

**UNIVERSIDADE FEDERAL DO PARÁ**  
**INSTITUTO DE TECNOLOGIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**THIAGO NICOLAU MAGALHÃES DE SOUZA CONTE**

**PREVISÃO DE SÉRIES TEMPORAIS NO SISTEMA ELÉTRICO BRASILEIRO**  
**UTILIZANDO PREDITORES BASEADOS EM APRENDIZADO DE MÁQUINA:**  
**UMA ANÁLISE EMPÍRICA**

TD: 07/2024

**BELÉM/PARÁ**

2024

THIAGO NICOLAU MAGALHÃES DE SOUZA CONTE

**PREVISÃO DE SÉRIES TEMPORAIS NO SISTEMA ELÉTRICO BRASILEIRO  
UTILIZANDO PREDITORES BASEADOS EM APRENDIZADO DE MÁQUINA:  
UMA ANÁLISE EMPÍRICA**

TD: 07/2024

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica do Instituto de Tecnologia da Universidade Federal do Pará como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica na área de Computação Aplicada.

Orientador: Prof. Dr. Roberto Célio Limão de Oliveira.

BELÉM/PARÁ

2024

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD  
Sistema de Bibliotecas da Universidade Federal do Pará  
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

---

- C761p Conte, Thiago Nicolau Magalhães de Souza.  
Previsão de Séries Temporais no Sistema Elétrico Brasileiro  
Utilizando Preditores Baseados em Aprendizado de Máquina : Uma  
Análise Empírica / Thiago Nicolau Magalhães de Souza Conte. —  
2024.  
XII, 116 f. : il. color.
- Orientador(a): Prof. Dr. Roberto Célio Limão de Oliveira  
Tese (Doutorado) - Universidade Federal do Pará, Instituto de  
Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica,  
Belém, 2024.
1. Preço de Liquidação das Diferenças. 2. Velocidade do  
Vento. 3. Sistema Elétrico Brasileiro. 4. Aprendizagem de  
Máquina. 5. Comitê de Máquinas. I. Título.

CDD 006.31

---

UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**“PREVISÃO DE SÉRIES TEMPORAIS NO SISTEMA ELÉTRICO BRASILEIRO  
UTILIZANDO PREDITORES BASEADOS EM APRENDIZADO DE MÁQUINA: UMA  
ANÁLISE EMPÍRICA”**

AUTOR: **THIAGO NICOLAU MAGALHÃES DE SOUZA CONTE**

TESE DE DOUTORADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO  
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO  
JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA  
ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 05/04/2024

BANCA EXAMINADORA:

Documento assinado digitalmente  
 **Roberto Celio Limao de Oliveira**  
Data: 08/04/2024 15:34:13-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. Roberto Célio Limão de Oliveira**

(Orientador - DDCEE /UEPA)

Documento assinado digitalmente

 **ADRIANA ROSA GARCEZ CASTRO**  
Data: 08/04/2024 18:29:19-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof.ª Dr.ª Adriana Rosa Garcez Castro**

(Avaliadora Interna - PPGEE/UFPA)

Documento assinado digitalmente

 **CAROLINA DE MATTOS AFFONSO**  
Data: 08/04/2024 16:57:54-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof.ª Dr.ª Carolina de Mattos Affonso**

(Avaliadora Interna - PPGEE/UFPA)

**ANDRE CRISTIANO SILVA** Assinado de forma digital por ANDRE  
**MELO:42355214204** CRISTIANO SILVA MELO:42355214204  
Dados: 2024.04.12 10:06:53 -03'00'

---

**Prof. Dr. André Cristiano Silva Melo**

(Avaliador Externo - UEPA)

Documento assinado digitalmente

 **HAROLDO GOMES BARROSO FILHO**  
Data: 10/04/2024 16:43:06-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. Haroldo Gomes Barroso Filho**

Documento assinado digitalmente

 **MARCUS DE BARROS BRAGA**  
Data: 11/04/2024 11:15:41-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. Marcus de Barros Braga**

(Avaliador Externo - UFRA)

VISTO:

---

**Prof. Dr. Diego Lisboa Cardoso**

(Coordenador do PPGEE/ITEC/UFPA)

Dedico este trabalho à minha mãe, Maria do Socorro, meu mais admirável exemplo de determinação, que sempre esteve ao meu lado, apoiando-me em todas as etapas da vida. Sua capacidade de transformar os momentos difíceis em razões para felicidade são fontes de inspiração constantes em minha jornada.

## AGRADECIMENTOS

À Universidade Federal do Pará, agradeço pela oportunidade concedida de participar do programa de pós-graduação em Engenharia Elétrica (PPGEE/UFPA), especializando-me na área de Computação Aplicada com foco em Inteligência Computacional.

Agradeço a todos os professores do programa de doutorado em Computação Aplicada, cuja dedicação e expertise foram fundamentais para transmitir seus valiosos conhecimentos durante minha jornada acadêmica.

Ao Prof. Dr. Roberto Limão, expresso minha sincera gratidão pelo apoio inestimável, orientação precisa e atenção dedicada durante a realização deste trabalho. Suas contribuições foram fundamentais para o meu sucesso e meu crescimento acadêmico.

Aos meus queridos pais, Nicolau e Maria do Socorro (in memoriam), expresso minha gratidão pelo incansável esforço, dedicação e compreensão que sempre demonstraram ao longo desta e de outras jornadas.

Agradeço aos meus irmãos Bruno e Hugo por estarem ao meu lado, me acompanhando e incentivando na busca dos meus objetivos. Sua presença e apoio foram essenciais em cada passo desta jornada.

À minha esposa, Analu Conte, agradeço imensamente pela compreensão, apoio e incentivo incondicionais. Também expresso minha gratidão à nossa filha, Luma, cuja presença renova constantemente minhas energias e inspira-me a seguir em frente.

Agradeço a todos que, de alguma maneira, contribuíram para a conclusão do meu curso de Pós-Graduação. Seja com palavras de encorajamento, apoio prático ou simples gestos de gentileza, cada contribuição foi valorizada e fundamental para o meu sucesso.

“O universo conspira a favor de quem tem a coragem de ser guiado pelo coração”.

(Jessica Behrens)

## RESUMO

O panorama da energia elétrica no Brasil é influenciado por uma variedade de fatores complexos e relações não lineares, o que torna a previsão desafiadora. Com o aumento da demanda por energia e a crescente preocupação ambiental, é crucial buscar soluções baseadas em práticas de energia limpa e renovável, visando tornar o mercado de energia mais sustentável. Essas práticas visam reduzir o desperdício e otimizar a eficiência dos processos envolvidos na operação das tecnologias de distribuição e geração de energia elétrica. Uma abordagem promissora para viabilizar a energia sustentável é a aplicação de técnicas de previsão para diversas variáveis do mercado energético. Esta pesquisa propõe uma análise empírica do uso de regressores para realizar previsões nas bases de dados do Preço de Liquidação das Diferenças (PLD) do mercado brasileiro e da velocidade do vento em aerogeradores do Nordeste do Brasil. Busca-se contribuir com informações significativas sobre as técnicas de aprendizagem de máquina, que podem ser empregadