelo, podendo incluir múltiplas camadas e diferentes tipos de conexões (Krose; Smagt, 1996; Kovács, 2006).

A Figura 10 fornece uma representação visual da arquitetura de uma RNA.

Figura 10 - Arquitetura de uma Rede Neural Artificial.



Fonte: Autoria própria.

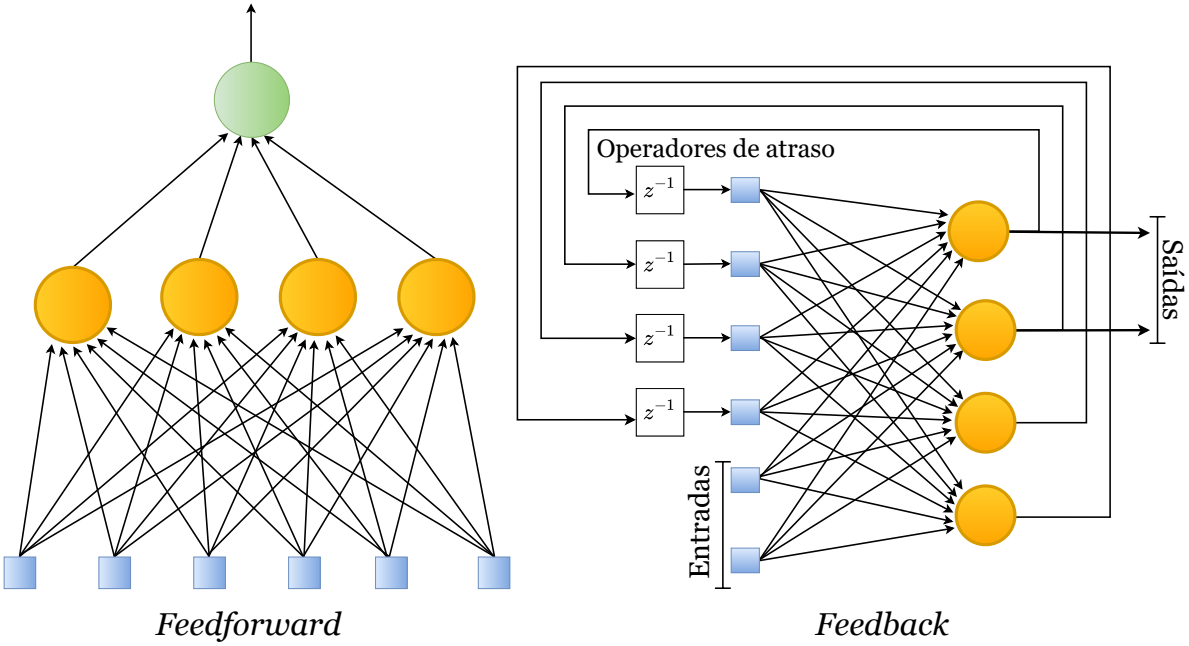
As unidades neurais são organizadas seguindo uma estrutura hierárquica que inclui a camada de entrada, na qual os dados são inicialmente introduzidos na rede, a camada intermediária, onde ocorre o processamento e a extração de características dos dados, e, por fim, a camada de saída, responsável por fornecer as respostas ou previsões da rede. Assim, as informações fornecidas à camada de entrada são processadas de forma sequencial pelas camadas intermediárias, culminando finalmente na camada de saída. As camadas intermediárias, comumente referidas como camadas ocultas, desempenham um papel essencial no processo de aprendizado, sendo responsáveis por extrair e transformar as características dos dados de entrada antes da interpretação final na camada de saída (Haykin, 2009; Kovács, 2006).

A estrutura hierárquica e sequencial das RNAs reflete a maneira como os sistemas biológicos processam informações, permitindo uma representação mais eficiente de dados complexos. Tal arranjo favorece tanto o aprendizado quanto a generalização de padrões, uma vez que as redes neurais conseguem extrair gradualmente características abstratas e significativas dos dados de entrada, assim como ocorre nos sistemas biológicos (Haykin, 2009; Kovács, 2006).

Em termos de organização estrutural, as redes neurais artificiais podem ser diferenciadas entre estáticas e dinâmicas, conforme categorização proposta por Haykin (2009). A distinção fundamental entre essas duas formas de organização reside na presença ou ausência de conexões que retroalimentam os neurônios dentro da arquitetura. Quando não há dependência das saídas

atuais em relação aos valores anteriores, as redes são consideradas estáticas. As representações gráficas dessas topologias podem ser observadas na Figura 11.

Figura 11 - Conexão *Feedforward* e Conexão *Feedback*.



Fonte: Adaptação (Haykin, 2009).

Portanto, levando em consideração tanto o método de propagação da informação através dos neurônios quanto a topologia estrutural das redes neurais, é possível fazer uma categorização das redes neurais artificiais com base nesses dois aspectos fundamentais:

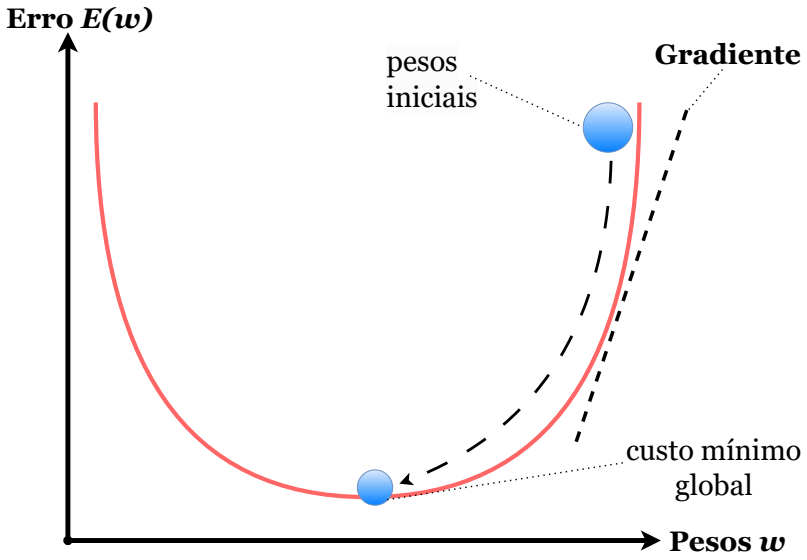
- Redes Neurais Não Recorrentes (*feedforward)*: Esses modelos neurais seguem uma propagação de informações unidirecional e direta, ou seja, os neurônios recebem informações apenas das camadas anteriores. Exemplos incluem redes as *Perceptron* (Rosenblatt, 1958) e *Perceptron* Multicamada (Rumelhart; Hinton; Williams, 1986);

- Redes Neurais Recorrentes (*feedback*): Essas abordagens possuem conexões que permitem o retrocesso de informações, ou seja, as interconexões entre neurônios formam ciclos, permitindo que resultados anteriores influenciem o processamento atual. Esse tipo de arquitetura é adequado para lidar com dados sequenciais ou temporais, como séries temporais e linguagem natural. Tais modelos desenvolvem uma forma de memória interna, como exemplificado pela rede auto-associativa de Hopfield (Hopfield, 1982).

4.3.3 Descida do Gradiente

A regra da descida do gradiente constitui um algoritmo de otimização amplamente aplicado na determinação dos valores dos parâmetros w (peso) e b (bias), comumente referidos como coeficientes. O objetivo principal desta técnica é minimizar uma função de custo associada ao modelo, visando aprimorar o desempenho. A descida do gradiente opera por meio de iterações, nas quais os parâmetros w e b são ajustados incrementalmente na direção do gradiente negativo da função de custo. Esta direção corresponde àquela que proporciona a maior redução na função de custo, conduzindo o modelo a se aproximar progressivamente de uma solução ótima (Bengio, 2012). A Figura 12 ilustra a descida do gradiente com um mínimo global.

Figura 12 - Exemplo de Descida do Gradiente.



Fonte: Autoria própria.

Para alcançar a minimização da função de custo, é comum inicializar os valores dos coeficientes w e b com pequenos valores aleatórios e, em seguida, iterar gradualmente em direção ao mínimo global. Cada passo dessa iteração ou ajuste nos parâmetros representa um avanço no gradiente em direção ao mínimo desejado (Bengio, 2012). Neste processo, a seleção inicial dos pesos desempenha um papel essencial no processo de descida do gradiente. Existem diversas estratégias de inicialização disponíveis, que podem ser aprofundadas na obra de Sousa (2016).

A regra de atualização dos pesos, conhecida como regra delta (Δw), visa ajustar os pesos de modo a minimizar os erros na rede neural. O erro é quantificado como a diferença entre a saída calculada pela rede e a saída desejada. Assim, os pesos são ajustados iterativamente durante o processo de aprendizado, com o objetivo de encontrar uma configuração que minimize o erro em todo o conjunto de treinamento da rede neural (Bengio, 2012). Matematicamente, a regra delta pode ser expressa conforme a equação (14).

$$\Delta w_{ij} = \eta \frac{\partial E}{\partial w_{ij}} \tag{14}$$

Em que:

Δw_{ij}: é a variação no peso da conexão entre os neurônios i e j;

η: simboliza a taxa de aprendizado;

$\frac{\partial E}{\partial w_{ij}}$: corresponde ao gradiente do erro em relação ao peso w_{ij};

E: refere-se a medida do erro entre a saída prevista e a saída desejada.

Durante o treinamento, a regra delta guia a atualização dos pesos na direção que reduz o erro, permitindo que a rede neural aprenda e melhore seu desempenho.

4.3.4 Funções de Ativação

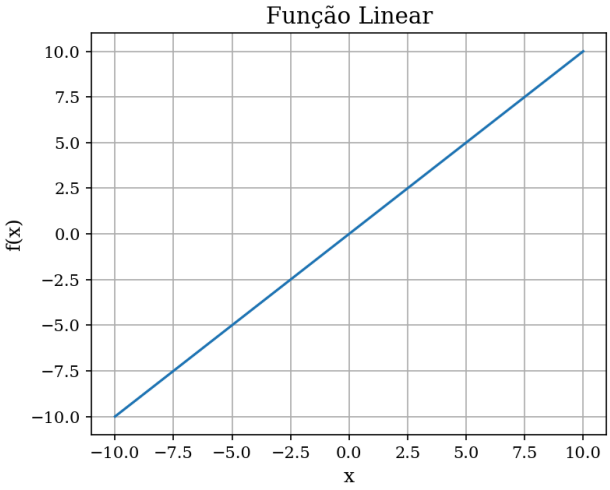
A função de ativação é um componente essencial na estrutura das redes neurais, que exerce influência direta sobre as saídas dos neurônios artificiais. Este elemento determina como os estímulos são processados em cada unidade neural, atuando como um limitador de saída. Nesse sentido, a seleção apropriada da função de ativação é, portanto, de fundamental importância, pois não apenas impacta a capacidade de aprendizado e generalização dos padrões, mas também afeta diretamente a precisão dos dados produzidos pela rede neural (Sharma; Sharma; Athaiya, 2017).

A literatura apresenta uma variedade de funções de ativação, permitindo a seleção adequada de acordo com as exigências específicas de diferentes tipos de problemas. O que possibilita a otimização do desempenho da rede neural para uma ampla gama de aplicações. A escolha apropriada da função de ativação é essencial para garantir que o modelo possa aprender e generalizar de forma eficaz em diversos contextos (Rasamoelina; Adjailia; Sinčák, 2020; Sharma; Sharma; Athaiya, 2017).

No âmbito deste trabalho, é priorizada a análise das principais funções de ativação reconhecidas na literatura. Além disso, são consideradas as funções que se enquadram no escopo definido para esta pesquisa, visando uma investigação alinhada aos objetivos propostos.

A função linear, representada na Figura 13, é caracterizada por preservar inalterada a relação entre as entradas e as saídas. Em outras palavras, não introduz transformações não lineares nos dados processados pelo neurônio. Esse tipo de função é particularmente útil quando se busca uma correspondência direta entre as variáveis de entrada e saída, sem a introdução de transformações não lineares.

Figura 13 - Exemplo de Função de Ativação Linear.



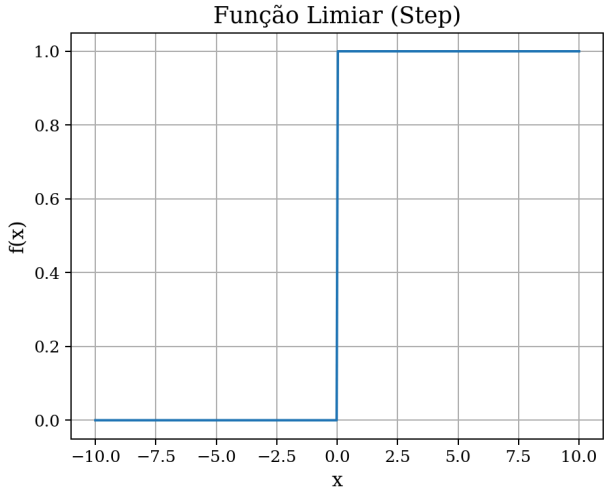
Fonte: Autoria própria.

Matematicamente, a função linear é representada pela equação (15):

$$f(x) = x \tag{15}$$

A função limiar, ilustrada na Figura 14, também conhecida como função de ativação com limite, age como um discriminador binário, determinando a saída com base em uma comparação com um valor limite previamente definido. Este valor limite pode ser ajustado para modificar a sensibilidade do discriminador.

Figura 14 - Exemplo de Função de Ativação Limiar.



Fonte: Autoria própria.

A expressão matemática correspondente a função de ativação limiar é representada pela equação (16):

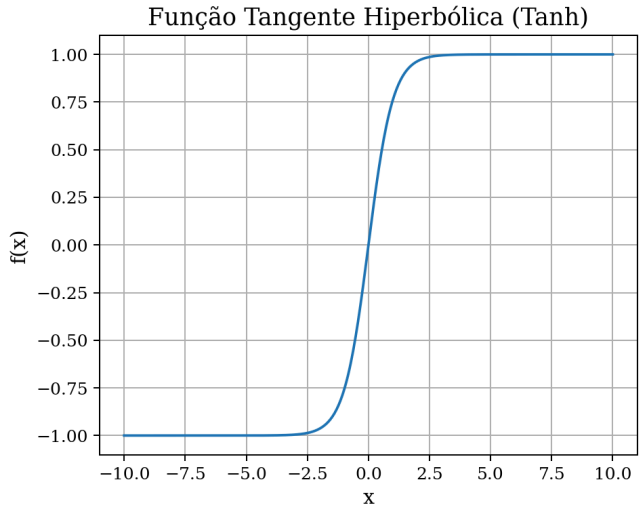
$$f(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases} \tag{16}$$

As funções de ativação que incorporam formas sigmoidais, como a tangente hiperbólica (Tanh) e a logística sigmoide, expressas respectivamente pelas equações (17) e (18), são amplamente utilizadas em redes neurais. A principal distinção entre estas reside no intervalo de valores que produzem como saída. Enquanto a função sigmoide, exemplificada na Figura 16, restringe as saídas a números positivos com sinais contínuos, a tangente hiperbólica permite uma faixa mais ampla de valores, variando entre -1 e 1, tal como retratado na Figura 15. A preferência por essas funções de ativação é motivada principalmente por suas propriedades de diferenciação e não linearidade, essenciais para o funcionamento dos algoritmos de descida do gradiente utilizados no treinamento das redes neurais.

$$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \tag{17}$$

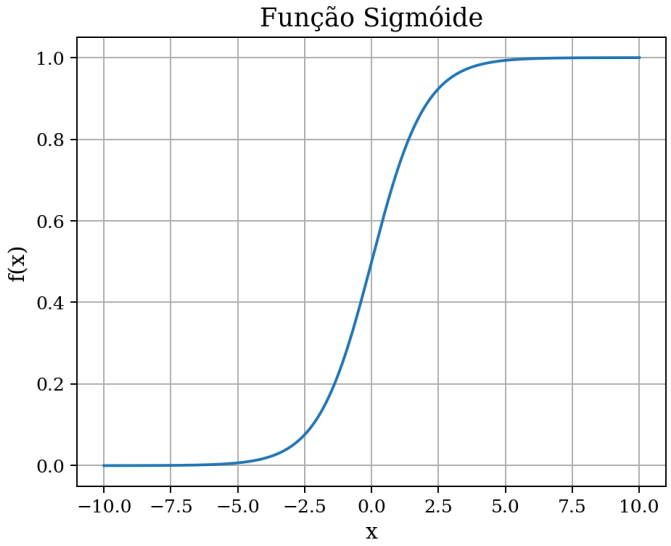
$$f(x) = \frac{1}{1 + e^{-x}} \tag{18}$$

Figura 15 - Exemplo de Função de Ativação Tangente Hiperbólica.



Fonte: Autoria própria.

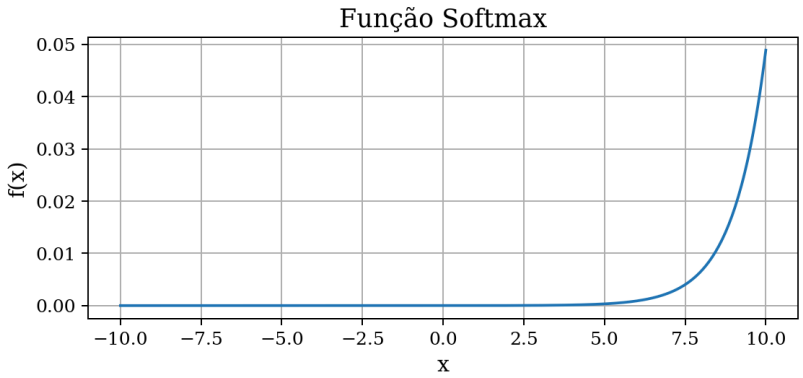
Figura 16 - Exemplo de Função de Ativação Sigmoide.



Fonte: Autoria própria.

Além das funções sigmoide e tangente hiperbólica, a função *softmax*, ilustrada na Figura 17, é outra ferramenta importante em redes neurais, especialmente em tarefas de classificação. Trata-se, portanto, de uma função amplamente aplicada em contextos nos quais se deseja obter uma distribuição de probabilidade sobre várias classes de saída. Ao ser aplicada, as saídas representam a probabilidade de um dado pertencer a cada classe pré-definida. Em outras palavras, essa função garante que a saída da rede consista em uma série de valores numéricos, onde a classe com a probabilidade mais alta é reconhecida como a classe prevista.

Figura 17 - Exemplo de Função de Ativação *Softmax*.



Fonte: Autoria própria.

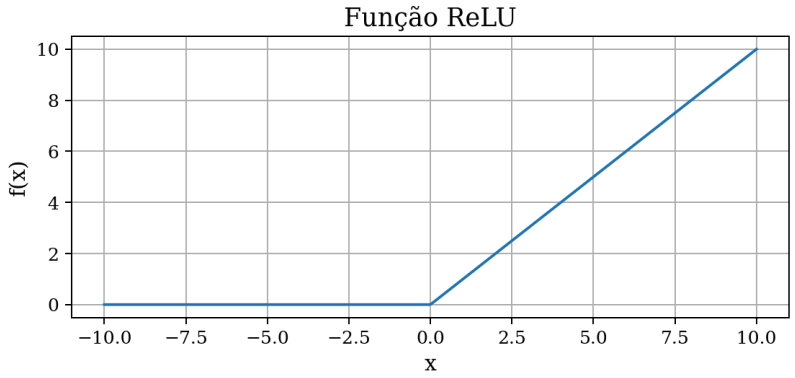
A função de ativação *softmax* é expressa matematicamente pela equação (19):

$$\text{Softmax}(x) = \frac{e^{x_j}}{\sum_{j=1}^{k} e^{x_j}} \tag{19}$$

Em que, a expressão $\text{Softmax}(x)$ denota o vetor resultante da aplicação da função *softmax* a um vetor de entrada x, composto por k elementos. Este vetor de saída representa a distribuição de probabilidade sobre as possíveis classes, onde cada elemento corresponde à probabilidade de que a entrada x pertença à j-ésima classe.

No âmbito do aprendizado profundo, a função de ativação ReLU (*Rectified Linear Unit*), ilustrada na Figura 18, emergiu como uma escolha predominante devido às vantagens que apresenta no processo de treinamento das redes neurais. Tal função é essencialmente linear para valores positivos de x e zero para valores negativos, resultando em uma não linearidade com simplicidade computacional. A preferência pela ReLU decorre, em parte, da propensão das funções de ativação sigmoidais a enfrentarem um desafio conhecido como *"vanishing gradient"* (Pascanu; Mikolov; Bengio, 2013). Esse fenômeno ocorre quando o gradiente tende a zero durante o treinamento, dificultando a atualização dos pesos da rede neural. Por outro lado, a função ReLU demonstra uma capacidade significativa de mitigar essa adversidade. A linearidade da ReLU para valores positivos promove um fluxo mais estável de gradientes durante a retropropagação, contribuindo para uma convergência mais rápida e estável (Nair; Hinton, 2010; Rasamoelina; Adjailia; Sinčák, 2020).

Figura 18 - Exemplo de Função de Ativação ReLU.



Fonte: Autoria própria.

A expressão matemática correspondente a função de ativação ReLU é definida pela equação (20):

$$\text{ReLU}(x) = \begin{cases} 0, & \text{se } x < 0 \\ x, & \text{se } x \geq 0 \end{cases} \tag{20}$$

4.3.5 Treinamento

Dada a necessidade de avaliar quão bem uma rede neural classifica corretamente uma saída, torna-se fundamental medir a disparidade entre a saída prevista pela RNA ($\hat{y}$) e o valor real (y). Essa diferença é quantificada por uma função de perda, descrita como $E(y, \hat{y})$. Tal função objetivo pode ser utilizada para calcular os gradientes descendentes das derivadas das funções de ativação em todas as camadas da rede, permitindo o ajuste dos pesos w de forma que as saídas $\hat{y}$ se aproximem ou correspondam aos valores desejados y. Esse procedimento é conhecido como treinamento neural (Goodfellow; Bengio; Courville, 2016).

A técnica de retropropagação, comumente referida pelo termo em inglês *backpropagation*, emergiu como um dos métodos mais adotados no treinamento das redes neurais artificiais. Essa significativa aceitação é atribuída à capacidade desse algoritmo de ajustar os pesos da rede, aprimorando a aprendizagem e garantindo que o modelo seja capaz de mapear com precisão as entradas para as respectivas saídas em múltiplas dimensões (Chauvin; Rumelhart, 2013). Quando aplicado a uma arquitetura com pelo menos uma camada oculta, este método resulta na abordagem conhecida como MLP (Rosenblatt, 1958), que serve como base para diversos modelos neurais (Haykin, 2009).

No processo de aprendizagem por *backpropagation*, os pesos da rede neural são ajustados iterativamente com base nos erros calculados pela função de perda, que quantifica a disparidade entre as saídas previstas e os valores reais. Esse método é fundamentado na aplicação repetida do que é conhecido como regra da cadeia, tal técnica é utilizada para propagar os erros da camada de saída até as camadas ocultas da rede, permitindo o ajuste dos pesos visando minimizar a imprecisão (Goodfellow; Bengio; Courville, 2016).

Conforme detalhado por Haykin (2009), o treinamento por *backpropagation* em redes neurais artificiais segue uma sequência de etapas bem definidas. A princípio, os pesos são inicializados com média zero e variância apropriada para evitar a saturação das funções de ativação. Durante o treinamento, cada iteração do algoritmo é chamada de época. Em cada época, um conjunto de exemplos de entrada, também conhecidos como *"batches"*, é apresentado à rede neural. A ordem dos *batches* pode ser relevante, especialmente quando a sequência dos dados é importante. Para cada apresentação, os valores de saída são computados através da propagação dos valores para frente (fase *forward*), seguida pela retropropagação dos erros com base na função de perda (fase *backward*).

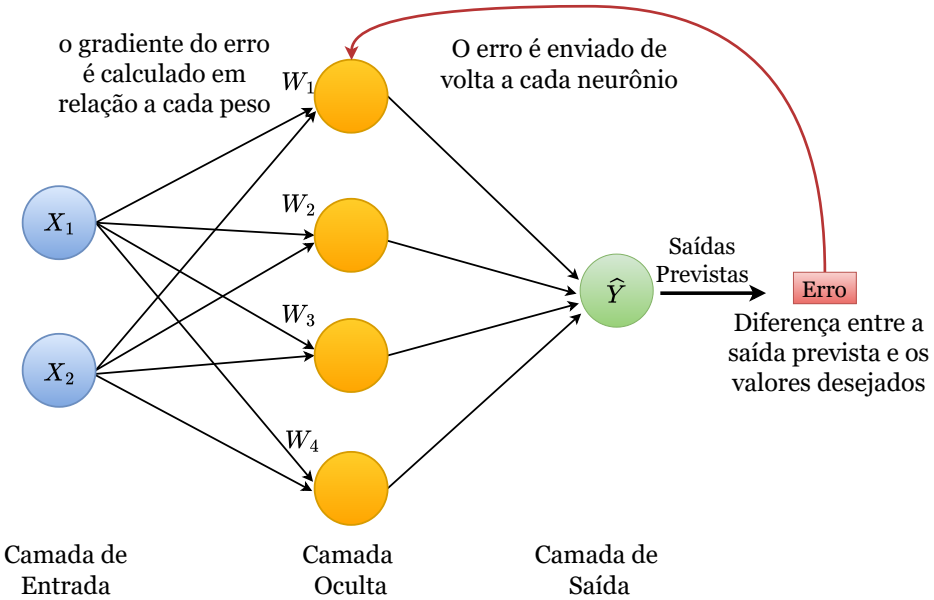
Durante a fase de propagação *forward*, os valores locais de ativação são computados em

cada camada. Este processo implica determinar a atuação de cada neurônio com base nas entradas, nos pesos correspondentes e nas funções de ativação associadas. Cada camada recebe como entrada os sinais provenientes da camada anterior e produz as próprias respostas como saída. Este procedimento é repetido até que as saídas correspondentes à camada final sejam calculadas.

Já na fase *backward*, os gradientes locais são calculados para a camada de saída e, em seguida, propagados de volta para as camadas ocultas. Essa etapa é realizada aplicando a regra da cadeia para calcular o gradiente do erro em relação aos pesos de cada camada. Após esse ciclo, os pesos da rede são atualizados usando uma técnica de otimização, como a regra delta. Esta técnica ajusta os pesos na direção que reduz o erro de previsão, com magnitude proporcional à taxa de aprendizado.

Os processos *forward* e *backward* são repetidos em cada época até que todos os exemplos no conjunto de treinamento tenham sido processados ou até que algum critério de parada seja alcançado. Os critérios de parada comuns incluem a tolerância ao erro, o número máximo de épocas ou a estagnação da convergência, onde o erro de validação não melhora significativamente ao longo das épocas. A Figura 19 ilustra o processo de aprendizagem por *backpropagation*.

Figura 19 - Treinamento por *Backpropagation*.



Fonte: Autoria própria.

4.4 Aprendizado Profundo

Nos últimos anos, o campo da inteligência artificial tem testemunhado um significativo avanço, impulsionado pelo desenvolvimento de novas tecnologias capazes de automatizar uma variedade de tarefas complexas. Tais atividades englobam desde o reconhecimento de padrões em imagens, processamento de fala, tradução de idiomas, diagnósticos médicos, análise preditiva, previsão de demanda, entre outros domínios. Este progresso substancial está diretamente ligado ao advento e ao aprimoramento das redes neurais artificiais de aprendizado profundo, as quais desempenham um papel fundamental em muitas inovações recentes.

Os avanços proporcionados pelos algoritmos de aprendizado de máquina têm sido notáveis em diversas áreas de aplicação, superando abordagens anteriores da inteligência artificial simbólica e probabilística, especialmente na resolução de problemas perceptivos. Contudo, é importante salientar que os algoritmos tradicionais de ML dependem significativamente da representação dos dados, e a seleção apropriada das características de entrada é de suma importância para o desempenho do sistema (Goodfellow; Bengio; Courville, 2016; LeCun; Bengio; Hinton, 2015). Em cenários caracterizados por grandes volumes de dados, como os do *big data*, surgem desafios adicionais na seleção de características, que transcendem a simples escolha de variáveis e podem envolver a inferência de informações a partir de dados não estruturados (Goodfellow; Bengio; Courville, 2016; LeCun; Bengio; Hinton, 2015).

Uma abordagem promissora para lidar com essa problemática é o aprendizado de representação, no qual o ML não apenas identifica o modelo ótimo, mas também a representação mais adequada dos dados. Tal estratégia além de aprimorar o desempenho do sistema, agiliza sua adaptação a novas tarefas, reduzindo a necessidade de intervenção humana. Além disso, algoritmos que incorporam aprendizado de representação tendem a ser mais robustos diante de novos dados, distinguindo automaticamente entre fontes de ruído e informações relevantes (Goodfellow; Bengio; Courville, 2016; LeCun; Bengio; Hinton, 2015).

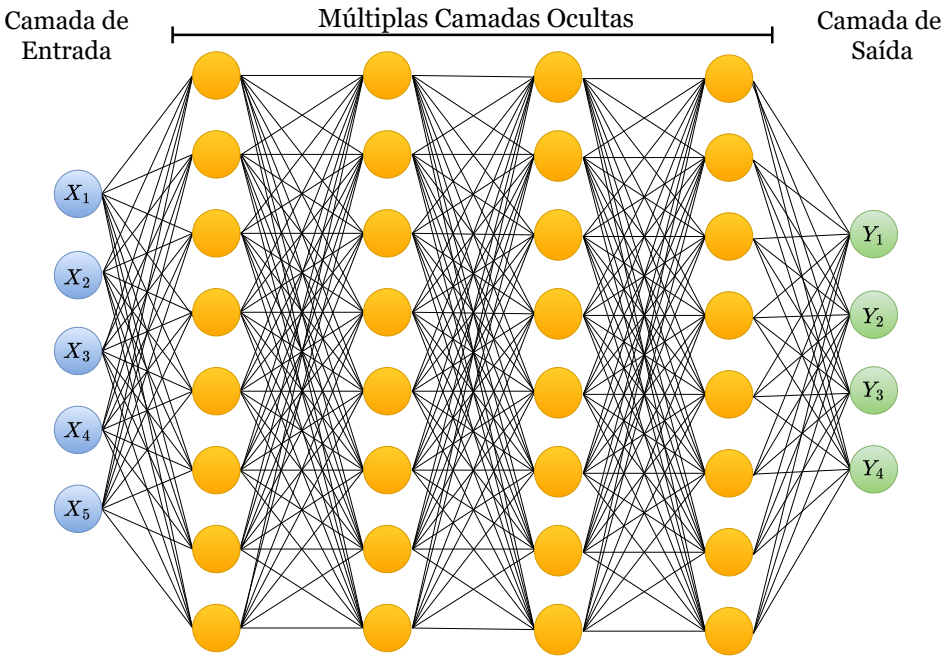
Entretanto, em contextos mais complexos, os *pipelines* multiestágio, que combinam algoritmos de aprendizado de representação com técnicas de ML, enfrentam desafios consideráveis. Nessas situações, o aprendizado profundo, também conhecido como *deep learning* (DL) em inglês, tem emergido como uma abordagem especialmente eficaz, particularmente em cenários nos quais a formalização de regras é impraticável, os conjuntos de dados são predominantemente não estruturados, a identificação de características relevantes é difícil e a gestão de *pipelines* é complexa (Chollet, 2021; Goodfellow; Bengio; Courville, 2016).

A estrutura de uma rede neural convencional compreende uma série de neurônios interconectados que geram ativações numéricas. Enquanto as unidades de entrada captam informações do ambiente através de sensores e são ativadas nesse processo, a ativação dos demais neurônios depende de conexões ponderadas com elementos previamente ativos. Alguns desses elementos

têm a capacidade de influenciar o ambiente, desencadeando ações específicas. Dessa forma, a atribuição de aprendizagem consiste em ajustar os pesos das conexões para que a rede neural demonstre os comportamentos desejados. Dependendo da natureza do problema e da organização das conexões neuronais, tais comportamentos podem exigir uma sequência extensa de etapas computacionais, onde cada estágio transforma, muitas vezes de forma não linear, a ativação geral da rede. O campo do aprendizado profundo se concentra na precisão da atribuição de aprendizagem ao longo desses estágios (Schmidhuber, 2015).

O aprendizado profundo é uma abordagem sofisticada da aprendizagem de máquina que se fundamenta em redes neurais artificiais com múltiplas camadas, conforme ilustra a Figura 20. Esta estrutura complexa atribui ao modelo a habilidade para discernir padrões complexos em conjuntos de dados extensos (Shrestha; Mahmood, 2019). Segundo Haykin (2009), quando o tamanho da camada de entrada é grande, a presença de múltiplas camadas ocultas torna-se uma abordagem especialmente valiosa, uma vez que essas camadas são capazes de extrair estatísticas de ordem superior, melhorando a capacidade de processamento e análise do modelo.

Figura 20 - Arquitetura de uma Rede Neural Profunda.



Fonte: Autoria própria.

Dessa evolução, diversas metodologias inovadoras emergiram, cada uma com suas características distintas e aplicações específicas. Entre essas arquiteturas, destacam-se as CNN (Bhatt *et al.*, 2021), no processamento de dados com estrutura espacial, como imagens. As RNNs (Yu *et al.*, 2019) são amplamente utilizadas em tarefas que envolvem sequências de dados, como processamento de linguagem natural e análise de séries temporais. As Redes Neurais de Grafos

(GNN) (Wu *et al.*, 2020) têm sido aplicadas em problemas que envolvem dados com estrutura de grafo, como modelagem molecular. Por último, os *Transformers* (Vaswani *et al.*, 2017) têm demonstrado eficácia em atividades que requerem o processamento extensivo de dados e a compreensão de relações de longo alcance, como exemplificado notavelmente pelo ChatGPT (*Generative Pre-trained Transformers*) da OpenAI. Os modelos derivados da arquitetura *Transformer* têm se destacado e impactado significativamente o domínio da inteligência artificial, propiciando o desenvolvimento de sistemas mais inteligentes e sofisticados.

Essas arquiteturas representam avanços significativos no campo da inteligência artificial, permitindo a resolução eficiente de problemas que anteriormente eram considerados desafiadores ou até mesmo impossíveis de serem abordados de forma automatizada. O contínuo desenvolvimento e aperfeiçoamento dessas metodologias abrem novas perspectivas e aplicações para a IA em uma ampla variedade de áreas.

4.5 Rede Neural Google *Transformer*

A atenção computacional fundamenta-se na premissa de que os seres humanos gerenciam de maneira eficiente seus recursos cognitivos para realizar tarefas que exigem foco em aspectos específicos da informação. Devido às limitações do cérebro humano, é inviável processar simultaneamente todas as informações disponíveis. Portanto, torna-se necessário concentrar-se em elementos específicos para superar tais gargalos de processamento (Ghazanfar; Schroeder, 2006; Wang; Tax, 2016).

Os modelos computacionais de atenção emergiram nos anos 1980, inicialmente aplicados em imagens e na robótica (Frintrop; Jensfelt, 2008). As Redes Neurais de Atenção, também conhecidas como *Attentional Neural Networks* (ANNs), foram pioneiras na incorporação dos mecanismos de atenção ao campo do aprendizado de máquina. Originalmente desenvolvidas para enfrentar desafios de processamento de linguagem natural (PLN), essas redes desempenharam um papel significativo no avanço dessa área (Galassi; Lippi; Torroni, 2020).

As arquiteturas de sequência a sequência (Seq2Seq) (Sutskever; Vinyals; Le, 2014; Cho *et al.*, 2014), também conhecidas como *encoder-decoder*, foram desenvolvidas inicialmente para abordar problemas de tradução de texto. A introdução de mecanismos de atenção nessa estrutura representou um avanço expressivo, pois permitiu a tradução de sequências mais longas sem perda de contexto ou informação relevante. A atenção proporcionou ao modelo a capacidade de focar em partes específicas da entrada durante o processo de geração da saída, melhorando substancialmente a qualidade das traduções (Galassi; Lippi; Torroni, 2020).

Desde então, novas arquiteturas têm sido desenvolvidas, e esses mecanismos vêm sendo aplicados a uma ampla gama de problemas, conduzido à formalização de diferentes categorias de atenção no campo do DL: *soft attention*, *hard attention* e *self-attention* (Luong; Pham;

Manning, 2015; Vaswani *et al.*, 2017).

- *Soft Attention*: Na técnica de atenção suave, a cada elemento do conjunto de entrada é atribuído um peso variando entre 0 e 1. Tais pesos indicam a quantidade de atenção que a rede neural deve dedicar a cada parte do conjunto. Essa atenção é aplicada tanto espacialmente quanto temporalmente. No aspecto espacial, separa-se as características mais relevantes das menos relevantes. No aspecto temporal, os pesos são ajustados considerando cada janela de tempo à medida que as amostras são apresentadas à rede. Esse processo permite uma avaliação personalizada para cada segmento temporal, pois os elementos de entrada podem ter importâncias diferentes em momentos distintos;
- *Hard Attention*: No método de atenção dura, a função é similar à *soft attention*, porém com uma abordagem diferente. Em vez de atribuir pesos contínuos, os valores são binários, sendo 0 ou 1. Isso implica que a rede determina se foca ou não em cada elemento do conjunto de entrada. Embora esse método reduza o esforço computacional, há uma limitação: nem todas as nuances do conjunto de entrada são capturadas pela atenção, o que pode levar à perda de informações importantes;
- *Self-attention*: Na técnica de autoatenção, além de identificar e ponderar os elementos relevantes em uma amostra, há também uma investigação e ponderação das interações entre esses elementos. Isso permite que as entradas relacionem entre si, contribuindo para uma compreensão mais abrangente do contexto. Esses cálculos são realizados de forma paralela, por meio de matrizes, evitando a necessidade de processamento sequencial.

As diversas categorias de mecanismos de atenção têm sido amplamente adotadas em uma variedade de modelos, impulsionando avanços em diferentes domínios. Entre essas abordagens, destaca-se o *Transformer*. A arquitetura *Transformer* representa um avanço significativo na evolução dos modelos de aprendizado profundo. Em contraste com abordagens tradicionais de transdução de sequência, que geralmente incorporam camadas recorrentes ou convolucionais, o *Transformer* utiliza exclusivamente mecanismos de atenção, estabelecendo assim um novo paradigma no campo da inteligência artificial (Niu; Zhong; Yu, 2021; Chitty-Venkata *et al.*, 2022).

Ao longo dos últimos anos, as redes baseadas em *Transformer* têm despertado um interesse considerável, impulsionado por sua eficácia na captura de dependências de longo alcance nos dados de entrada e por alcançarem desempenho de ponta em uma variedade de tarefas de processamento de linguagem natural. Esta abordagem se beneficia do mecanismo de autoatenção para modelar as relações globais nos dados da sequência de entrada. Em comparação com seu predecessor, o LSTM, o *Transformer*, baseado em autoatenção, demonstra alta escalabilidade, capacidade de modelagem e habilidade de processamento paralelo (Niu; Zhong; Yu, 2021; Chitty-Venkata *et al.*, 2022).

A RNA *Transformer* adota uma arquitetura *encoder-decoder* aprimorada por um sofisticado mecanismo de autoatenção. Portanto, os processos desta abordagem são sustentados por diversas operações, cuja base teórica e matemática é apresentada nesta pesquisa conforme com os trabalhos de Vaswani *et al.* (2017), Lim *et al.* (2021) e Luong, Pham e Manning (2015).

Inicialmente, a sequência de entrada (X), é submetida a uma transformação que resulta na criação de três vetores distintos: o vetor de consulta (Q), o vetor chave (K) e o vetor de valor (V). Essa operação visa a reconfiguração dos elementos do conjunto original em novos espaços, os quais facilitam a aplicação do mecanismo de autoatenção. O vetor de consulta (Q) desempenha o papel de avaliar a relevância dos demais elementos na sequência, mediante a comparação com os vetores chave (K). A relevância calculada, por conseguinte, é utilizada para ponderar os vetores de valor (V), permitindo que o modelo concentre sua atenção nas partes mais significativas do conjunto de entradas. Essa transformação é matematicamente expressa através das equações (21), (22) e (23).

$$Q = XW^{(Q)}, \tag{21}$$

$$K = XW^{(K)}, \tag{22}$$

$$V = XW^{(V)}. \tag{23}$$

Em que:

$X \in \mathbb{R}^{T \times d_x}$: representa a sequência de entrada;

T: simboliza o comprimento da sequência;

d_x: corresponde a dimensionalidade dos vetores de entrada;

$W^{(Q)} \in \mathbb{R}^{d_x \times d_k}$, $W^{(K)} \in \mathbb{R}^{d_x \times d_k}$, e $W^{(V)} \in \mathbb{R}^{d_x \times d_v}$: são as matrizes de pesos;

X: denota o elemento projetado nos espaços dos vetores de consulta, chave e valor;

d_k: refere-se as dimensões dos vetores de consulta e chave;

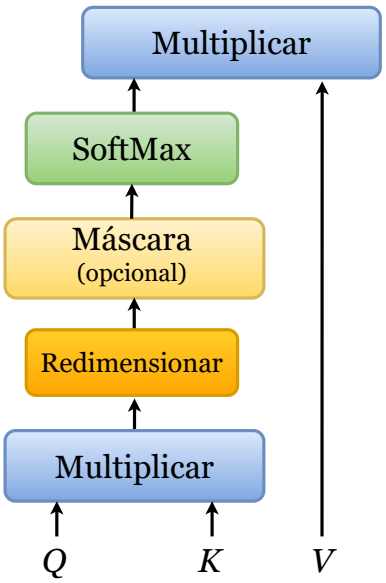
d_v: corresponde a dimensão do vetor de valor.

Posteriormente, conforme descrito na equação (24), os vetores Q e K são combinados através da operação de produto escalar. Os valores resultantes são então normalizados e submetidos à função *softmax*. Assim, o resultado da operação *softmax* é multiplicado pelo vetor V para produzir a saída final.

$$\text{Atenção}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{24}$$

O esquema do mecanismo de autoatenção pode ser observado por meio da Figura 21, que ilustra o processo do produto escalar normalizado realizado pelo modelo *Transformer*. No contexto da referida ilustração, o estágio inicial compreende a operação de produto escalar entre os vetores Q e V. O resultado dessa etapa é encaminhado para o estágio de redimensionamento, onde é dividido por d_k. A etapa de "Máscara"é facultativa, aplicada apenas quando há sequências de comprimentos diversos. Nesses casos, as sequências mais curtas são expandidas até alcançarem o mesmo comprimento das mais longas através do preenchimento com *tokens*. O estágio subsequente, *"Softmax"*, normaliza todos os valores resultantes para o intervalo entre 0 e 1. Posteriormente, ocorre uma nova operação de multiplicação entre o vetor resultante e o vetor V, a fim de realizar uma atualização ponderada dos vetores de valor.

Figura 21 - Mecanismo de Autoatenção.



Fonte: Adaptação (Vaswani *et al.*, 2017).

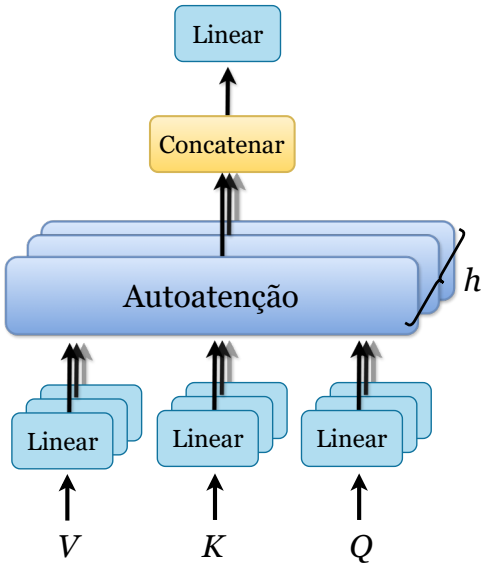
Este processo permite ao modelo atribuir diferentes pesos às diversas partes da entrada, possibilitando a captura de dependências de longo alcance e interações complexas dentro da sequência original. A combinação de autoatenção e transformação linear possibilita que o *Transformer* aprenda as representações e a contextualidade das informações.

Ao identificar as vantagens da projeção linear de vetores com dimensões diversas, Vaswani *et al.* (2017) propuseram uma abordagem denominada autoatenção multicabeça ou autoatenção multifacetada. Essa estratégia possibilita que o modelo processe informações provenientes de

diferentes subespaços de representação em múltiplas posições simultaneamente, permitindo a compreensão de diferentes aspectos das entradas de forma paralela.

No contexto representado pela Figura 22, o mecanismo de autoatenção é executado em paralelo múltiplas vezes, com cada iteração sendo denominada de "cabeça de atenção". As matrizes Q, K e V são particionadas em h partes, correspondendo ao número de cabeças. Em cada cabeça de atenção, é aplicado o produto escalar normalizado e, posteriormente, os resultados são concatenados, permitindo a agregação das informações adquiridas em cada aplicação do produto escalar normalizado.

Figura 22 - Autoatenção Multifacetada.



Fonte: Adaptação (Vaswani *et al.*, 2017).

A camada de autoatenção multicabeça pode ser definida matematicamente conforme a equação (25).

$$\text{MultiCabeça}(Q,K,V) = \text{Concat}(H_1,H_2,\ldots,H_h)W^O \tag{25}$$

Em que:

h: denota o número de cabeças de atenção;

Concat: representa a concatenação dos resultados das h cabeças de atenção;

W^O: corresponde a uma matriz de pesos linear aprendida a partir dos dados;

$W^O \in \mathbb{R}^{h_{dv} \times d_{model}}$.

4.6 *Temporal Fusion Transformer*

O TFT (Lim *et al.*, 2021) é uma arquitetura avançada baseada em autoatenção, uma extensão do modelo *Transformer* (Vaswani *et al.*, 2017). Enquanto o *Transformer* é amplamente reconhecido por sua eficácia em tarefas de processamento de linguagem natural, o TFT expande seu escopo para a análise de séries temporais. Ao empregar o mecanismo de *multi-head attention*, o TFT é capaz de discernir padrões complexos em dados sequenciais, permitindo a captura de relações de longo prazo entre os elementos da série temporal em diferentes escalas de tempo.

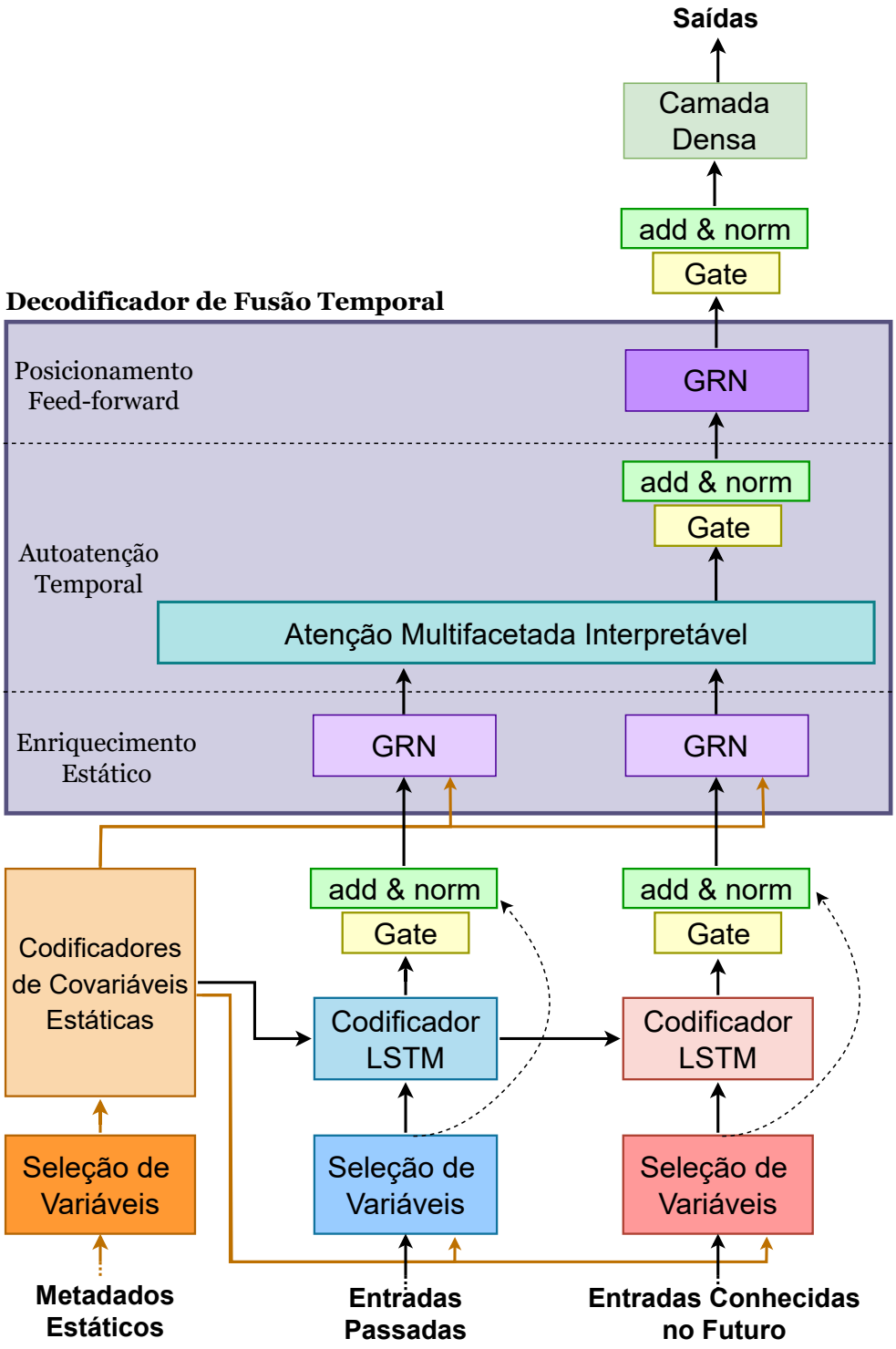
O diferencial do TFT reside em sua habilidade de lidar com séries temporais multi-horizonte (Wen *et al.*, 2017) de forma simultânea, uma capacidade fundamental para enfrentar desafios em diversas áreas, onde a previsão de múltiplos passos à frente é crucial. Ao incorporar esse recurso específico, o TFT oferece uma solução robusta para a modelagem e previsão de séries temporais em diversas aplicações. Além disso, essa arquitetura demonstra flexibilidade nos resultados, fornecendo saídas tanto na forma de valores pontuais quanto na forma de *quantis* probabilísticos, o que é particularmente vantajoso em cenários com incertezas, possibilitando uma análise mais abrangente e informada dos dados (Lim *et al.*, 2021). A Figura 23 ilustra a arquitetura do *Temporal Fusion Transformer*.

Através da Figura 23, é possível visualizar o fluxo de dados no *Temporal Fusion Transformer*, cuja arquitetura integra dados estáticos, além de covariáveis passadas e conhecidas no futuro. Em cada passo de tempo, essas informações são fornecidas a diferentes redes de seleção de variáveis, agrupadas de acordo com os pesos atribuídos a cada tipo de covariável. Essa abordagem é necessária devido à heterogeneidade das informações nas aplicações do mundo real. Frequentemente, em modelos preditivos, os dados estáticos são negligenciados ou combinados com as entradas futuras, o que não permite uma representação adequada da complexidade dos dados temporais (Lim *et al.*, 2021).

Com o intuito de aprimorar a precisão da modelagem, o TFT foi concebido para reconhecer a diversidade das informações e lidar com distintas categorias de dados. Assim, considerando as particularidades temporais, o *Temporal Fusion Transformer* categoriza as covariáveis em três grupos: covariáveis estáticas, covariáveis futuras conhecidas e covariáveis futuras desconhecidas. As covariáveis estáticas consistem em elementos que mantêm seus valores inalterados ao longo do tempo. Por sua vez, as covariáveis futuras conhecidas referem-se a recursos cujos valores variam com o tempo e são conhecidos com antecedência. Por fim, a terceira categoria engloba os dados dependentes do tempo, cujos valores futuros não são conhecidos previamente.

A estrutura do TFT é composta principalmente por cinco componentes fundamentais que desempenham papéis cruciais no modelo, a saber: 1) redes de seleção de variáveis; 2) mecanismos de *gating*; 3) codificadores de covariáveis estáticas; 4) processamento temporal; e 5) definição de intervalos de previsão. A fundamentação teórica e matemática que sustenta cada

Figura 23 - Arquitetura do *Temporal Fusion Transformer*.



Fonte: Adaptação (Lim *et al.*, 2021).

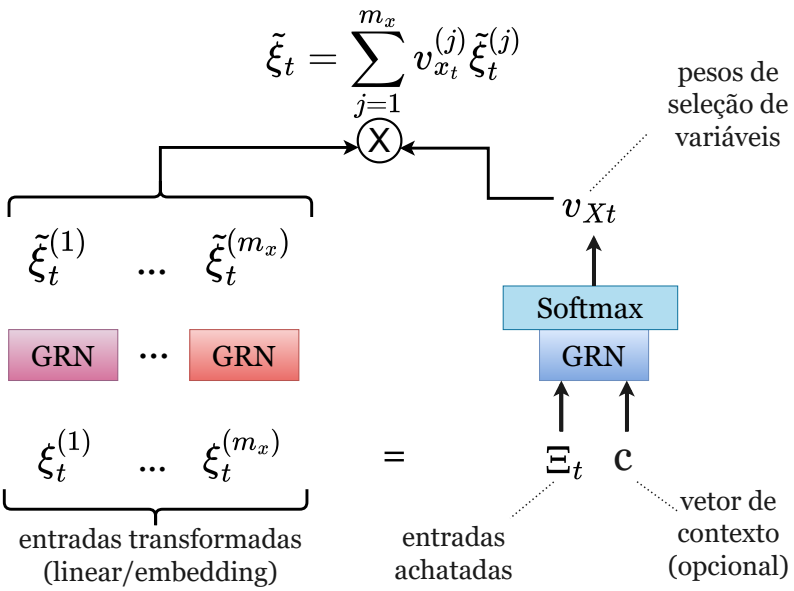
uma dessas técnicas é descrita nesta pesquisa conforme a obra de (Lim *et al.*, 2021).

4.6.1 Redes de Seleção de Variáveis

Este mecanismo desempenha o papel fundamental de selecionar os recursos de entrada mais relevantes para cada intervalo de tempo, permitindo ao modelo descartar entradas ruidosas e irrelevantes que possam impactar negativamente as previsões. Além disso, esta componente possibilita atribuir um grau de contribuição a cada entrada em relação à saída da rede neural, fornecendo informações sobre os elementos mais significativos na geração das previsões, o que permite a identificação de oportunidades de melhoria ou ajustes nos recursos utilizados.

A estrutura das redes de seleção de variáveis, apresentada na Figura 24, utiliza o processo de *embedding* (Gal; Ghahramani, 2016) para variáveis categóricas e transformações lineares para variáveis contínuas. Isso resulta em uma codificação vetorial ξ_t^j de dimensão d_{model}, correspondente ao tamanho dos estados ocultos na arquitetura. Em cada passo de tempo, as entradas transformadas são combinadas linearmente, formando um vetor achatado. Este vetor é então submetido a um mecanismo de bloqueio do tipo GRN (*Gate Residual Network*), seguido por uma camada da função *softmax*, juntamente com o vetor de contexto c obtido pela codificação das covariáveis estáticas.

Figura 24 - Rede de Seleção de Variáveis



Fonte: Adaptação (Lim *et al.*, 2021).

Assim, os pesos que determinam a seleção das variáveis v_{x_t} são calculados, enquanto as características processadas $\tilde{\xi}_t^{(j)}$ passam por sua própria rede neural GRN e os recursos são combinados e acumulados ao longo do tempo, conforme descrito nas equações (26), (27) e (28).

$$v_{x_t} = \text{Softmax}(GRN_{v_x}(\Xi_t, c_s)), \tag{26}$$

$$\tilde{\xi}_t^{(j)} = GRN_{\tilde{\xi}(j)}(\xi_t^{(j)}), \tag{27}$$

$$\tilde{\xi}_t = \sum_{j=1}^{m_x} v_{x_t}^{(j)} \tilde{\xi}_t^{(j)}. \tag{28}$$

4.6.2 Mecanismos de Bloqueio

Os mecanismos de *gating* são elementos do TFT responsáveis por lidar com dois aspectos fundamentais da modelagem preditiva. O primeiro, diz respeito à interação entre as entradas exógenas e a variável-alvo, cuja relação frequentemente é desconhecida, tornando um grande desafio identificar quais recursos de entrada são mais relevantes para a aplicação. O segundo aspecto consiste em determinar o nível de processamento não linear necessário para o problema em questão, identificando situações em que modelos reduzidos podem ser mais vantajosos. Neste contexto, o TFT utiliza dois módulos de bloqueio principais, que desempenham papéis essenciais na adaptabilidade e no controle do fluxo de informações no modelo, sendo: as GLUs (*Gated Linear Units*) e o GRN.

O módulo de *gating* GLU capacita o TFT a regular a propagação dos dados na arquitetura, suprimindo seletivamente certas entradas conforme necessário. Isso permite que o GLU ajuste a influência das entradas exógenas, destacando as variáveis mais relevantes. Esse processo implica na aplicação de uma transformação linear no vetor de entrada, utilizando pesos e vieses, seguida pela função sigmoide para mapear os valores para o intervalo entre 0 e 1. Valores próximos de 0 indicam inibição, enquanto valores próximos de 1 indicam ativação. Posteriormente, ocorre uma segunda transformação linear, seguida pelo produto de Hadamard (Bocci; Carlini; Kileel, 2016), elemento por elemento, conforme representado na equação (29).

$$\text{GLU}\,\omega(\gamma) = \sigma(W_{4,\omega}\gamma + b_{4,\omega}) \odot (W_{5,\omega}\gamma + b_{5,\omega}) \tag{29}$$

Em que:

$\sigma(\cdot)$: corresponde a função de ativação sigmoide;

$W_{(.)}$: refere-se aos pesos;

$W_{(.)} \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$;

$b_{(.)}$: representa os vieses;

$b_{(.)} \in \mathbb{R}^{d_{\text{model}}}$;

d_{model}: simboliza o tamanho dos estados ocultos;

$\odot$: está relacionado ao produto de Hadamard.

As GRNs representam um mecanismo fundamental que confere ao TFT a capacidade de aplicar processamento não linear de maneira adaptativa, conforme necessário. Esse aspecto garante que a rede neural seja capaz de ajustar-se de forma otimizada a diversas condições e conjuntos de dados. As camadas GRN implementam o bloqueio de componentes, suprimindo blocos da arquitetura do TFT que são considerados dispensáveis com base na complexidade da aplicação específica. Nesse processo, as GLUs são empregadas para controlar o fluxo de transformações não lineares. O bloco GRN do TFT recebe duas entradas: uma entrada primária a e um vetor de contexto opcional c. Esse processo segue as equações (30), (31) e (32).

$$\text{GRN}_{\omega}(a,c) = \text{LayerNorm}(a + \text{GLU}_{\omega}(\eta_1)), \tag{30}$$

$$\eta_1 = W_{1,\omega}\eta_2 + b_{1,\omega}, \tag{31}$$

$$\eta_2 = \text{ELU}\left(W_{2,\omega}a + W_{3,\omega}c + b_{2,\omega}\right). \tag{32}$$

Em que:

GLU: corresponde a Unidades Lineares com *Gate* (Dauphin *et al.*, 2017);

LayerNorm: refere-se a uma camada de normalização, conforme (Ba; Kiros; Hinton, 2016);

η_1 e η_2: representa as camadas intermediárias densas;

η_1 e $\eta_2 \in \mathbb{R}^{d_{\text{model}}}$;

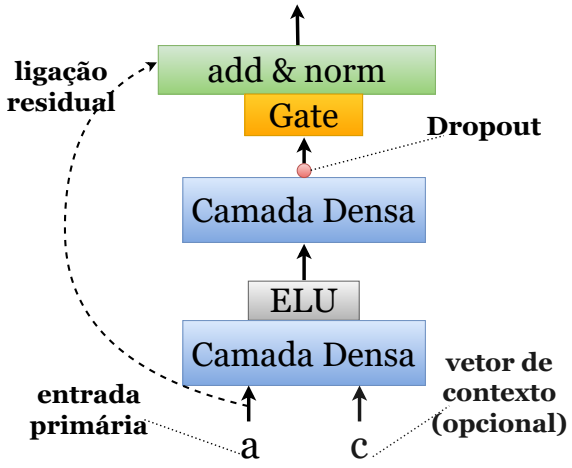
d_{model}: designa o tamanho dos estados ocultos;

ω: simboliza o índice que denota o compartilhamento de pesos;

ELU (*Exponential Linear Unit*): corresponde a função de ativação.

A Figura 25 detalha o mecanismo de bloqueio GRN. No contexto da referida ilustração, a primeira camada densa corresponde a η_2 (32), enquanto a segunda está relacionada a η_1 (31). Durante o treinamento, a técnica de *dropout* (Srivastava *et al.*, 2014) é aplicada antes da camada de normalização (LayerNorm), conforme descrito na equação (30).

Figura 25 - *Gate Residual Network* (GRN).



Fonte: Adaptação (Lim *et al.*, 2021).

4.6.3 *Encoders* de Covariáveis Estáticas

Ao analisar a Figura 23, nota-se que no *Temporal Fusion Transformer*, as informações estáticas são submetidas a um tratamento distinto em comparação com as covariáveis observadas e futuras. Com o objetivo de aprimorar a modelagem do problema e integrar metadados globais em várias partes da estrutura, o TFT emprega codificadores de covariáveis estáticas para produzir quatro vetores de contexto: c_s, c_e, c_c e c_h.

Estes vetores são gerados por diferentes codificadores GRN e alimentam componentes específicos do TFT. Assim, vetor de contexto c_s é utilizado na seleção de variáveis, enquanto o vetor c_e é designado para enriquecer as características temporais com informações estáticas. Os vetores c_c e c_h são aplicados no processamento das características temporais em nível local e são utilizados pelo primeiro codificador LSTM.

4.6.4 Processamento Temporal

Com o propósito de discernir as relações temporais tanto de curto quanto de longo prazo, considerando as entradas observadas e conhecidas ao longo do tempo, o *Temporal Fusion Transformer* emprega principalmente duas camadas para esta finalidade: a camada de atenção multicabeça interpretável e a camada de Seq2Seq.

4.6.4.1 Atenção Multicabeça Interpretável

Através da análise da Figura 23, torna-se evidente que a camada de atenção interpretável de múltiplas cabeças (*interpretable multi-head attention*) assume uma posição central no Decodificador de Fusão Temporal. Este componente do TFT desempenha um papel fundamental

no processo de decodificação, pois é capaz de discernir as relações de longo prazo e os padrões nos dados de entrada. Apresentando uma contribuição significativa desempenho global desta arquitetura.

A técnica de autoatenção, envolve a interação entre os vetores Q, K e V, que representam respectivamente, a consulta, a chave e o valor de cada vetor de incorporação. Nessa abordagem, a saída é obtida por meio da soma ponderada dos vetores de valor, após a realização do cálculo de similaridade que relaciona a consulta e a chave, bem como, após a aplicação da função *softmax*. No que se refere à atenção multicabeça (*multi-head attention*), essa técnica utiliza vários mecanismos de atenção para processar o mesmo conjunto de entradas. Assim, ao empregar diferentes cabeças de atenção com perspectivas distintas, esse método pode processar dados em diversos subespaços e posições, facilitando a identificação de informações relevantes e de padrões de longo prazo. Desta forma, nessa abordagem, os resultados das cabeças de atenção são concatenados e, posteriormente, combinados linearmente através de uma matriz de pesos. Esses mecanismos são detalhados na seção 4.5 deste trabalho.

A atenção multicabeça interpretável (*interpretable multi-head attention*), do TFT, apresenta uma abordagem aprimorada em relação à concatenação dos resultados das cabeças de atenção. Nesse caso, como ilustrado na Figura 26, ao invés de concatenar os resultados, realiza-se a média dos valores obtidos, seguida de uma transformação linear. Além disso, nessa técnica os pesos são compartilhados por todas as cabeças de atenção. Essas melhorias proporcionaram uma maior interpretabilidade e eficiência ao modelo TFT. Nas equações (33), (34), (35) e (36) é possível observar a modelagem deste mecanismo.

$$\text{MultiHeadInterpretável}(Q,K,V) = \tilde{H}W_H, \tag{33}$$

$$\tilde{H} = \tilde{A}(Q,K)VW_V, \tag{34}$$

$$= \frac{1}{m_H}\sum_{h=1}^{m_H} A(QW_Q^{(h)}, KW_K^{(h)})VW_V, \tag{35}$$

$$= \frac{1}{m_H}\sum_{h=1}^{m_H} \text{Atenção}(QW_Q^{(h)}, KW_K^{(h)}, VW_V). \tag{36}$$

Em que:

W_K e W_Q:simboliza os pesos específicos para chave e consulta, respectivamente;

W_K e $W_Q \in \mathbb{R}^{d_{\text{model}} \times d_{\text{attn}}}$;

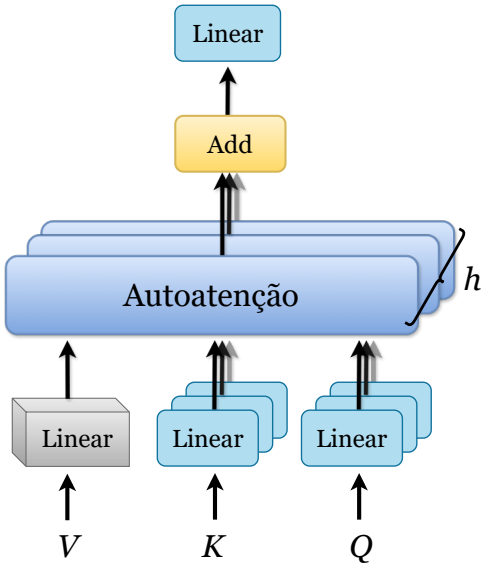
W_V: corresponde aos pesos compartilhados pelas cabeças de atenção;

$W_V \in \mathbb{R}^{d_{\text{model}} \times d_v}$;

W_H: é responsável pelo mapeamento linear;

$W_H \in \mathbb{R}^{d_{\text{attn}} \times d_{\text{model}}}$.

Figura 26 - Atenção Multicabeça Interpretável.



Fonte: Adaptação (Vaswani *et al.*, 2017).

4.6.4.2 Camada de Sequência a Sequência

Nas análises de séries temporais, é recorrente deparar-se com anomalias, padrões cíclicos, pontos de transição e outros comportamentos que se destacam em relação aos valores adjacentes. A captura destas informações contextuais é crucial para a otimização do desempenho de estruturas fundamentadas em mecanismos de atenção. Neste contexto, a camada de Seq2Seq, implementada no TFT, emprega um bloco de codificador-decodificador baseado em unidades de LSTM, visando identificar e capturar padrões locais presentes nos dados de entrada. Além disso, este mecanismo pode ser considerado como uma alternativa à codificação posicional convencional, uma vez que proporciona informações sobre a ordenação temporal dos dados integrados à rede neural.

A camada Seq2Seq adota uma estrutura de LSTM *encoder-decoder* para examinar padrões locais nos dados de entrada, gerando atributos que alimentam o decodificador de fusão temporal, $\tilde{\phi}(t,n)$. Para incorporar metadados estáticos, vetores de contexto são empregados para inicializar os estados da célula e o estado oculto da primeira camada LSTM. Durante esse processo de transformação da sequência original, ocorre normalização, seleção de uma parte específica da

entrada $\tilde{\xi}_{t+n}$ e aplicação de uma unidade GLU. Essa operação é detalhada conforme a equação (37).

$$\tilde{\phi}(t,n) = \text{LayerNorm}(\tilde{\xi}_{t+n} + \text{GLU}_{\tilde{\phi}}(\phi(t,n))). \tag{37}$$

4.6.4.3 Camadas Adicionais

O decodificador de fusão temporal não apenas implementa as camadas de *multi-head attention* interpretável e de sequência a sequência, mas também integra outros três estágios com o objetivo de otimizar o desempenho do modelo. Estes estágios adicionais são projetados para habilitar a detecção e a internalização das relações temporais presentes nos dados de entrada, compreendendo a camada de enriquecimento estático, a camada de auto-atenção temporal e a camada de *position-wise feed-forward*.

Inicialmente, a camada de enriquecimento estático é aplicada com o objetivo de aprimorar os aspectos temporais através da adição de informações estáticas. Nesse contexto, para cada índice de uma posição específica n, essa melhoria assume a estrutura apresentada na equação (38).

$$\theta(t,n) = \text{GRN}_{\theta}(\tilde{\phi}(t,n), c_e) \tag{38}$$

Em que:

GRN_{θ}: corresponde aos pesos utilizados globalmente dentro da camada;

c_e: representa um vetor contextual proveniente de um codificador estático.

Na sequencia, uma camada de auto-atenção temporal é utilizada para capturar as relações de longo prazo, processando tanto as características cronológicas quanto os recursos estáticos por meio do módulo de atenção multicabeça interpretável. Conforme expresso na equação (39).

$$B(t) = \text{MultiCabeçaInterpretável}(\Theta(t), \Theta(t), \Theta(t)) \tag{39}$$

Em que, $\Theta(t)$ corresponde uma matriz que agrupa os recursos temporais enriquecidos com informações estáticas.

Neste processo, o módulo de atenção multicabeça interpretável é designado a cada instante de previsão, com $N = \tau_{\max} + k + 1$. Além disso, após a etapa de autoatenção, outra camada de bloqueio é adicionada para simplificar o processo de treinamento, conforme descrito na equação (40).

$$\delta(t,n) = \text{LayerNorm}(\theta(t,n) + \text{GLU}_{\delta}(\beta(t,n))). \tag{40}$$

Por fim, a camada de *Position-wise Feed-forward* introduz um processamento não linear adicional às saídas da camada de autoatenção, de maneira análoga à camada de enriquecimento estático. Conforme descrito na equação (41).

$$\psi(t,n) = \text{GRN}_{\psi}(\delta(t,n)) \tag{41}$$

Onde, GRN_{ψ} corresponde aos os pesos de GRN que são compartilhados por todas as camadas.

Nesta etapa, uma conexão residual interna também é utilizada, visando criar um atalho direto para a camada de sequência a sequência e assim, evitar uma complexidade adicional quando não é necessário, resultando em um modelo mais simplificado. Esse processo é exemplificado na equação (42).

$$\tilde{\psi}(t,n) = \text{LayerNorm}(\tilde{\phi}(t,n) + \text{GLU}_{\tilde{\psi}}(\psi(t,n))) \tag{42}$$

Essas camadas desempenham um papel essencial no aprimoramento das informações temporais, capacitando a rede neural a absorver informações estáticas e compreender de maneira mais abrangente a dimensão temporal.

4.6.5 Intervalos de Previsão

O TFT tem a capacidade de criar possibilidades de probabilidade em vez de simplesmente fornecer valores de previsão específicos. Nesse modelo, a previsão quantílica é empregada para estabelecer faixas de valores prováveis para a variável-alvo ao longo do horizonte de previsão. Assim, em cada ponto no tempo, o modelo pode prever simultaneamente valores para diferentes percentuais. Desta forma, as variações quantílicas são produzidas através de uma transformação linear aplicada à saída do decodificador de fusão temporal, conforme expresso na equação (43).

$$\hat{y}(q,t,\tau) = W_q \tilde{\psi}(t,\tau) + b_q, \tag{43}$$

Em que:

$W_q \in \mathbb{R}^{1\times d}$ e $b_q \in \mathbb{R}$: são coeficientes lineares ajustados para um *quantil* específico q.

Neste contexto, no processo de treinamento TFT, o objetivo é minimizar a perda quantílica, que é calculada pela soma das discrepâncias entre as variações e os valores reais em todos os

quantis. Esse processo é expresso matematicamente por meio das equações (44) e (45).

$$\mathscr{L}(\Omega, W) = \sum_{y_t \in \Omega} \sum_{q \in \mathscr{Q}} \sum_{\tau=1}^{\tau_{\max}} \frac{QL(y_t, \hat{y}(q, t-\tau, \tau), q)}{M_{\tau_{\max}}} \tag{44}$$

$$QL(y, \hat{y}, q) = q(y - \hat{y})_+ + (1-q)(\hat{y} - y)_+ \tag{45}$$

Em que:

Ω: corresponde ao domínio dos dados de treinamento;

M: denota a quantidade de amostras;

W: representa os pesos da arquitetura;

$\mathscr{Q}$: simboliza o conjunto de *quantis* de saída.

4.7 Síntese do Capítulo

Este capítulo explora os conceitos essenciais das redes neurais de aprendizado profundo, abordando tópicos fundamentais para a compreensão dos métodos aplicados nesta pesquisa. As subseções apresentadas destacam a relevância das redes neurais no campo da inteligência artificial e contextualizam os objetivos deste estudo. Além disso, discutem o aprendizado de máquina e suas abordagens, com ênfase no aprendizado supervisionado.

Na subseção sobre redes neurais artificiais, detalham-se o neurônio artificial, a arquitetura das redes, a descida do gradiente, as funções de ativação e o processo de treinamento, componentes essenciais para o funcionamento das redes neurais. O aprendizado profundo é apresentado como um subcampo que utiliza redes neurais com múltiplas camadas para modelar complexidades inerentes aos dados, diferenciando-se das demais abordagens de aprendizado de máquina.

A subseção sobre Redes Neurais Google *Transformer* enfatiza as inovações significativas no processamento de sequências, detalhando os mecanismos de atenção. Ao final do capítulo, o TFT é abordado no campo da previsão temporal. Em suma, este capítulo fornece uma visão detalhada e abrangente das redes neurais de aprendizado profundo, estabelecendo uma base abrangente para a compreensão das técnicas discutidas nesta pesquisa.

5 PREVISÃO DE DEMANDA ELÉTRICA COM TFT

Este capítulo delineia a metodologia adotada neste estudo para a previsão de demandas elétricas utilizando a rede neural TFT. Portanto, esta seção visa contextualizar os procedimentos metodológicos aplicados, proporcionando uma compreensão aprofundada das técnicas e conceitos subjacentes ao desenvolvimento dos métodos propostos nesta pesquisa.

Na primeira Subseção, 5.1, discute-se a utilização de múltiplas séries temporais no contexto dos sistemas de energia elétrica. Na Subseção 5.2, são descritas as bases de dados utilizadas, com foco nas características específicas de cada conjunto de dados e no processo de pré-processamento necessário. A Subseção 5.3 aborda a escolha do modelo de rede neural empregado, detalhando as configurações do algoritmo e abordando aspectos como a seleção de hiperparâmetros.

A Subseção 5.4 trata dos sistemas de previsão desenvolvidos, incluindo a integração de múltiplas séries temporais e a modelagem matemática utilizada para representar as dinâmicas do consumo de energia no contexto da previsão. Na Subseção 5.5, são descritos os procedimentos e definições do arranjo experimental, abordando a divisão dos dados, utilização de variáveis exógenas, critérios de avaliação, processo de treinamento do modelo e o ambiente de desenvolvimento utilizado na implementação do TFT, fornecendo detalhes sobre os recursos computacionais empregados. Finalmente, a Subseção 5.6 sintetiza os pontos principais discutidos no capítulo, reforçando os conceitos metodológicos apresentados.

5.1 Introdução

As STs no setor energético são caracterizadas por interdependências dinâmicas, que abrangem desde o consumo de energia até as condições climáticas. Neste contexto, os dados de demanda frequentemente apresentam correlações. Por exemplo, em uma rede inteligente (*smart grid*), diferentes consumidores dentro de uma mesma área geográfica, mesmo que possuam características distintas, podem ter padrões de consumo que se influenciam mutuamente, especialmente durante eventos específicos, como ondas de calor ou feriados. Essas correlações são essenciais para as previsões de demanda, pois representam as interações e os impactos mútuos entre as variáveis que influenciam diretamente a dinâmica do consumo de energia na região, permitindo uma modelagem mais precisa das condições futuras.

Tradicionalmente, as previsões de demanda são realizadas de forma isolada, exigindo a configuração de um modelo exclusivo para cada série temporal. Isso implica na definição de parâmetros específicos e ajustes personalizados para cada conjunto de dados individualmente, aumentando significativamente o tempo necessário para a modelagem. A personalização e otimização de um modelo para cada série temporal demandam consideráveis recursos, tornando o

processo de previsão mais complexo e prolongado. Além disso, essa abordagem pode limitar a precisão das estimativas devido à falta de integração e análise conjunta das diferentes séries temporais que compõem o sistema.

A utilização de múltiplas séries temporais na previsão de consumo de energia apresenta vantagens significativas. Primeiramente, ao considerar simultaneamente diversos conjuntos de dados relacionados, os modelos preditivos conseguem capturar de forma mais eficaz os padrões temporais e as relações não lineares entre as variáveis. Além disso, a integração de múltiplas séries temporais otimiza o uso dos dados disponíveis, ampliando a quantidade de informações extraídas de cada variável. Outro aspecto importante reside na simplificação do processo de modelagem, pois a adoção de um modelo integrado reduz consideravelmente a complexidade envolvida.

Partindo da premissa de que as correlações são essenciais para projeções mais precisas, este estudo sugere que a PMST oferece uma perspectiva mais vantajosa para os sistemas de energia. Ao integrar as séries temporais correlacionadas no sistema elétrico, esta pesquisa propõe que tal abordagem captura as interdependências entre variáveis e melhora a capacidade de resposta do modelo a variações e tendências. Assim, a combinação dos dados integrados de consumo com técnicas avançadas, como o *Temporal Fusion Transformer*, tem o potencial de melhorar significativamente a precisão das previsões, contribuindo para uma gestão mais eficaz e uma operação mais robusta dos sistemas de energia.

5.2 Descrição dos Dados

Para evidenciar os resultados desta pesquisa foram conduzidos testes abrangentes em diferentes níveis do sistema elétrico de potência. Para isso, utilizou-se duas bases de dados distintas, uma referente ao nível de distribuição e outra em nível de transmissão. Esta abordagem visa validar de forma ampla os resultados obtidos, demonstrando a aplicabilidade do método proposto em diferentes contextos da infraestrutura elétrica. Portanto, os dados utilizados foram obtidos de duas fontes principais: a Base de Dados A, referente ao nível de distribuição; e a Base de Dados B, que corresponde ao nível de transmissão. Ambas incluem dados sobre o consumo de energia e variáveis exógenas que refletem o comportamento das cargas.

5.2.1 Base de Dados A

A Base de Dados A (Monteiro *et al.*, 2024) contém informações sobre o consumo de energia de 172 edifícios localizados na região de Loureiro, Portugal. Estas habitações são propriedade de uma cooperativa local e estão situadas em proximidade geográfica umas das outras. Os dados foram coletados utilizando medidores inteligentes (*smart meters*), que registraram a demanda energética a cada 15 minutos ao longo do período de maio de 2022 a setembro de 2023. O

conjunto de dados é composto por um total de 46,608 registros, onde cada registro corresponde a um edifício específico.

Além das medições de consumo de energia, a Base de Dados A também inclui dados meteorológicos coletados com a mesma frequência e intervalo de tempo. Essas informações foram obtidas da estação meteorológica de Aveiro, localizada nas proximidades de Loureiro. As colunas disponíveis no conjunto de dados abrangem:

- Data e hora da medição registrada;
- Consumo de energia em intervalos de 15 minutos (kWh);
- Temperatura média do ar a 1,5 metros acima do solo (°C);
- Umidade relativa média (%);
- Direção média do vento (de 0 a 360°);
- Velocidade média do vento (m/s);
- Velocidade instantânea máxima do vento (m/s);
- Temperatura instantânea do ar a 1,5 metros acima do solo (°C);
- Quantidade de precipitação (mm);
- Intensidade máxima de precipitação instantânea (mm/h);
- Radiação solar global total (KJ/m2).

Alguns registros apresentam valores ausentes, identificados como "NaN"(*Not a Number*), que necessitaram de tratamento durante a análise e modelagem dos dados. A coleta simultânea dos dados energéticos e meteorológicos resultou em conjuntos de dados com o mesmo número de linhas e datas correspondentes, permitindo uma análise integrada e contextualizada do consumo de energia em relação às condições climáticas.

5.2.2 Base de Dados B

A Base de Dados B (CDS, 2010) contém informações históricas de consumo de energia de um subsistema de distribuição de energia elétrica localizado na região de Waikato, Nova Zelândia. Este subsistema abrange duas usinas termoelétricas e sete subestações. Os dados abrangem o período de janeiro de 2007 a março de 2009, totalizando 39,408 registros coletados a cada 30 minutos. Além das informações sobre a demanda de energia, o conjunto de dados também inclui variáveis exógenas, tais como:

- Hora;
- Dia da semana;
- Mês;
- Ano;
- Feriados;
- Horário de verão.

A base de dados continha alguns registros não preenchidos, os quais precisaram ser ajustados durante a análise e elaboração do modelo.

5.2.3 Pré-processamento dos Dados

O pré-processamento dos dados envolveu várias etapas para garantir a qualidade e a integridade das informações utilizadas nos modelos preditivos propostos:

- Tratamento de Valores Ausentes: Para lidar com valores ausentes identificados como "NaN", utilizou-se a técnica de interpolação das observações adjacentes, assegurando a continuidade das séries temporais;
- Normalização: Todas as séries temporais foram normalizadas para o intervalo [0, 1], facilitando o treinamento do modelo. É relevante mencionar que essa etapa foi executada pela própria rede neural TFT, garantindo a consistência no processo de preparação dos dados;
- Análise Exploratória: Gráficos de série temporal, histogramas e análises de correlação foram gerados para uma melhor compreensão das características e das inter-relações dos dados.

5.3 Arquitetura Neural Utilizada

Neste estudo, a escolha do TFT como modelo preditivo é justificada por sua capacidade de integrar de forma coordenada múltiplas fontes de dados. O TFT inclui uma camada de fusão temporal que permite a combinação de informações provenientes de diferentes séries temporais e de variáveis exógenas distintas em um único vetor de características. Além disso, essa rede neural demonstra a habilidade de lidar com a complexidade e a natureza não linear das séries temporais, especialmente em cenários com significativa influência de variáveis exógenas sobre a variável-alvo, como é comum em sistemas de energia.

Assim, ao escolher o TFT como modelo para a previsão de múltiplas séries temporais, esta pesquisa não apenas visa explorar suas capacidades avançadas de modelagem, mas também destacar suas vantagens distintivas na abordagem dos complexos desafios que envolvem as projeções da demanda energética. O TFT representa uma fusão poderosa entre redes de atenção e redes neurais recorrentes, proporcionando uma estrutura robusta para capturar dependências temporais e interações entre variáveis. Detalhes sobre a arquitetura do *Temporal Fusion Transformer* podem ser encontrados na Subseção 4.6.

5.3.1 Configurações do Modelo

A rede neural TFT depende da configuração de diferentes hiperparâmetros para atingir um desempenho ótimo. Portanto, a compreensão detalhada e o ajuste preciso dos hiperparâmetros apresentados a seguir são essenciais para maximizar a performance deste modelo.

1. Taxa de Aprendizado (*Learning Rate*): Controla a magnitude dos ajustes nos parâmetros do modelo durante o processo de otimização. Uma taxa de aprendizado baixa pode resultar em ajustes mais precisos, pois cada passo de atualização é pequeno, o que pode ser vantajoso para a convergência estável do modelo. No entanto, isso também pode aumentar o tempo necessário para o treinamento. Por outro lado, uma taxa de aprendizado alta pode acelerar o treinamento, mas corre o risco de oscilações ou até mesmo de falhas na convergência para uma solução ótima;

2. Tamanho dos Neurônios Ocultos (*Hidden Size*): Este parâmetro define o número de neurônios em cada camada oculta da arquitetura. A escolha do *Hidden Size* afeta diretamente a capacidade de aprendizado do modelo. Um tamanho maior pode aumentar a capacidade de representação do modelo, permitindo-lhe aprender padrões complexos nos dados, mas também pode aumentar o risco de *overfitting* se não for ajustado com técnicas adequadas de regularização;

3. Tamanho das Cabeças de Atenção (*Attention Head Size*): As cabeças de atenção determinam quantas maneiras diferentes o modelo pode focar simultaneamente em diferentes partes dos dados durante a etapa de atenção. Um número maior de cabeças de atenção pode aumentar a capacidade do modelo de capturar correlações complexas e padrões interdependentes nos dados, melhorando assim o desempenho em tarefas que exigem modelagem de longo alcance e dependências contextuais;

4. Taxa de *Dropout* (*Dropout*): A técnica de *dropout* é usada para regularizar o modelo, ajudando a reduzir o *overfitting* ao desativar aleatoriamente um percentual de neurônios durante o treinamento. Isso força o modelo a não depender excessivamente de nenhum conjunto específico de neurônios, tornando-o mais robusto e capaz de generalizar melhor para novos dados;

5. Tamanho dos Neurônios Contínuos Ocultos (*Hidden Continuous Size*): O tamanho dos neurônios ocultos contínuos define o espaço dimensional das representações vetoriais das variáveis contínuas no modelo. Um tamanho maior pode capturar nuances mais detalhadas nos dados contínuos, possibilitando uma modelagem mais refinada e previsões mais precisas;

6. Função de Perda (*Loss Function*): Quantifica quão bem o modelo está realizando a tarefa desejada durante o treinamento ao comparar as previsões feitas pelo modelo com os valores reais dos dados. Esta métrica descreve o quão distantes as previsões do modelo estão dos valores verdadeiros e é fundamental na direção da atualização dos pesos (ou parâmetros do modelo) durante o processo de otimização;

7. Valor de *Clipping* de Gradiente (*Gradient Clip Value*): Este parâmetro define um limite superior para os gradientes calculados durante o processo de retropropagação (*backpropagation*) e é essencial para prevenir a ocorrência de explosões de gradientes. Em modelos de aprendizado profundo ou complexos, os gradientes podem aumentar rapidamente de magnitude, o que pode causar instabilidades durante o treinamento e prejudicar a convergência do modelo;

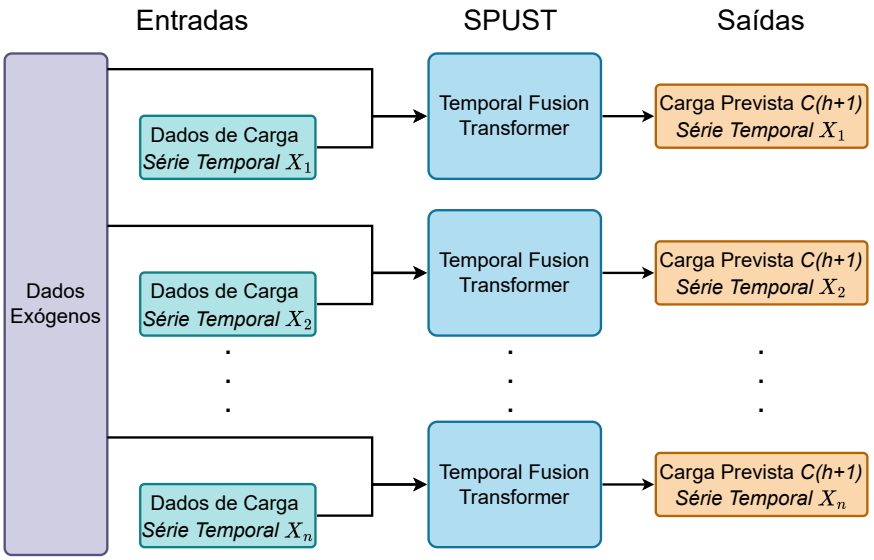
8. Tamanho do Lote (*Batch Size*): Define quantas amostras são processadas simultaneamente pelo modelo em cada iteração durante o treinamento. Um *Batch Size* maior pode acelerar o treinamento aproveitando melhor os recursos computacionais disponíveis, enquanto um tamanho menor pode permitir atualizações mais frequentes dos pesos do modelo e, às vezes, pode ajudar na convergência;

9. Épocas (*Epochs*): Indica quantas vezes o modelo percorre integralmente o conjunto de dados de treinamento. O treinamento do modelo por múltiplas épocas permite que ele aprenda continuamente a partir dos dados, aprimorando gradualmente seu desempenho até que não ocorram mais melhorias significativas.

5.4 Desenvolvimento dos Modelos de Previsão Propostos

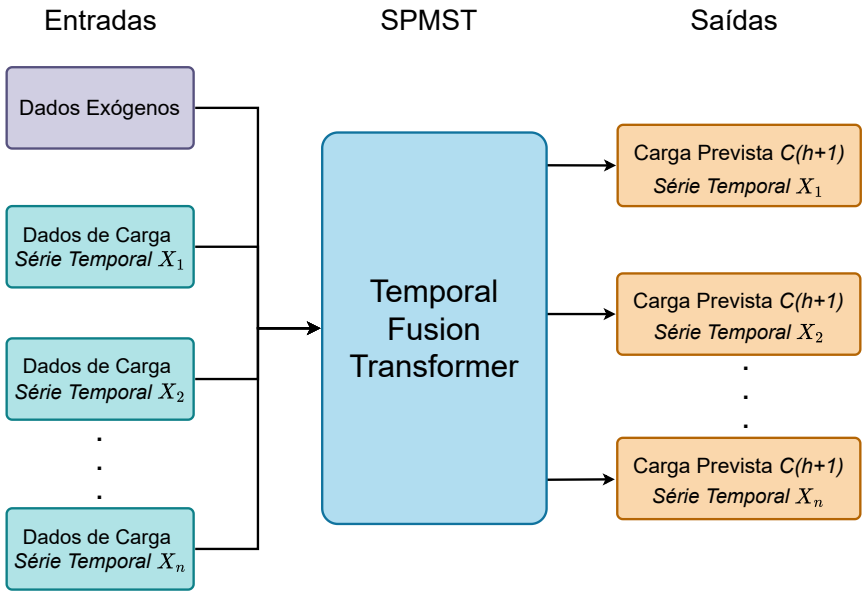
Neste estudo, foram elaborados dois sistemas de previsão baseados em RNA do tipo TFT. O primeiro, denominado SPUST, foi concebido para prever o comportamento futuro de uma única sequência de dados ao longo do tempo. Em contraste, a segunda abordagem, denominada SPMST, foi projetada para integrar múltiplas séries temporais em um modelo unificado. A estrutura do SPUST está ilustrada na Figura 27, enquanto a arquitetura do SPMST é apresentada na Figura 28.

Figura 27 - Sistema Previsor de Única Série Temporal (SPUST).



Fonte: Autoria própria.

Figura 28 - Sistema Previsor de Múltiplas Séries Temporais (SPMST).



Fonte: Autoria própria.

O SPUST, representado na Figura 27, concentra-se exclusivamente nas projeções isoladas obtidas para cada série temporal. Nesse processo, os módulos do SPUST recebem como entrada informações exógenas e os dados referentes às cargas elétricas. Em seguida, um bloco da Rede Neural Artificial *Temporal Fusion Transformer* realiza o processamento individualizado

para cada série temporal. Essa abordagem permite a geração de resultados baseados na análise específica de cada sequência histórica.

Já o SPMST, ilustrado na Figura 28, foi desenvolvido com o propósito de integrar diversas STs em um único modelo, visando explorar as interações entre diferentes conjuntos de dados e melhorar a precisão das previsões. Para implementar o SPMST, adotou-se uma estratégia de fusão de dados na qual cada sequência temporal foi tratada como uma variável de entrada no modelo TFT. Esta abordagem permite que o SPMST utilize informações exógenas e um conjunto de séries temporais como entrada simultaneamente. Assim, os dados são processados de forma integrada pela rede neural TFT, que gera previsões individuais para cada ST.

Os dois sistemas de previsão desenvolvidos, SPUST e SPMST, apresentam abordagens distintas para a previsão de séries temporais no contexto de sistemas de energia. Enquanto o SPUST se concentra exclusivamente em uma única série temporal, o SPMST utiliza informações provenientes de múltiplas séries temporais. Ao explorar essas abordagens distintas, este estudo não apenas compara o desempenho preditivo dos dois sistemas, mas também investiga como a integração de múltiplas séries temporais pode influenciar a precisão das previsões, identificando as vantagens e limitações específicas de cada método no contexto das projeções de demandas elétricas.

5.4.1 Modelagem Matemática da Previsão

Considerando que as amostras de dados foram obtidas em intervalos específicos de tempo, as previsões correspondem a instantes temporais denotados por h, onde cada h representa um ponto de dados na série temporal. Por exemplo, na Base de Dados A, as informações foram coletadas a cada 15 minutos, enquanto na Base de Dados B, os intervalos de coleta foram de 30 minutos. Neste contexto, a projeção da carga elétrica para o próximo momento após h é representada por $C(h+1)$. Assim, para realizar as previsões, foram utilizados três conjuntos de informações:

1. X_{carga}: Descreve as propriedades da carga elétrica até o momento h;
2. $X_{\text{exógeno}}$: Engloba as características externas que podem influenciar a carga elétrica, como variáveis meteorológicas, eventos sazonais, entre outros;
3. t: Contém informações temporais dos dados históricos até h, possibilitando ao modelo entender padrões sazonais, tendências ou ciclos.

Desta forma, a função de previsão f_{TFT} utiliza dados históricos de demanda X_{carga}, informações externas $X_{\text{exógeno}}$, e a dimensão temporal t para modelar e prever a carga elétrica no

próximo intervalo de tempo $(h+1)$ após h. Matematicamente, essa função pode ser expressa da seguinte forma:

$$C(h+1) = f_{TFT}(X_{\text{carga}}, X_{\text{exógeno}}, t). \tag{46}$$

No contexto do SPMST, embora a função de previsão seja idêntica, é importante considerar as seguintes definições:

1. Séries Temporais (ST): Sejam $S_1, S_2, \ldots, S_n$ as múltiplas séries temporais a serem previstas. Cada série S_i é composta por observações ao longo do tempo, representadas como $S_i = \{s_{i1}, s_{i2}, \ldots, s_{iT}\}$, onde T é o número de pontos temporais na série;

2. Dados Exógenos: Sejam $x_1, x_2, \ldots, x_k$ os dados exógenos que podem influenciar as séries temporais. Estes dados são organizados como $X_{\text{exógeno}} = x_1, x_2, \ldots, x_k$, onde k é o número de variáveis externas.

Portanto, na implementação do SPMST, cada série temporal S_i é tratada de forma integrada. Isso implica na combinação de todas as séries temporais em um único vetor de entrada para o modelo neural. Essa representação pode ser formalizada da seguinte maneira:

$$X_{\text{carga}} = [S_1, S_2, \ldots, S_n] \tag{47}$$

É importante ressaltar que no contexto do SPMST, as variáveis exógenas são comuns a todas as séries de carga incluídas no modelo. Isso implica que cada série temporal utiliza, de forma compartilhada, o mesmo conjunto de variáveis exógenas para suas previsões. Ao considerar $X_{\text{exógeno}} = x_1, x_2, \ldots, x_k$ como um conjunto unificado de informações externas, o SPMST integra esses dados de forma consistente, visando capturar e incorporar os efeitos desses fatores nas previsões individuais de cada série temporal.

5.5 Procedimentos e Definições do Arranjo Experimental

Nesta subseção, são apresentados os procedimentos adotados e as definições fundamentais que delineiam o arranjo experimental utilizado neste estudo.

5.5.1 Divisão dos Dados

Nesta pesquisa, foram selecionadas oito séries temporais para as simulações. Quatro dessas sequências foram obtidas da Base de Dados A, enquanto as outras quatro foram extraídas

da Base de Dados B. A Tabela 4 resume as principais informações dos conjuntos de dados utilizados.

Tabela 4 - Principais Características das Bases de Dados.

Característica	**Base de Dados A**	**Base de Dados B**
Localização	Loureiro, Portugal	Waikato, Nova Zelândia
Entidades	172 edifícios	2 termoelétricas e 7 subestações
Coleta	Maio/2022 a Setembro/2023	Janeiro/2007 a março/2009
Frequência	A cada 15 minutos	A cada 30 minutos
Registros	46,608	39,408

Fonte:Autoria própria.

A seleção das séries temporais utilizadas baseou-se na qualidade dos dados, priorizando aquelas com consistência, mínimo de lacunas e anomalias, garantindo a robustez da análise comparativa dos resultados. De modo similar, na escolha dos períodos para a investigação experimental, a uniformidade dos dados disponíveis foi priorizada. Conforme a Tabela 5, para a base de dados A, determinou-se o intervalo de 07/05/2022 a 07/05/2023 como o mais adequado, pois as análises indicaram maior estabilidade nas condições de coleta. No caso da base de dados B, adotou-se o conjunto completo de dados. Embora tenham sido identificadas inconsistências, estratégias de tratamento e correção, detalhadas na Subseção 5.2.3, foram implementadas para mitigar potenciais distorções. Assim, garantiu-se que os dados utilizados fossem suficientemente robustos para não comprometer a integridade e a validade das conclusões desta pesquisa.

Tabela 5 - Período da Análise Experimental,

Base de Dados	**Período**	**Amostras**
A	07/05/2022 a 07/05/2023	35,040
B	01/01/2007 a 31/03/2009	39,408

Fonte:Autoria própria.

Para facilitar a referência ao longo do estudo, as séries temporais foram designadas conforme a nomenclatura apresentada na Tabela 6. Essa terminologia será utilizada ao longo desta investigação para referenciar cada série temporal especificamente, favorecendo a compreensão e a análise dos resultados. Assim, as séries da Base de Dados A serão referidas como ST1A a ST4A, enquanto as da Base de Dados B serão denominadas ST1B a ST4B.

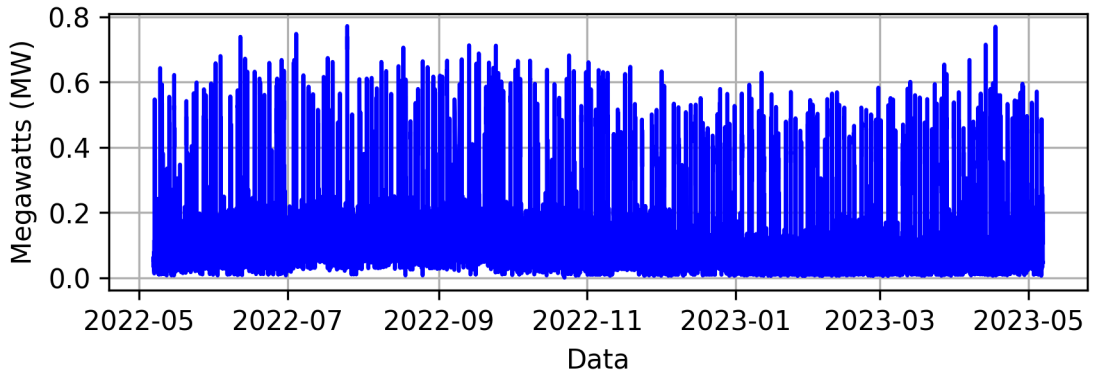
Tabela 6 - Nomenclatura das Séries Temporais.

Base de Dados	**Série Temporal 1**	**Série Temporal 2**	**Série Temporal 3**	**Série Temporal 4**
A	ST1A	ST2A	ST3A	ST4A
B	ST1B	ST2B	ST3B	ST4B

Fonte:Autoria própria.

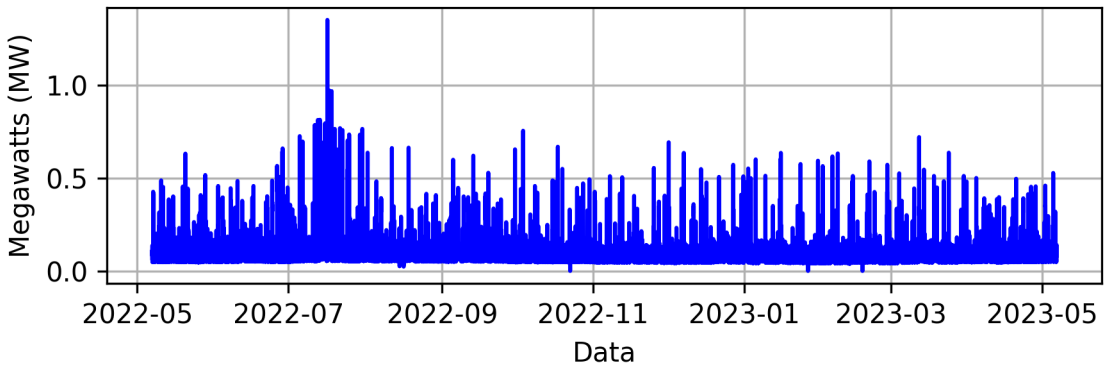
As Figuras 29 a 36 apresentam as séries temporais utilizadas, proporcionando uma visualização detalhada dos dados empregados nas análises desta pesquisa.

Figura 29 - Visualização da Série Temporal ST1A.



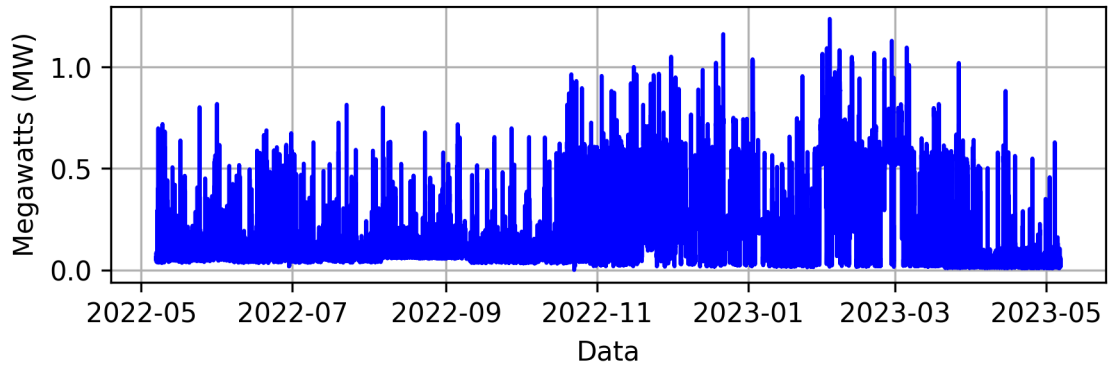
Fonte: Autoria própria.

Figura 30 - Visualização da Série Temporal ST2A.



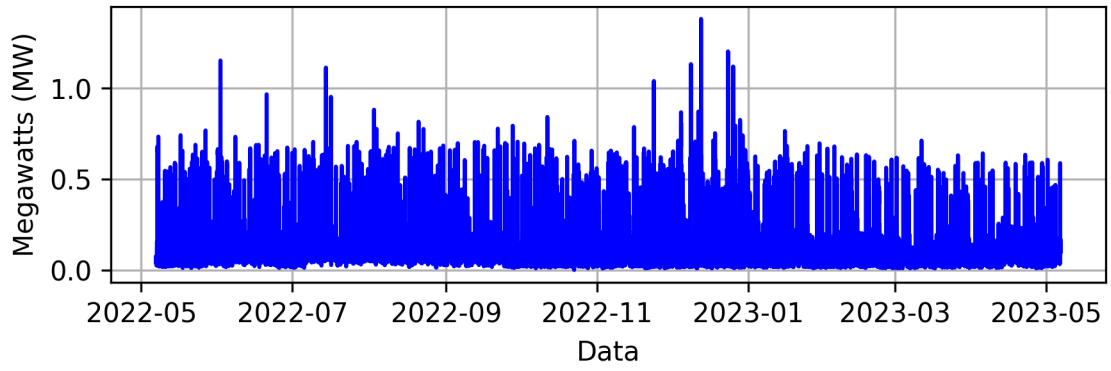
Fonte: Autoria própria.

Figura 31 - Visualização da Série Temporal ST3A.



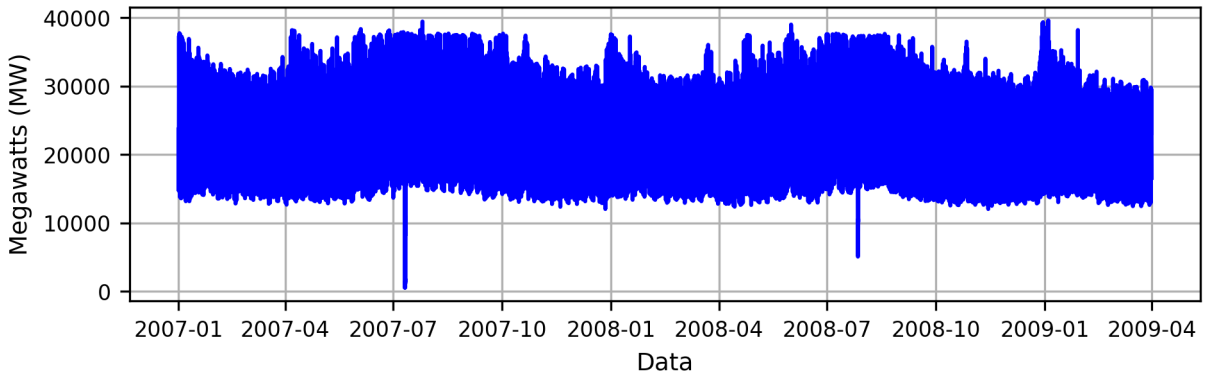
Fonte: Autoria própria.

Figura 32 - Visualização da Série Temporal ST4A.



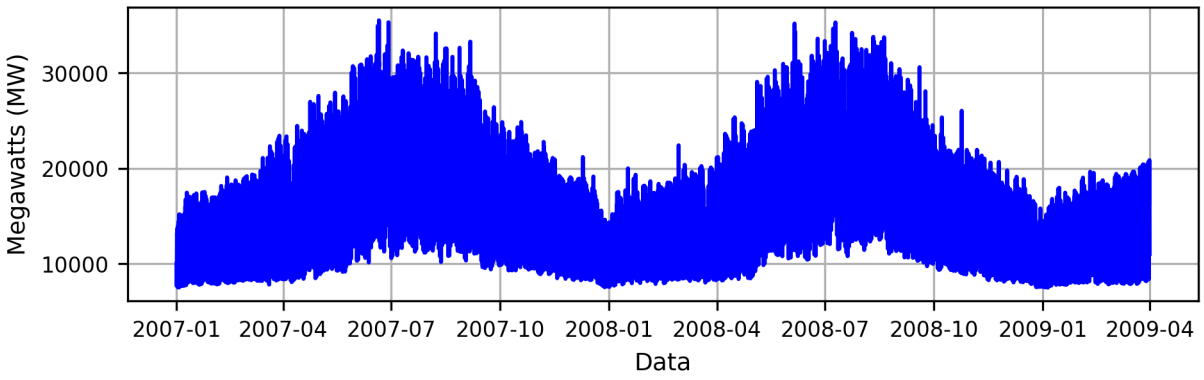
Fonte: Autoria própria.

Figura 33 - Visualização da Série Temporal ST1B.



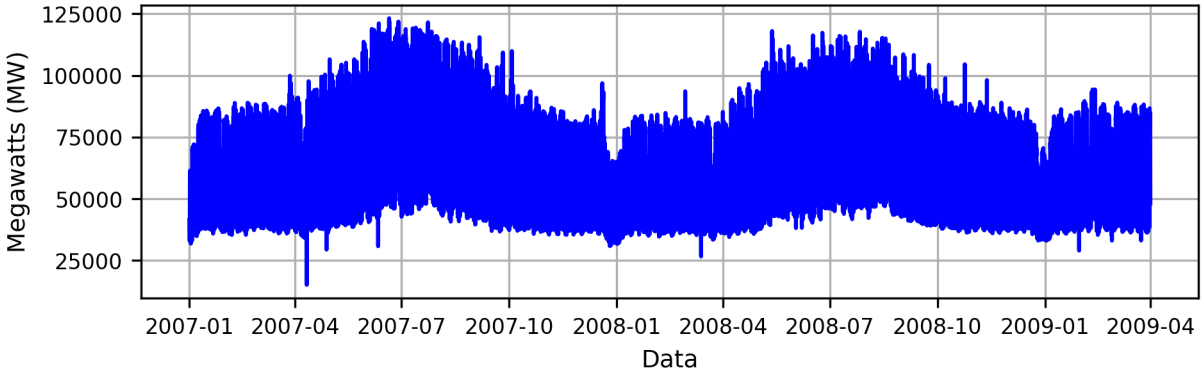
Fonte: Autoria própria.

Figura 34 - Visualização da Série Temporal ST2B.



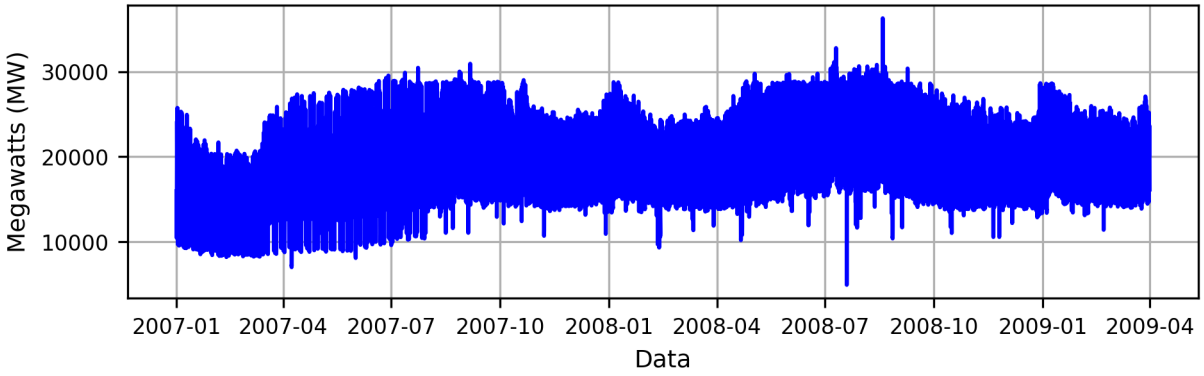
Fonte: Autoria própria.

Figura 35 - Visualização da Série Temporal ST3B.



Fonte: Autoria própria.

Figura 36 - Visualização da Série Temporal ST4B.



Fonte: Autoria própria.

A análise gráfica das séries temporais revela uma tendência geral de instabilidade, caracterizada por flutuações contínuas ao longo do tempo. Contudo, observam-se comportamentos distintos entre as séries da Base de Dados A e da Base de Dados B. A Base de Dados A, que se refere ao nível de distribuição de energia elétrica, tende a apresentar maior instabilidade, evidenciada por um comportamento com numerosos picos. Essa volatilidade sugere que a distribuição de energia, no contexto do sistema analisado, é altamente sensível a diversos fatores externos, como eventos climáticos e períodos de feriados prolongados.

Por outro lado, as séries temporais da Base de Dados B, relacionadas ao nível de transmissão de energia elétrica, exibem um comportamento mais atenuado. Embora ainda apresentem variações ao longo do tempo, estas são menos pronunciadas e mais regulares em comparação com as séries da Base de Dados A. A distribuição, sendo o estágio final antes do consumo, enfrenta desafios mais imediatos e diretos devido a variações diárias e sazonais. A transmissão, por sua vez, operando em uma escala mais ampla, consegue manter uma maior estabilidade, apesar das flutuações ainda presentes.

Dada a natureza altamente volátil das séries temporais da Base de Dados A, é plausível que a previsão para este conjunto de dados represente um desafio maior. A elevada sensibilidade a fatores externos torna a modelagem preditiva mais complexa, exigindo que os algoritmos de previsão se adaptem rapidamente às mudanças e à alta frequência de picos observados. Em contraste, a maior regularidade e menor volatilidade das séries temporais da Base de Dados B tornam as previsões mais manejáveis, ainda que não isentas de desafios.

Para a implementação da abordagem proposta, cada ST foi dividida em partes distintas. A primeira, designada para treinar a rede neural, permitindo o aprendizado de padrões e recursos presentes nos dados de entrada. A segunda parte, por sua vez, foi reservada para a validação e avaliação, proporcionando uma análise do desempenho da estratégia sugerida. A segmentação de cada serie temporal utilizada pode ser observada na Tabela 7.

Tabela 7 - Distribuição dos Dados.

Amostras	**Base de Dados A**	**Base de Dados B**
Janela Prospectiva	96 (24 horas)	48 (24 horas)
Janela Retrospectiva	1,344 (14 Dias)	672 (14 Dias)
Treinamento	33,600	38,688

Fonte:Autoria própria.

No TFT, os dados são frequentemente segmentados em janelas deslizantes (*sliding windows*) para criar múltiplos exemplos de treinamento. Essa abordagem permite que cada janela inclua uma parte retrospectiva (histórica) e uma parte prospectiva (futura):

- Janela Retrospectiva (*Lookback Window*): Também conhecida como janela de entrada, esta parte dos dados compreende um intervalo de tempo no passado utilizado para observar e aprender padrões históricos. Por exemplo, para prever 24 horas à frente, a janela retrospectiva pode consistir de dados dos 14 dias anteriores;

- Janela Prospectiva (*Prediction Horizon*): Corresponde ao período futuro para o qual o modelo faz previsões. Por exemplo, para prever 24 horas à frente, a janela prospectiva seria de 24 horas no futuro.

Como poder ser observado na Tabela 7, durante a fase de treinamento, a maior parte dos dados é utilizada para instruir o modelo. Nessa etapa, as janelas deslizantes são aplicadas para criar exemplos de treinamento variados, abrangendo diferentes períodos de tempo. Além disso, uma parte dos dados é reservada pelo próprio TFT para validar o desempenho do modelo ao longo do treinamento. Para avaliar o desempenho final da rede neural, emprega-se um conjunto separado de dados, que também inclui uma janela retrospectiva e uma janela prospectiva, o que permite avaliar a capacidade de generalização do modelo em relação a dados não observados.

Portanto, a distribuição apresentada na Tabela 7 foi aplicada a cada série temporal das Bases de Dados A e B. Foram selecionadas 96 amostras das séries temporais com intervalos de 15 minutos (Base de Dados A) e 48 amostras das séries temporais com intervalos de 30 minutos (Base de Dados B), totalizando um período prospectivo de 24 horas nos respectivos intervalos. Além disso, adotou-se uma janela retrospectiva de 14 dias, correspondente a 1,344 amostras na Base de Dados A e a 672 amostras na Base de Dados B. Após essa segmentação e extração de amostras destinadas à previsão, restaram 33,600 amostras para o treinamento na Base de Dados A e 38,688 amostras para o treinamento das séries temporais pertencentes à Base de Dados B.

A escolha de uma janela retrospectiva de 14 dias é fundamentada na necessidade de capturar padrões sazonais e variações de curto prazo que influenciam significativamente o consumo energético. Esse período permite ao modelo incorporar informações sobre flutuações diárias, que tendem a ser regulares ao longo da semana, como o aumento do consumo durante os dias úteis e a redução nos finais de semana. Além disso, uma janela de duas semanas é suficientemente longa para capturar eventos esporádicos, como mudanças climáticas e eventos especiais, sem introduzir dados excessivamente antigos que possam diluir a relevância das informações mais recentes. Outro ponto importante é a eficiência computacional. Uma janela de 14 dias oferece um volume de dados gerenciável em termos de processamento, permitindo que os modelos sejam atualizados de maneira mais ágil e com menor custo computacional. Portanto, a escolha de uma janela retrospectiva de 14 dias se mostra eficaz para captar variações relevantes de curto prazo, incorporar eventos significativos e manter a eficiência computacional.

É fundamental ressaltar que a divisão apresentada na Tabela 7 foi aplicada, individualmente, a cada série temporal analisada nesta pesquisa. Em outras palavras, para cada conjunto

de dados proveniente tanto da Base de Dados A quanto da Base de Dados B, foi realizada uma divisão isolada. Neste contexto, conforme a modelagem matemática da previsão descrita na Subseção 5.4.1, a implementação do SPMST trata as sequências temporais de forma integrada, o que implica na combinação de diferentes séries temporais em um único vetor de entrada. Consequentemente, devido à variação na quantidade de séries temporais combinadas nos experimentos conduzidos nesta pesquisa, o volume de dados utilizados no treinamento variou conforme demonstrado na Tabela 8.

Tabela 8 - Informações de Treinamento por Teste Experimental.

Análise	**Sistema Previsor**	**Base de Dados**	**Número de STs**	**Amostras**
01	SPUST	A	1	33,600
02	SPMST	A	3	100,800
03	SPMST	A	4	134,400
04	SPUST	B	1	38,688
05	SPMST	B	3	116,064
06	SPMST	B	4	154,752

Fonte:Autoria própria.

Portanto, cada teste experimental listado representa uma configuração específica de estudo aplicado a diferentes sequências de dados provenientes das Bases de Dados A e B. Cada análise foi conduzida com um número diferente de STs, resultando em volumes distintos de amostras de dados utilizadas para treinar os modelos neurais. Neste contexto, é relevante destacar que as janelas retrospectiva e prospectiva permaneceram constantes, conforme definido na Tabela 7, variando apenas no número de amostras utilizadas para o treinamento.

5.5.2 Variáveis Exógenas Utilizadas

O TFT foi projetado com o objetivo de identificar diferentes tipos de recursos, classificados como: covariáveis estáticas, conhecidas no futuro e desconhecidas no futuro. Nos conjuntos de dados utilizados nesta pesquisa, não há elementos estáticos, e, portanto, esses não foram considerados nas simulações realizadas. Desta forma, neste estudo, foram utilizados dois tipos de metadados: aqueles conhecidos no futuro e aqueles desconhecidos no futuro. É importante destacar que, embora o TFT seja capaz de lidar com diferentes categorias de dados, não é obrigatório que todas sejam utilizadas.

As Tabelas 9 e 10 apresentam, respectivamente, as covariáveis exógenas utilizadas nas Bases de Dados A e B.

Tabela 9 - Covariáveis Utilizadas na Base de Dados A.

Variáveis Conhecidas Futuras	**Variáveis Desconhecidas Futuras**
Hora	Temperatura média
Mês	Umidade relativa média
Ano	Velocidade do vento média
Dia do ano	Carga média
Dia da semana	Carga mínima
	Carga máxima

Fonte:Autoria própria.

Tabela 10 - Covariáveis Utilizadas na Base de Dados B.

Variáveis Conhecidas Futuras	**Variáveis Desconhecidas Futuras**
Hora	Temperatura mínima
Mês	Temperatura máxima
Ano	Cargas médias
Dia do ano	Cargas mínimas
Dia da semana	Cargas máximas
Dias atípicos	
Feriados	
Horário de verão	

Fonte:Autoria própria.

As informações relacionadas às cargas médias, mínimas e máximas foram obtidas conforme as equações (48), (49) e (50), respectivamente.

$$C_{\text{med}} = \frac{C(h_k) + C(h)}{k+1}, \tag{48}$$

$$C_{\text{min}} = \min\{C(h_k), C(h)\}, \tag{49}$$

$$C_{\text{max}} = \max\{C(h_k), C(h)\}, \tag{50}$$

em que, $C(h_k)$ representa as k amostras anteriores ao momento atual $C(h)$.

Para a Base de Dados A, onde as amostras são registradas a cada 15 minutos, foi adotado

$k = 4$, pois uma hora contém 4 intervalos de 15 minutos. Para a Base de Dados B, na qual as amostras são registradas a cada 30 minutos, foi considerado $k = 2$, visto que uma hora é composta por 2 intervalos de 30 minutos. Essa abordagem é melhor explicada a seguir:

- **Base de Dados A** ($k = 4$)

  As últimas 4 amostras anteriores ao momento atual $C(h)$ são consideradas:

  $$C(h_k) = \{C(h-4), C(h-3), C(h-2), C(h-1)\}$$

- **Base de Dados B** ($k = 2$)

  As últimas 2 amostras anteriores ao momento atual $C(h)$ são consideradas:

  $$C(h_k) = \{C(h-2), C(h-1)\}$$

Portanto, a carga média C_{med} no instante atual h, foi calculada utilizando-se a média aritmética das amostras consecutivas, incluindo $C(h)$. Por outro lado, as cargas elétricas mínima C_{min} e máxima C_{max} no momento h, foram determinadas pelos valores mínimo e máximo dessas amostras.

5.5.3 Critérios de Avaliação

Para avaliar a performance dos modelos de previsão desenvolvidos, foram escolhidas métricas amplamente adotadas na avaliação de tais modelos. O MAPE foi selecionado por sua capacidade de expressar os erros relativos em termos percentuais, oferecendo uma compreensão direta da acurácia das previsões. O MAE foi adotado por proporcionar uma medida direta da magnitude das discrepâncias entre as previsões e os dados observados. Por fim, a RMSE foi escolhida devido à sua habilidade em capturar e quantificar as discrepâncias entre os valores estimados e os valores reais. A Subseção 3.4.1 fornece uma explicação detalhada sobre esses indicadores.

5.5.4 Definições de Treinamento do Modelo Neural

Para a implementação da rede neural *Temporal Fusion Transformer*, os hiperparâmetros foram ajustados de acordo com a configuração detalhada na Tabela 11, os quais foram determinados com base na análise de experimentos preliminares.

Tabela 11 - Configuração dos Hiperparâmetros.

Hiperparâmetro	**Valor**
Learning rate	0,0001
Hidden size	128
Attention head size	4
Dropout	0,1
Hidden continuous size	128
Loss	QuantileLoss()
Gradient clip val	0,1
Batch size	32
Epochs	50

Fonte:Autoria própria.

Nesta investigação, a configuração dos hiperparâmetros do modelo TFT foi elaborada com o objetivo de otimizar o desempenho e a acurácia das projeções. A taxa de aprendizado foi definida em 0,0001 para permitir ajustes mais precisos e estáveis, além de assegurar uma convergência controlada do modelo e evitar grandes oscilações durante o treinamento. O tamanho dos neurônios ocultos foi estabelecido em 128, buscando equilibrar a capacidade de modelagem com a eficiência computacional, permitindo que o modelo seja suficientemente complexo sem ser excessivamente oneroso. Foram selecionadas quatro cabeças de atenção com o intuito de captar diferentes aspectos das dependências temporais nos dados. A taxa de *dropout* foi fixada em 0,1, indicando que 10% dos neurônios serão desativados em cada atualização, o que auxilia na prevenção do *overfitting* e promove a generalização do modelo.

Para modelar as variáveis contínuas, o *hidden continuous size* foi definido em 128, proporcionando uma capacidade suficiente para capturar nuances complexas nos dados. A função de perda QuantileLoss foi escolhida por sua habilidade de estimar intervalos de confiança nas previsões. Um valor de *clipping* de gradiente de 0,1 foi selecionado para manter a estabilidade durante o treinamento, evitando a explosão de gradientes que pode comprometer o aprendizado. O tamanho do lote foi estabelecido em 32, com o intuito de proporcionar um equilíbrio entre a estabilidade do gradiente e a eficiência computacional, visando atualizações frequentes e balanceadas. Finalmente, 50 épocas foram escolhidas para que o modelo tenha tempo suficiente para aprender padrões complexos sem incorrer em *overfitting*, oportunizando um desempenho robusto e generalizável.

Neste contexto, é fundamental ressaltar que todos os experimentos realizados neste estudo, tanto para a Base de Dados A quanto para a Base de Dados B, foram conduzidos utilizando a mesma configuração de hiperparâmetros, conforme detalhado nesta subseção. Esta abordagem

permitiu que a arquitetura proposta fosse aplicada de maneira uniforme e sistemática em todas as fases de testes e análises realizadas. Ao manter uma consistência nas especificações da rede neural ao longo dos experimentos, foi possível assegurar resultados comparáveis. Assim, essa uniformidade permitiu uma avaliação precisa do desempenho dos modelos desenvolvidos e uma investigação aprofundada dos efeitos das variações nas condições experimentais. Além disso, proporcionou uma base sólida para validar as hipóteses formuladas neste estudo, fortalecendo as conclusões derivadas desta pesquisa.

5.5.5 Ferramentas e Ambiente de Desenvolvimento

Os métodos apresentados na proposta deste estudo foram desenvolvidos utilizando a linguagem de programação Python, conhecida por sua flexibilidade e pela ampla gama de bibliotecas especializadas em ciência de dados. A escolha do *framework* PyTorch® e do pacote PyTorch *Lightning* baseou-se na capacidade de fornecer uma infraestrutura eficiente para a construção e treinamento de modelos de aprendizado profundo. Em particular, a implementação da rede neural TFT foi realizada utilizando a biblioteca PyTorch *Forecasting*, que simplifica o processo de desenvolvimento e otimização de modelos para séries temporais complexas.

Para o ambiente de desenvolvimento, optou-se pelo Google *Colaboratory*, uma plataforma que disponibiliza recursos computacionais avançados e integra diversas bibliotecas de *machine learning*. Assim, os modelos foram treinados utilizando os recursos de aceleração gráfica oferecidos pela referida plataforma. Nesse contexto, especificamente uma única GPU NVIDIA Tesla T4 foi utilizada nas simulações realizadas neste trabalho.

5.6 Síntese do Capítulo

Este capítulo descreve a metodologia desenvolvida para prever demandas elétricas utilizando a rede neural TFT. As subseções oferecem uma compreensão detalhada das técnicas e conceitos fundamentais aplicados nesta investigação. Inicialmente, discute-se a utilização de múltiplas séries temporais no contexto dos sistemas de energia elétrica, enfatizando a importância de capturar correlações entre variáveis para aprimorar a precisão das previsões. Em seguida, são apresentadas as bases de dados utilizadas neste estudo, com destaque para as características específicas de cada conjunto de dados e para o processo de pré-processamento aplicado.

Na subseção dedicada ao modelo neural, discute-se a escolha do TFT, detalhando as configurações do algoritmo e a seleção de hiperparâmetros. Na sequência, abordam-se os sistemas de previsão desenvolvidos, incluindo a modelagem matemática e a integração de múltiplas séries temporais. A subseção sobre procedimentos experimentais descreve os critérios de avaliação, os recursos exógenos utilizados, a divisão dos dados, o processo de treinamento do modelo e o ambiente de desenvolvimento utilizado para a implementação da abordagem sugerida.

6 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos a partir da metodologia proposta, estruturando-se de forma a evidenciar os achados dos experimentos realizados com os sistemas preditivos desenvolvidos para a previsão de séries temporais elétricas. A Subseção 6.1 fornece o contexto necessário para a compreensão das observações experimentais e descreve os procedimentos de análise adotados. A Subseção 6.2 detalha os resultados e explora o desempenho dos modelos SPUST e SPMST nas Base de Dados A e B.

A Subseção 6.3 discute e avalia a eficácia das diferentes abordagens de previsão, comparando modelos que utilizam uma única série temporal com aqueles que incorporam múltiplas séries temporais. A Subseção 6.4 concentra-se na análise de interpretabilidade, explorando as informações que podem ser extraídas para fornecer uma melhor compreensão do comportamento dos modelos. Finalmente, a Subseção 6.5 oferece uma síntese das principais conclusões e *insights* do capítulo.

6.1 Introdução

Na presente pesquisa, cada experimento foi projetado para avaliar a precisão das abordagens propostas em diferentes configurações e contextos, conforme detalhado no Capítulo 5. A estrutura da investigação segue a ordem descrita na Tabela 12, o que facilita a comparação entre os diferentes cenários analisados.

Tabela 12 - Configurações Experimentais.

Sistema Previsor	Procedimento	Fonte de Dados	Número de STs
SPUST	Experimento 1	A	1
	Experimento 2	B	1
SPMST	Experimento 3	A	3
	Experimento 4	A	4
	Experimento 5	B	3
	Experimento 6	B	4

Fonte:Autoria própria.

Para validar a eficácia dos modelos nas diferentes condições experimentais investigadas, a análise dos resultados é estruturada com base nos seguintes critérios:

1. Gráficos Comparativos: Apresentam-se representações visuais das séries temporais observadas e preditas, incluindo a janela retrospectiva de 14 dias utilizada na análise;

2. Gráficos de Dispersão: Ilustram a relação entre os dados reais e previstos, permitindo a visualização da correlação e da dispersão dos valores preditos em comparação com os valores observados;

3. Métricas de Desempenho: A precisão das previsões geradas pelos modelos é avaliada em relação aos valores reais de demanda, utilizando os indicadores MAE, MAPE e RMSE.

Portanto, cada experimento é analisado de maneira abrangente, combinando gráficos comparativos e de dispersão com métricas de desempenho. Essa abordagem permite uma avaliação visual e quantitativa, destacando tanto a acurácia das previsões quanto a correlação entre os valores reais e previstos.

6.2 Análise dos Resultados

Esta seção apresenta os resultados dos experimentos conduzidos com os Sistemas Previsores de Única Série Temporal (SPUST) e de Múltiplas Séries Temporais (SPMST). As configurações variam em função do número de sequências históricas e do conjunto de dados utilizado. Portanto, para facilitar a comparação entre os diferentes procedimentos, a análise é estruturada em subseções específicas.

6.2.1 Resultados Experimentais do SPUST

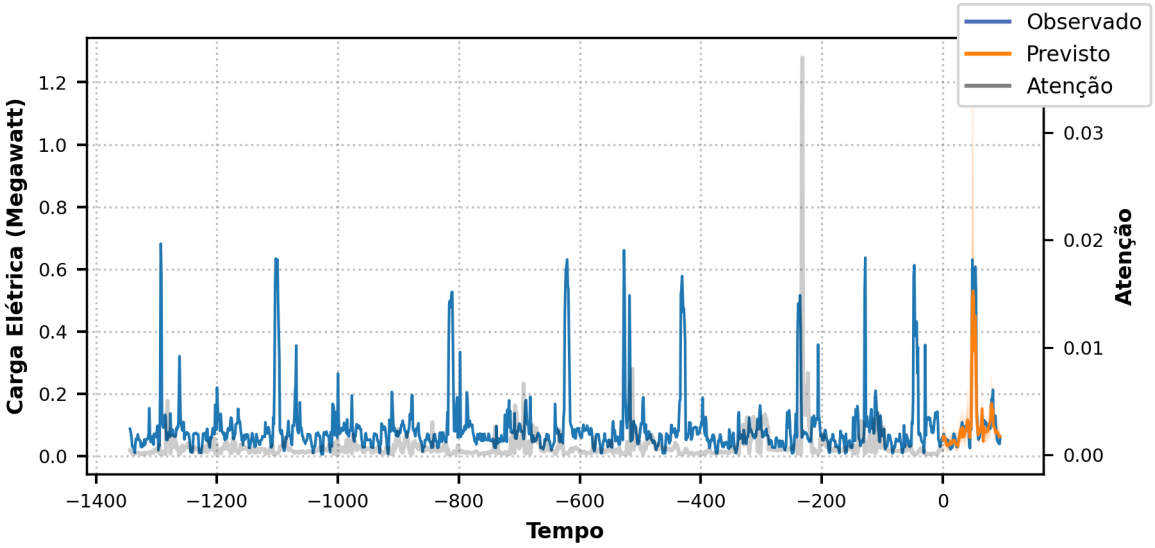
Nos procedimentos descritos nesta subseção, a metodologia utilizada visou avaliar a eficácia do Sistema Previsor de Única Série Temporal (SPUST) na previsão de séries temporais de forma isolada. Para isso, as sequências históricas foram analisadas sem a inclusão de dados provenientes de outras séries correlacionadas, focando-se exclusivamente nas informações da série temporal em questão.

6.2.1.1 Experimento 1: Base de Dados A

No Experimento 1, aplicou-se o SPUST para realizar previsões de demanda energética com base na Base de Dados A. Nesta investigação, as sequências históricas ST1A, ST2A e ST3A foram analisadas de forma independente. Portanto, cada uma dessas séries temporais foi tratada individualmente, permitindo uma análise focada em seus comportamentos específicos. Para isso, foram processadas 33,600 amostras em cada análise individual realizada pelo modelo.

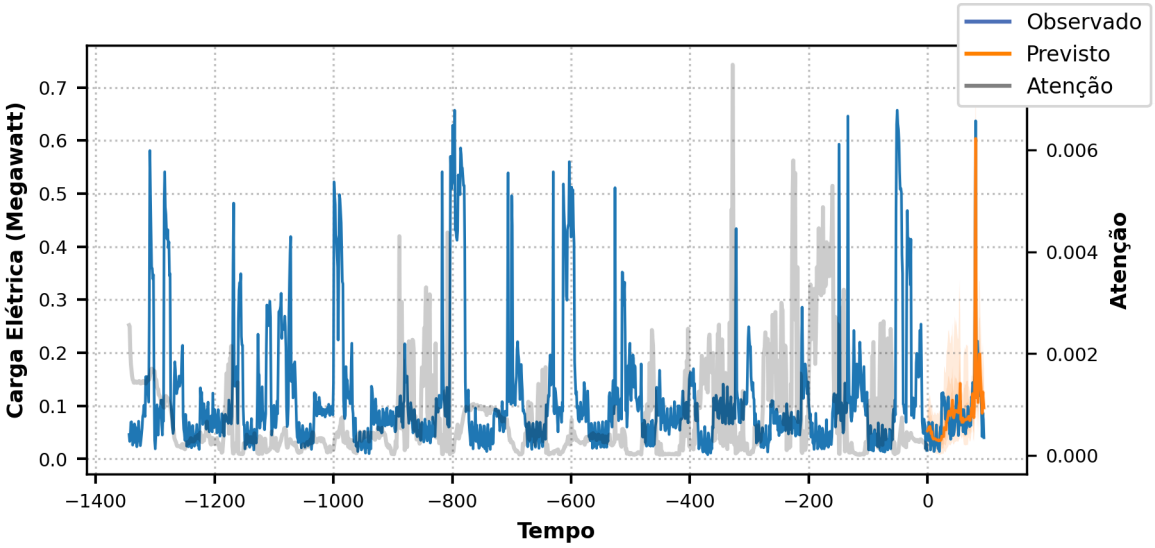
As Figuras 37 a 39 exibem a janela retrospectiva e as previsões obtidas no Experimento 1.

Figura 37 - Experimento 1 - Janela Retrospectiva e Previsão da ST1A.



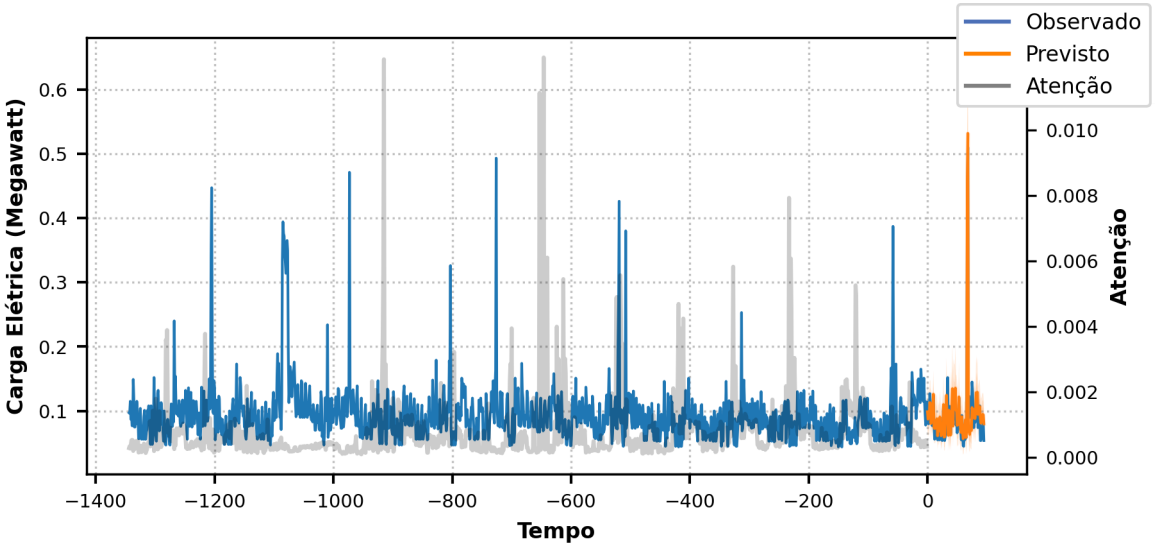
Fonte: Autoria própria.

Figura 38 - Experimento 1 - Janela Retrospectiva e Previsão da ST2A.



Fonte: Autoria própria.

Figura 39 - Experimento 1 - Janela Retrospectiva e Previsão da ST3A.

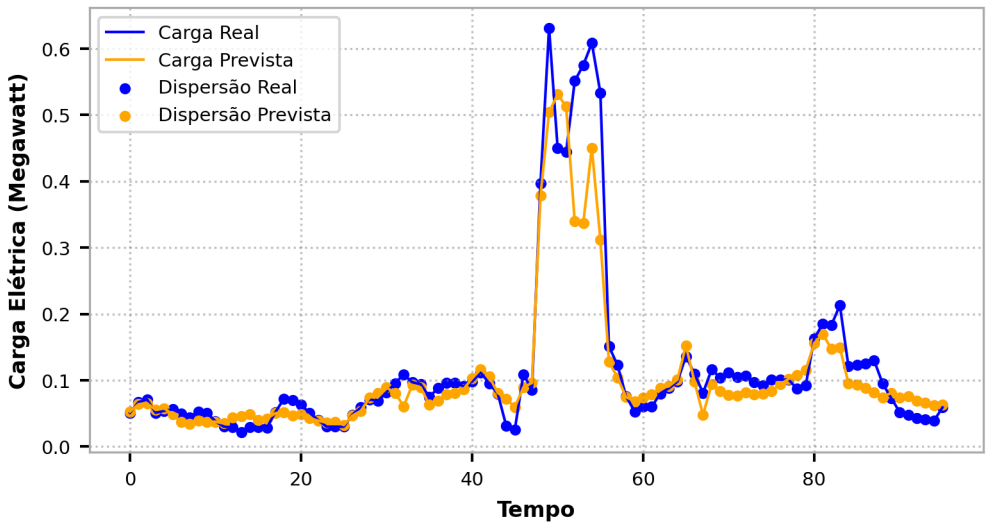


Fonte: Autoria própria.

Nas ilustrações apresentadas, o eixo temporal varia de 0 a -1400, que corresponde a 1,344 amostras em intervalos de 15 minutos. No último intervalo, são exibidos os resultados das previsões. A linha azul representa os valores reais da carga, a linha laranja reflete as previsões obtidas, e a linha cinza indica os pesos atribuídos pelo mecanismo de atenção.

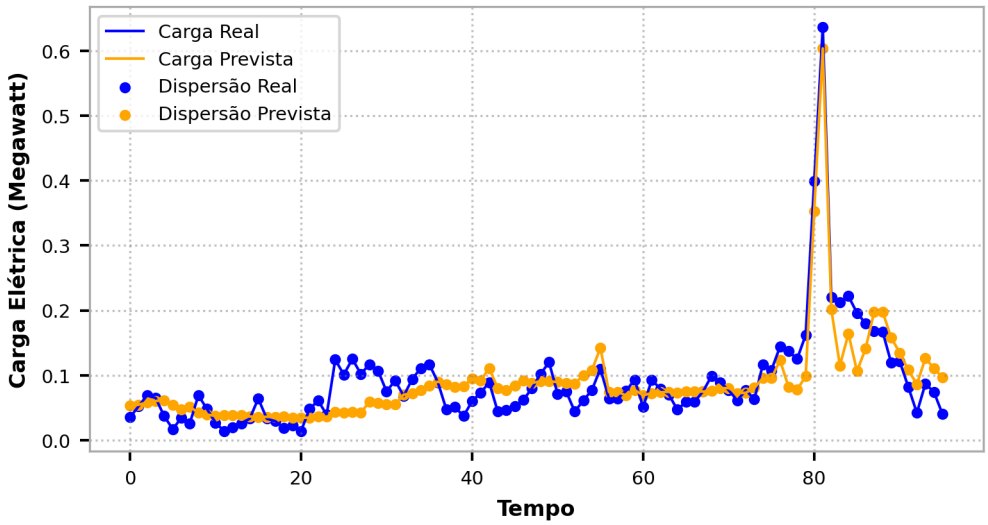
As Figuras 40, 41 e 42 fornecem uma visualização detalhada das previsões obtidas.

Figura 40 - Experimento 1: Curvas de Carga Real e Prevista da ST1A.



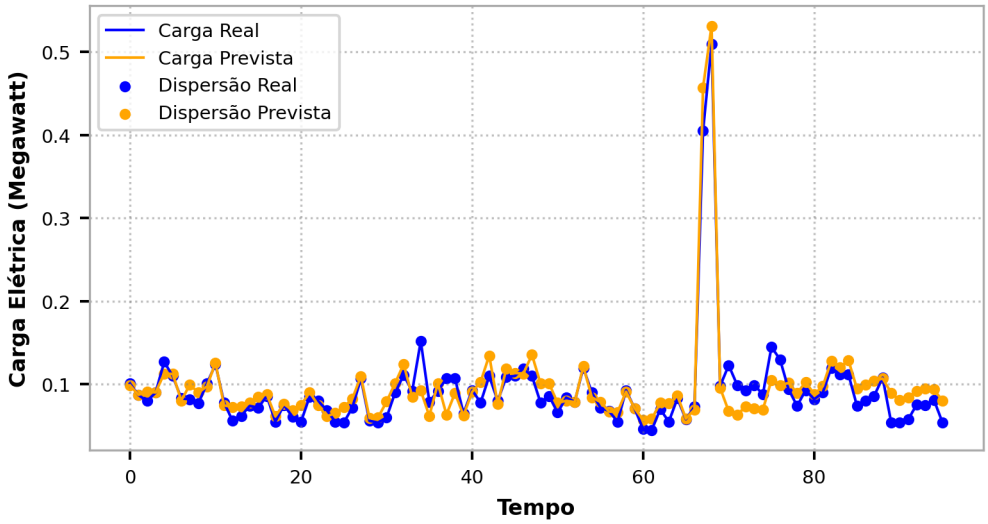
Fonte: Autoria própria.

Figura 41 - Experimento 1: Curvas de Carga Real e Prevista da ST2A.



Fonte: Autoria própria.

Figura 42 - Experimento 1: Curvas de Carga Real e Prevista da ST3A.



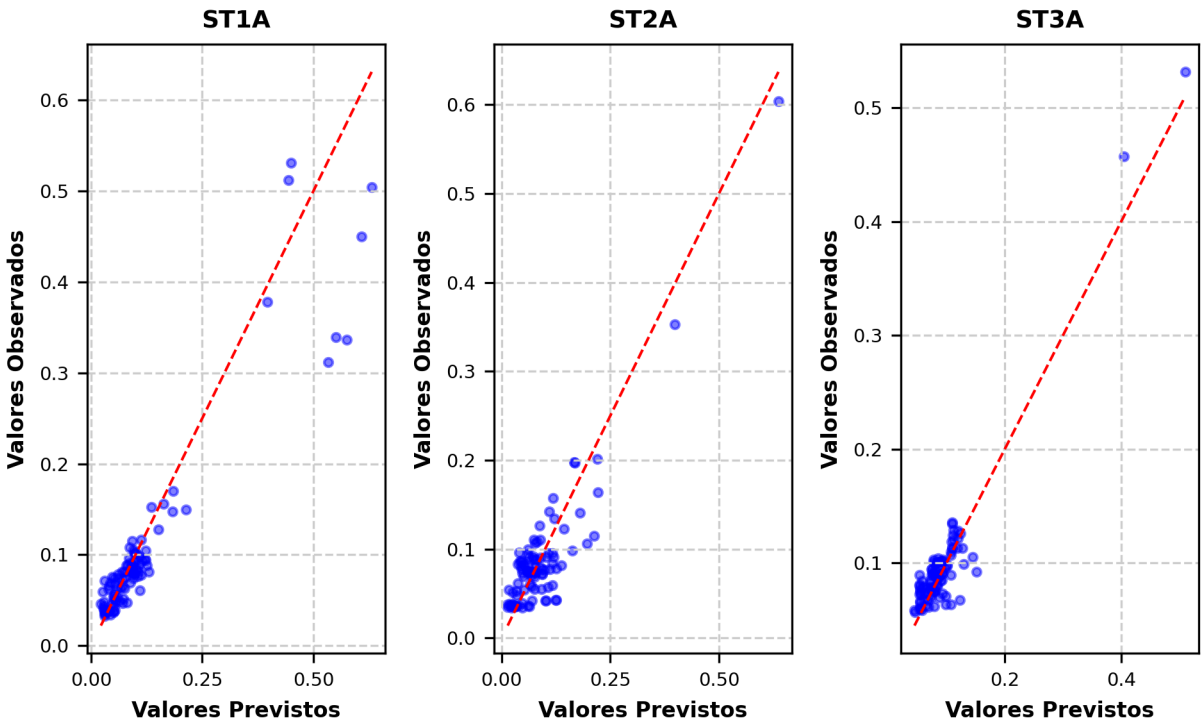
Fonte: Autoria própria.

Nas visualizações fornecidas, o eixo temporal abrange o intervalo de 0 a 100, correspondendo a 96 observações em intervalos de 15 minutos ao longo de 24 horas. A linha azul denota os valores reais da carga, enquanto a linha laranja representa as previsões. Também, é possível visualizar os pontos de dispersão entre essas duas curvas.

O gráfico de dispersão do Experimento 1, apresentado na Figura 43, permite uma análise abrangente dos resultados obtidos para as séries temporais ST1A, ST2A e ST3A. Cada ponto

representa um par de valores, sendo um correspondente aos dados observados e o outro às previsões geradas. A linha vermelha em estilo tracejado destaca as igualdades entre os valores reais e as projeções obtidas.

Figura 43 - Gráfico de Dispersão do Experimento 1.



Fonte: Autoria própria.

A Tabela 13 apresenta os valores obtidos no Experimento 1, baseados nos indicadores MAPE, MAE e RMSE.

Tabela 13 - Métricas de Desempenho do Experimento 1.

Métricas	ST1A	ST2A	ST3A
MAPE(%)	23,54	27,15	17,18
MAE(MW)	0,03	0,02	0,01
RMSE(MW)	0,05	0,02	0,02

Fonte:Autoria própria.

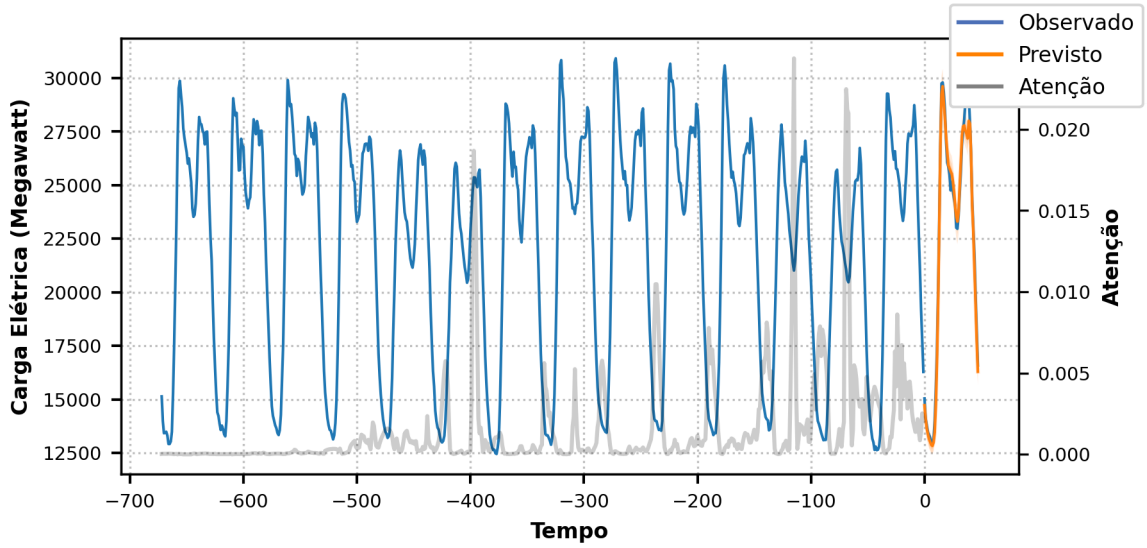
6.2.1.2 Experimento 2: Base de Dados B

No Experimento 2, o método SPUST foi utilizado para projetar o consumo de energia, aplicando a Base de Dados B. Neste contexto, as sequências ST1B, ST2B e ST3B foram analisadas

de forma individualizada, resultando em um total de 38,688 amostras para cada processo de treinamento.

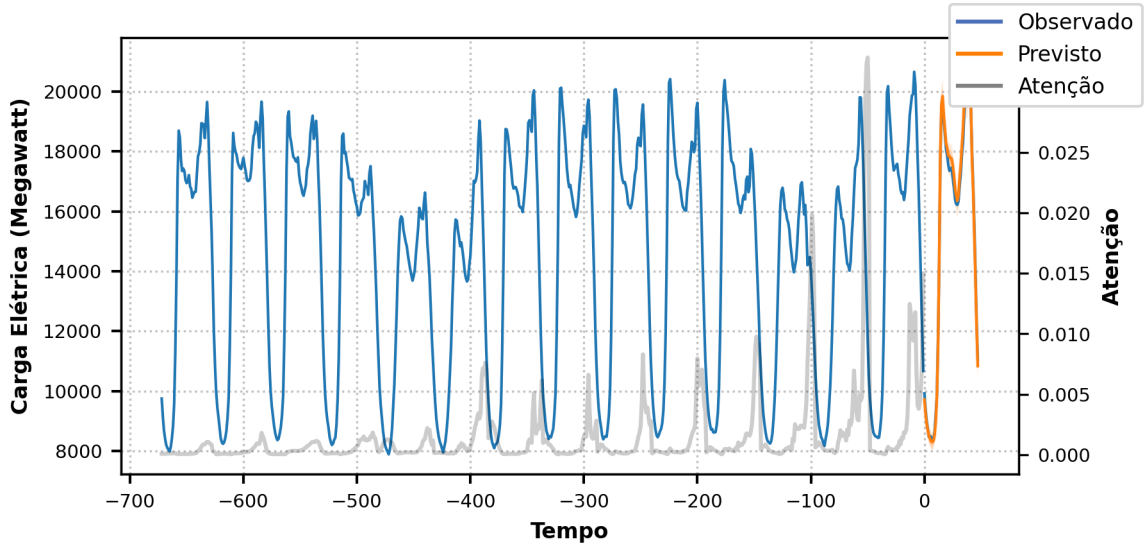
As Figuras 44 a 46 ilustram a janela retrospectiva e as projeções obtidas no Experimento 2.

Figura 44 - Experimento 2 - Janela Retrospectiva e Previsão da ST1B.



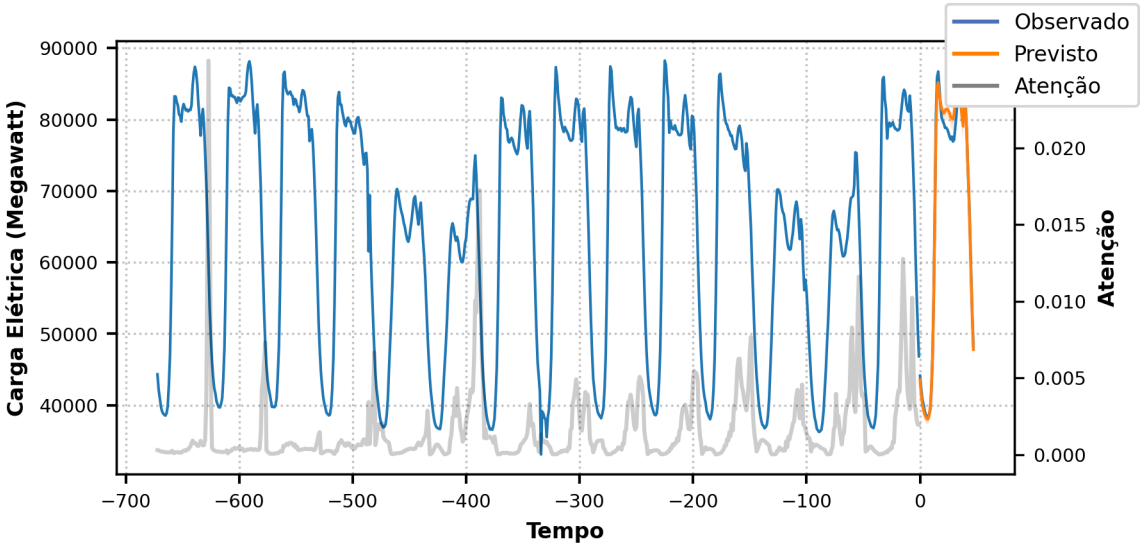
Fonte: Autoria própria.

Figura 45 - Experimento 2 - Janela Retrospectiva e Previsão da ST2B.



Fonte: Autoria própria.

Figura 46 - Experimento 2 - Janela Retrospectiva e Previsão da ST3B.

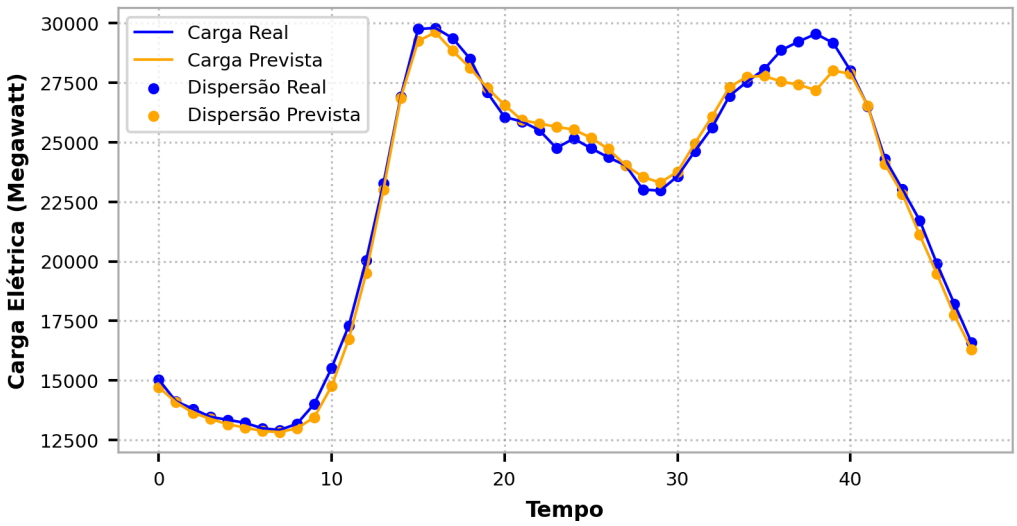


Fonte: Autoria própria.

O conjunto de representações visuais apresentado demonstra o eixo temporal variando de 0 a -700, o que corresponde a 672 amostras coletadas a cada 30 minutos. No último intervalo de tempo, são apresentados os resultados das previsões. Nos gráficos, a linha azul denota os valores reais de carga, enquanto a linha laranja representa as predições realizadas. A linha cinza, por sua vez, ilustra os pesos do mecanismo de atenção.

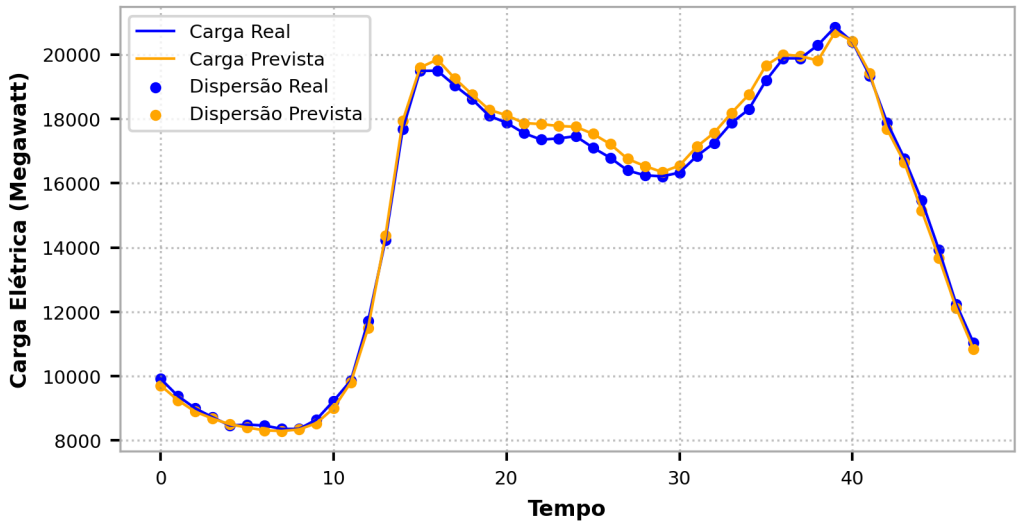
As previsões são mais claramente ilustradas nas Figuras 47, 48 e 49.

Figura 47 - Experimento 2: Curvas de Carga Real e Prevista da ST1B.



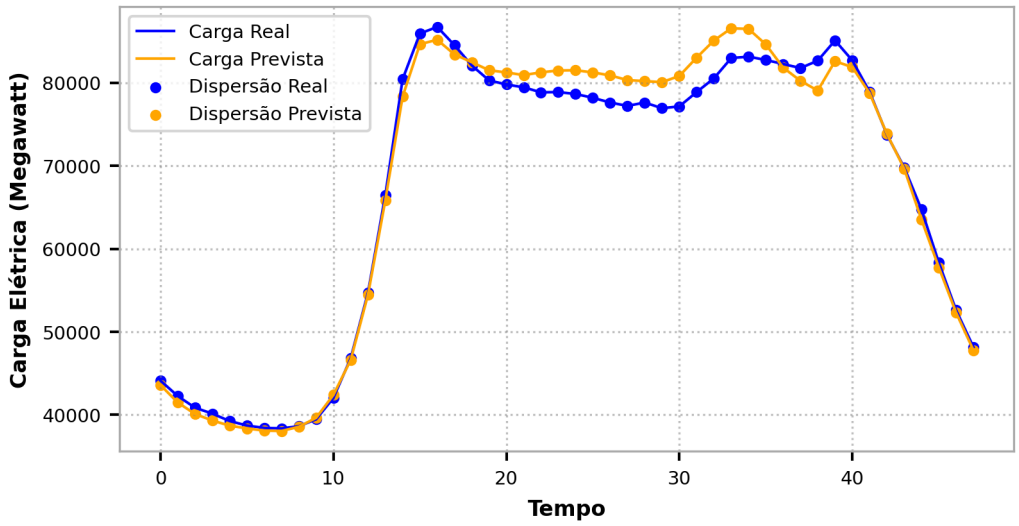
Fonte: Autoria própria.

Figura 48 - Experimento 2: Curvas de Carga Real e Prevista da ST2B.



Fonte: Autoria própria.

Figura 49 - Experimento 2: Curvas de Carga Real e Prevista da ST3B

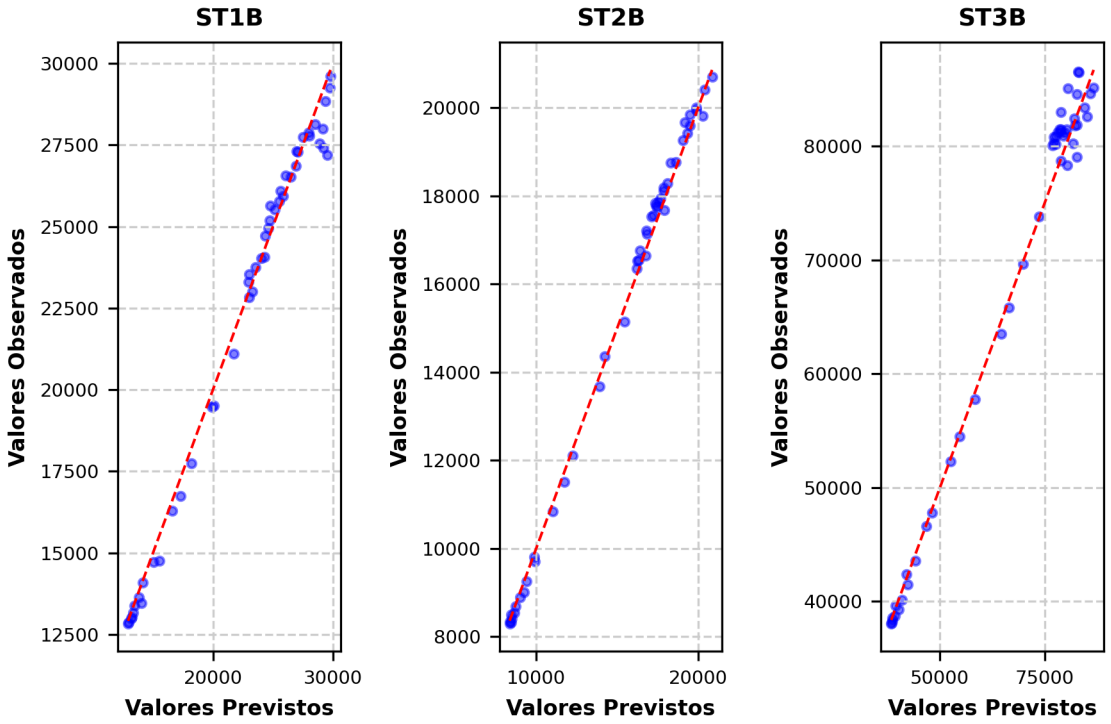


Fonte: Autoria própria.

No contexto das ilustrações apresentadas, o eixo temporal cobre o intervalo de 0 a 50, representando 48 instâncias de dados, que correspondem a intervalos de 30 minutos ao longo de 24 horas. Nessas representações gráficas, a linha azul denota os valores reais de carga, enquanto a linha laranja retrata as previsões. Além disso, é possível observar os pontos de discrepância entre essas duas curvas.

A Figura 50 apresenta o gráfico de dispersão do Experimento 2, permitindo uma análise mais detalhada dos resultados obtidos para as séries temporais ST1B, ST2B e ST3B.

Figura 50 - Gráfico de Dispersão do Experimento 2.



Fonte: Autoria própria.

No gráfico de dispersão apresentado, cada ponto representa um par de valores: um é o valor real e o outro é a previsão obtida. A linha tracejada vermelha destaca os casos em que os valores reais coincidem com as previsões, indicando a precisão exata do modelo nesses pontos.

A Tabela 14 apresenta os valores obtidos no Experimento 2, com base nas métricas MAPE, MAE e RMSE.

Tabela 14 - Métricas de Desempenho do Experimento 2.

Métricas	**ST1B**	**ST2B**	**ST3B**
MAPE(%)	1,86	1,41	2,05
MAE(MW)	428,45	217,74	1514,03
RMSE(MW)	614,87	253,88	1985,77

Fonte:Autoria própria.

6.2.2 Resultados Experimentais do SPMST

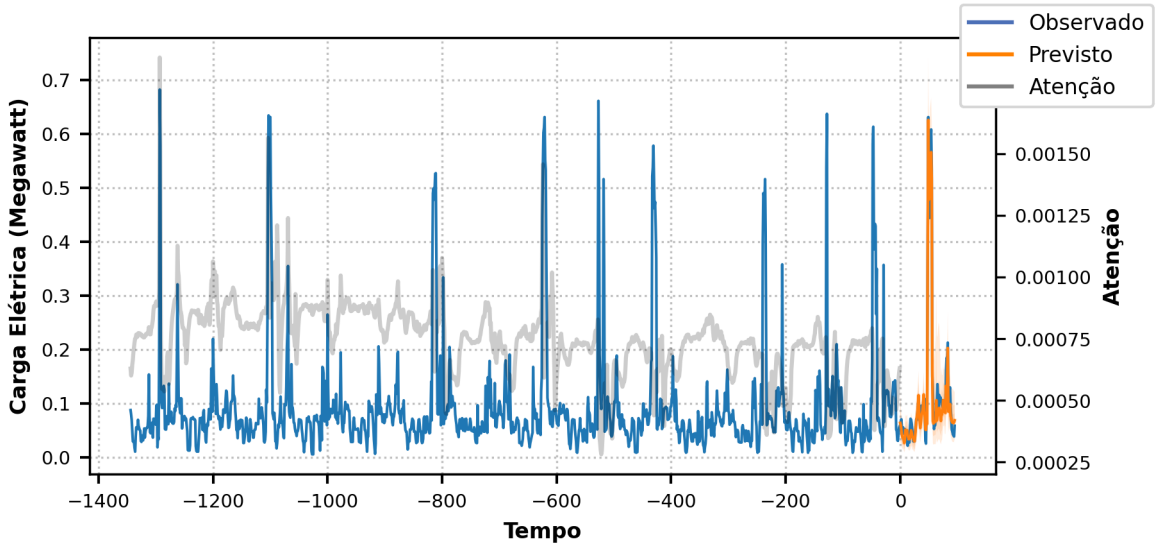
Nos experimentos descritos nesta subseção, a metodologia adotada visou investigar a integração de séries temporais correlacionadas e avaliar o potencial de melhoria na precisão das previsões do SPMST. Para esse fim, foi implementada uma abordagem que integra as séries temporais utilizadas, possibilitando a análise do impacto dessa fusão na qualidade das previsões.

6.2.2.1 Experimento 3: Base de Dados A (3 STs)

No contexto do Experimento 3, o SPMST foi implementado para realizar previsões de demanda utilizando a Base de Dados A como fonte de informações. Neste cenário, as séries temporais ST1A, ST2A e ST3A foram analisadas de forma integrada, resultando em um total de 100,800 amostras para o treinamento do modelo.

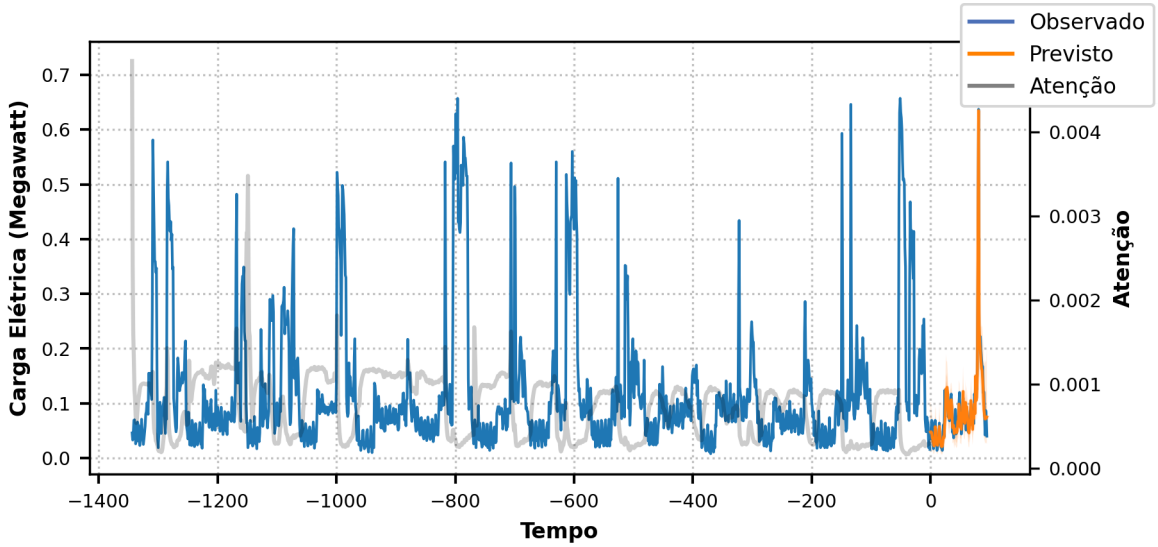
As Figuras 51 a 53 apresentam a janela retrospectiva e as previsões geradas durante o experimento. Nessas representações gráficas, o eixo temporal abrange de 0 a -1400, correspondendo a 1,344 amostras coletadas a cada 15 minutos. Os resultados das previsões são exibidos no último intervalo de tempo. A linha azul indica os valores reais da carga, a linha laranja mostra as previsões geradas e a linha cinza representa os pesos atribuídos pelo mecanismo de atenção.

Figura 51 - Experimento 3 - Janela Retrospectiva e Previsão da ST1A.



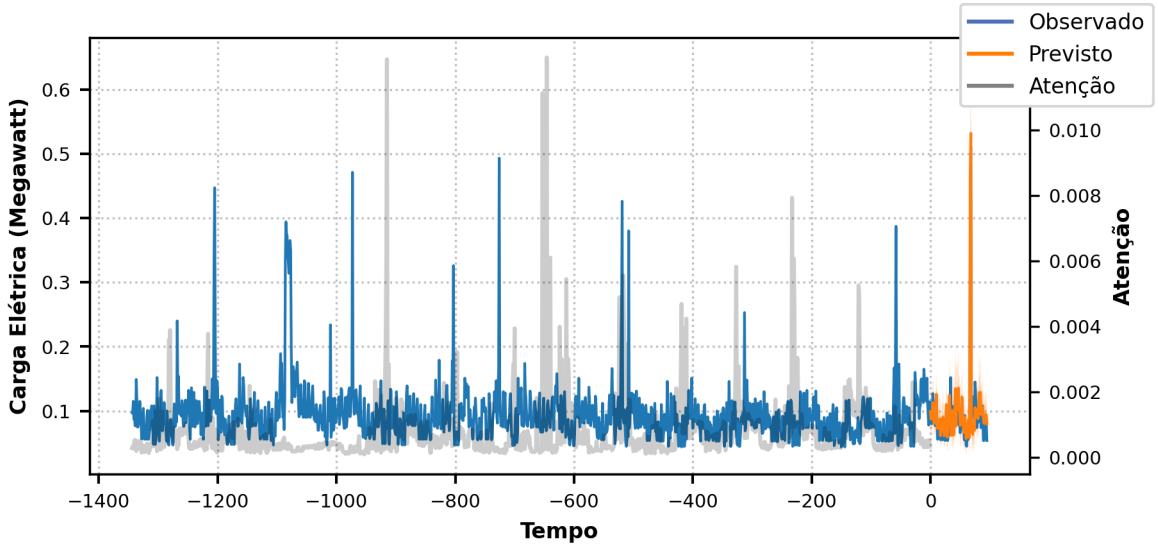
Fonte: Autoria própria.

Figura 52 - Experimento 3 - Janela Retrospectiva e Previsão da ST2A.



Fonte: Autoria própria.

Figura 53 - Experimento 3 - Janela Retrospectiva e Previsão da ST3A.

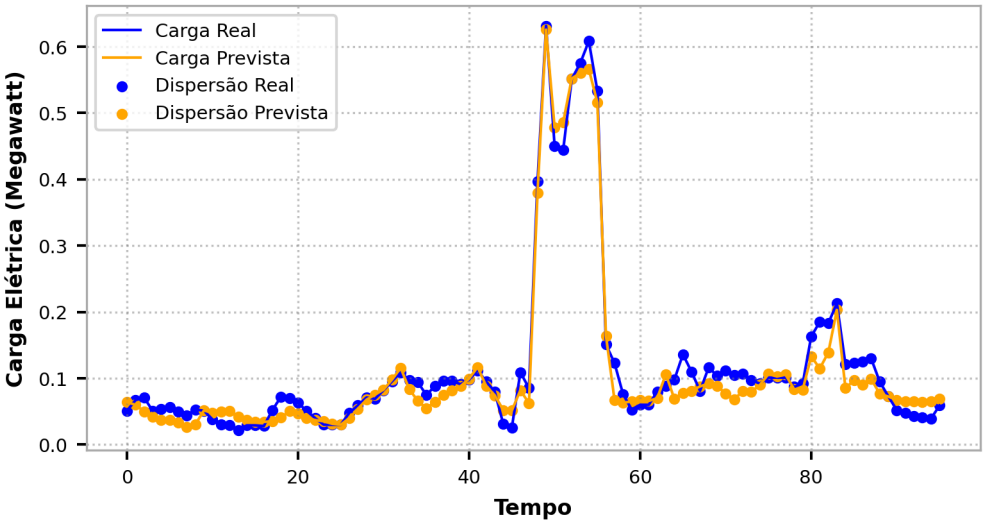


Fonte: Autoria própria.

Uma visualização detalhada das previsões é fornecida pelas Figuras 54, 55 e 56. Nesses gráficos, o eixo temporal abrange o intervalo de 0 a 100, refletindo 96 amostras registradas a cada 15 minutos durante um ciclo de 24 horas. Os valores reais da carga são indicados pela

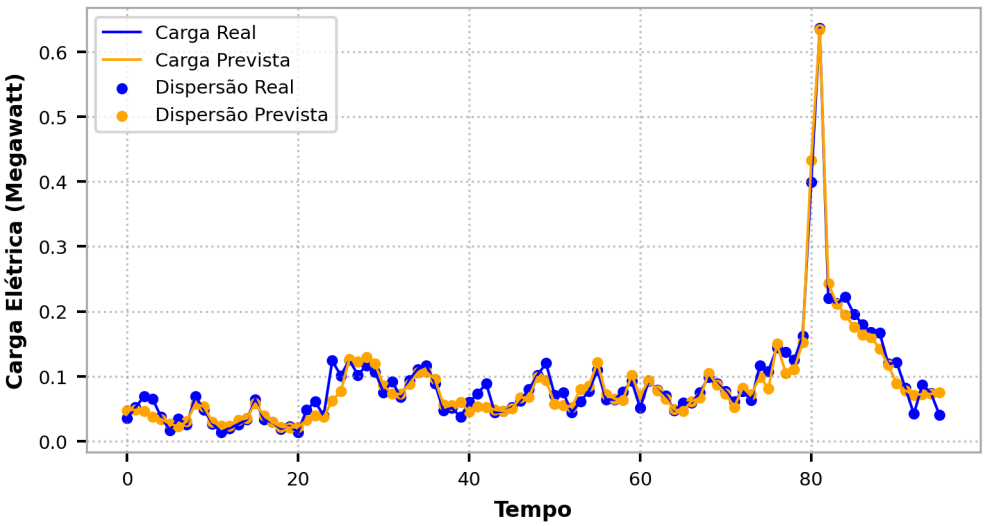
linha azul, com as previsões representadas pela linha laranja. Também, é possível notar as discrepâncias entre essas curvas.

Figura 54 - Experimento 3: Curvas de Carga Real e Prevista da ST1A.



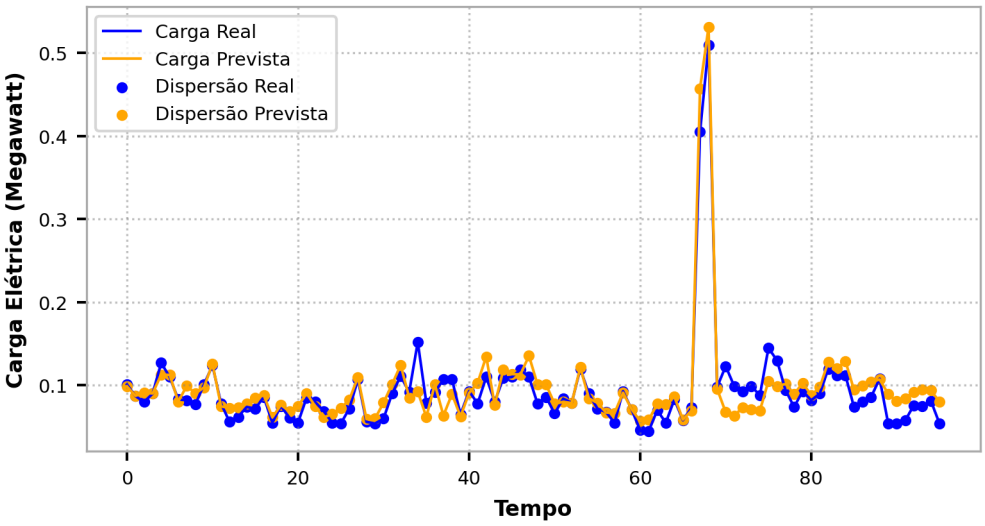
Fonte: Autoria própria.

Figura 55 - Experimento 3: Curvas de Carga Real e Prevista da ST2A.



Fonte: Autoria própria.

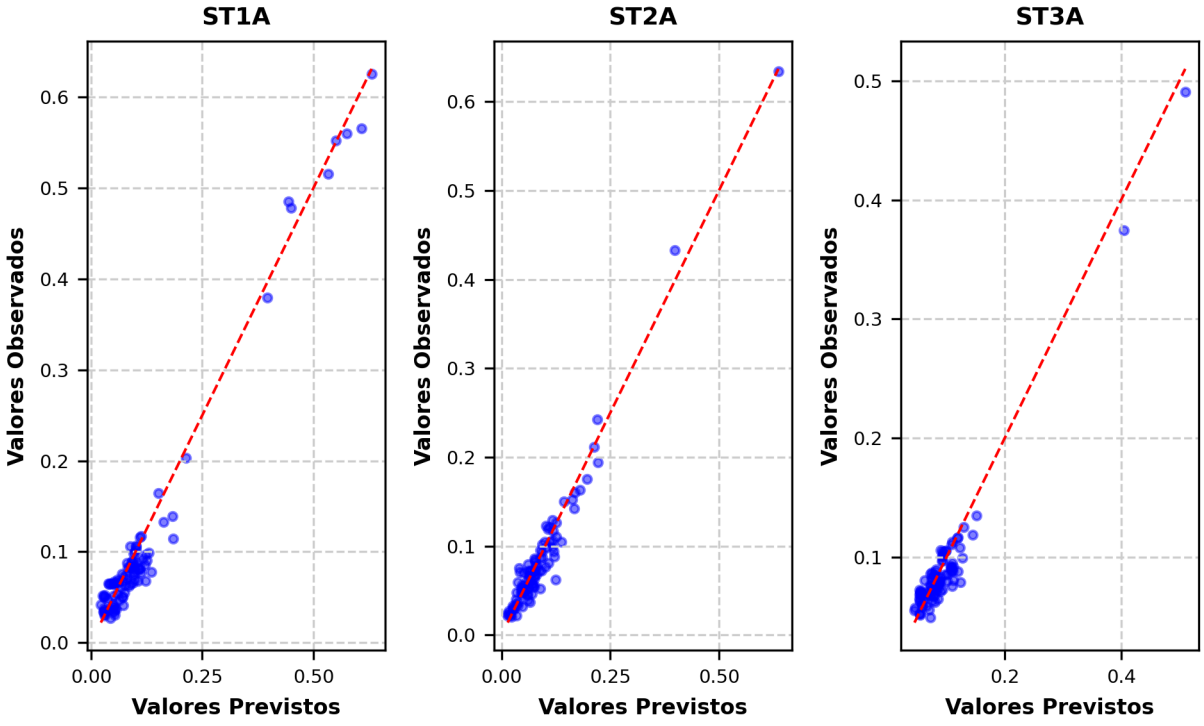
Figura 56 - Experimento 3: Curvas de Carga Real e Prevista da ST3A.



Fonte: Autoria própria.

A Figura 57 exibe o gráfico de dispersão referente ao Experimento 3, apresentando os resultados obtidos para cada série temporal analisada.

Figura 57 - Gráfico de Dispersão do Experimento 3.



Fonte: Autoria própria.

No gráfico apresentado, cada ponto representa um par de valores, sendo um valor dos dados reais e o outro das previsões geradas. A linha vermelha tracejada destaca os pontos em que os valores reais correspondem às previsões.

Os resultados do Experimento 3, conforme os indicadores MAPE, MAE e RMSE, estão detalhados na Tabela 15.

Tabela 15 - Métricas de Desempenho do Experimento 3.

Métricas	**ST1A**	**ST2A**	**ST3A**
MAPE(%)	22,05	23,67	15,67
MAE(MW)	0,02	0,02	0,01
RMSE(MW)	0,02	0,02	0,02

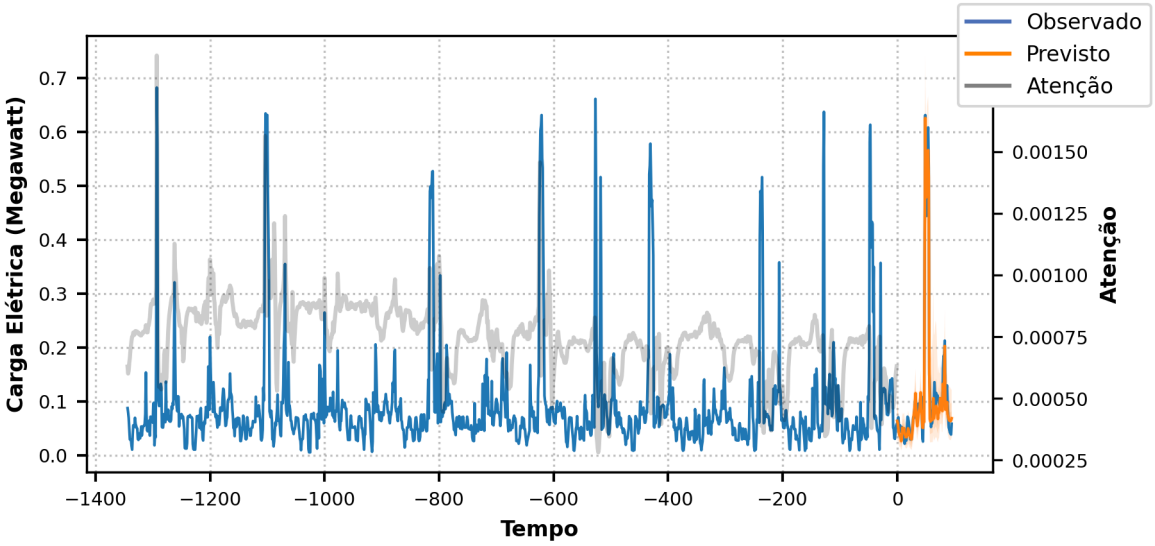
Fonte:Autoria própria.

6.2.2.2 Experimento 4: Base de Dados A (4 STs)

No Experimento 4, foi aplicado o SPMST com o objetivo de prever a demanda energética. Nesta análise, foram integradas quatro séries temporais pertencentes à Base de Dados A: ST1A, ST2A, ST3A e ST4A. Desta forma, o treinamento do modelo foi realizado utilizando um total de 134,400 amostras.

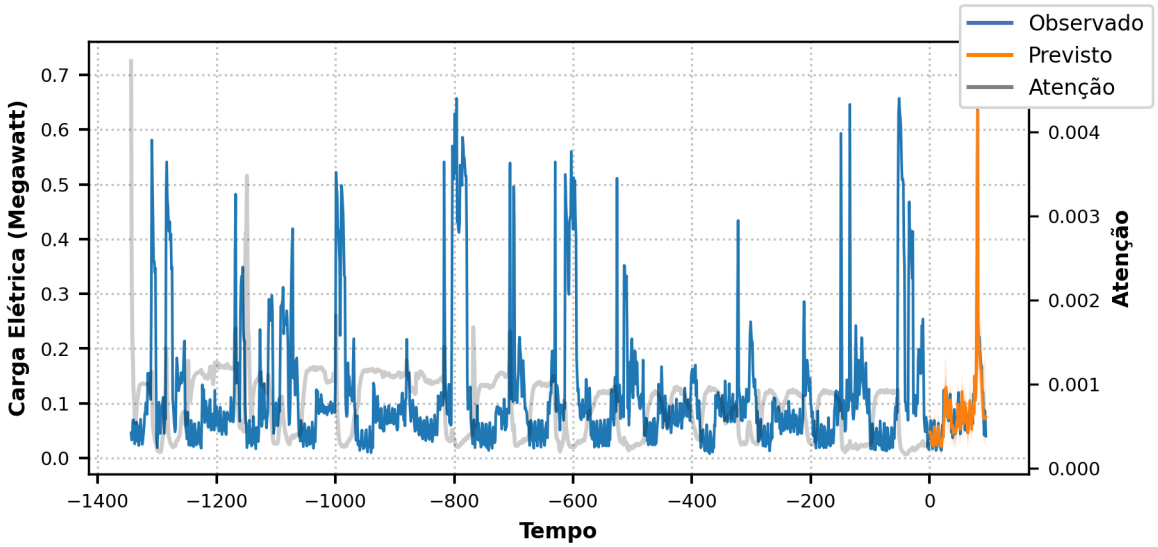
A janela retrospectiva e as previsões associadas a este experimento são detalhadas nas Figuras 58 a 61.

Figura 58 - Experimento 4 - Janela Retrospectiva e Previsão da ST1A.



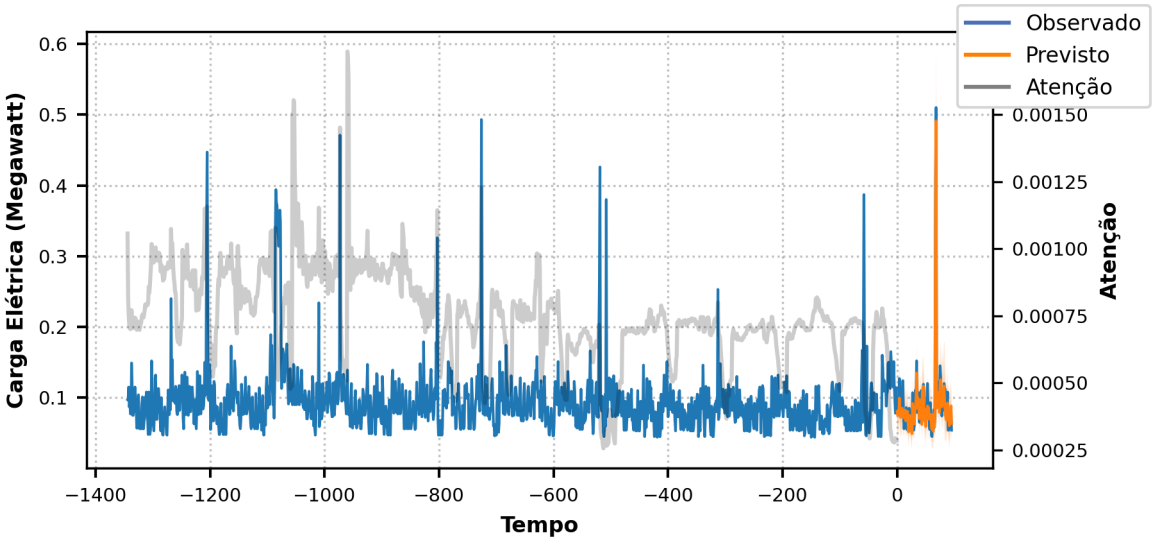
Fonte: Autoria própria.

Figura 59 - Experimento 4 - Janela Retrospectiva e Previsão da ST2A.



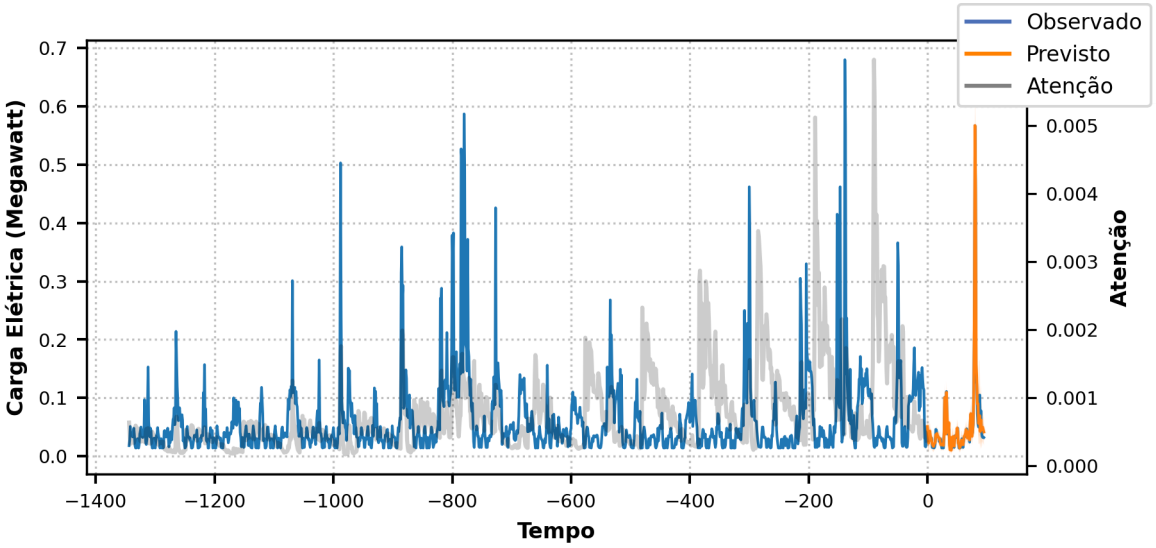
Fonte: Autoria própria.

Figura 60 - Experimento 4 - Janela Retrospectiva e Previsão da ST3A.



Fonte: Autoria própria.

Figura 61 - Experimento 4 - Janela Retrospectiva e Previsão da ST4A.

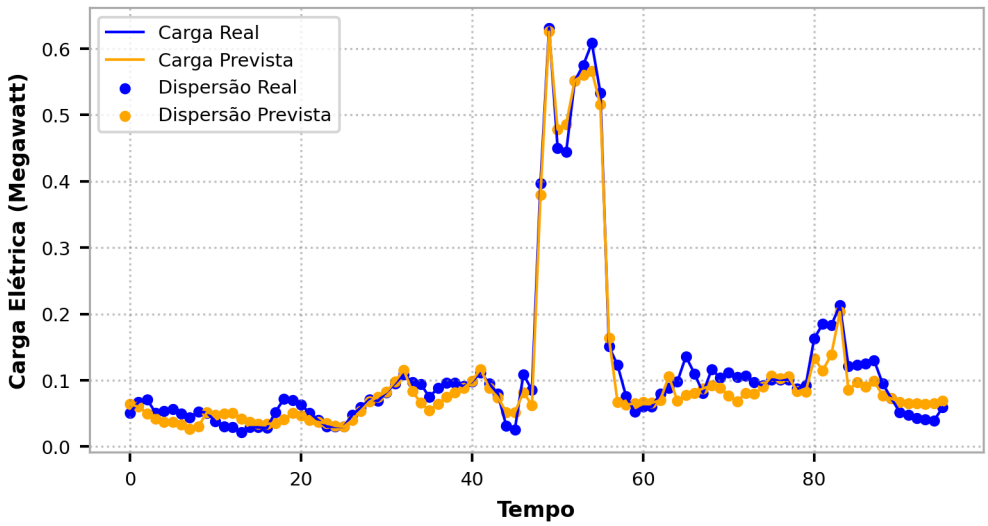


Fonte: Autoria própria.

Nas ilustrações apresentadas, o eixo temporal estende-se de 0 a -1400, representando 1,344 amostras obtidas a cada 15 minutos. No último intervalo, são apresentados os resultados das previsões. A linha azul indica os valores reais de carga, a linha laranja demonstra as previsões obtidas e a linha cinza destaca os pesos atribuídos pelo mecanismo de atenção.

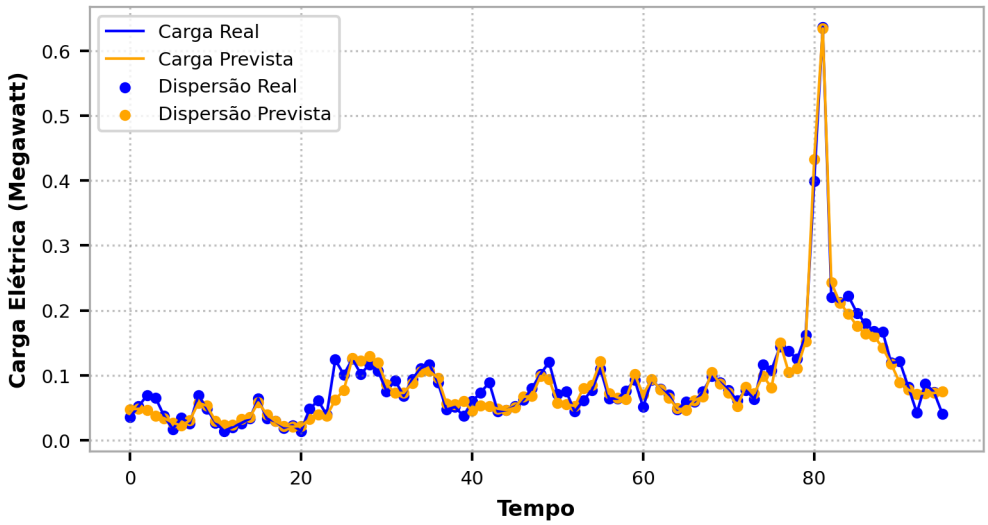
As Figuras 40, 41 e 42 oferecem uma visualização mais clara das previsões geradas.

Figura 62 - Experimento 4: Curvas de Carga Real e Prevista da ST1A.



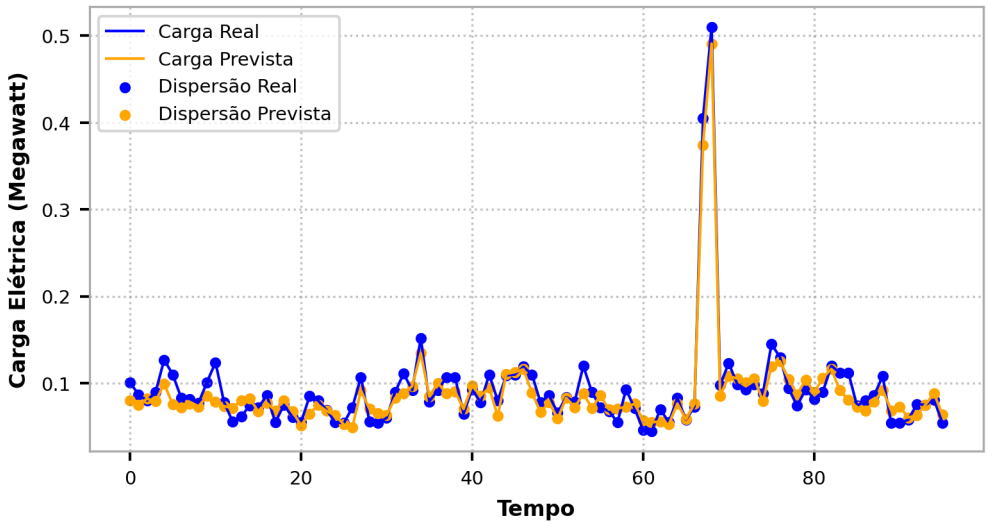
Fonte: Autoria própria.

Figura 63 - Experimento 4: Curvas de Carga Real e Prevista da ST2A.



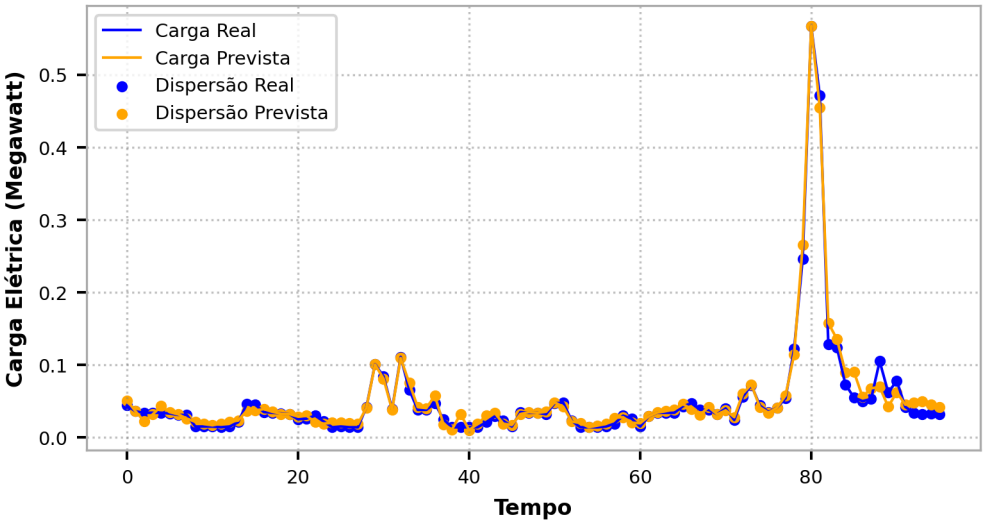
Fonte: Autoria própria.

Figura 64 - Experimento 4: Curvas de Carga Real e Prevista da ST3A.



Fonte: Autoria própria.

Figura 65 - Experimento 4: Curvas de Carga Real e Prevista da ST4A.



Fonte: Autoria própria.

Nas representações fornecidas, o eixo temporal abrange o intervalo de 0 a 100, correspondendo a 96 pontos de dados em intervalos de 15 minutos ao longo de 24 horas. A linha azul denota os valores observados, enquanto a linha laranja representa as previsões. Os pontos de dispersão ao longo dessas duas curvas também podem ser observados.

Os valores obtidos no Experimento 4, baseados nos indicadores MAPE, MAE e RMSE, estão apresentados na Tabela 18.

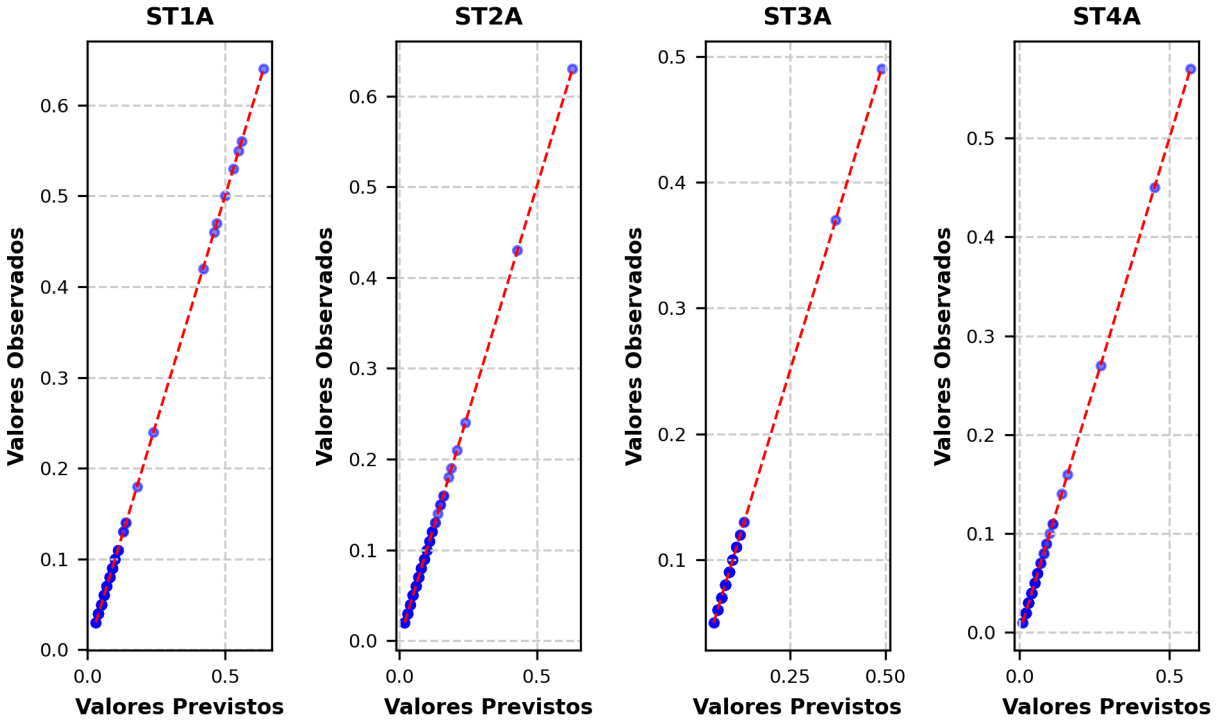
Tabela 16 - Métricas de Desempenho do Experimento 4.

Métricas	**ST1A**	**ST2A**	**ST3A**	**ST4A**
MAPE(%)	17,96	17,40	13,01	18,05
MAE(MW)	0,02	0,01	0,01	0,01
RMSE(MW)	0,02	0,02	0,01	0,02

Fonte:Autoria própria.

A Figura 66 mostra o gráfico de dispersão do Experimento 4, permitindo uma análise abrangente dos resultados das sequências temporais ST1A, ST2A, ST3A e ST4A. Cada ponto no gráfico representa um par de valores, sendo um valor correspondente aos dados observados e o outro às previsões geradas pelo modelo. A linha vermelha tracejada destaca os pares em que há uma correspondência entre os dados reais e as previsões.

Figura 66 - Gráfico de Dispersão do Experimento 4.



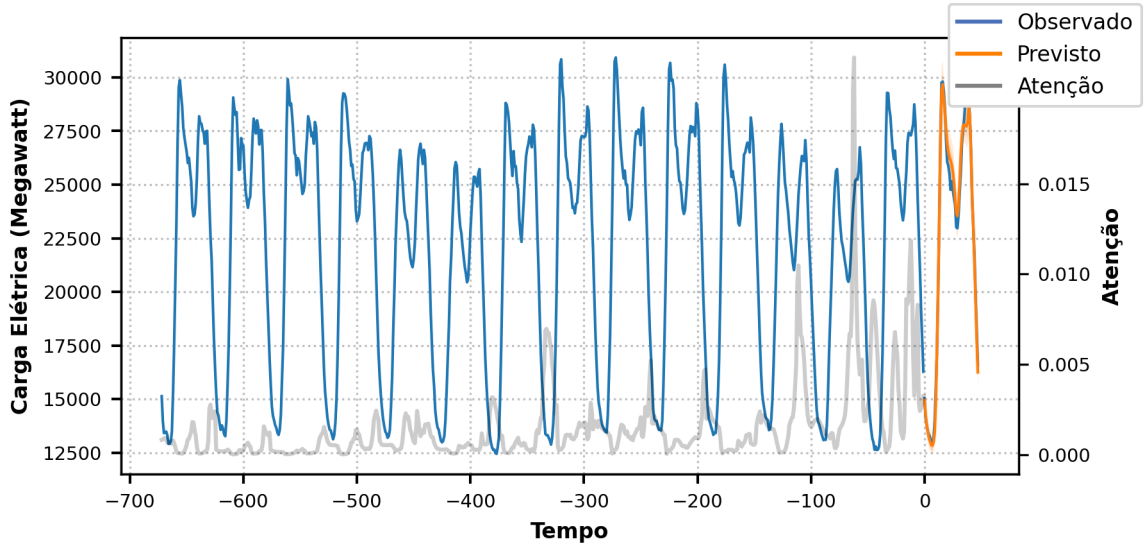
Fonte: Autoria própria.

6.2.2.3 Experimento 5: Base de Dados B (3 STs)

No Experimento 5, o SPMST foi empregado para estimar a demanda de eletricidade com base na Base de Dados B. Neste cenário, foram integradas as séries temporais ST1B, ST2B e ST3B, com o objetivo de melhorar a precisão das previsões e capturar variações sazonais e dinâmicas de curto e médio prazo da demanda. Como resultado dessa integração, o modelo foi treinado utilizando um total de 116,064 amostras.

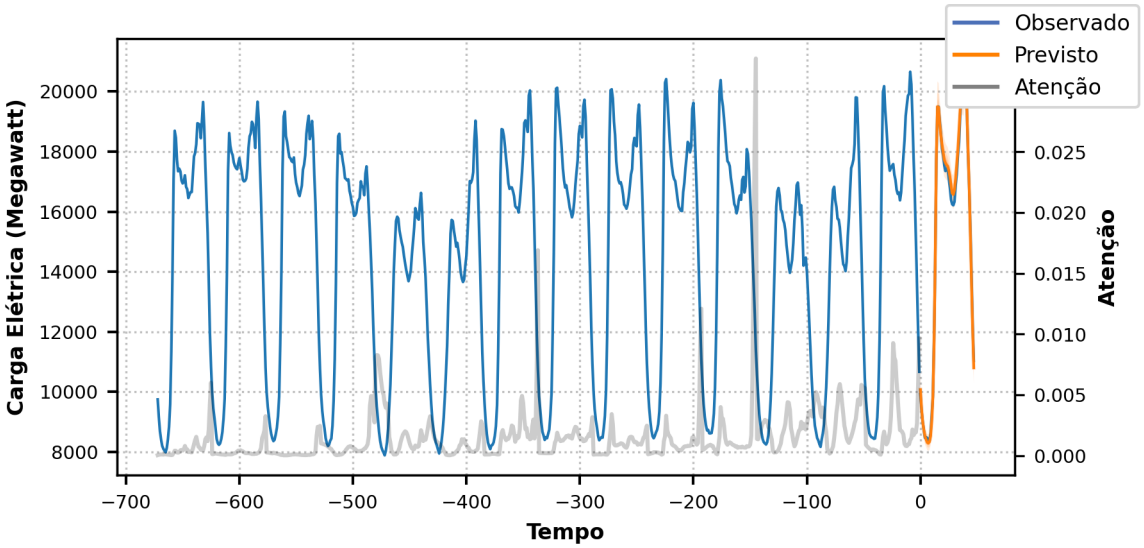
As Figuras 67 a 69 ilustram a janela retrospectiva de 14 dias, cujo eixo temporal se estende de 0 a -700, abrangendo um total de 672 amostras coletadas a cada 30 minutos. No último intervalo de tempo, são apresentados os resultados das previsões. Nos gráficos, a linha azul representa os valores reais da carga elétrica, enquanto a linha laranja exibe as previsões geradas pelo modelo, evidenciando o alinhamento entre as previsões e os dados observados. Além disso, a linha cinza demonstra os pesos atribuídos pelo mecanismo de atenção, permitindo uma interpretação mais detalhada sobre quais momentos históricos tiveram maior relevância para as previsões.

Figura 67 - Experimento 5 - Janela Retrospectiva e Previsão da ST1B.



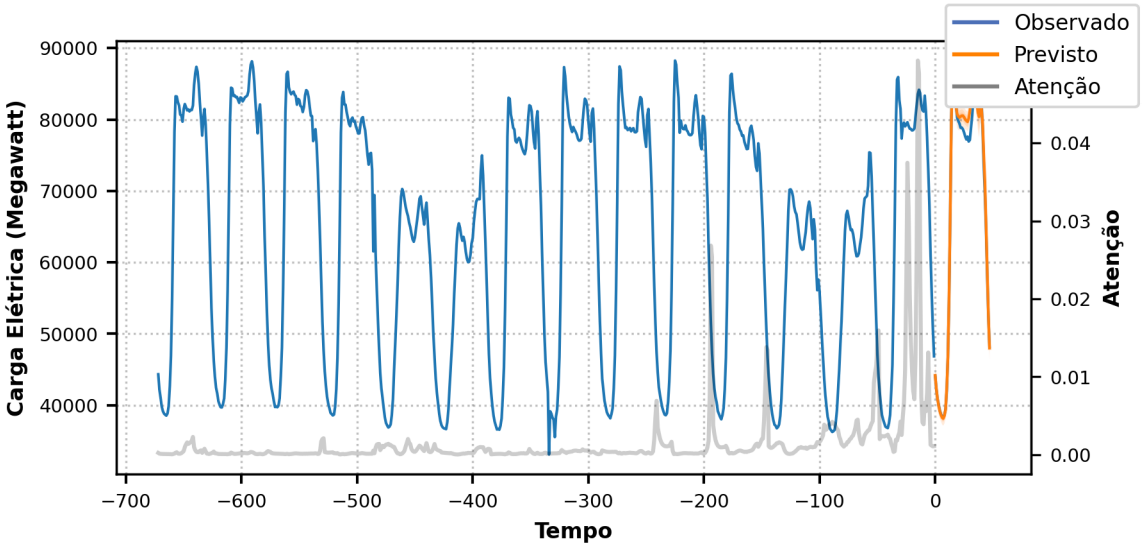
Fonte: Autoria própria.

Figura 68 - Experimento 5 - Janela Retrospectiva e Previsão da ST2B.



Fonte: Autoria própria.

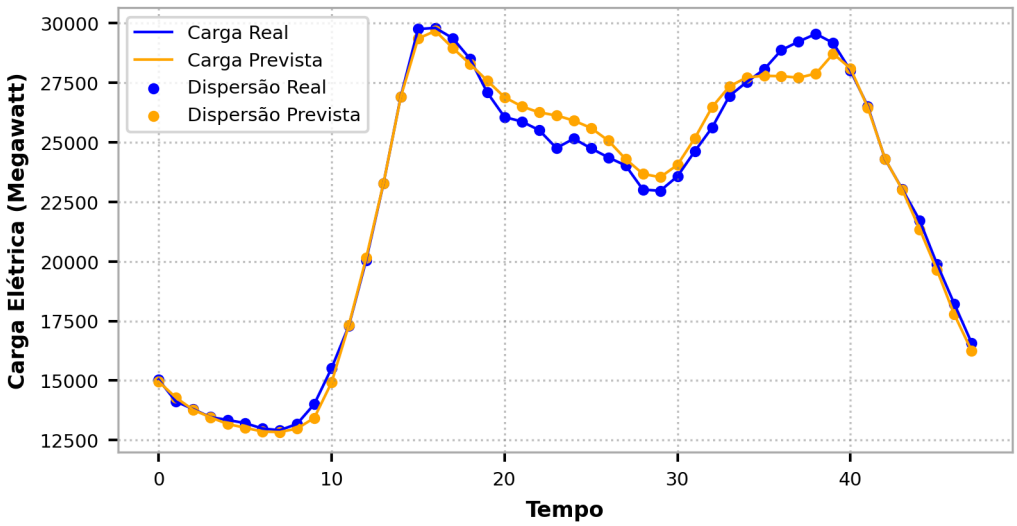
Figura 69 - Experimento 5 - Janela Retrospectiva e Previsão da ST3B.



Fonte: Autoria própria.

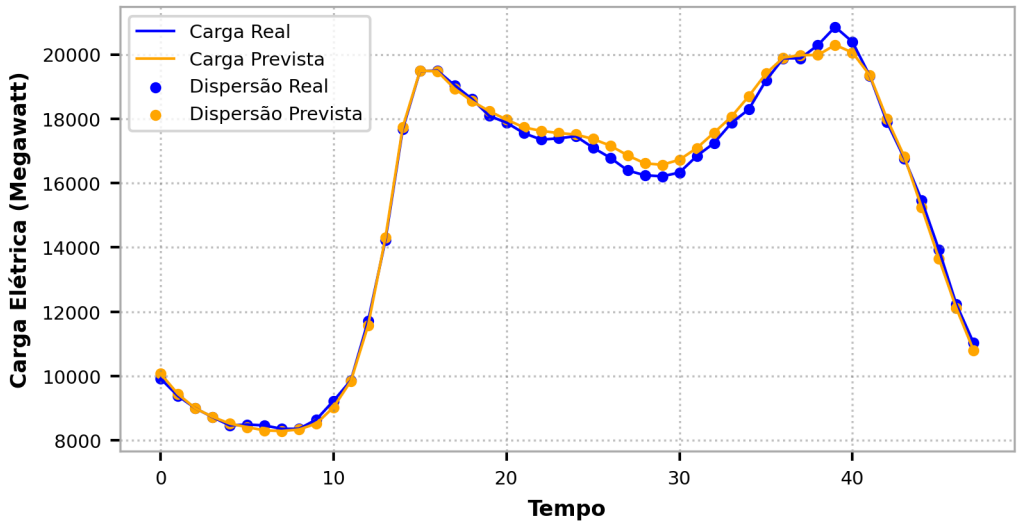
As previsões são detalhadamente ilustradas nas Figuras 70, 71 e 72, nas quais o eixo temporal abrange o intervalo de 0 a 50, correspondendo a 48 registros históricos em intervalos de 30 minutos ao longo de um período de 24 horas. Nestes gráficos, a linha azul representa os valores reais de carga elétrica, enquanto a linha laranja exibe as previsões geradas pelo modelo. Além disso, permitem a identificação visual dos pontos de discrepância entre as curvas de valores reais e as previsões.

Figura 70 - Experimento 5: Curvas de Carga Real e Prevista da ST1B.



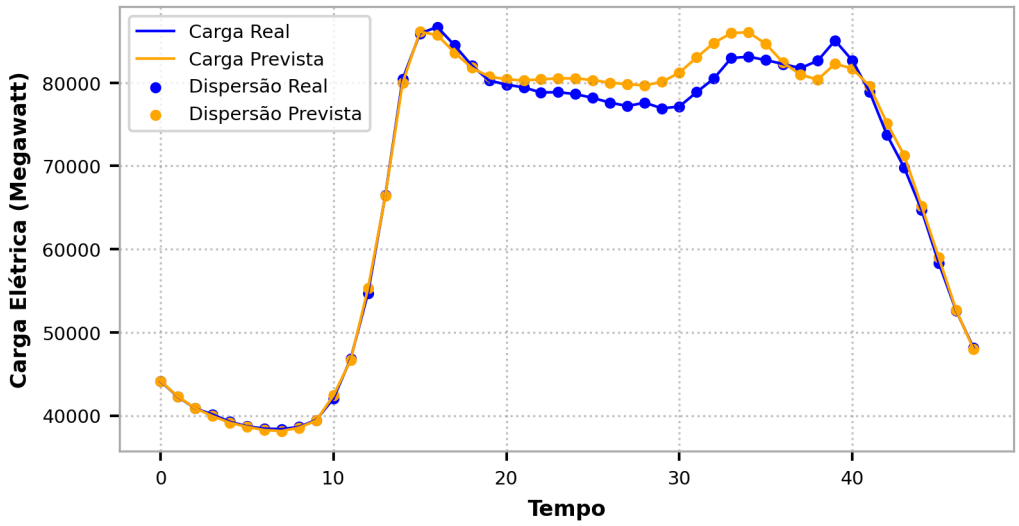
Fonte: Autoria própria.

Figura 71 - Experimento 5: Curvas de Carga Real e Prevista da ST2B.



Fonte: Autoria própria.

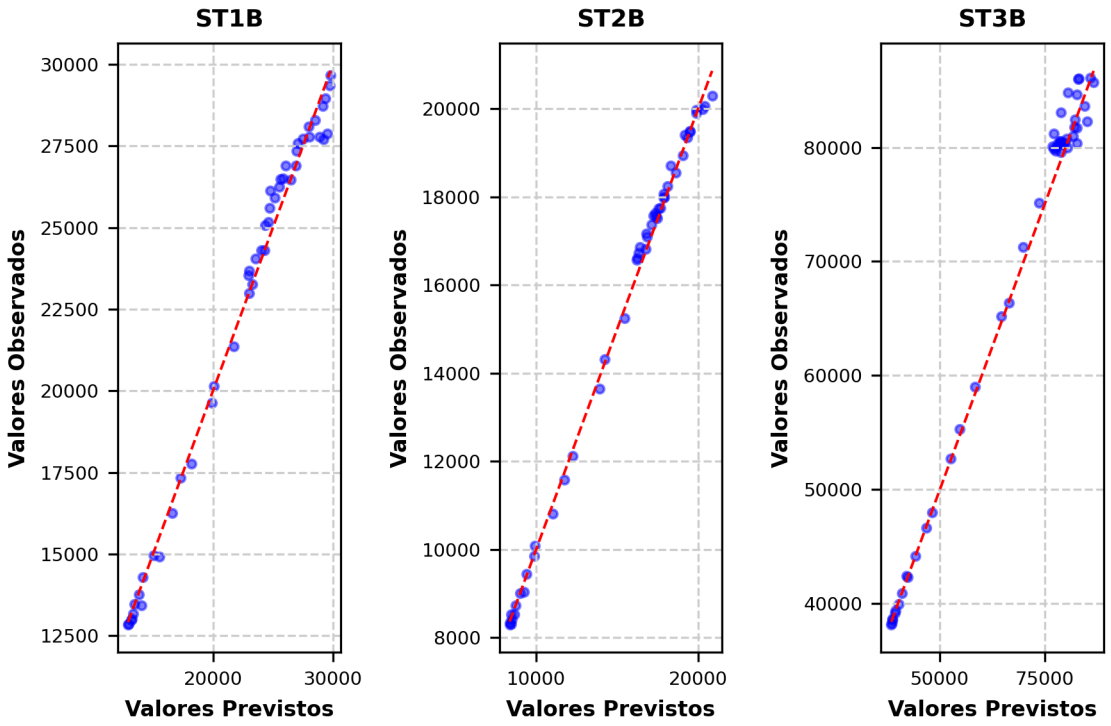
Figura 72 - Experimento 5: Curvas de Carga Real e Prevista da ST3B



Fonte: Autoria própria.

O gráfico de dispersão do Experimento 5 está apresentado na Figura 73, proporcionando uma visão mais detalhada dos resultados das séries temporais ST1B, ST2B e ST3B.

Figura 73 - Gráfico de Dispersão do Experimento 5.



Fonte: Autoria própria.

No gráfico, os pontos representam pares de valores, correspondendo aos dados reais e às previsões geradas. A linha tracejada vermelha sublinha os casos em que os valores reais coincidem exatamente com as previsões.

Para uma análise mais detalhada dos resultados, a Tabela 17 apresenta os valores obtidos no Experimento 5 por meio dos indicadores MAPE, MAE e RMSE.

Tabela 17 - Métricas de Desempenho do Experimento 5.

Métricas	**ST1B**	**ST2B**	**ST3B**
MAPE(%)	1,81	1,13	1,55
MAE(MW)	427,26	173,58	1185,57
RMSE(MW)	581,11	221,28	1696,42

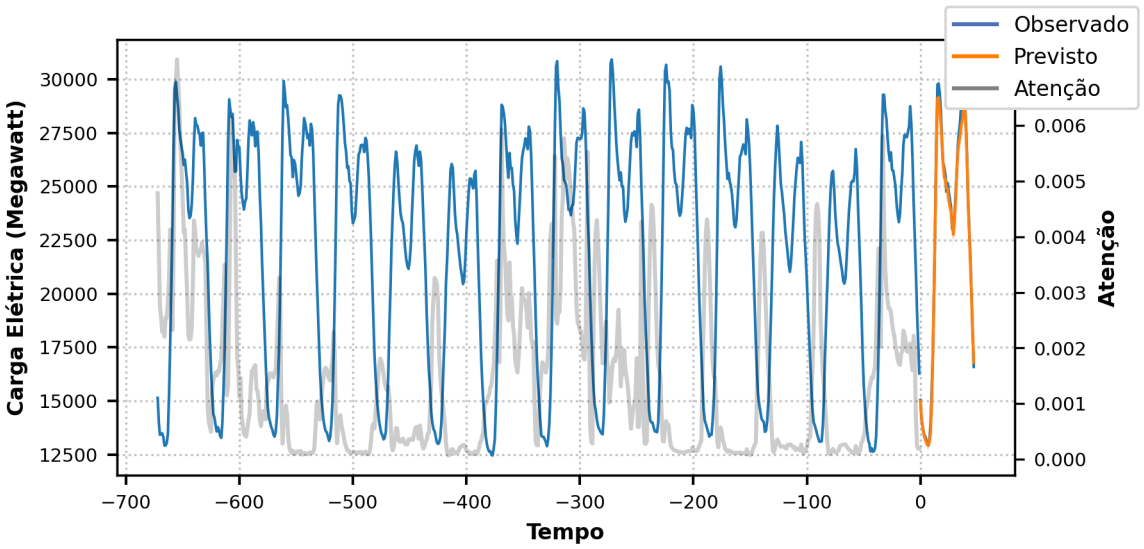
Fonte:Autoria própria.

6.2.2.4 Experimento 6: Base de Dados B (4 STs)

No Experimento 6, o SPMST foi aplicado para estimar a demanda elétrica utilizando a Base de Dados B. Neste cenário, foram integradas as séries temporais ST1B, ST2B, ST3B e ST4B,

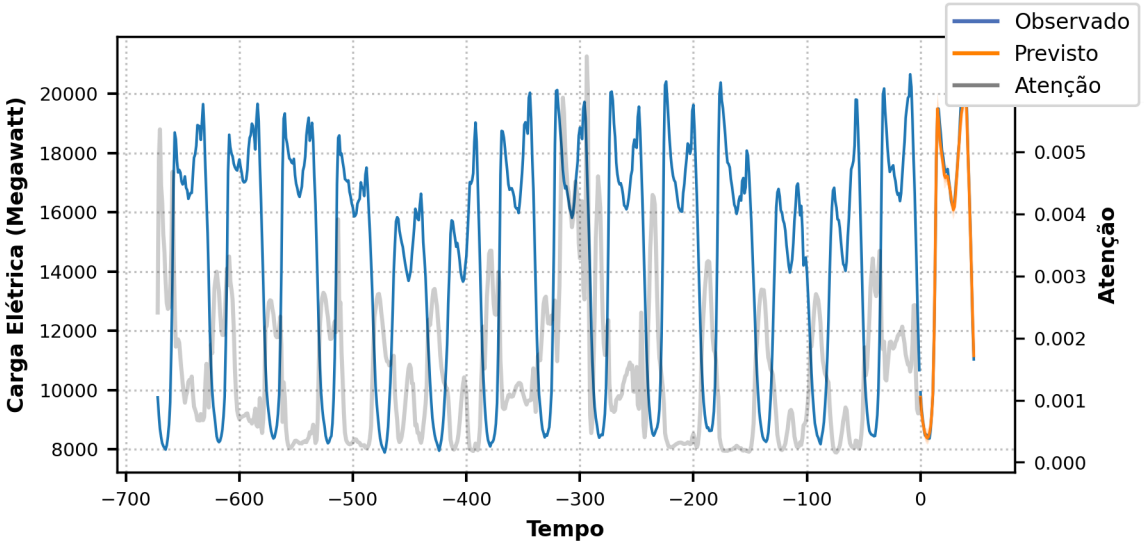
resultando em um total de 154,752 amostras utilizadas no treinamento do modelo. As Figuras 74 a 77 ilustram a janela retrospectiva e as previsões obtidas neste experimento.

Figura 74 - Experimento 6 - Janela Retrospectiva e Previsão da ST1B.



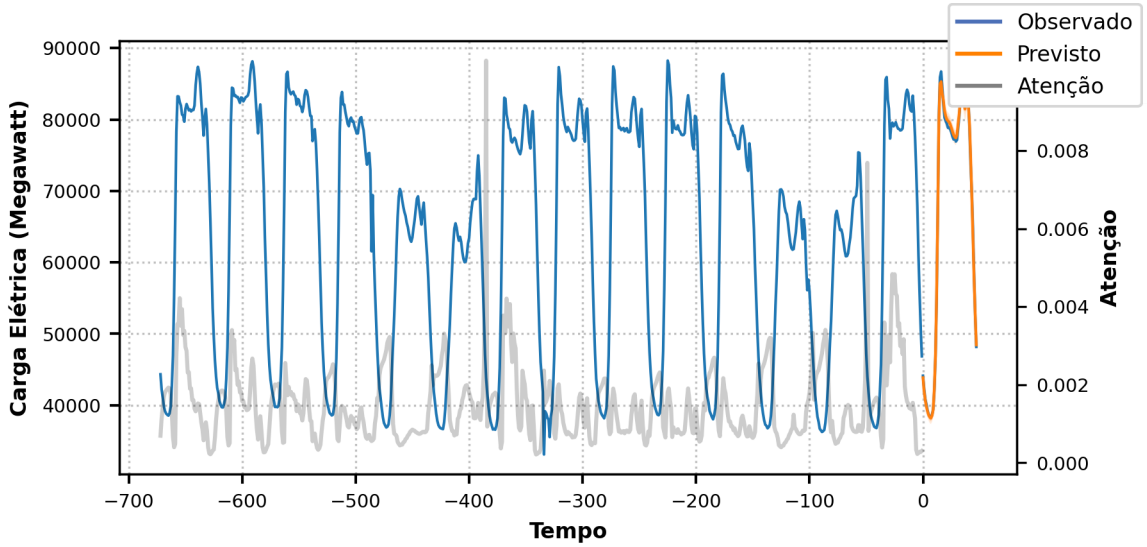
Fonte: Autoria própria.

Figura 75 - Experimento 6 - Janela Retrospectiva e Previsão da ST2B.



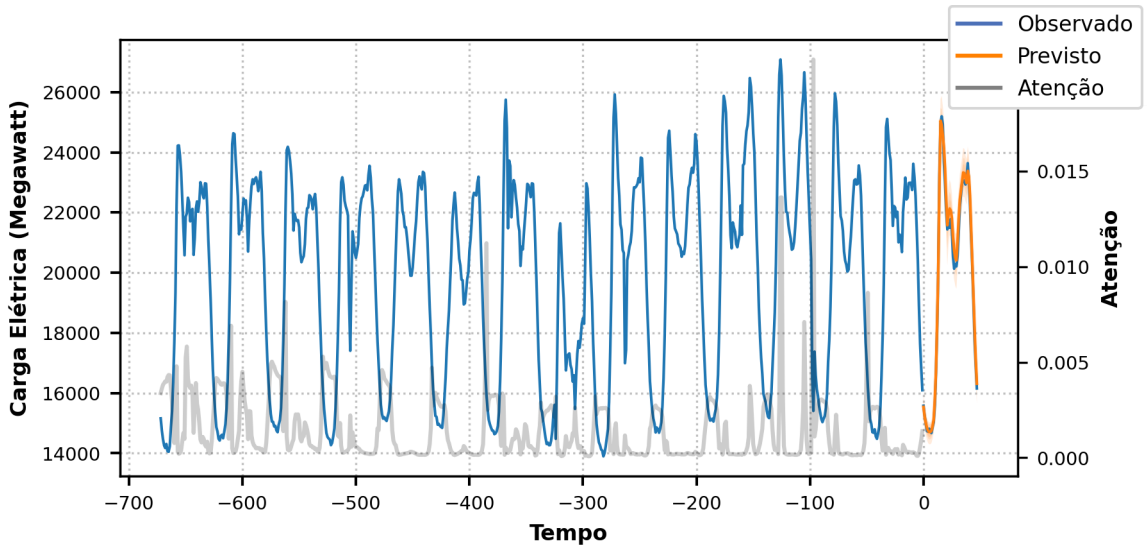
Fonte: Autoria própria.

Figura 76 - Experimento 6 - Janela Retrospectiva e Previsão da ST3B.



Fonte: Autoria própria.

Figura 77 - Experimento 6 - Janela Retrospectiva e Previsão da ST4B.



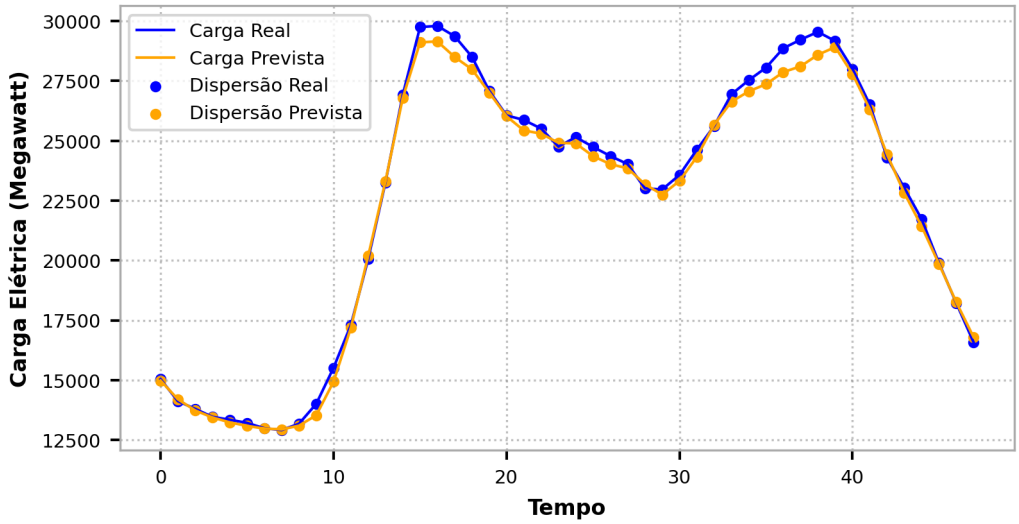
Fonte: Autoria própria.

Nas imagens gráficas fornecidas, o eixo temporal se estende de 0 a -700, abrangendo 672 amostras obtidas a cada 30 minutos. No último intervalo, estão disponíveis os resultados das previsões. A linha azul representa os valores reais da carga, a linha laranja corresponde às previsões, e a linha cinza denota os pesos do mecanismo de atenção.

As Figuras 78, 79, 80 e 81 oferecem uma perspectiva detalhada das previsões. Nesses

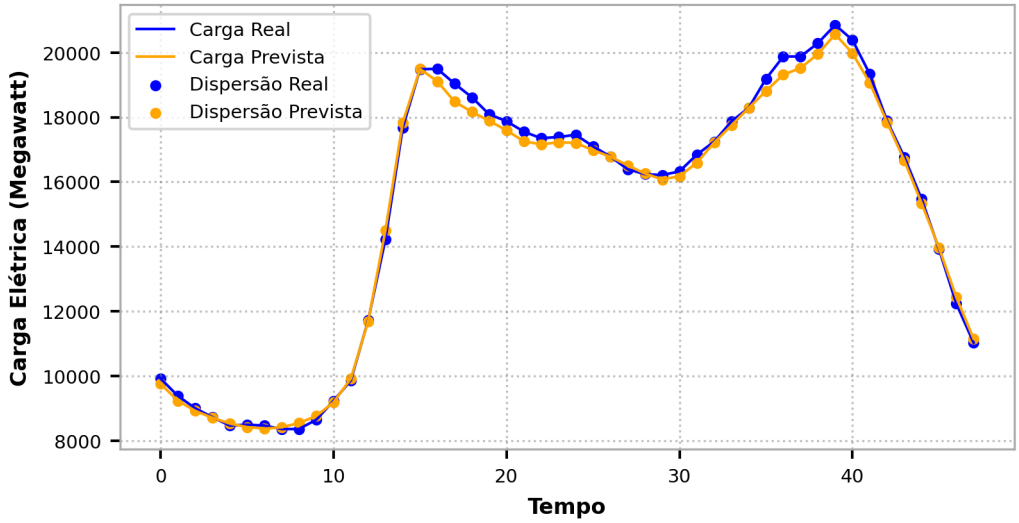
gráficos, o eixo temporal varia de 0 a 50, representando 48 instâncias em intervalos de 30 minutos ao longo de 24 horas. A linha azul representa os dados observados de carga, enquanto a linha laranja indica os valores previstos. Além disso, as variações nos pontos de dispersão entre essas duas curvas também podem ser observadas.

Figura 78 - Experimento 6: Curvas de Carga Real e Prevista da ST1B.



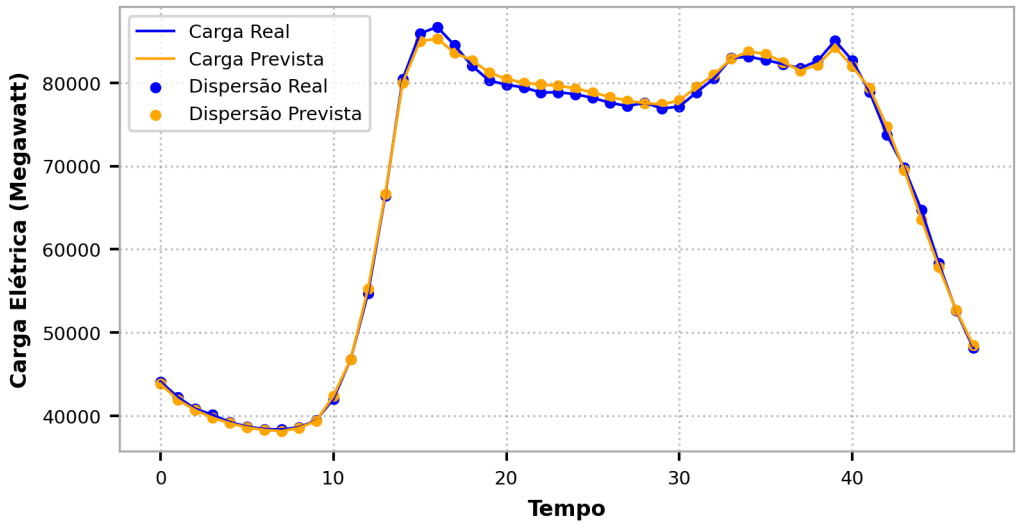
Fonte: Autoria própria.

Figura 79 - Experimento 6: Curvas de Carga Real e Prevista da ST2B.



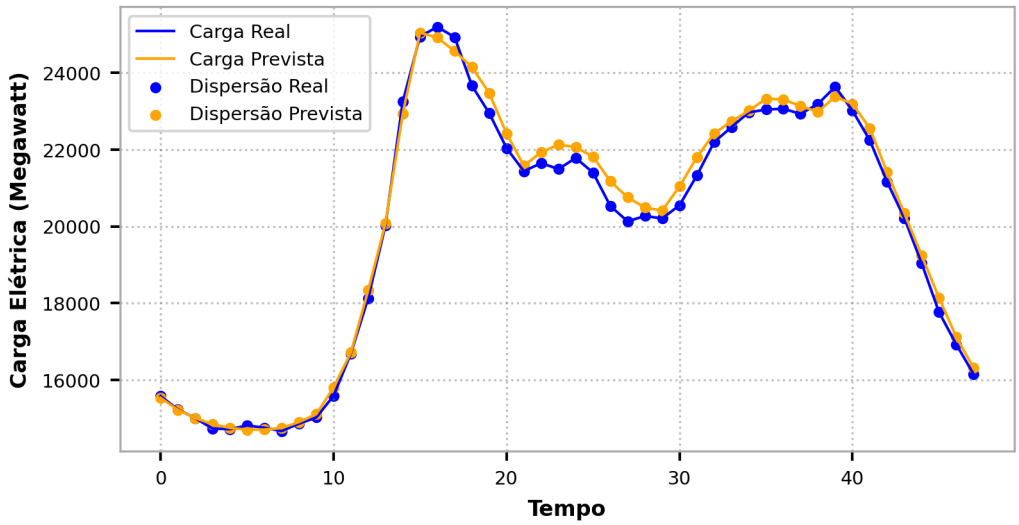
Fonte: Autoria própria.

Figura 80 - Experimento 6: Curvas de Carga Real e Prevista da ST3B



Fonte: Autoria própria.

Figura 81 - Experimento 6: Curvas de Carga Real e Prevista da ST4B

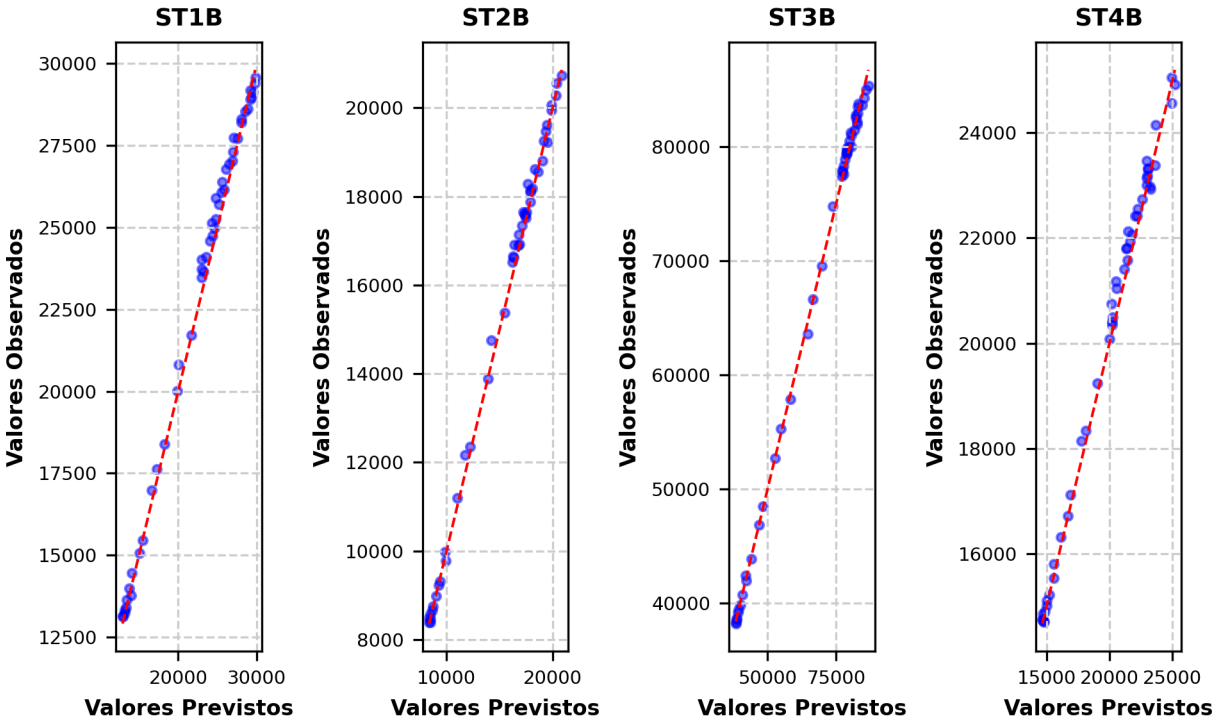


Fonte: Autoria própria.

O gráfico de dispersão do Experimento 6, apresentado na Figura 82, oferece uma perspectiva pormenorizada dos resultados obtidos para as séries temporais ST1B, ST2B e ST3B, com a adição da ST4B. Neste contexto, cada ponto no gráfico representa um par de valores, no qual um deles corresponde aos dados reais e o outro às previsões geradas pelo modelo. A linha tracejada vermelha ressalta os casos em que há uma correspondência significativa entre os valores reais e as previsões, permitindo uma visualização clara das áreas em que o modelo demonstrou

um desempenho satisfatório.

Figura 82 - Gráfico de Dispersão do Experimento 6.



Fonte: Autoria própria.

Os resultados do Experimento 6, obtidos por meio dos indicadores MAPE (*Mean Absolute Percentage Error*), MAE (*Mean Absolute Error*) e RMSE (*Root Mean Square Error*), estão detalhados na Tabela 18, que apresenta os valores alcançados por cada sequência temporal analisada.

Tabela 18 - Métricas de Desempenho do Experimento 6.

Métricas	**ST1B**	**ST2B**	**ST3B**	**ST4B**
MAPE(%)	1,20	1,12	0,69	1,13
MAE(MW)	290,58	178,14	496,76	236,05
RMSE(MW)	399,82	227,69	594,12	290,16

Fonte:Autoria própria.

6.3 Discussão dos Resultados

Nesta discussão, serão analisados os resultados dos seis experimentos realizados, com ênfase nas diferenças de desempenho entre os métodos SPUST e SPMST, bem como no impacto da quantidade de séries temporais na precisão das previsões. Portanto, a análise visa oferecer uma compreensão aprofundada das variáveis envolvidas e examinar como cada abordagem afeta a acurácia dos modelos preditivos.

6.3.1 Considerações sobre os Experimentos com a Base de Dados A

No Experimento 1, aplicou-se o SPUST às sequências históricas ST1A, ST2A e ST3A. Os resultados mostraram variações significativas nas métricas de desempenho entre as diferentes sequências. A ST3A apresentou o menor erro percentual absoluto médio, de 17,18%, indicando um desempenho superior na previsão em comparação com as outras séries. Em contraste, a ST2A revelou o maior MAPE, de 27,15%, o que pode ser atribuído a características específicas ou a uma maior variabilidade na demanda desta sequência em particular. A ST1A demonstrou um MAPE intermediário, refletindo uma precisão de previsão que se situou entre os valores observados para ST2A e ST3A.

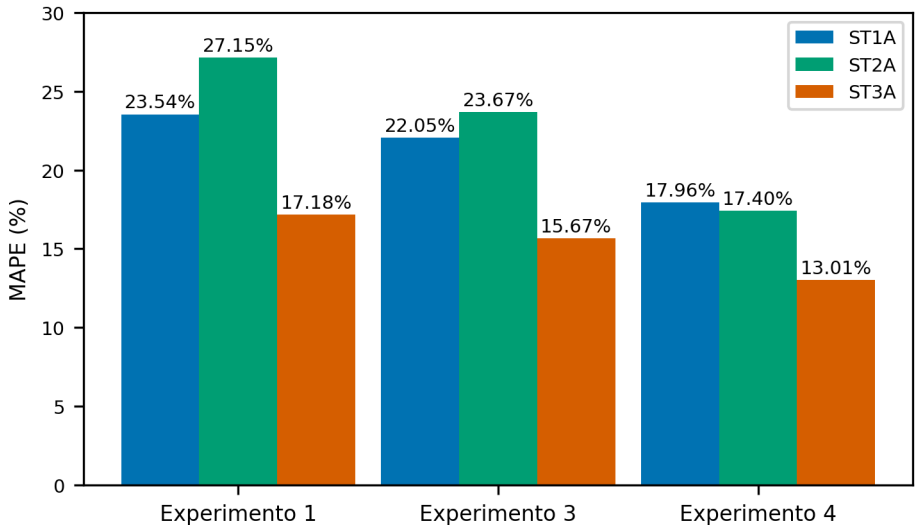
No Experimento 3, aplicou-se o SPMST para integrar as três sequências anteriormente mencionadas. Essa combinação resultou em uma leve melhoria na precisão das previsões, com o MAPE para a ST2A diminuindo para 23,67% em comparação ao Experimento 1. Este resultado sugere que a fusão das séries temporais proporcionou benefícios adicionais, especialmente para a ST2A, que apresentava o desempenho individual menos satisfatório. Embora as demais sequências também tenham mostrado melhorias, estas foram menos expressivas. Em contraste, no Experimento 4, a incorporação da quarta série temporal (ST4A) resultou em uma minimização geral dos valores de MAPE. A inclusão da ST4A trouxe um equilíbrio mais uniforme ao modelo, com reduções no erro para as séries ST1A, ST2A e ST3A, indicando que a integração de uma série temporal adicional aprimorou a capacidade do modelo de capturar padrões complexos e reduzir erros nas previsões.

A Figura 83 ilustra a comparação do MAPE entre os três experimentos conduzidos com as séries temporais ST1A, ST2A e ST3A. O gráfico de barras exibe a precisão das previsões em termos de MAPE para cada série temporal ao longo dos experimentos. Os resultados são representados por barras coloridas, cada uma correspondendo a uma série temporal específica e categorizada conforme o experimento. A série ST1A é representada pela cor azul, a série ST2A pela cor verde, e a série ST3A pela cor laranja.

Na referida ilustração gráfica, é possível observar que a ST3A apresenta consistentemente o menor MAPE, evidenciando um desempenho superior na previsão em todos os experimentos

comparados. Em contraste, a ST2A exibe o maior MAPE, especialmente no Experimento 1, sugerindo maior variabilidade ou características específicas que impactam a precisão das previsões. A ST1A, com um MAPE intermediário, demonstra um desempenho de previsão situado entre os valores extremos observados para a ST2A e a ST3A. A análise das barras indica que a inclusão de mais séries temporais tende a aprimorar a precisão das previsões, como evidenciado pela redução geral do MAPE em experimentos subsequentes.

Figura 83 - Comparação do MAPE entre Experimentos com a Base de Dados A.



Fonte: Autoria própria.

A Tabela 19 fornece uma visão detalhada das melhorias percentuais observadas ao longo dos experimentos, com base na métrica MAPE.

Tabela 19 - Percentuais de Melhoria nos Experimentos da Base de Dados A.

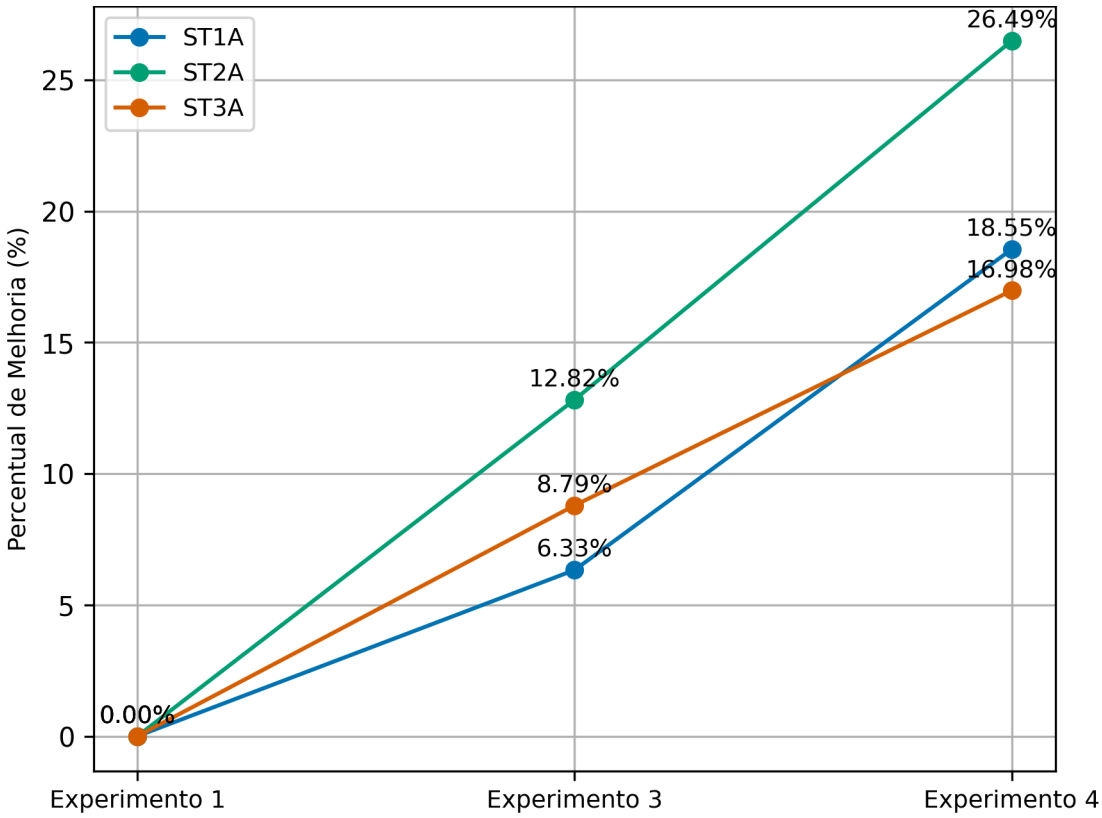
Melhoria (%)	**ST1A**	**ST2A**	**ST3A**
Exp. 1 - Exp. 3	6,33%	12,82%	8,79%
Exp. 3 - Exp. 4	18,55%	26,49%	16,98%
Exp. 1 - Exp. 4	23,33%	35,53%	24,58%

Fonte: Autoria própria.

A análise global entre o Experimento 1 e o Experimento 4 revela melhorias significativas. Para a ST1A, o refinamento da precisão foi de 23,33%, para a série ST2A foi de 35,53% e para a ST3A foi de 24,58%. Esta redução sistemática nos erros sugere que a inclusão de múltiplas séries temporais contribui efetivamente para uma maior precisão das previsões, corroborando a eficácia da metodologia aplicada.

Uma visualização mais detalhada das melhorias obtidas ao longo dos experimentos pode ser observada na Figura 84. No gráfico, cada linha representa uma série temporal específica, e os pontos marcados indicam os valores percentuais de melhoria.

Figura 84 - Percentuais de Melhoria Entre Experimentos com a Base de Dados A.



Fonte: Autoria própria.

Na representação gráfica fornecida, a série ST1A, representada pela linha azul, mostrou um aumento na acurácia de aproximadamente 6,33% entre o Experimento 1 e o Experimento 3. A melhoria foi de cerca de 18,55% do Experimento 3 para o Experimento 4. Para a série ST2A, representada pela linha verde, o aumento na exatidão foi de aproximadamente 12,82% do Experimento 1 para o Experimento 3 e de 26,49% do Experimento 3 para o Experimento 4. Por sua vez, a série ST3A, representada pela linha laranja, exibiu um refinamento da precisão de cerca de 8,79% do Experimento 1 para o Experimento 3 e uma melhoria adicional de aproximadamente 16,98% do Experimento 3 para o Experimento 4.

Na avaliação de modelos preditivos, o MAE e o RMSE são frequentemente utilizados para avaliar a precisão. O MAE representa a média dos erros absolutos, enquanto o RMSE mede a dispersão dos erros. Portanto, valores menores para ambos os critérios sugerem uma maior exatidão do modelo. Neste contexto, considerando que as magnitudes dos dados reais das ST1A, ST2A e ST3A excedem 0,6 megawatts, a análise dos valores de MAE e RMSE nos três

experimentos revela um aprimoramento significativo na precisão das previsões à medida que se aumenta a quantidade de séries temporais integradas no modelo previsor. No Experimento 1, utilizando o SPUST, o MAE variou entre 0,01 MW e 0,03 MW, enquanto o RMSE oscilou entre 0,02 MW e 0,05 MW. Com a aplicação do SPMST no Experimento 3, esses valores mostraram uma redução, com o MAE fixando-se consistentemente em 0,01 MW a 0,02 MW e o RMSE diminuindo para uma faixa mais estreita de 0,02 MW. Esta tendência de melhoria continua no Experimento 4, onde a integração da quarta série temporal resultou em MAE entre 0,01 MW e 0,02 MW e um RMSE variando de 0,01 MW a 0,02 MW.

As análises gráficas apresentadas na Subseção 6.2 mostraram que, no Experimento 1, houve discrepâncias expressivas entre as cargas reais e as previstas, com ST1A e ST2A apresentando os maiores erros. Os gráficos de dispersão demonstraram que a variabilidade das previsões em relação aos valores reais foi maior para essas séries, enquanto ST3A exibiu uma dispersão menor, sugerindo previsões mais precisas. No Experimento 3, os gráficos não indicaram uma melhoria significativa na previsão em relação aos resultados obtidos com SPUST, com uma redução de erro mais pronunciada apenas para a ST2A. No entanto, no Experimento 4, as representações gráficas evidenciaram uma aproximação mais consistente das curvas de carga prevista em relação às curvas de carga real. Os gráficos de dispersão mostraram uma redução significativa na dispersão das previsões, com os pontos agrupados ao redor da linha tracejada vermelha, indicando uma alta correspondência entre as previsões e os dados reais.

De maneira geral, os resultados demonstram uma melhoria significativa na precisão das previsões para as séries temporais da Base de Dados A. As análises indicam uma redução consistente nos erros de previsão ao longo dos experimentos, evidenciando que a abordagem metodológica que incorporou múltiplas séries temporais proporcionou um aprimoramento na acurácia das previsões nos diferentes cenários avaliados.

6.3.2 Considerações sobre os Experimentos com a Base de Dados B

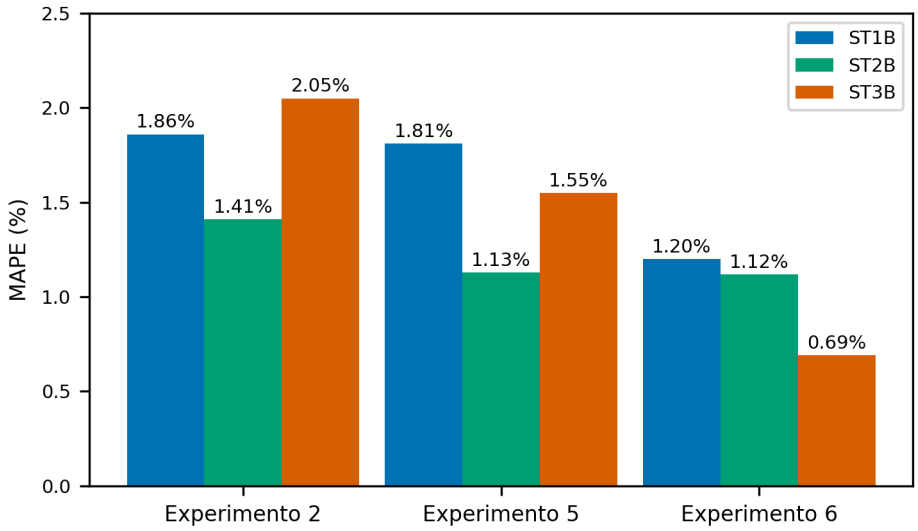
No Experimento 2, o SPUST foi implementado para as sequências históricas ST1B, ST2B e ST3B. Os resultados indicaram que a ST2B apresentou o melhor desempenho, com um MAPE de 1,41%, denotando a maior precisão entre as séries. Em contraste, a ST3B revelou o maior erro percentual, de 2,05%, sugerindo desafios significativos na previsão desta sequência. A ST1B, por outro lado, exibiu um MAPE intermediário, posicionando-se entre os valores registrados para ST2B e ST3B. Os resultados individuais demonstram que a ST2B é mais previsível em relação à ST3B e à ST1B, possivelmente devido a padrões de demanda menos complexos ou mais estáveis.

No Experimento 5, o SPMST foi implementado para combinar as três sequências discutidas anteriormente. Os resultados demonstraram uma redução no erro de precisão para a ST2B,

com um MAPE de 1,13%, em comparação ao Experimento 2. Por outro lado, para a ST3B, que anteriormente apresentava um desempenho individual inferior, a combinação das séries temporais resultou em melhorias significativas, com o MAPE reduzido para 1,55%. Embora a ST1A também tenha registrado avanços, esses não foram significativos. Em contraste, no Experimento 6, a inclusão da quarta série temporal (ST4B) resultou em uma melhoria mais expressiva para a ST1A, com um MAPE de 1,20%. Além disso, na ST3B, o erro percentual reduziu significativamente para 0,69%. Embora a ST2B também tenha apresentado progressos, esses resultados não foram tão relevantes em comparação com as melhorias observadas nas ST1B e ST3B.

A Figura 85 apresenta uma comparação visual do MAPE entre os três experimentos realizados com as séries temporais ST1B, ST2B e ST3B. O gráfico de barras oferece uma visão detalhada da precisão das previsões, expressa em termos de MAPE, para cada série temporal ao longo dos experimentos. Cada barra colorida no gráfico representa uma série temporal específica e é categorizada conforme o experimento correspondente. As séries temporais são distinguidas pelas seguintes cores: azul para ST1B, verde para ST2B e laranja para ST3B.

Figura 85 - Comparação do MAPE entre Experimentos com a Base de Dados B.



Fonte: Autoria própria.

Na ilustração gráfica fornecida, observa-se que a série temporal ST3B apresenta a redução de MAPE mais significativa ao longo dos experimentos, atingindo o menor valor no Experimento 6. Em contraste, a série ST1B demonstra valores de MAPE relativamente elevados, embora também exiba uma tendência de redução ao longo dos experimentos, indicando uma melhoria gradual no desempenho. Por outro lado, a série ST2B exibe consistentemente baixos valores de MAPE, destacando-se como a série com o melhor desempenho geral em termos de precisão de previsão. Esse padrão sugere que a ST2B possui características que facilitam a

previsibilidade, contribuindo para um menor erro percentual. A análise das barras indica uma adaptação e um refinamento do modelo SPMST nos diferentes experimentos, o que contribuiu para a melhoria da precisão das previsões, conforme evidenciado pela redução dos valores de MAPE da ST3B.

A Tabela 20 proporciona uma avaliação detalhada das melhorias percentuais registradas durante os experimentos, com base na métrica MAPE.

Tabela 20 - Percentuais de Melhoria nos Experimentos da Base de Dados B.

Melhoria (%)	**ST1B**	**ST2B**	**ST3B**
Exp. 2 - Exp. 5	2,69%	19,86%	24,39%
Exp. 5 - Exp. 6	33,70%	0,88%	55,48%
Exp. 2 - Exp. 6	36,13%	20,74%	84,25%

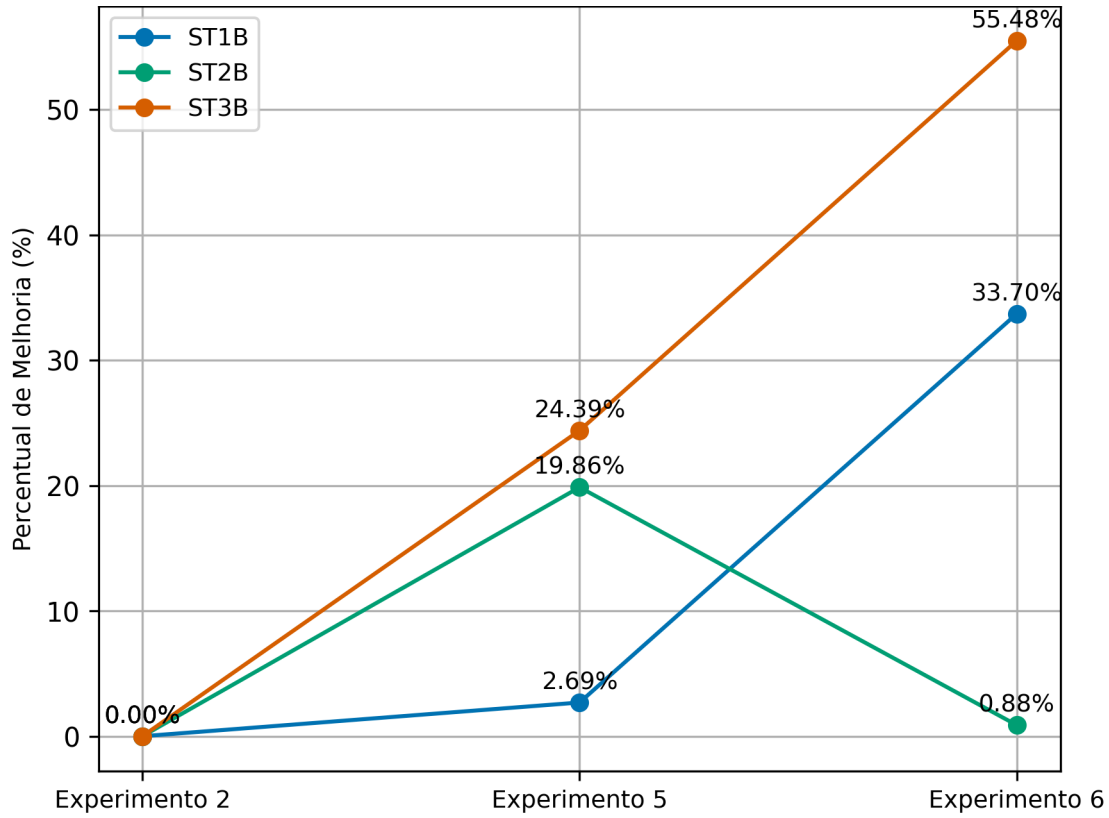
Fonte: Autoria própria.

A análise global dos experimentos demonstra um aprimoramento significativo na acurácia das previsões. Para a ST1B, a melhoria na precisão entre o Experimento 2 e o Experimento 6 foi de 36,13%. No contexto da ST2B, o avanço observado no mesmo intervalo foi de 20,74%. A série ST3B apresentou a maior melhoria, com um incremento de 84,25% na precisão das previsões. Esses resultados indicam que a metodologia aplicada, que integra múltiplas séries temporais, contribui substancialmente para a redução dos erros de previsão, como evidenciado pelo aumento consistente na precisão ao longo dos experimentos.

A Figura 86 oferece uma visualização mais detalhada das melhorias alcançadas ao longo dos experimentos realizados. No gráfico, cada linha representa uma série temporal específica, permitindo a comparação entre diferentes conjuntos de dados. Os pontos destacados indicam os valores percentuais de melhoria, facilitando a interpretação das variações de desempenho ao longo dos experimentos. Isso permite uma análise mais profunda do impacto das intervenções implementadas, evidenciando como essas modificações influenciaram os resultados obtidos.

Na referida ilustração, é possível observar que a série ST1B, representada pela linha azul, mostra um aperfeiçoamento na acurácia de aproximadamente 2,69% entre o Experimento 2 e o Experimento 5, e uma melhoria adicional de cerca de 33,70% do Experimento 5 para o Experimento 6. Com relação à ST2B, representada pela linha verde, o aumento na exatidão foi de aproximadamente 19,86% do Experimento 2 para o Experimento 5, e de 0,88% do Experimento 5 para o Experimento 6. Por sua vez, a série ST3B, representada pela linha laranja, exibiu um aprimoramento na precisão de cerca de 24,39% do Experimento 2 para o Experimento 5, e uma melhoria adicional de aproximadamente 55,48% do Experimento 5 para o Experimento 6.

Figura 86 - Percentuais de Melhoria Entre Experimentos com a Base de Dados B.



Fonte: Autoria própria.

Na análise dos modelos preditivos, destacam-se o MAE e o RMSE como métricas fundamentais para avaliar a precisão. O MAE reflete a média dos erros absolutos, enquanto o RMSE fornece uma medida da dispersão dos erros. Neste contexto, valores menores para ambos os critérios sugerem uma maior exatidão do modelo. No Experimento 2, os valores de MAE variam de 217,74 MW a 1514,03 MW; no Experimento 5, de 173,58 MW a 1185,57 MW; e no Experimento 6, de 178,14 MW a 496,76 MW. Considerando que as séries temporais utilizadas podem atingir demandas de energia superiores a 80,000 MW, esses indicadores são relativamente baixos, representando uma fração reduzida da capacidade máxima, como, por exemplo, aproximadamente 0,62% no caso do maior MAE do Experimento 6. O RMSE também apresenta valores baixos em comparação com a capacidade total das séries históricas. No Experimento 2, os valores de RMSE variam de 253,88 MW a 1985,77 MW; no Experimento 5, de 221,28 MW a 1696,42 MW; e no Experimento 6, de 227,69 MW a 594,12 MW, sendo que o maior RMSE representa cerca de 0,74% da capacidade máxima de 80,000 MW. A análise dos valores de MAE e RMSE nos três experimentos revela um aprimoramento significativo na precisão das previsões à medida que se aumenta a quantidade de séries temporais integradas no modelo previsor.

Por meio das análises gráficas apresentadas na Subseção 6.2, é possível interpretar os resultados obtidos nos Experimentos 2, 5 e 6. No Experimento 2, a curva prevista mostra certa correspondência com a curva real, especialmente para a ST2B, enquanto as séries ST1B e ST3B apresentam maiores divergências. A análise do gráfico de dispersão corrobora esses achados, revelando que os pontos de previsão estão mais distantes da linha de referência dos dados reais, especialmente para ST1B e ST3B. No Experimento 5, as curvas de previsão demonstram uma aproximação mais estreita em relação às curvas reais para ST1B e ST3B, refletindo uma redução na discrepância das previsões. Os gráficos de dispersão para este experimento mostram uma diminuição na variabilidade das previsões, com os pontos se agrupando de forma mais coesa ao redor da linha de referência dos dados reais. A ST2B continua a se destacar com uma precisão superior, evidenciada pela menor dispersão dos pontos de previsão. No Experimento 6, as previsões para ST3B apresentam uma correspondência muito próxima à curva real, e as curvas para ST1B e ST2B também estão bastante alinhadas com os valores observados, indicando uma boa precisão das previsões. No gráfico de dispersão, ST3B exibe a menor dispersão dos pontos, com a maioria deles situando-se na linha de identidade. ST1B e ST2B também apresentam uma boa concentração de pontos próximos à linha de referência, mas com uma leve dispersão, indicando erros menores.

Em termos amplos, observou-se uma redução consistente nos erros de previsão ao longo dos experimentos conduzidos com as séries temporais da Base de Dados B. No Experimento 6, verificou-se uma diminuição significativa nos erros de previsão para as séries ST1B e ST3B. É particularmente relevante ressaltar que a série ST3B, que havia exibido o maior erro quando avaliada individualmente com o SPUST, apresentando a menor previsibilidade em comparação com as demais séries, mostrou uma redução substancial nos indicadores de erro no Experimento 6. Essa melhoria expressiva reforça a hipótese de que a integração de séries temporais correlacionadas pode proporcionar ganhos consideráveis na acurácia das previsões, evidenciando a eficácia da abordagem múltipla na generalização dos modelos preditivos.

6.3.3 Considerações Gerais

A análise comparativa entre as Bases de Dados A e B revela *insights* importantes sobre a robustez e a generalização dos sistemas previsores no âmbito do setor elétrico. A Base de Dados B, relacionada ao nível de transmissão, apresentou melhores resultados em termos de MAPE, indicando uma maior precisão das previsões, especialmente quando utilizadas múltiplas séries temporais com o SPMST. Em contraste, a Base de Dados A, que se refere ao nível de distribuição, apresentou MAPEs mais altos, demonstrando um desafio maior quanto à previsibilidade. Essas diferenças de desempenho sugerem que o contexto geográfico e os padrões de consumo inerentes a cada cenário operacional podem influenciar significativamente a eficácia dos modelos preditivos. Esses resultados reforçam a importância de avaliar as abordagens de previsão

em diferentes níveis de operação do sistema elétrico para garantir a robustez e a capacidade de generalização.

Neste contexto, é importante destacar que o TFT mostrou-se eficaz em capturar interdependências e correlações entre diferentes sequências de dados, resultando em previsões mais precisas quando múltiplas séries foram integradas. A eficácia dessa abordagem neural é particularmente expressiva nos experimentos conduzidos com a Base de Dados A, onde os padrões de consumo são mais complexos e variáveis. Nesses cenários, as correlações entre diferentes séries temporais são menos acentuadas; contudo, o TFT foi capaz de explorar essas relações de maneira eficiente, proporcionando uma redução sistemática nos erros das projeções. A habilidade do TFT em lidar com tais complexidades e maximizar o uso das informações temporais disponíveis destaca sua robustez e capacidade de generalização nos desafiadores cenários do setor elétrico. Além disso, esses resultados sublinham a importância de utilizar modelos avançados como o TFT para melhorar a precisão preditiva em contextos onde a variabilidade e a complexidade dos dados representam desafios significativos.

Quanto ao impacto do número de séries temporais no sistema previsor, a abordagem múltipla demonstrou uma melhoria significativa na precisão das previsões. Os Experimentos 4 e 6, que utilizaram quatro séries temporais simultaneamente, apresentaram uma redução considerável nos valores de MAPE em comparação com os experimentos que empregaram séries isoladas. Essa melhoria pode ser explicada pela maior capacidade do modelo de captar variáveis contextuais e padrões complexos presentes em dados mais extensos. A fusão das sequências históricas oferece uma visão mais abrangente e detalhada do comportamento da demanda energética, permitindo ao modelo lidar melhor com variações e correlações. Em contraste, o tratamento isolado das séries temporais no SPUST limitou a capacidade de aprimorar a precisão das previsões, especialmente em contextos onde as interdependências entre diferentes séries são significativas. Por exemplo, as séries ST3A e ST3B frequentemente exibiram um desempenho superior quando combinadas com sequências correlacionadas, sugerindo que essas séries podem conter padrões que se beneficiam mais da integração com outros elementos.

Diante do exposto, é importante destacar que a integração de múltiplas séries temporais permite ao modelo capturar uma gama mais ampla de padrões e dependências temporais, resultando em previsões mais robustas e precisas. No entanto, essa complexidade adicional também aumenta a dimensionalidade dos dados e requer um tempo de processamento mais longo em comparação com os modelos baseados em uma única série temporal. Essa questão pode representar um desafio em contextos onde os recursos computacionais são limitados ou onde há restrições de tempo para o treinamento da abordagem preditiva.

No contexto desta investigação, a infraestrutura utilizada não demandou recursos computacionais substanciais. Em todos os experimentos, foram empregadas as ferramentas e o ambiente descritos na Subseção 5.5.5. A única variação observada foi no tempo de processamento, que

aumentou proporcionalmente à complexidade das análises. Nas simulações realizadas com a Base de Dados A, o tempo computacional médio por época foi de 5 minutos no SPUST. Em contraste, para o SPMST, o tempo médio necessário foi de 12 minutos para 3 séries e 16 minutos para 4 sequências de dados. Foram realizados testes que compreenderam entre 14 e 27 épocas. Para a Base de Dados B, o tempo computacional médio por época no SPUST foi de 6 minutos. No SPMST, o tempo médio foi de 17 minutos para 3 séries temporais e 22 minutos para 4 sequências de informações, com os testes variando entre 16 e 29 épocas.

Dado que o processamento nesta pesquisa foi realizado utilizando uma única GPU, o custo computacional pode ser considerado relativamente baixo, especialmente quando comparado ao treinamento que envolve múltiplas GPUs ou *clusters* de computadores. Além disso, é importante ressaltar que os tempos computacionais apresentados referem-se exclusivamente à fase de treinamento dos modelos. Durante essa etapa, a rede neural ajusta seus parâmetros para minimizar o erro nas previsões com base nos dados históricos fornecidos, o que é computacionalmente intensivo. No entanto, uma vez que o modelo está treinado, a geração de previsões pode ocorrer em um intervalo de poucos segundos, independentemente do número de séries temporais. Isso ocorre porque o processo de previsão envolve a aplicação direta dos parâmetros ajustados durante o treinamento aos dados atuais, o que é significativamente menos complexo do que o processo de aprendizado. Portanto, embora o ajuste do algoritmo possa demandar mais tempo e recursos, a aplicação do modelo para obter previsões é eficiente e rápida, permitindo sua utilização prática em ambientes onde a velocidade de resposta é fundamental.

6.4 Análise de Interpretabilidade em Nível de Recursos

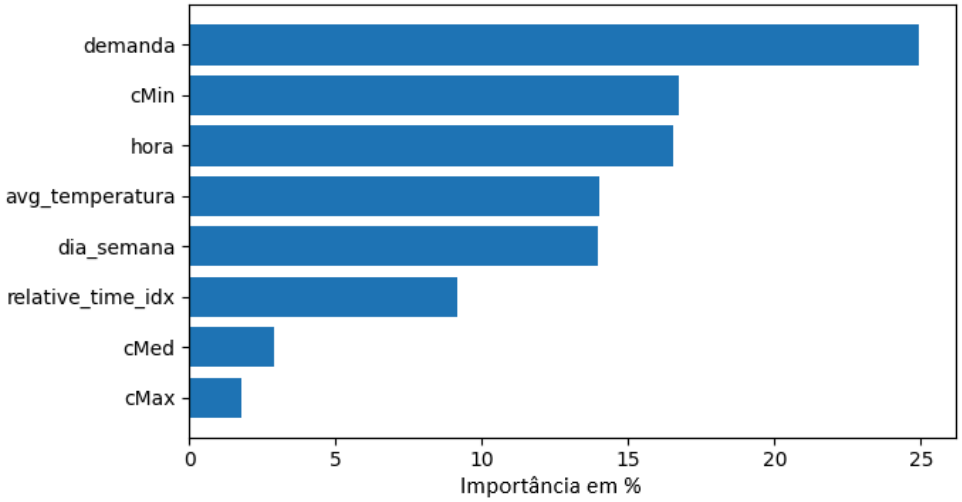
As Redes de Seleção de Variáveis do modelo TFT são projetadas para avaliar e estimar a importância de cada recurso de entrada, fornecendo um indicador para a compreensão do impacto das covariáveis nas previsões. Esse processo de avaliação é uma forma de interpretabilidade, que se refere à capacidade de entender e explicar como um modelo toma decisões e quais fatores influenciam suas previsões. As Redes de Seleção de Variáveis do TFT avaliam dinamicamente a relevância de cada entrada ao longo do tempo, permitindo uma interpretação detalhada dos fatores que mais influenciaram as projeções. Esse recurso é particularmente valioso para identificar variáveis-chave em conjuntos de dados complexos, contribuindo para a melhoria da precisão do modelo e oferecendo *insights* para a tomada de decisões baseadas em dados. Além disso, a interpretabilidade facilita a compreensão dos mecanismos internos do modelo, promovendo maior transparência e confiança nas previsões geradas.

Nesta seção, para cada experimento, são apresentados dois gráficos relacionados ao grau de importância das covariáveis: um referente ao codificador (*encoder*) e outro ao decodificador (*decoder*). O gráfico do codificador ilustra a importância atribuída a cada variável na fase de codificação, na qual o modelo analisa e processa as entradas temporais e contextuais para

extrair características relevantes. Nessa etapa, o TFT utiliza mecanismos de atenção para determinar quais fatores têm maior influência na formação da representação interna dos dados, a qual é utilizada para as previsões futuras. O gráfico do decodificador, por outro lado, demonstra a importância dos recursos na fase de decodificação. Durante essa etapa, o modelo emprega a representação interna dos dados gerada pelo codificador, juntamente com as entradas subsequentes, para produzir as projeções finais. Nesse contexto, o TFT novamente aplica mecanismos de atenção para identificar quais covariáveis são mais críticas para a precisão da previsão em um momento específico. Esses gráficos são fundamentais para a análise e interpretação dos resultados, pois permitem visualizar e comparar a relevância de diferentes características em distintas etapas do processo preditivo.

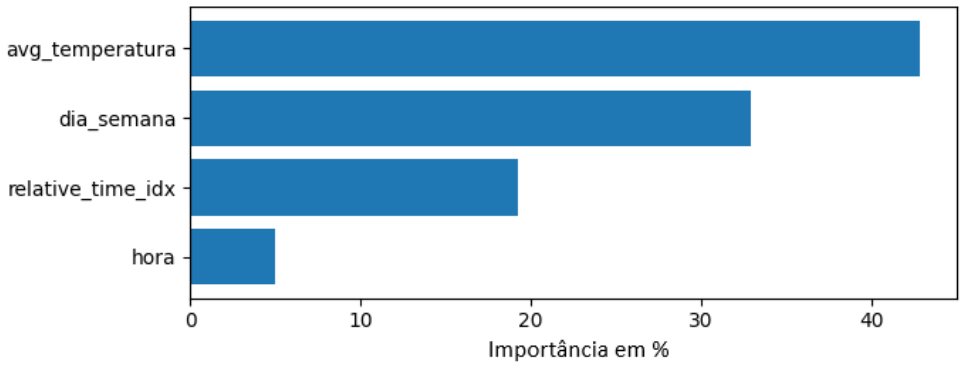
A análise da interpretabilidade dos experimentos realizados com a Base de Dados A possibilita uma compreensão mais profunda do impacto das variáveis nos resultados obtidos, considerando diferentes configurações das séries temporais. As Figuras 87 a 92 oferecem uma visão clara dos efeitos individuais e combinados das entradas sobre o desempenho dos Experimentos 1, 3 e 4.

Figura 87 - Experimento 1: Importância dos Recursos - Codificador.



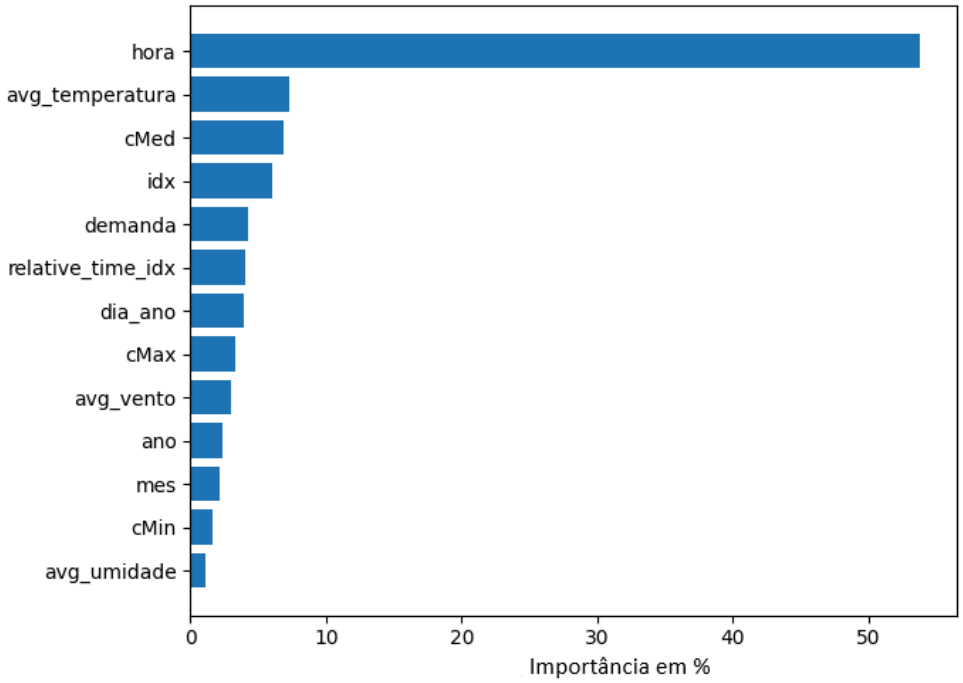
Fonte: Autoria própria.

Figura 88 - Experimento 1: Importância dos Recursos - Decodificador.



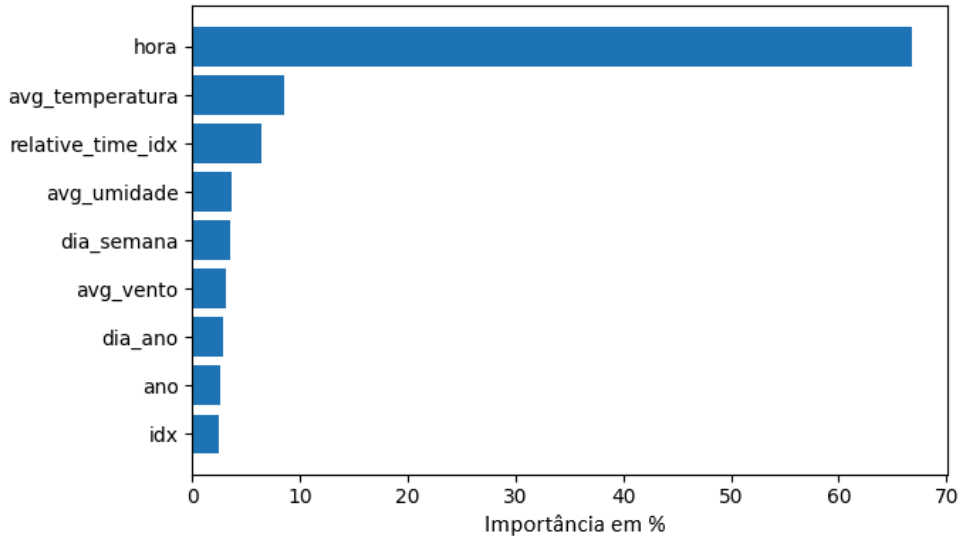
Fonte: Autoria própria.

Figura 89 - Experimento 3: Importância dos Recursos - Codificador.



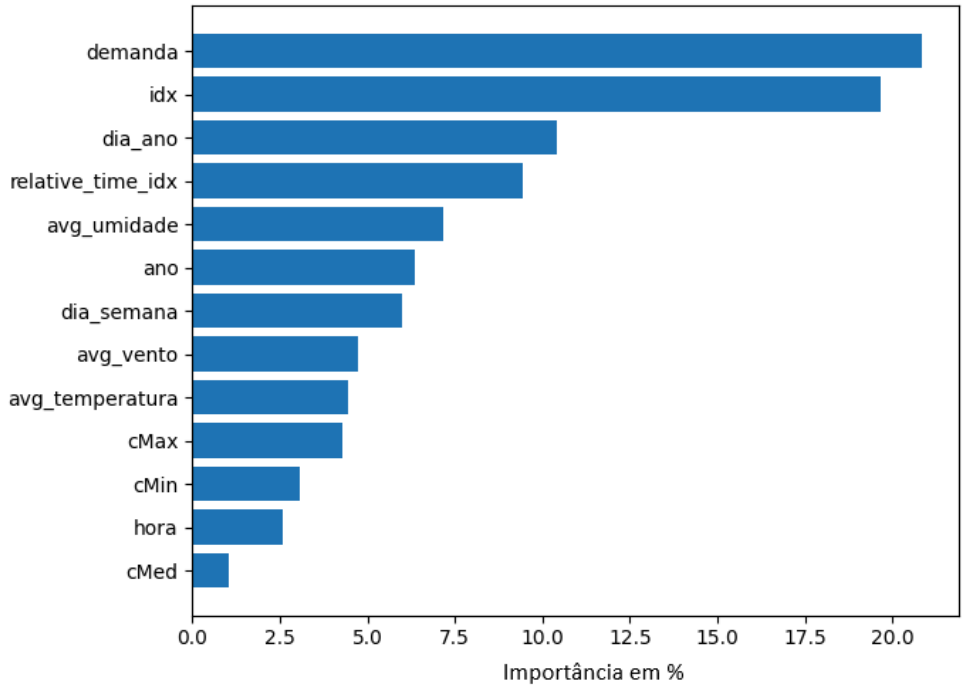
Fonte: Autoria própria.

Figura 90 - Experimento 3: Importância dos Recursos - Decodificador.



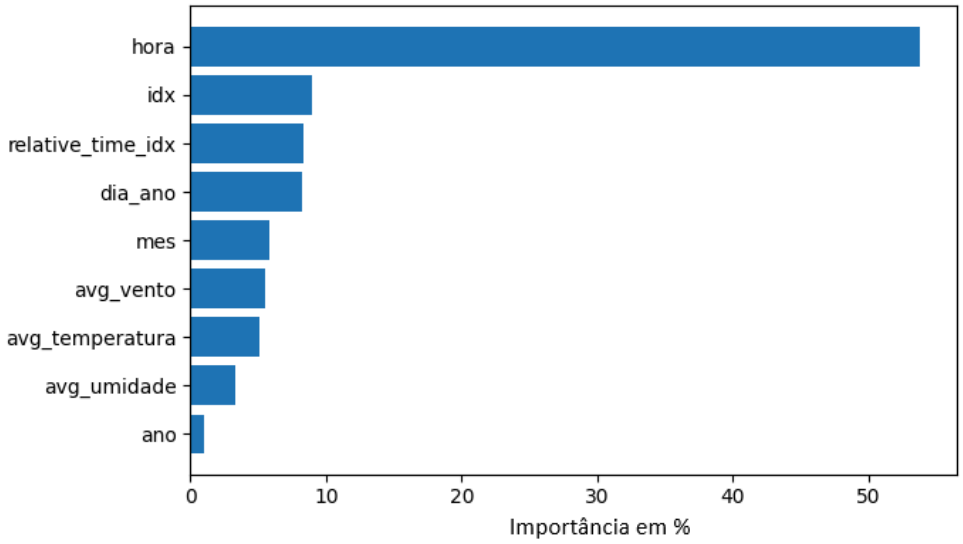
Fonte: Autoria própria.

Figura 91 - Experimento 4: Importância dos Recursos - Codificador.



Fonte: Autoria própria.

Figura 92 - Experimento 4: Importância dos Recursos - Decodificador.



Fonte: Autoria própria.

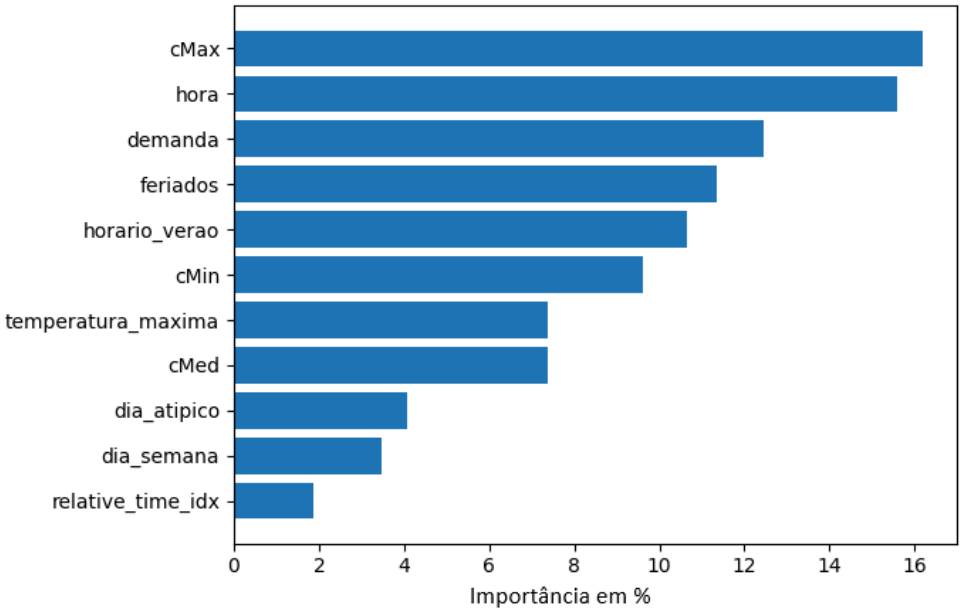
Nos gráficos referentes ao Experimento 1, que utiliza uma série temporal isolada, observa-se que o codificador priorizou variáveis diretamente relacionadas à demanda elétrica, incluindo demanda, carga mínima (cMin), hora, temperatura média (avg_temperatura), dia da semana (dia_semana), carga média (cMed) e carga máxima (cMax). O decodificador, por sua vez, deu maior ênfase a fatores temporais e ambientais, como temperatura média (avg_temperatura), dia da semana (dia_semana) e hora. Este experimento sugere que, no contexto da Base de Dados A e em um cenário de uma única série temporal, as características associadas diretamente à carga e à demanda elétrica são particularmente destacadas pelo modelo preditivo, sendo essenciais para a precisão das previsões.

Já nos Experimentos 3 e 4, que integram múltiplas séries temporais, observa-se uma diversificação nas variáveis de importância tanto para o codificador quanto para o decodificador, bem como uma mudança expressiva na distribuição dos recursos. No Experimento 3, o codificador prioriza uma gama mais ampla de fatores em comparação ao Experimento 1, incluindo hora, temperatura média (avg_temperatura), carga média (cMed), demanda, dia do ano (dia_ano), carga máxima (cMax), velocidade média do vento (avg_vento), ano, mês, carga mínima (cMin) e umidade média (avg_umidade). O decodificador, por sua vez, concentra-se em fatores como hora, temperatura média (avg_temperatura), umidade média (avg_umidade), dia da semana (dia_semana), velocidade média do vento (avg_vento), dia do ano (dia_ano) e ano. No Experimento 4, observa-se uma ligeira mudança na prioridade das variáveis. O codificador prioriza a demanda, dia do ano (dia_ano), umidade média (avg_umidade), ano, dia da semana (dia_semana), velocidade média do vento (avg_vento), temperatura média (avg_temperatura), carga máxima (cMax), carga mínima (cMin), hora e carga média (cMed), enquanto o decodifi-

cador enfatiza os elementos como hora, dia do ano (dia_ano), mês, velocidade média do vento (avg_vento), temperatura média (avg_temperatura), umidade média (avg_umidade) e ano.

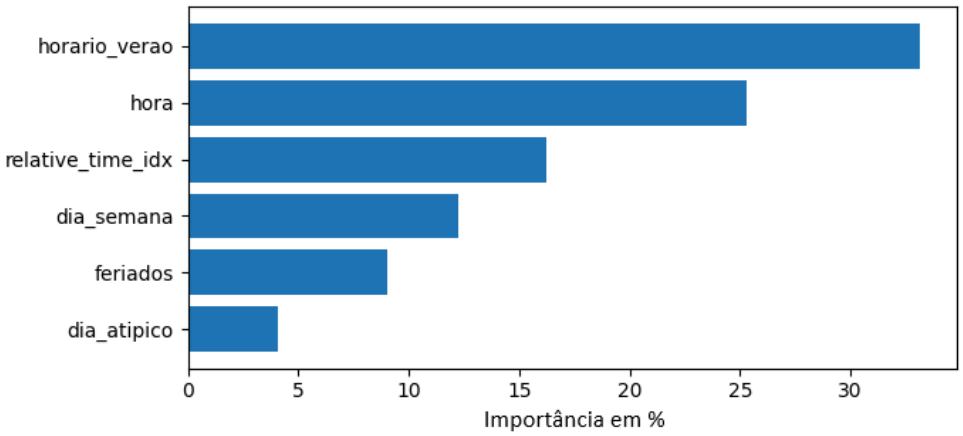
A análise da interpretabilidade dos experimentos conduzidos com a Base de Dados B permite explorar o efeito das entradas sobre os resultados obtidos para este conjunto de dados, considerando as diferentes configurações das séries temporais. As Figuras 93 a 98 ilustram como os recursos utilizados influenciaram as projeções obtidas nos Experimentos 2, 5 e 6.

Figura 93 - Experimento 2: Importância dos Recursos - Codificador.



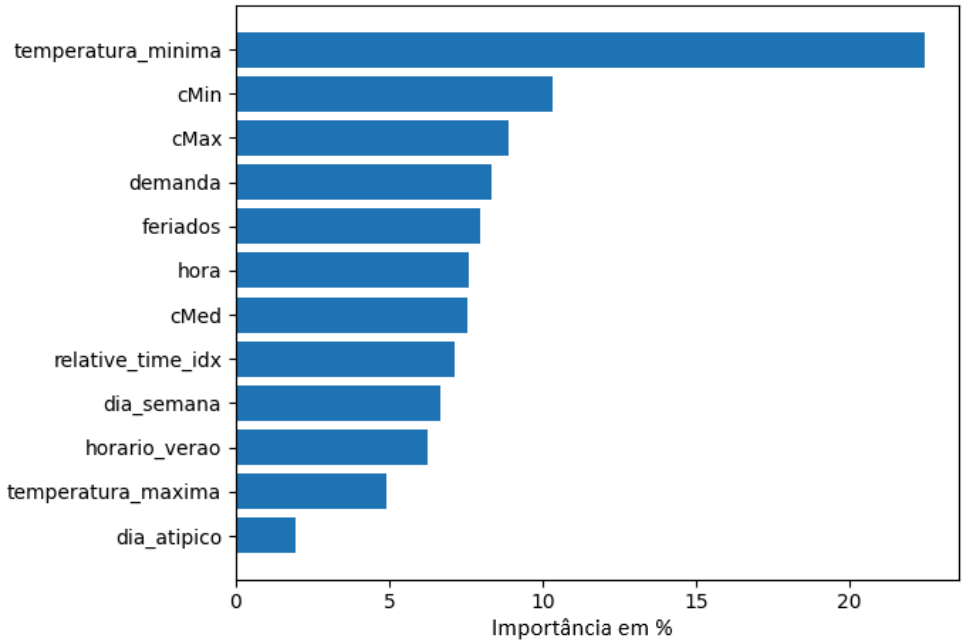
Fonte: Autoria própria.

Figura 94 - Experimento 2: Importância dos Recursos - Decodificador.



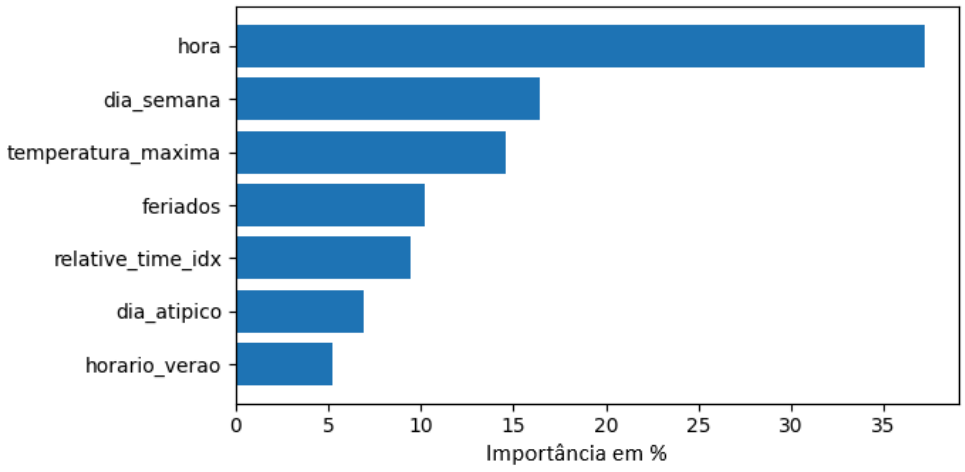
Fonte: Autoria própria.

Figura 95 - Experimento 5: Importância dos Recursos - Codificador.



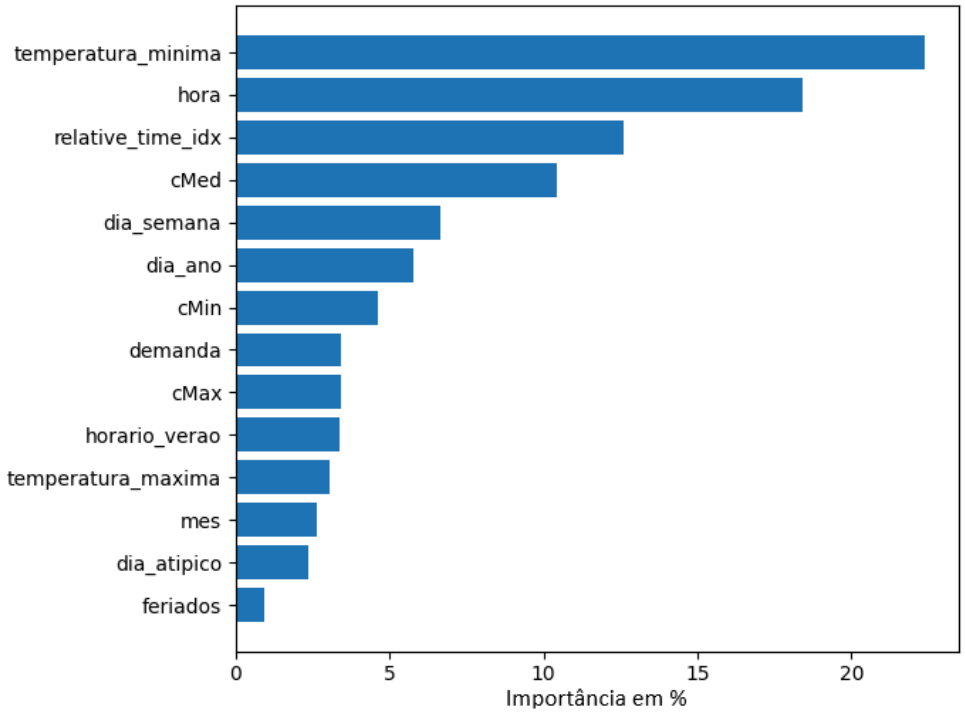
Fonte: Autoria própria.

Figura 96 - Experimento 5: Importância dos Recursos - Decodificador.



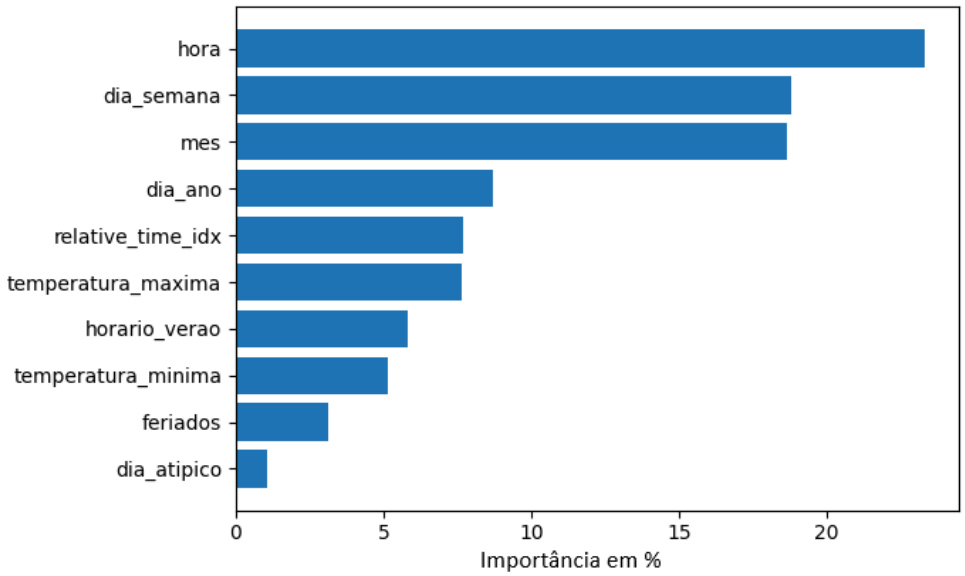
Fonte: Autoria própria.

Figura 97 - Experimento 6: Importância dos Recursos - Codificador.



Fonte: Autoria própria.

Figura 98 - Experimento 6: Importância dos Recursos - Decodificador.



Fonte: Autoria própria.

Nos gráficos de interpretabilidade do Experimento 2, que utiliza uma série temporal isolada,

observa-se que o codificador prioriza variáveis diretamente relacionadas à carga elétrica e a fatores temporais. Entre os elementos mais relevantes destacam-se a carga máxima (cMax), a hora, a demanda, os feriados, o horário de verão, a carga mínima (cMin), a temperatura máxima, a carga média (cMed), dias atípicos e o dia da semana (dia_semana). Por outro lado, o decodificador enfatiza predominantemente no horário de verão, na hora, no dia da semana, nos feriados e nos dias atípicos. Esses resultados sugerem que, em cenários que envolvem uma única série temporal, as características associadas diretamente à carga elétrica e a eventos especiais, como feriados e horário de verão, desempenham um papel mais significativo nas previsões da Base de Dados B.

Ao utilizar múltiplas séries temporais, observa-se uma alteração significativa na importância das variáveis consideradas pelo modelo. No Experimento 5, o codificador destaca variáveis como temperatura mínima, carga mínima (cMin), carga máxima (cMax), demanda, feriados, hora, carga média (cMed), dia da semana (dia_semana), horário de verão, temperatura máxima e dia atípico. Em contraste, o decodificador prioriza hora, dia da semana, temperatura máxima, feriados, dia atípico e horário de verão. Já no Experimento 6, a prioridade do codificador se estende a temperatura mínima, hora, carga média (cMed), dia da semana (dia_semana), dia do ano (dia_ano), carga mínima (cMin), demanda, carga máxima (cMax), horário de verão, temperatura máxima, mês, dia atípico e feriados. O decodificador, por sua vez, enfatiza a hora, dia da semana (dia_semana), mês, dia do ano (dia_ano), temperatura máxima, horário de verão, temperatura mínima, feriados e dia atípico. Esses experimentos revelam que, ao considerar um contexto integrado, a previsão na Base de Dados B passa a levar em conta tanto os padrões temporais quanto as variáveis climáticas e eventos especiais, demonstrando uma adaptação abrangente às complexidades dos dados.

A maior diversificação observada nos Experimentos 3, 4, 5 e 6, realizados com o SPMST, em comparação com os Experimentos 1 e 2, conduzidos com o SPUST, sugere que a inclusão de múltiplas séries temporais amplia a capacidade do modelo ao incorporar uma gama mais extensa de fatores que influenciam a demanda elétrica. A integração de múltiplas fontes de dados enriquece o modelo com informações adicionais e permite uma compreensão mais abrangente das inter-relações entre variáveis climáticas, sazonais e eventos especiais e a demanda elétrica. Dessa forma, a diversidade de dados não só melhora a precisão das previsões, mas também proporciona *insights* mais profundos sobre como diferentes fatores interagem para afetar a demanda, evidenciando a importância de uma abordagem multivariada na modelagem e previsão das séries temporais de energia elétrica.

Para as Bases de Dados A e B, as análises de interpretabilidade indicam que, ao utilizar uma única série temporal, tanto os codificadores quanto os decodificadores priorizaram atributos diretamente relacionados à demanda elétrica e a características temporais. Na Base de Dados A, a integração de múltiplas séries temporais expandiu o conjunto de variáveis consideradas pelo

codificador, englobando aspectos climáticos e sazonais, enquanto o decodificador continuou a enfatizar fatores temporais. Na Base de Dados B, a inclusão de várias séries também diversificou os atributos considerados pelo codificador, mantendo uma ênfase constante em variáveis climáticas e sazonais, com o decodificador priorizando predominantemente características temporais e sazonais. Esta análise revela que, independentemente da base de dados, a integração de múltiplas séries temporais permitiu uma modelagem mais rica e complexa, refletida na maior diversidade de atributos considerados pelos codificadores e decodificadores.

6.5 Síntese do Capítulo

Este capítulo apresenta uma análise detalhada dos resultados obtidos nos experimentos realizados e discute suas implicações no contexto da previsão de séries temporais de energia elétrica. As seções fornecem uma visão abrangente do desempenho dos modelos propostos, destacando os principais achados e oferecendo uma interpretação crítica dos dados. Na subseção de Análise dos Resultados, são discutidos os achados experimentais obtidos com os previsores SPUST e SPMST, no contexto de dois conjuntos de dados distintos: Base de Dados A e Base de Dados B. O desempenho dos modelos é avaliado por meio de análises gráficas e métricas de desempenho, permitindo uma comparação aprofundada entre as diferentes abordagens.

A Discussão dos Resultados é organizada em várias subseções, começando com uma análise dos experimentos realizados com a Base de Dados A, contextualizando os resultados dentro do cenário específico desse conjunto de dados. Em seguida, são abordados os experimentos conduzidos com a Base de Dados B, permitindo uma comparação direta entre as diferentes análises implementadas. A subseção de Considerações Gerais sintetiza os principais pontos discutidos anteriormente e oferece uma visão global da investigação realizada. Por fim, uma análise de interpretabilidade em nível de recursos é discutida, destacando os aspectos mais relevantes e seus impactos nas projeções obtidas.

7 CONCLUSÃO

Este estudo investigou a aplicação de previsões múltiplas no setor de energia elétrica, utilizando a rede neural TFT. A pesquisa foi motivada pela crescente complexidade das redes elétricas modernas e pela necessidade de ferramentas preditivas mais sofisticadas, capazes de lidar com a diversidade e as correlações presentes nos dados gerados por esses sistemas. A metodologia adotada baseou-se na fusão de séries históricas correlacionadas, visando capturar as interações entre variáveis e adaptar o modelo às variações e tendências dos cenários analisados. O estudo concentrou-se no aprimoramento das previsões de demanda energética com um horizonte preditivo de 24 horas, utilizando dados de consumo provenientes de dois contextos operacionais da infraestrutura elétrica e técnicas avançadas de aprendizado de máquina.

O desenvolvimento dos sistemas preditivos — SPUST e SPMST — possibilitou uma análise aprofundada da eficácia de cada abordagem em diferentes contextos. Os resultados indicam que tanto o SPUST quanto o SPMST apresentam vantagens e limitações específicas, refletindo a complexidade e os desafios inerentes à previsão de demanda energética.

O SPUST demonstrou uma eficácia significativa ao lidar com séries temporais individuais, como evidenciado pelos resultados obtidos na Base de Dados B, onde os valores de MAPE foram relativamente baixos. Esses resultados sugerem que o SPUST é particularmente adequado para cenários em que há uma única série temporal relevante e onde a variabilidade pode ser bem modelada a partir dessa fonte de dados. Entretanto, os erros observados na Base de Dados A destacam uma limitação importante do SPUST em prever períodos de alta demanda energética. Os elevados valores de erro indicam que o modelo pode não capturar adequadamente as variações extremas ou eventos raros, como aqueles que ocorrem em contextos de alta demanda. Nesses cenários, o consumo de energia atinge níveis excepcionalmente elevados, geralmente devido a fatores como condições climáticas extremas, eventos especiais ou picos repentinos na atividade industrial.

Em contraste, o SPMST demonstrou que a integração de múltiplas séries temporais pode melhorar significativamente a precisão das previsões. A análise dos Experimentos 3 e 4, bem como dos Experimentos 5 e 6, indicou que a inclusão de um maior número de séries temporais correlacionadas resulta em uma modelagem mais rica e robusta dos fatores que influenciam a demanda energética. A capacidade do SPMST de capturar interações e correlações entre diferentes séries temporais contribuiu para uma precisão aprimorada nas previsões. No entanto, a necessidade de processamento intensivo pode se tornar um obstáculo, especialmente em ambientes com recursos computacionais limitados.

Portanto, ao comparar o SPUST e o SPMST, é essencial considerar não apenas a capacidade preditiva de cada abordagem, mas também o contexto específico da demanda energética, a disponibilidade de dados e os recursos computacionais disponíveis. Para aplicações práticas,

o SPUST pode ser mais adequado em cenários com uma única série temporal de alta qualidade e onde a demanda permanece relativamente estável. Nessas situações, a simplicidade e a eficiência do SPUST oferecem vantagens significativas, possibilitando previsões precisas sem a necessidade de processamento intensivo.

Por outro lado, o SPMST emerge como uma abordagem altamente promissora quando se dispõe de múltiplas séries temporais correlacionadas. A capacidade de integrar diversas fontes de dados permite uma modelagem mais sofisticada e detalhada, resultando em previsões mais precisas em contextos complexos. Ao capturar interações e correlações entre variáveis, o SPMST oferece uma visão abrangente e adaptada às nuances do setor energético. Embora essa abordagem exija um tempo de treinamento maior, os benefícios potenciais superam amplamente os desafios. O aprimoramento na precisão das previsões e a capacidade de adaptação a cenários dinâmicos e variáveis justificam o investimento adicional, tornando o SPMST uma escolha estratégica para enfrentar os desafios do segmento elétrico.

Neste contexto, é importante destacar que as séries temporais de energia elétrica tendem a ser instáveis e suscetíveis a flutuações causadas por condições climáticas, eventos sazonais e variáveis econômicas. Além disso, os dados de demanda frequentemente estão correlacionados e refletem uma complexa rede de fatores que influenciam o consumo. A análise de interpretabilidade revelou que a inclusão de múltiplas séries temporais no SPMST oferece uma visão mais detalhada dessas influências. Com a inclusão de séries temporais adicionais, os codificadores e decodificadores conseguiram identificar uma gama mais ampla de variáveis, abrangendo aspectos climáticos, sazonais e outras características temporais relevantes. Em contraste, o SPUST, que se baseia em uma única série temporal, apresentou uma interpretação mais restrita, focando principalmente em atributos diretamente relacionados à demanda elétrica e às características temporais básicas. Essa diferença ressalta a importância da integração de múltiplas fontes de dados no setor energético, pois a fusão dessas fontes permite uma modelagem mais robusta, capaz de proporcionar uma compreensão mais completa da dinâmica da demanda energética.

Diante dos resultados obtidos neste trabalho, pesquisas futuras poderão explorar:

1. Expansão do Número de Séries Temporais: Investigar a eficácia do modelo TFT ao incorporar um maior número de séries temporais, com o objetivo de avaliar se a inclusão de dados adicionais pode resultar em melhorias significativas na precisão das previsões e na capacidade de generalização do modelo;

2. Avaliação de Séries Temporais Não Correlacionadas: Analisar o impacto da integração de séries temporais não correlacionadas para determinar se a inclusão de dados variados pode aprimorar ou comprometer a performance preditiva do modelo;

3. Integração de Dados Externos e Contextuais: Explorar a inclusão de séries temporais de dados externos e contextuais, como eventos econômicos ou tendências de mercado, para

avaliar se esses fatores adicionais podem aprimorar a capacidade preditiva no contexto abordado nesta pesquisa;

4. Estudos de Caso em Outros Cenários Operacionais: Aplicar a metodologia proposta em outros contextos de demanda para avaliar a flexibilidade e a aplicabilidade do modelo no setor elétrico;

5. Análise de Impacto Econômico e Operacional: Avaliar o impacto econômico e operacional da adoção de modelos que utilizam múltiplas séries temporais, incluindo uma análise de custo-benefício e o efeito sobre os processos de tomada de decisão.

Este estudo evidenciou que a previsão de múltiplas séries temporais, utilizando o *Temporal Fusion Transformer*, representa uma abordagem promissora para a previsão de demanda energética. Os aprimoramentos observados na precisão das previsões contribuem significativamente para uma gestão mais eficiente e sustentável dos sistemas elétricos de potência, destacando a relevância de técnicas avançadas de aprendizado de máquina no desenvolvimento de soluções para os desafios contemporâneos enfrentados pelo setor energético.

Esta pesquisa propõe uma perspectiva inovadora para o uso de modelos preditivos no contexto dos sistemas de energia elétrica, explorando a fusão de informações contextuais e a análise das interdependências entre séries temporais para aprofundar a compreensão das dinâmicas do consumo. Essa abordagem tem o potencial de melhorar a precisão na previsão de flutuações na demanda e de apoiar decisões estratégicas, contribuindo para uma alocação mais eficiente de recursos. Além disso, o estudo sugere caminhos para práticas operacionais mais ágeis e informadas, que podem fortalecer a adaptabilidade das redes elétricas diante das mudanças nas condições operacionais e das exigências do mercado energético.

REFERÊNCIAS

AL-KANDARI, A.; SOLIMAN, S.; EL-HAWARY, M. Fuzzy short-term electric load forecasting. *International Journal of Electrical Power & Energy Systems*, Elsevier, Amsterdam, v. 26, n. 2, p. 111–122, 2004.

ALPAYDIN, E. *Introduction to machine learning*. Cambridge, Massachusetts: MIT press, 2020.

ANAND, D.; KHALAF, O. I.; ABDULSAHIB, G. M.; CHANDRA, G. R. Original research article identification of meningioma tumor using recurrent neural networks. *Journal of Autonomous Intelligence*, *[S. l.]*, v. 7, n. 2, 2024.

ANDRADE, G. N. de; SANT'ANNA, A. P. Estimativa do impacto das variações de temperatura sobre o consumo residencial de energia elétrica no rio de janeiro. *Simpósio brasileiro de Pesquisa Operacional*, Rio de Janeiro, p. 949–260, 2013.

ARMSTRONG, J. S. *Principles of forecasting: a handbook for researchers and practitioners*. Berlin: Springer, 2001. v. 30.

ARNAOUT, R.; CURRAN, L.; ZHAO, Y.; LEVINE, J. C.; CHINN, E.; MOON-GRADY, A. J. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nature medicine*, Nature Publishing Group, London, v. 27, n. 5, p. 882–891, 2021.

AYACHI, R.; SAID, Y.; ABDELAALI, A. B. Pedestrian detection based on light-weighted separable convolution for advanced driver assistance systems. *Neural Processing Letters*, Springer, Berlin, v. 52, n. 3, p. 2655–2668, 2020.

BA, J. L.; KIROS, J. R.; HINTON, G. E. Layer normalization. *arXiv preprint arXiv:1607.06450*, *[S. l.]*, 2016.

BADŽA, M. M.; BARJAKTAROVIĆ, M. Č. Classification of brain tumors from mri images using a convolutional neural network. *Applied Sciences*, MDPI, Basel, v. 10, n. 6, p. 1999, 2020.

BAI, Y.; XIE, J.; LIU, C.; TAO, Y.; ZENG, B.; LI, C. Regression modeling for enterprise electricity consumption: A comparison of recurrent neural network and its variants. *International Journal of Electrical Power & Energy Systems*, Elsevier, Amsterdam, v. 126, p. 106612, 2021.

BALAKRISHNAN, R.; GEETHA, V.; KUMAR, M. R.; LEUNG, M.-F. Reduction in residential electricity bill and carbon dioxide emission through renewable energy integration using an adaptive feed-forward neural network system and mppt technique. *Sustainability*, MDPI, Basel, v. 15, n. 19, p. 14088, 2023.

BANDARA, K.; BERGMEIR, C.; HEWAMALAGE, H. Lstm-msnet: Leveraging forecasts on sets of related time series with multiple seasonal patterns. *IEEE transactions on neural networks and learning systems*, IEEE, New York, v. 32, n. 4, p. 1586–1599, 2020.

BASHEER, I. A.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, Elsevier, Amsterdam, v. 43, n. 1, p. 3–31, 2000.

BECKEN, S.; FRAMPTON, C.; SIMMONS, D. Energy consumption patterns in the accommodation sector—the new zealand case. *Ecological economics*, Elsevier, Amsterdam, v. 39, n. 3, p. 371–386, 2001.

BENGIO, Y. Practical recommendations for gradient-based training of deep architectures. In: *Neural networks: Tricks of the trade: Second edition*. Berlin: Springer, 2012. p. 437–478.

BHATT, D.; PATEL, C.; TALSANIA, H.; PATEL, J.; VAGHELA, R.; PANDYA, S.; MODI, K.; GHAYVAT, H. Cnn variants for computer vision: History, architecture, application, challenges and future scope. *Electronics*, MDPI, Basel, v. 10, n. 20, p. 2470, 2021.

BHAVSAR, P.; SAFRO, I.; BOUAYNAYA, N.; POLIKAR, R.; DERA, D. Machine learning in transportation data analytics. In: *Data analytics for intelligent transportation systems*. Amsterdam: Elsevier, 2017. p. 283–307.

BOCCI, C.; CARLINI, E.; KILEEL, J. Hadamard products of linear spaces. *Journal of Algebra*, Elsevier, Amsterdam, v. 448, p. 595–617, 2016.

BUNN, D.; FARMER, E. D. Comparative models for electrical load forecasting. John Wiley and Sons Inc., New York, NY, 1985.

CARBONELL, J. G.; MICHALSKI, R. S.; MITCHELL, T. M. An overview of machine learning. *Machine learning*, Elsevier, Amsterdam, p. 3–23, 1983.

CDS. *CDS-Centralized Dataset*. 2010. Electricity commission for the load dataset. New Zealand.

CHAE, Y. T.; HORESH, R.; HWANG, Y.; LEE, Y. M. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy and Buildings*, Elsevier, Amsterdam, v. 111, p. 184–194, 2016.

CHANDRASEKARAN, K.; SELVARAJ, J.; AMALADOSS, C. R.; VEERAPAN, L. Hybrid renewable energy based smart grid system for reactive power management and voltage profile enhancement using artificial neural network. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, Taylor & Francis, London, v. 43, n. 19, p. 2419–2442, 2021.

CHANG, J.-H.; TSENG, C.-Y. Analysis of correlation between secondary pm2. 5 and factory pollution sources by using ann and the correlation coefficient. *Ieee Access*, IEEE, New York, v. 5, p. 22812–22822, 2017.

CHATFIELD, C. *Time-series forecasting*. Boca Raton, FL: Chapman and Hall/CRC, 2000.

CHAUVIN, Y.; RUMELHART, D. E. *Backpropagation: theory, architectures, and applications*. New York: Psychology press, 2013.

CHEN, J. I.-Z.; LAI, K.-L. Deep convolution neural network model for credit-card fraud detection and alert. *Journal of Artificial Intelligence*, *[S. l.]*, v. 3, n. 02, p. 101–112, 2021.

CHEN, S.; DONG, J.; HA, P.; LI, Y.; LABI, S. Graph neural network and reinforcement learning for multi-agent cooperative control of connected autonomous vehicles. *Computer-Aided Civil and Infrastructure Engineering*, Wiley Online Library, Hoboken, New Jersey, v. 36, n. 7, p. 838–857, 2021.

CHENG, D.; YANG, F.; XIANG, S.; LIU, J. Financial time series forecasting with multi-modality graph neural network. *Pattern Recognition*, Elsevier, Amsterdam, v. 121, p. 108218, 2022.

CHITALIA, G.; PIPATTANASOMPORN, M.; GARG, V.; RAHMAN, S. Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks. *Applied Energy*, Elsevier, Amsterdam, v. 278, p. 115410, 2020.

CHITTY-VENKATA, K. T.; EMANI, M.; VISHWANATH, V.; SOMANI, A. K. Neural architecture search for transformers: A survey. *IEEE Access*, IEEE, New York, v. 10, p. 108374–108412, 2022.

CHO, K.; MERRIËNBOER, B. V.; GULCEHRE, C.; BAHDANAU, D.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, *[S. l.]*, 2014.

CHOLLET, F. *Deep learning with Python*. New York: Simon and Schuster, 2021.

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

DAGUM, E. B. Time series modeling and decomposition. *Statistica*, Bologna, v. 70, n. 4, p. 433–457, 2010.

DAMBORG, M.; EL-SHARKAWI, M.; AGGOUNE, M.; MARKS, R. Potential of artificial neural networks in power system operation. In: IEEE. *IEEE International Symposium on Circuits and Systems*. New York, 1990. p. 2933–2937.

DAUPHIN, Y. N.; FAN, A.; AULI, M.; GRANGIER, D. Language modeling with gated convolutional networks. In: PMLR. *International conference on machine learning*. *[S. l.]*, 2017. p. 933–941.

DESAI, M.; SHAH, M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (mlp) and convolutional neural network (cnn). *Clinical eHealth*, Elsevier, Amsterdam, v. 4, p. 1–11, 2021.

DHANJAL, A. S.; SINGH, W. A comprehensive survey on automatic speech recognition using neural networks. *Multimedia Tools and Applications*, Springer, Berlin, v. 83, n. 8, p. 23367–23412, 2024.

FAUSETT, L. V. *Fundamentals of neural networks: architectures, algorithms and applications*. Delhi: Pearson Education India, 2006.

FEKRI, M. N.; PATEL, H.; GROLINGER, K.; SHARMA, V. Deep learning for load forecasting with smart meter data: Online adaptive recurrent neural network. *Applied Energy*,

Elsevier, Amsterdam, v. 282, p. 116177, 2021.

FENG, Q.; CHEN, H.; JIANG, R. Analysis of early warning of corporate financial risk via deep learning artificial neural network. *Microprocessors and Microsystems*, Elsevier, Amsterdam, v. 87, p. 104387, 2021.

FRINTROP, S.; JENSFELT, P. Attentional landmarks and active gaze control for visual slam. *IEEE Transactions on Robotics*, IEEE, New York, v. 24, n. 5, p. 1054–1065, 2008.

FÜEGI, J.; FRANCIS, J. Lovelace & babbage and the creation of the 1843'notes'. *ACM Inroads*, ACM New York, NY, USA, v. 6, n. 3, p. 78–86, 2015.

GAL, Y.; GHAHRAMANI, Z. A theoretically grounded application of dropout in recurrent neural networks. *Advances in neural information processing systems*, *[S. l.]*, v. 29, 2016.

GALASSI, A.; LIPPI, M.; TORRONI, P. Attention in natural language processing. *IEEE transactions on neural networks and learning systems*, IEEE, New York, v. 32, n. 10, p. 4291–4308, 2020.

GALIĆ, D.; STOJANOVIĆ, Z.; ČAJIĆ, E. Application of neural networks and machine learning in image recognition. *Tehnički vjesnik*, Strojarski fakultet u Slavonskom Brodu; Fakultet elektrotehnike, računarstva ..., *[S. l.]*, v. 31, n. 1, p. 316–323, 2024.

GAO, J.; CHEN, Y.; HU, W.; ZHANG, D. An adaptive deep-learning load forecasting framework by integrating transformer and domain knowledge. *Advances in Applied Energy*, Elsevier, Amsterdam, v. 10, p. 100142, 2023.

GAO, Y.; WEI, X.; ZHOU, L. Personalized qoe improvement for networking video service. *IEEE Journal on Selected Areas in Communications*, IEEE, New York, v. 38, n. 10, p. 2311–2323, 2020.

GASPARIN, A.; LUKOVIC, S.; ALIPPI, C. Deep learning for time series forecasting: The electric load case. *CAAI Transactions on Intelligence Technology*, Wiley Online Library, Hoboken, New Jersey, v. 7, n. 1, p. 1–25, 2022.

GHADAMI, N.; GHEIBI, M.; KIAN, Z.; FARAMARZ, M. G.; NAGHEDI, R.; EFTEKHARI, M.; FATHOLLAHI-FARD, A. M.; DULEBENETS, M. A.; TIAN, G. Implementation of solar energy in smart cities using an integration of artificial neural network, photovoltaic system and classical delphi methods. *Sustainable Cities and Society*, Elsevier, Amsterdam, v. 74, p. 103149, 2021.

GHAZANFAR, A. A.; SCHROEDER, C. E. Is neocortex essentially multisensory? *Trends in cognitive sciences*, Elsevier, Amsterdam, v. 10, n. 6, p. 278–285, 2006.

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

GROSS, G.; GALIANA, F. D. Short-term load forecasting. *Proceedings of the IEEE*, IEEE, New York, v. 75, n. 12, p. 1558–1573, 1987.

GURNEY, K. *An introduction to neural networks*. *[S. l.]*: CRC press, 2018.

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

HAYKIN, S. *Neural networks and learning machines, 3/E*. London: Pearson, 2009.

HELBING, G.; RITTER, M. Deep learning for fault detection in wind turbines. *Renewable and Sustainable Energy Reviews*, Elsevier, Amsterdam, v. 98, p. 189–198, 2018.

HIREŠ, M.; GAZDA, M.; DROTÁR, P.; PAH, N. D.; MOTIN, M. A.; KUMAR, D. K. Convolutional neural network ensemble for parkinson's disease detection from voice recordings. *Computers in biology and medicine*, Elsevier, Amsterdam, v. 141, p. 105021, 2022.

HONG, T. Energy forecasting: Past, present, and future. *Foresight: The International Journal of Applied Forecasting*, *[S. l.]*, n. 32, 2014.

HONG, T.; PINSON, P.; FAN, S.; ZAREIPOUR, H.; TROCCOLI, A.; HYNDMAN, R. J. *Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond*. Amsterdam: Elsevier, 2016. 896–913 p.

HONG, T.; PINSON, P.; WANG, Y.; WERON, R.; YANG, D.; ZAREIPOUR, H. Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, IEEE, New York, v. 7, p. 376–388, 2020.

HONG, T.; WANG, P. Fuzzy interaction regression for short term load forecasting. *Fuzzy optimization and decision making*, Springer, Berlin, v. 13, p. 91–103, 2014.

HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, National Acad Sciences, Washington, v. 79, n. 8, p. 2554–2558, 1982.

HU, L.; WANG, J.; GUO, Z.; ZHENG, T. Load forecasting based on lvmd-dbfcm load curve clustering and the cnn-ivia-blstm model. *Applied Sciences*, MDPI, Basel, v. 13, n. 12, p. 7332, 2023.

HUANG, B.; WANG, J. Applications of physics-informed neural networks in power systems-a review. *IEEE Transactions on Power Systems*, IEEE, New York, v. 38, n. 1, p. 572–588, 2022.

HUY, P. C.; MINH, N. Q.; TIEN, N. D.; ANH, T. T. Q. Short-term electricity load forecasting based on temporal fusion transformer model. *Ieee Access*, IEEE, New York, v. 10, p. 106296–106304, 2022.

HYNDMAN, R. J. *et al.* Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, *[S. l.]*, v. 4, n. 4, p. 43–46, 2006.

INNAN, N.; SAWAIKA, A.; DHOR, A.; DUTTA, S.; THOTA, S.; GOKAL, H.; PATEL, N.; KHAN, M. A.-Z.; THEODONIS, I.; BENNAI, M. Financial fraud detection using quantum graph neural networks. *Quantum Machine Intelligence*, Springer, Berlin, v. 6, n. 1, p. 1–18, 2024.

IVANKO, D.; RYUMIN, D.; KASHEVNIK, A.; AXYONOV, A.; KARNOV, A. Visual speech recognition in a driver assistance system. In: IEEE. *2022 30th European Signal Processing Conference (EUSIPCO)*. New York, 2022. p. 1131–1135.

JITTANON, S.; MENSIN, Y.; TERMRITTHIKUN, C. Intelligent forecasting of energy consumption using temporal fusion transformer model. In: IEEE. *2023 IEEE International Conference on Cybernetics and Innovations (ICCI)*. New York, 2023. p. 1–5.

JOVANOVIĆ, R. Ž.; SRETENOVIĆ, A. A.; ŽIVKOVIĆ, B. D. Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*, Elsevier, Amsterdam, v. 94, p. 189–199, 2015.

KAMAEV, V.; SHCHERBAKOV, M.; PANCHENKO, D.; SHCHERBAKOVA, N.; BREBELS, A. Using connectionist systems for electric energy consumption forecasting in shopping centers. *Automation and Remote Control*, Springer, Berlin, v. 73, n. 6, p. 1075–1084, 2012.

KANG, K.-M.; CHOI, B.-Y.; LEE, H.; AN, C.-G.; KIM, T.-G.; LEE, Y.-S.; KIM, M.; YI, J.; WON, C.-Y. Energy management method of hybrid ac/dc microgrid using artificial neural network. *Electronics*, MDPI, Basel, v. 10, n. 16, p. 1939, 2021.

KAUR, K.; KUMAR, Y.; KAUR, S. Artificial intelligence and machine learning in financial services to improve the business system. In: *Computational Intelligence for Modern Business Systems: Emerging Applications and Strategies*. Berlin: Springer, 2023. p. 3–30.

KHAN, M. A.-Z.; INNAN, N.; GALIB, A. A. O.; BENNAI, M. Brain tumor diagnosis using quantum convolutional neural networks. *arXiv preprint arXiv:2401.15804*, *[S. l.]*, 2024.

KHANDAY, N. Y.; SOFI, S. A. Deep insight: Convolutional neural network and its applications for covid-19 prognosis. *Biomedical Signal Processing and Control*, Elsevier, Amsterdam, v. 69, p. 102814, 2021.

KHATOON, S.; SINGH, A. K. *et al.* Effects of various factors on electric load forecasting: An overview. In: IEEE. *2014 6th IEEE Power India International Conference (PIICON)*. New York, 2014. p. 1–5.

KHODAYAR, M.; WANG, J.; MANTHOURI, M. Interval deep generative neural network for wind speed forecasting. *IEEE Transactions on Smart Grid*, IEEE, New York, v. 10, n. 4, p. 3974–3989, 2018.

KONG, K. G. H.; HOW, B. S.; TENG, S. Y.; LEONG, W. D.; FOO, D. C.; TAN, R. R.; SUNARSO, J. Towards data-driven process integration for renewable energy planning. *Current Opinion in Chemical Engineering*, Elsevier, Amsterdam, v. 31, p. 100665, 2021.

KOPRINSKA, I.; WU, D.; WANG, Z. Convolutional neural networks for energy time series forecasting. In: IEEE. *2018 international joint conference on neural networks (IJCNN)*. New York, 2018. p. 1–8.

KOVÁCS, Z. L. *Redes neurais artificiais*. São Paulo: Editora Livraria da Fisica, 2006.

KROSE, B.; SMAGT, P. v. d. *An introduction to neural networks*. Amsterdam: The University of Amsterdam, 1996.

KUBAT, M.; BRATKO, I.; MICHALSKI, R. S. A review of machine learning methods. *Machine learning and data mining: methods and applications*, Citeseer, *[S. l.]*, p. 3–69, 1998.

KUMAR, K.; PANDE, S. V.; KUMAR, T. C. A.; SAINI, P.; CHATURVEDI, A.; REDDY, P. C. S.; SHAH, K. B. Intelligent controller design and fault prediction using machine learning model. *International Transactions on Electrical Energy Systems*, Wiley Online Library, Hoboken, New Jersey, v. 2023, n. 1, p. 1056387, 2023.

LAGO, J.; BRABANDERE, K. D.; RIDDER, F. D.; SCHUTTER, B. D. Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data. *Solar Energy*, Elsevier, Amsterdam, v. 173, p. 566–577, 2018.

LAI, K. K.; YU, L.; WANG, S.; ZHOU, L. Credit risk analysis using a reliability-based neural network ensemble model. In: SPRINGER. *International Conference on Artificial Neural Networks*. Berlin, 2006. p. 682–690.

LAITSOS, V.; VONTZOS, G.; BARGIOTAS, D.; DASKALOPULU, A.; TSOUKALAS, L. H. Enhanced automated deep learning application for short-term load forecasting. *Mathematics*, MDPI, Basel, v. 11, n. 13, p. 2912, 2023.

LÄNGKVIST, M.; KARLSSON, L.; LOUTFI, A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern recognition letters*, Elsevier, Amsterdam, v. 42, p. 11–24, 2014.

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

LI, H.; SUN, J.; LIAO, X. A novel short-term load forecasting model by tcn-lstm structure with attention mechanism. In: IEEE. *2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*. New York, 2022. p. 178–182.

LI, H.; YEO, J. H.; BORNSHEUER, A. L.; OVERBYE, T. J. The creation and validation of load time series for synthetic electric power systems. *IEEE Transactions on Power Systems*, IEEE, New York, v. 36, n. 2, p. 961–969, 2020.

LI, L.; YANG, L.; ZENG, Y. Improving sentiment classification of restaurant reviews with attention-based bi-gru neural network. *Symmetry*, MDPI, Basel, v. 13, n. 8, p. 1517, 2021.

LI, X.; WANG, J.; YANG, C. Risk prediction in financial management of listed companies based on optimized bp neural network under digital economy. *Neural Computing and Applications*, Springer, Berlin, v. 35, n. 3, p. 2045–2058, 2023.

LIAO, H.; RADHAKRISHNAN, K. K. Short-term load forecasting with temporal fusion transformers for power distribution networks. In: IEEE. *2022 IEEE Sustainable Power and Energy Conference (iSPEC)*. New York, 2022. p. 1–5.

LIAO, W.; BAK-JENSEN, B.; PILLAI, J. R.; WANG, Y.; WANG, Y. A review of graph neural networks and their applications in power systems. *Journal of Modern Power Systems and Clean Energy*, SGEPRI, Nanjing, v. 10, n. 2, p. 345–360, 2021.

LIM, B.; ARIK, S. Ö.; LOEFF, N.; PFISTER, T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, Elsevier, Amsterdam, v. 37, n. 4, p. 1748–1764, 2021.

LIM, B.; ZOHREN, S. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, The Royal Society Publishing, London, v. 379, n. 2194, p. 20200209, 2021.

LIU, Z.; WANG, X.; XING, J.; REN, M.; XU, X. Short-term power load forecasting based on iwoa-attention-bilstm. In: IEEE. *2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS)*. New York, 2022. p. 444–450.

LUO, Y.; CHENG, Q.; YAN, S.; YANG, D. Situation awareness method of the distribution network based on emd-svd and elman neural network. *Energy Reports*, Elsevier, Amsterdam, v. 8, p. 632–639, 2022.

LUONG, M.-T.; PHAM, H.; MANNING, C. D. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, *[S. l.]*, 2015.

LYASHENKO, V.; LAARIEDH, F.; SOTNIK, S.; AYAZ, A. M. Recognition of voice commands based on neural network. TEM Journal, *[S. l.]*, 2021.

MARCOS, I. P.; JÚNIOR, A. P. P. Previsão do consumo de energia elétrica na região nordeste do brasil. *Revista de Engenharia e Pesquisa Aplicada*, *[S. l.]*, v. 6, n. 3, p. 21–30, 2021.

MARQUES, G.; AGARWAL, D.; DÍEZ, I. De la T. Automated medical diagnosis of covid-19 through efficientnet convolutional neural network. *Applied soft computing*, Elsevier, Amsterdam, v. 96, p. 106691, 2020.

MASON, E. S. *Energy requirements and economic growth*. Washington, D.C.: National planning association, 1955.

MCCULLOCH, W. S.; PITTS, W. A. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, New York, v. 5, p. 115–133, 1943.

MICHALSKI, R. S. A theory and methodology of inductive learning. In: *Machine learning*. Amsterdam: Elsevier, 1983. p. 83–134.

MILLER, A.; BLOTT, B.; HAMES, T. Review of neural network applications in medical imaging and signal processing. *Medical and Biological Engineering and Computing*, Springer, Berlin, v. 30, p. 449–464, 1992.

MISHRA, M.; NAYAK, J.; NAIK, B.; ABRAHAM, A. Deep learning in electrical utility industry: A comprehensive review of a decade of research. *Engineering Applications of Artificial Intelligence*, Elsevier, Amsterdam, v. 96, p. 104000, 2020.

MOGHRAM, I.; RAHMAN, S. Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on power systems*, IEEE, New York, v. 4, n. 4, p. 1484–1491, 1989.

MOHSENIMANESH, A.; ENTCHEV, E.; BOSNJAK, F. Hybrid model based on an sd selection, ceemdan, and deep learning for short-term load forecasting of an electric vehicle fleet. *Applied Sciences*, Basel, v. 12, n. 18, 2022. ISSN 2076-3417.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas*

inteligentes-Fundamentos e aplicações, Barueri, v. 1, n. 1, p. 32, 2003.

MONTEIRO, F.; OLIVEIRA, R.; ALMEIDA, J.; GONÇALVES, P.; BARTOLOMEU, P.; NETO, J.; DEUS, R. Electricity consumption dataset of a local energy cooperative. *Data in Brief*, Elsevier, Amsterdam, v. 54, p. 110373, 2024.

MORAL-CARCEDO, J.; VICÉNS-OTERO, J. Modelling the non-linear response of spanish electricity demand to temperature variations. *Energy economics*, Elsevier, Amsterdam, v. 27, n. 3, p. 477–494, 2005.

MORETTIN, P. A.; TOLOI, C. M. d. C. Análise de séries temporais. *[S. l.]*, 2022.

MOTIE, S.; RAAHEMI, B. Financial fraud detection using graph neural networks: A systematic review. *Expert Systems With Applications*, Elsevier, Amsterdam, p. 122156, 2023.

MOTLAGH, O.; BERRY, A.; O'NEIL, L. Clustering of residential electricity customers using load time series. *Applied energy*, Elsevier, Amsterdam, v. 237, p. 11–24, 2019.

MÜLLER, A. C.; GUIDO, S. *Introduction to machine learning with Python: a guide for data scientists*. Sebastopol, California: "O'Reilly Media, Inc.", 2016.

MURPHY, K. P. *Machine learning: a probabilistic perspective*. Cambridge, Massachusetts: MIT press, 2012.

MURTO, P. Neural network models for short-term load forecasting. *[S. l.]*, 1998.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. *[S. l.]*: [s.n.], 2010. p. 807–814.

NAZIR, A.; SHAIKH, A. K.; SHAH, A. S.; KHALIL, A. Forecasting energy consumption demand of customers in smart grid using temporal fusion transformer (tft). *Results in Engineering*, Elsevier, Amsterdam, v. 17, p. 100888, 2023.

NIELSEN, M. A. *Neural networks and deep learning*. San Francisco, CA: Determination press San Francisco, CA, USA, 2015. v. 25.

NILSSON, N. J. *Principles of artificial intelligence*. New York: Springer Science & Business Media, 1982.

NIU, Z.; ZHONG, G.; YU, H. A review on the attention mechanism of deep learning. *Neurocomputing*, Elsevier, Amsterdam, v. 452, p. 48–62, 2021.

O'DONOVAN, T. M. Short term forecasting: An introduction to the box-jenkins approach. *(No Title)*, *[S. l.]*, 1983.

ONS. *Operador Nacional do Sistema Elétrico - Glossário*. 2024. Acesso em: 31 maio 2024. Disponível em: <https://www.ons.org.br/paginas/conhecimento/glossario>.

OYEWOLE, A. T.; ADEOYE, O. B.; ADDY, W. A.; OKOYE, C. C.; OFODILE, O. C.; UGOCHUKWU, C. E. Predicting stock market movements using neural networks: a review and application study. *Computer Science & IT Research Journal*, *[S. l.]*, v. 5, n. 3, p. 651–670,

2024.

PARLOS, A. G.; OUFI, E.; MUTHUSAMI, J.; PATTON, A. D.; ATIYA, A. F. Development of an intelligent long-term electric load forecasting system. In: IEEE. *Proceedings of international conference on intelligent system application to power systems*. New York, 1996. p. 288–292.

PASCANU, R.; MIKOLOV, T.; BENGIO, Y. On the difficulty of training recurrent neural networks. In: PMLR. *International conference on machine learning*. *[S. l.]*, 2013. p. 1310–1318.

QIN, B.; GAO, X.; LI, F.; LIU, D.; ZHANG, Z.; DING, T.; HUANG, R. Cnn-gru-attention based short-term load forecasting of distribution networks. *Authorea Preprints*, Authorea, 2022.

RÁCZ, L.; SZABÓ, D.; GÖCSEI, G.; NÉMETH, B. Grid management technology for the integration of renewable energy sources into the transmission system. In: IEEE. *2018 7th International Conference on Renewable Energy Research and Applications (ICRERA)*. New York, 2018. p. 612–617.

RAGAB, D. A.; SHARKAS, M.; MARSHALL, S.; REN, J. Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ*, PeerJ Inc., San Mateo, California, v. 7, p. e6201, 2019.

RASAMOELINA, A. D.; ADJAILIA, F.; SINČÁK, P. A review of activation function for artificial neural network. In: IEEE. *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. New York, 2020. p. 281–286.

RAVINDRAN, R.; SANTORA, M. J.; JAMALI, M. M. Multi-object detection and tracking, based on dnn, for autonomous vehicles: A review. *IEEE Sensors Journal*, IEEE, New York, v. 21, n. 5, p. 5668–5677, 2020.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, Washington, D.C., v. 65, n. 6, p. 386, 1958.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning internal representations by error propagation, parallel distributed processing, explorations in the microstructure of cognition, ed. de rumelhart and j. mcclelland. vol. 1. 1986. *Biometrika*, *[S. l.]*, v. 71, p. 599–607, 1986.

RUSSELL, S.; NORVIG, P. *Inteligência Artificial*. 3. ed. Rio de Janeiro: GEN LTC, 2013. 1016 p. ISBN 8535237011.

RUSSELL, S. J.; NORVIG, P. *Artificial intelligence: a modern approach*. London: Pearson, 2016.

SADI, M. Z.; DEEBA, S. R.; SIDDIQUE, R. H. *et al.* Temperature sensitivity forecasting of electrical load. In: IEEE. *2010 4th International Power Engineering and Optimization Conference (PEOCO)*. New York, 2010. p. 244–248.

SANTOS, M. L.; GARCÍA-SANTIAGO, X.; CAMARERO, F. E.; GIL, G. B.; ORTEGA, P. C. Application of temporal fusion transformer for day-ahead pv power forecasting. *Energies*, MDPI, Basel, v. 15, n. 14, p. 5232, 2022.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural networks*, Elsevier, Amsterdam, v. 61, p. 85–117, 2015.

SFETSOS, A. Short-term load forecasting with a hybrid clustering algorithm. *IEE Proceedings-Generation, Transmission and Distribution*, IET, London, v. 150, n. 3, p. 257–262, 2003.

SHARMA, S.; SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. *Towards Data Sci*, *[S. l.]*, v. 6, n. 12, p. 310–316, 2017.

SHEPHERD, G. M. *The synaptic organization of the brain*. Oxford: Oxford university press, 2003.

SHRESTHA, A.; MAHMOOD, A. Review of deep learning algorithms and architectures. *IEEE access*, IEEE, New York, v. 7, p. 53040–53065, 2019.

SOUSA, C. A. de. An overview on weight initialization methods for feedforward neural networks. In: IEEE. *2016 International Joint Conference on Neural Networks (IJCNN)*. New York, 2016. p. 52–59.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, JMLR. org, *[S. l.]*, v. 15, n. 1, p. 1929–1958, 2014.

SUTSKEVER, I.; VINYALS, O.; LE, Q. V. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *[S. l.]*, v. 27, 2014.

SUTTON, R. S.; BARTO, A. G. *Reinforcement learning: An introduction*. Cambridge, Massachusetts: MIT press, 2018.

TIAN, Y. Artificial intelligence image recognition method based on convolutional neural network algorithm. *Ieee Access*, IEEE, New York, v. 8, p. 125731–125744, 2020.

TURING, A. M. Computing machinery and intelligence. *Mind*, *[S. l.]*, LIX, n. 236, p. 433–460, 1950. Disponível em: <https://doi.org/10.1093/mind/LIX.236.433>.

VALENCIA, V. A. N.; SANCHEZ-GALAN, J. E. Use of attention-based neural networks to short-term load forecasting in the republic of panama. In: IEEE. *2022 IEEE 40th Central America and Panama Convention (CONCAPAN)*. New York, 2022. p. 1–6.

VASKOVSKY, A. M.; CHVANOVA, M. S.; REBEZOV, M. B. Creation of digital twins of neural network technology of personalization of food products for diabetics. In: IEEE. *2020 4th Scientific School on Dynamics of Complex Networks and their Application in Intellectual Robotics (DCNAIR)*. New York, 2020. p. 251–253.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. Attention is all you need. *Advances in neural information*

processing systems, *[S. l.]*, v. 30, 2017.

VISCONDI, G. de F.; ALVES-SOUZA, S. N. A systematic literature review on big data for solar photovoltaic electricity generation forecasting. *Sustainable Energy Technologies and Assessments*, Elsevier, Amsterdam, v. 31, p. 54–63, 2019.

WANG, F.; TAX, D. M. Survey on the attention based rnn model and its applications in computer vision. *arXiv preprint arXiv:1601.06823*, *[S. l.]*, 2016.

WANG, H.; YI, H.; PENG, J.; WANG, G.; LIU, Y.; JIANG, H.; LIU, W. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy conversion and management*, Elsevier, Amsterdam, v. 153, p. 409–422, 2017.

WANG, N.; XUE, S.; LI, Z. Short-term load forecasting based on vmd and combined deep learning model. *IEEJ Transactions on Electrical and Electronic Engineering*, Wiley Online Library, Hoboken, New Jersey, v. 18, n. 7, p. 1067–1075, 2023.

WANG, R. Electricity load model forecasting research based on woabilstm-attention algorithm. In: SPIE. *Second International Conference on Algorithms, Microchips, and Network Applications (AMNA 2023)*. Bellingham, 2023. v. 12635, p. 64–69.

WANG, Y.-L.; JAHANSHAHI, H.; BEKIROS, S.; BEZZINA, F.; CHU, Y.-M.; ALY, A. A. Deep recurrent neural networks with finite-time terminal sliding mode control for a chaotic fractional-order financial system with market confidence. *Chaos, Solitons & Fractals*, Elsevier, Amsterdam, v. 146, p. 110881, 2021.

WANG, Z.; JIA, L.; REN, C. Attention-bidirectional lstm based short term power load forecasting. In: IEEE. *2021 Power System and Green Energy Conference (PSGEC)*. New York, 2021. p. 171–175.

WASSERMAN, P. D. *Neural computing: theory and practice*. New York: Van Nostrand Reinhold Co., 1989.

WEN, R.; TORKKOLA, K.; NARAYANASWAMY, B.; MADEKA, D. A multi-horizon quantile recurrent forecaster. *arXiv preprint arXiv:1711.11053*, *[S. l.]*, 2017.

WU, B.; YANG, L.; WANG, T. Dual attention lstm for building power load forecasting based on feature selection. In: SPIE. *International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2023)*. Bellingham, 2023. v. 12645, p. 1311–1320.

WU, Z.; PAN, S.; CHEN, F.; LONG, G.; ZHANG, C.; PHILIP, S. Y. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, IEEE, New York, v. 32, n. 1, p. 4–24, 2020.

XIAO, Y.; ZHANG, X.; XU, X.; LIU, X.; LIU, J. Deep neural networks with koopman operators for modeling and control of autonomous vehicles. *IEEE transactions on intelligent vehicles*, IEEE, New York, v. 8, n. 1, p. 135–146, 2022.

XIONG, J.; ZHANG, Y. A unifying framework of attention-based neural load forecasting. *IEEE Access*, IEEE, New York, 2023.

XIONG, J.; ZHOU, P.; CHEN, A.; ZHANG, Y. Attention-based neural load forecasting: A dynamic feature selection approach. In: IEEE. *2021 IEEE Power & Energy Society General Meeting (PESGM)*. New York, 2021. p. 01–05.

YAN, K.; LI, W.; JI, Z.; QI, M.; DU, Y. A hybrid lstm neural network for energy consumption forecasting of individual households. *Ieee Access*, IEEE, New York, v. 7, p. 157633–157642, 2019.

YANG, R.; YU, Y. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. *Frontiers in oncology*, Frontiers Media SA, Lausanne, v. 11, p. 638182, 2021.

YANG, S.; ZHU, K.; LI, F.; WENG, L.; CHENG, L. Mfamnet: Multi-scale feature attention mixture network for short-term load forecasting. *Applied Sciences*, MDPI, Basel, v. 13, n. 5, p. 2998, 2023.

YANG, W.; SPARROW, S. N.; WALLOM, D. C. A generalised multi-factor deep learning electricity load forecasting model for wildfire-prone areas. *arXiv preprint arXiv:2304.10686*, *[S. l.]*, 2023.

YU, Y.; SI, X.; HU, C.; ZHANG, J. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , Cambridge, Massachusetts, v. 31, n. 7, p. 1235–1270, 2019.

YUE, X. Application of ai technology in personalized recommendation system for financial services. *Applied Mathematics and Nonlinear Sciences*, Berlin, v. 9, n. 1, 2023.

ZANKE, P.; SONTAKKE, D.; HASSAN, A. Personalization in insurance and banking services: Ai and ml applications. *Journal of Artificial Intelligence Research and Applications*, *[S. l.]*, v. 4, n. 1, p. 39–55, 2024.

ZARNOWITZ, V.; OZYILDIRIM, A. Time series decomposition and measurement of business cycles, trends and growth cycles. *Journal of Monetary Economics*, Elsevier, Amsterdam, v. 53, n. 7, p. 1717–1739, 2006.

ZHANG, J.; LIN, T.; XIONG, B.; YE, L.; HUANG, Z.; ZHOU, N. Short-term load forecasting of power system based on attention mechanism cnn-bilstm. In: IEEE. *2022 China Automation Congress (CAC)*. New York, 2022. p. 1443–1446.

ZHANG, K.; LU, F.; YU, C.; DAI, C.; HUAYUN, Z.; ZHANG, T.; CHEN, X.; LU, J.; LIN, Z. Short-term electrical load forecasting based on attention-gru networks. In: IEEE. *2023 IEEE 6th International Electrical and Energy Conference (CIEEC)*. New York, 2023. p. 3338–3343.

ZHENG, H.; YUAN, J.; CHEN, L. Short-term load forecasting using emd-lstm neural networks with a xgboost algorithm for feature importance evaluation. *Energies*, MDPI, Basel, v. 10, n. 8, p. 1168, 2017.

ZHOU, X.; WU, X. Load prediction model based on lstm and attention mechanism. In: *Proceedings of the 5th International Conference on Computer Science and Software Engineering*. *[S. l.]*: [s.n.], 2022. p. 487–491.

ZOR, K.; TIMUR, O.; TEKE, A. A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting. In: IEEE. *2017 6th international youth conference on energy (IYCE)*. New York, 2017. p. 1–7.

ZULQARNAIN, M.; GHAZALI, R.; GHOUSE, M. G.; HASSIM, Y. M. M.; JAVID, I. Predicting financial prices of stock market using recurrent convolutional neural networks. *International Journal of Intelligent Systems and Applications (IJISA)*, *[S. l.]*, v. 12, n. 6, p. 21–32, 2020.

APÊNDICE A - VALORES OBSERVADOS E PREVISTOS

Neste apêndice, são fornecidos os dados observados e os resultados previstos provenientes dos experimentos realizados nesta pesquisa. A Tabela 21 apresenta os valores reais da Base de Dados B, enquanto a Tabela 22 exibe os dados observados da Base de Dados A. As Tabelas 23 a 28 detalham os valores previstos obtidos em cada um dos seis experimentos conduzidos.

Tabela 21 - Valores Observados da Base de Dados B.

ST	**Valores Observados**
ST1B	15042, 14122, 13800, 13466, 13346, 13214, 12988, 12916, 13178, 14012, 15520, 17302, 20050, 23264, 26920, 29752, 29800, 29370, 28510, 27088, 26060, 25870, 25510, 24758, 25152, 24746, 24364, 24028, 23014, 22966, 23574, 24626, 25618, 26944, 27530, 28056, 28866, 29226, 29550, 29168, 28008, 26514, 24292, 23026, 21722, 19894, 18222, 16584
ST2B	9920, 9384, 9002, 8724, 8466, 8494, 8466, 8360, 8360, 8648, 9232, 9864, 11722, 14230, 17678, 19488, 19496, 19038, 18616, 18090, 17878, 17554, 17352, 17390, 17458, 17102, 16786, 16394, 16242, 16212, 16328, 16844, 17246, 17870, 18290, 19190, 19880, 19880, 20292, 20856, 20388, 19334, 17888, 16768, 15466, 13924, 12238, 11032
ST3B	44130, 42268, 40864, 40148, 39260, 38716, 38430, 38374, 38630, 39462, 42038, 46822, 54724, 66466, 80412, 85910, 86712, 84508, 82072, 80268, 79782, 79438, 78836, 78866, 78636, 78178, 77606, 77204, 77606, 76918, 77148, 78866, 80526, 82960, 83132, 82732, 82244, 81758, 82646, 85080, 82674, 78894, 73740, 69788, 64720, 58334, 52606, 48168
ST4B	15578, 15234, 15004, 14736, 14718, 14814, 14756, 14670, 14862, 15024, 15588, 16688, 18130, 20024, 23246, 24948, 25196, 24928, 23664, 22948, 22040, 21438, 21648, 21494, 21782, 21400, 20530, 20132, 20270, 20208, 20550, 21324, 22194, 22586, 22968, 23044, 23064, 22930, 23188, 23638, 23016, 22242, 21162, 20214, 19038, 17768, 16926, 16152

Fonte:Autoria própria.

Tabela 22 - Valores Observados da Base de Dados A.

ST	**Valores Observados**
ST1A	0,05, 0,07, 0,07, 0,05, 0,05, 0,06, 0,05, 0,04, 0,05, 0,05, 0,04, 0,03, 0,03, 0,02, 0,03, 0,03, 0,03, 0,05, 0,07, 0,07, 0,06, 0,05, 0,04, 0,03, 0,03, 0,03, 0,05, 0,06, 0,07, 0,07, 0,08, 0,1, 0,11, 0,1, 0,09, 0,08, 0,09, 0,1, 0,1, 0,09, 0,1, 0,11, 0,09, 0,08, 0,03, 0,03, 0,11, 0,09, 0,4, 0,63, 0,45, 0,44, 0,55, 0,57, 0,61, 0,53, 0,15, 0,12, 0,08, 0,05, 0,06, 0,06, 0,08, 0,09, 0,1, 0,14, 0,11, 0,08, 0,12, 0,1, 0,11, 0,1, 0,11, 0,1, 0,09, 0,1, 0,1, 0,1, 0,09, 0,09, 0,16, 0,19, 0,18, 0,21, 0,12, 0,12, 0,12, 0,13, 0,09, 0,07, 0,05, 0,05, 0,04, 0,04, 0,04, 0,06
ST2A	0,04, 0,05, 0,07, 0,06, 0,04, 0,02, 0,03, 0,03, 0,07, 0,05, 0,03, 0,01, 0,02, 0,03, 0,03, 0,06, 0,03, 0,03, 0,02, 0,02, 0,01, 0,05, 0,06, 0,04, 0,12, 0,1, 0,12, 0,1, 0,12, 0,11, 0,08, 0,09, 0,07, 0,09, 0,11, 0,12, 0,09, 0,05, 0,05, 0,04, 0,06, 0,07, 0,09, 0,04, 0,05, 0,05, 0,06, 0,08, 0,1, 0,12, 0,07, 0,08, 0,04, 0,06, 0,08, 0,11, 0,06, 0,06, 0,08, 0,09, 0,05, 0,09, 0,08, 0,07, 0,05, 0,06, 0,06, 0,08, 0,1, 0,09, 0,08, 0,06, 0,08, 0,06, 0,12, 0,11, 0,14, 0,14, 0,12, 0,16, 0,4, 0,64, 0,22, 0,21, 0,22, 0,2, 0,18, 0,17, 0,17, 0,12, 0,12, 0,08, 0,04, 0,09, 0,07, 0,04
ST3A	0,1, 0,09, 0,08, 0,09, 0,13, 0,11, 0,08, 0,08, 0,08, 0,1, 0,12, 0,08, 0,06, 0,06, 0,07, 0,07, 0,09, 0,05, 0,08, 0,06, 0,05, 0,09, 0,08, 0,07, 0,05, 0,05, 0,07, 0,11, 0,06, 0,05, 0,06, 0,09, 0,11, 0,09, 0,15, 0,08, 0,09, 0,11, 0,11, 0,06, 0,09, 0,08, 0,11, 0,08, 0,11, 0,11, 0,12, 0,11, 0,08, 0,09, 0,07, 0,08, 0,08, 0,12, 0,09, 0,07, 0,07, 0,05, 0,09, 0,07, 0,05, 0,05, 0,07, 0,05, 0,08, 0,06, 0,07, 0,41, 0,51, 0,1, 0,12, 0,1, 0,09, 0,1, 0,09, 0,14, 0,13, 0,09, 0,07, 0,09, 0,08, 0,09, 0,12, 0,11, 0,11, 0,07, 0,08, 0,09, 0,11, 0,05, 0,05, 0,06, 0,08, 0,08, 0,08, 0,05
ST4A	0,04, 0,04, 0,03, 0,03, 0,03, 0,03, 0,03, 0,03, 0,01, 0,01, 0,01, 0,01, 0,01, 0,02, 0,05, 0,05, 0,04, 0,03, 0,03, 0,03, 0,03, 0,03, 0,03, 0,02, 0,01, 0,01, 0,01, 0,01, 0,04, 0,10, 0,08, 0,04, 0,11, 0,07, 0,04, 0,04, 0,05, 0,03, 0,01, 0,01, 0,01, 0,01, 0,02, 0,03, 0,02, 0,01, 0,04, 0,03, 0,03, 0,03, 0,05, 0,05, 0,02, 0,01, 0,01, 0,01, 0,01, 0,02, 0,03, 0,03, 0,01, 0,03, 0,03, 0,03, 0,03, 0,04, 0,05, 0,04, 0,04, 0,03, 0,04, 0,02, 0,06, 0,07, 0,04, 0,04, 0,04, 0,05, 0,12, 0,25, 0,57, 0,47, 0,13, 0,12, 0,07, 0,05, 0,05, 0,05, 0,10, 0,06, 0,08, 0,04, 0,03, 0,03, 0,03, 0,03

Fonte:Autoria própria.

Tabela 23 - Valores Previstos no Experimento 1.

ST	**Valores Previstos**
ST1A	0,05, 0,06, 0,07, 0,06, 0,06, 0,05, 0,04, 0,03, 0,04, 0,04, 0,04, 0,03, 0,04, 0,05, 0,05, 0,04, 0,04, 0,05, 0,05, 0,05, 0,05, 0,04, 0,04, 0,04, 0,04, 0,03, 0,05, 0,05, 0,07, 0,08, 0,09, 0,08, 0,06, 0,09, 0,09, 0,06, 0,07, 0,08, 0,08, 0,09, 0,10, 0,12, 0,11, 0,08, 0,07, 0,06, 0,09, 0,10, 0,38, 0,50, 0,53, 0,51, 0,34, 0,34, 0,45, 0,31, 0,13, 0,10, 0,07, 0,07, 0,07, 0,08, 0,09, 0,09, 0,10, 0,15, 0,10, 0,05, 0,09, 0,08, 0,08, 0,08, 0,08, 0,08, 0,08, 0,08, 0,09, 0,10, 0,11, 0,12, 0,16, 0,17, 0,15, 0,15, 0,10, 0,09, 0,09, 0,08, 0,07, 0,08, 0,07, 0,08, 0,07, 0,07, 0,06, 0,06
ST2A	0,03, 0,03, 0,04, 0,04, 0,04, 0,04, 0,04, 0,04, 0,04, 0,04, 0,04, 0,03, 0,04, 0,04, 0,04, 0,04, 0,04, 0,03, 0,03, 0,03, 0,03, 0,04, 0,06, 0,04, 0,08, 0,10, 0,10, 0,08, 0,10, 0,06, 0,05, 0,06, 0,06, 0,07, 0,08, 0,06, 0,07, 0,06, 0,06, 0,06, 0,07, 0,09, 0,08, 0,07, 0,06, 0,08, 0,08, 0,10, 0,10, 0,10, 0,07, 0,07, 0,06, 0,08, 0,08, 0,10, 0,06, 0,07, 0,08, 0,10, 0,07, 0,09, 0,08, 0,07, 0,07, 0,07, 0,08, 0,09, 0,12, 0,09, 0,07, 0,08, 0,06, 0,06, 0,07, 0,07, 0,09, 0,09, 0,11, 0,13, 0,42, 0,62, 0,18, 0,16, 0,20, 0,21, 0,19, 0,19, 0,15, 0,10, 0,09, 0,09, 0,07, 0,07, 0,06, 0,06
ST3A	0,08, 0,09, 0,08, 0,09, 0,10, 0,11, 0,09, 0,09, 0,08, 0,09, 0,11, 0,08, 0,07, 0,07, 0,08, 0,09, 0,09, 0,07, 0,08, 0,08, 0,08, 0,08, 0,07, 0,07, 0,07, 0,06, 0,07, 0,07, 0,07, 0,07, 0,07, 0,07, 0,11, 0,07, 0,08, 0,07, 0,07, 0,07, 0,08, 0,08, 0,07, 0,07, 0,11, 0,12, 0,12, 0,09, 0,08, 0,09, 0,08, 0,09, 0,06, 0,09, 0,09, 0,08, 0,08, 0,06, 0,07, 0,07, 0,07, 0,07, 0,07, 0,07, 0,07, 0,08, 0,07, 0,41, 0,51, 0,08, 0,12, 0,10, 0,10, 0,11, 0,09, 0,10, 0,13, 0,10, 0,09, 0,09, 0,08, 0,09, 0,08, 0,11, 0,11, 0,09, 0,09, 0,11, 0,11, 0,10, 0,09, 0,10, 0,10, 0,09, 0,09, 0,09, 0,10, 0,10

Fonte:Autoria própria.

Tabela 24 - Valores Previstos no Experimento 2.

ST	**Valores Previstos**
ST1B	14720,74, 14082,18, 13631,63, 13384,67, 13155,78, 13019,11, 12872,35, 12831,48, 12986,31, 13451,71, 14756,32, 16737,24, 19503,88, 23013,45, 26854,66, 29248,53, 29615,97, 28839,41, 28127,88, 27282,93, 26562,95, 25929,82, 25785,51, 25646,89, 25530,72, 25186,66, 24714,03, 24027,18, 23541,97, 23302,39, 23762,93, 24950,37, 26083,23, 27316,60, 27738,65, 27785,67, 27543,42, 27415,98, 27185,30, 28003,30, 27874,10, 26522,13, 24068,09, 22822,84, 21108,08, 19464,84, 17747,59, 16292,66
ST2B	9711,95, 9251,58, 8896,56, 8682,32, 8503,80, 8404,83, 8316,87, 8295,04, 8343,54, 8536,32, 9001,61, 9808,19, 11503,82, 14363,20, 17938,73, 19592,61, 19842,43, 19258,81, 18759,54, 18278,32, 18111,83, 17867,46, 17835,29, 17776,89, 17751,42, 17531,43, 17211,87, 16752,54, 16530,36, 16344,75, 16547,01, 17133,76, 17563,98, 18178,65, 18748,81, 19661,71, 20000,51, 19947,69, 19812,17, 20699,24, 20412,65, 19422,91, 17674,07, 16640,15, 15143,71, 13672,31, 12109,28, 10833,92
ST3B	43591,05, 41486,62, 40110,84, 39316,87, 38705,16, 38361,63, 38108,82, 38055,66, 38565,28, 39641,59, 42400,14, 46609,84, 54477,27, 65828,30, 78313,02, 84626,23, 85166,76, 83396,21, 82462,69, 81494,51, 81225,09, 80912,49, 81238,51, 81464,90, 81494,42, 81226,83, 80876,41, 80304,76, 80188,45, 80058,27, 80811,28, 83014,45, 85079,20, 86543,30, 86473,80, 84599,39, 81814,40, 80221,10, 79046,47, 82586,02, 81859,53, 78722,88, 73832,26, 69633,05, 63495,44, 57755,25, 52298,29, 47789,09

Fonte:Autoria própria.

Tabela 25 - Valores Previstos no Experimento 3.

ST	**Valores Previstos**
ST1A	0,06, 0,06, 0,05, 0,04, 0,04, 0,04, 0,03, 0,03, 0,03, 0,05, 0,05, 0,05, 0,05, 0,04, 0,04, 0,03, 0,03, 0,04, 0,04, 0,05, 0,05, 0,04, 0,04, 0,04, 0,03, 0,03, 0,04, 0,05, 0,07, 0,08, 0,08, 0,10, 0,12, 0,08, 0,07, 0,05, 0,06, 0,07, 0,08, 0,09, 0,10, 0,12, 0,09, 0,07, 0,05, 0,05, 0,08, 0,06, 0,38, 0,63, 0,48, 0,49, 0,55, 0,56, 0,57, 0,52, 0,16, 0,07, 0,06, 0,06, 0,07, 0,07, 0,07, 0,11, 0,07, 0,08, 0,08, 0,09, 0,09, 0,09, 0,08, 0,07, 0,08, 0,08, 0,09, 0,11, 0,10, 0,11, 0,08, 0,08, 0,13, 0,11, 0,14, 0,20, 0,09, 0,10, 0,09, 0,10, 0,08, 0,07, 0,07, 0,07, 0,06, 0,06, 0,07, 0,07
ST2A	0,04, 0,05, 0,05, 0,05, 0,04, 0,04, 0,04, 0,04, 0,04, 0,04, 0,04, 0,03, 0,03, 0,03, 0,02, 0,03, 0,04, 0,04, 0,03, 0,03, 0,03, 0,04, 0,04, 0,05, 0,08, 0,08, 0,08, 0,07, 0,10, 0,12, 0,08, 0,10, 0,06, 0,08, 0,12, 0,13, 0,10, 0,07, 0,07, 0,05, 0,07, 0,08, 0,10, 0,07, 0,08, 0,06, 0,06, 0,09, 0,06, 0,07, 0,05, 0,06, 0,05, 0,05, 0,06, 0,07, 0,06, 0,05, 0,05, 0,05, 0,05, 0,09, 0,08, 0,09, 0,06, 0,06, 0,06, 0,08, 0,10, 0,09, 0,08, 0,06, 0,06, 0,06, 0,10, 0,09, 0,16, 0,13, 0,11, 0,11, 0,42, 0,65, 0,21, 0,18, 0,23, 0,15, 0,15, 0,14, 0,16, 0,11, 0,11, 0,09, 0,07, 0,07, 0,08, 0,07
ST3A	0,10, 0,09, 0,09, 0,09, 0,11, 0,11, 0,08, 0,10, 0,09, 0,10, 0,13, 0,07, 0,07, 0,07, 0,08, 0,09, 0,09, 0,06, 0,08, 0,07, 0,07, 0,09, 0,07, 0,06, 0,07, 0,07, 0,08, 0,11, 0,06, 0,06, 0,08, 0,10, 0,12, 0,09, 0,09, 0,06, 0,10, 0,06, 0,09, 0,06, 0,09, 0,10, 0,13, 0,08, 0,12, 0,11, 0,11, 0,14, 0,10, 0,10, 0,08, 0,08, 0,08, 0,12, 0,08, 0,08, 0,07, 0,07, 0,09, 0,07, 0,06, 0,06, 0,08, 0,08, 0,09, 0,06, 0,07, 0,46, 0,53, 0,10, 0,07, 0,06, 0,07, 0,07, 0,07, 0,11, 0,10, 0,10, 0,09, 0,10, 0,09, 0,10, 0,13, 0,12, 0,13, 0,09, 0,10, 0,10, 0,11, 0,09, 0,08, 0,08, 0,09, 0,09, 0,09, 0,08

Fonte:Autoria própria.

Tabela 26 - Valores Previstos no Experimento 4.

ST	**Valores Previstos**
ST1A	0,04, 0,04, 0,04, 0,04, 0,05, 0,05, 0,05, 0,05, 0,05, 0,05, 0,04, 0,04, 0,04, 0,03, 0,03, 0,03, 0,03, 0,04, 0,05, 0,07, 0,07, 0,05, 0,04, 0,04, 0,04, 0,03, 0,04, 0,04, 0,05, 0,05, 0,06, 0,07, 0,08, 0,08, 0,09, 0,08, 0,09, 0,11, 0,10, 0,10, 0,11, 0,09, 0,07, 0,06, 0,05, 0,05, 0,09, 0,09, 0,42, 0,64, 0,47, 0,46, 0,53, 0,55, 0,56, 0,50, 0,14, 0,08, 0,06, 0,06, 0,06, 0,06, 0,06, 0,06, 0,06, 0,07, 0,07, 0,07, 0,08, 0,09, 0,08, 0,08, 0,09, 0,10, 0,10, 0,09, 0,10, 0,09, 0,09, 0,08, 0,09, 0,14, 0,18, 0,24, 0,13, 0,13, 0,11, 0,11, 0,08, 0,06, 0,06, 0,06, 0,05, 0,05, 0,05, 0,05
ST2A	0,05, 0,05, 0,05, 0,04, 0,03, 0,03, 0,02, 0,03, 0,06, 0,05, 0,03, 0,02, 0,02, 0,03, 0,04, 0,06, 0,04, 0,03, 0,02, 0,02, 0,02, 0,03, 0,04, 0,04, 0,06, 0,08, 0,13, 0,12, 0,13, 0,12, 0,09, 0,07, 0,07, 0,09, 0,11, 0,11, 0,10, 0,06, 0,06, 0,06, 0,04, 0,05, 0,05, 0,05, 0,05, 0,07, 0,07, 0,10, 0,09, 0,06, 0,06, 0,05, 0,08, 0,08, 0,12, 0,07, 0,07, 0,06, 0,10, 0,07, 0,09, 0,08, 0,06, 0,05, 0,05, 0,06, 0,07, 0,10, 0,09, 0,07, 0,05, 0,08, 0,07, 0,10, 0,08, 0,15, 0,11, 0,11, 0,15, 0,43, 0,63, 0,24, 0,21, 0,19, 0,18, 0,16, 0,16, 0,14, 0,12, 0,09, 0,08, 0,07, 0,07, 0,07, 0,08
ST3A	0,08, 0,08, 0,08, 0,08, 0,10, 0,08, 0,07, 0,08, 0,07, 0,09, 0,08, 0,07, 0,07, 0,08, 0,08, 0,07, 0,08, 0,07, 0,08, 0,07, 0,05, 0,06, 0,07, 0,07, 0,06, 0,05, 0,05, 0,09, 0,07, 0,07, 0,06, 0,08, 0,09, 0,10, 0,13, 0,08, 0,10, 0,09, 0,09, 0,07, 0,10, 0,09, 0,09, 0,06, 0,11, 0,11, 0,12, 0,09, 0,07, 0,08, 0,06, 0,08, 0,07, 0,09, 0,07, 0,09, 0,07, 0,07, 0,07, 0,08, 0,06, 0,06, 0,06, 0,05, 0,08, 0,06, 0,08, 0,37, 0,49, 0,09, 0,11, 0,11, 0,10, 0,11, 0,08, 0,12, 0,13, 0,10, 0,09, 0,10, 0,09, 0,11, 0,12, 0,09, 0,08, 0,07, 0,07, 0,08, 0,09, 0,07, 0,07, 0,06, 0,06, 0,08, 0,09, 0,06
ST4A	0,05, 0,04, 0,02, 0,03, 0,04, 0,03, 0,03, 0,03, 0,02, 0,02, 0,02, 0,02, 0,02, 0,02, 0,04, 0,04, 0,04, 0,04, 0,03, 0,03, 0,03, 0,03, 0,02, 0,02, 0,02, 0,02, 0,02, 0,04, 0,10, 0,08, 0,04, 0,11, 0,08, 0,04, 0,04, 0,06, 0,02, 0,01, 0,03, 0,01, 0,02, 0,03, 0,03, 0,02, 0,02, 0,03, 0,03, 0,03, 0,04, 0,05, 0,04, 0,02, 0,02, 0,01, 0,02, 0,02, 0,03, 0,03, 0,02, 0,02, 0,03, 0,03, 0,04, 0,04, 0,05, 0,04, 0,03, 0,04, 0,03, 0,04, 0,03, 0,06, 0,07, 0,04, 0,03, 0,04, 0,06, 0,11, 0,27, 0,57, 0,45, 0,16, 0,14, 0,09, 0,09, 0,06, 0,07, 0,07, 0,04, 0,06, 0,05, 0,05, 0,05, 0,05, 0,04

Fonte:Autoria própria.

Tabela 27 - Valores Previstos no Experimento 5.

ST	**Valores Previstos**
ST1B	14958,35, 14286,02, 13769,65, 13458,82, 13173,19, 13022,12, 12858,88, 12840,77, 12982,45, 13433,32, 14920,45, 17328,40, 20141,44, 23274,45, 26900,74, 29354,24, 29672,73, 28960,45, 28288,54, 27586,84, 26892,08, 26505,45, 26254,94, 26127,28, 25911,58, 25601,65, 25070,88, 24297,41, 23674,69, 23535,22, 24054,20, 25172,37, 26477,70, 27345,58, 27723,14, 27787,60, 27777,12, 27710,37, 27881,72, 28714,37, 28101,51, 26457,36, 24310,96, 23000,03, 21351,63, 19635,64, 17775,46, 16255,08
ST2B	10074,35, 9442,95, 8999,46, 8733,26, 8522,23, 8413,14, 8311,84, 8291,58, 8338,90, 8522,20, 9031,90, 9845,84, 11580,21, 14308,39, 17745,37, 19495,85, 19476,18, 18934,89, 18546,61, 18236,04, 17974,41, 17732,03, 17624,37, 17555,04, 17511,77, 17371,14, 17163,75, 16861,84, 16613,38, 16569,94, 16726,94, 17089,94, 17567,90, 18064,00, 18698,72, 19407,60, 19900,37, 19978,72, 19982,27, 20296,38, 20058,07, 19352,59, 17991,18, 16819,88, 15242,21, 13643,39, 12114,86, 10800,40
ST3B	44166,43, 42299,96, 40882,06, 39942,49, 39142,67, 38644,68, 38258,92, 38141,63, 38533,89, 39389,76, 42400,87, 46659,00, 55280,07, 66390,74, 79971,95, 86122,86, 85735,04, 83635,59, 81793,44, 80748,47, 80417,73, 80272,16, 80425,36, 80522,77, 80528,11, 80320,72, 79995,78, 79840,52, 79683,84, 80112,38, 81209,35, 83046,81, 84791,88, 85969,20, 86071,76, 84663,86, 82429,72, 80965,85, 80384,85, 82277,09, 81711,73, 79592,04, 75154,91, 71262,23, 65203,96, 58979,74, 52711,00, 47995,75

Fonte:Autoria própria.

Tabela 28 - Valores Previstos no Experimento 6.

ST	**Valores Previstos**
ST1B	14720,74, 14082,18, 13631,63, 13384,67, 13155,78, 13019,11, 12872,35, 12831,48, 12986,31, 13451,71, 14756,32, 16737,24, 19503,88, 23013,45, 26854,66, 29248,53, 29615,97, 28839,41, 28127,88, 27282,93, 26562,95, 25929,82, 25785,51, 25646,89, 25530,72, 25186,66, 24714,03, 24027,18, 23541,97, 23302,39, 23762,93, 24950,37, 26083,23, 27316,60, 27738,65, 27785,67, 27543,42, 27415,98, 27185,30, 28003,30, 27874,10, 26522,13, 24068,09, 22822,84, 21108,08, 19464,84, 17747,59, 16292,66
ST2B	9711,95, 9251,58, 8896,56, 8682,32, 8503,80, 8404,83, 8316,87, 8295,04, 8343,54, 8536,32, 9001,61, 9808,19, 11503,82, 14363,20, 17938,73, 19592,61, 19842,43, 19258,81, 18759,54, 18278,32, 18111,83, 17867,46, 17835,29, 17776,89, 17751,42, 17531,43, 17211,87, 16752,54, 16530,36, 16344,75, 16547,01, 17133,76, 17563,98, 18178,65, 18748,81, 19661,71, 20000,51, 19947,69, 19812,17, 20699,24, 20412,65, 19422,91, 17674,07, 16640,15, 15143,71, 13672,31, 12109,28, 10833,92
ST3B	43856,63, 41942,75, 40704,57, 39804,16, 39130,23, 38610,72, 38317,37, 38187,20, 38539,71, 39402,41, 42392,08, 46820,39, 55263,76, 66630,55, 80001,24, 84992,59, 85295,73, 83613,78, 82712,03, 81234,43, 80468,30, 80007,34, 79792,80, 79658,02, 79321,43, 78836,28, 78293,65, 77849,03, 77558,80, 77413,47, 77885,31, 79507,26, 80988,27, 82893,95, 83755,67, 83461,06, 82483,40, 81458,17, 82174,74, 84264,06, 81966,62, 79358,14, 74756,20, 69516,38, 63585,55, 57862,81, 52696,52, 48455,90
ST4B	15533,41, 15221,11, 15006,94, 14847,70, 14752,54, 14704,28, 14708,53, 14747,97, 14903,38, 15108,18, 15802,52, 16714,24, 18332,04, 20075,87, 22934,92, 25046,86, 24915,46, 24564,99, 24147,88, 23468,69, 22421,01, 21579,60, 21930,51, 22129,19, 22064,01, 21816,04, 21176,82, 20756,01, 20491,82, 20410,41, 21039,85, 21799,53, 22420,26, 22733,82, 23010,87, 23326,24, 23305,98, 23135,18, 22984,43, 23380,27, 23191,74, 22556,65, 21413,07, 20342,59, 19241,99, 18143,74, 17120,44, 16311,57

Fonte:Autoria própria.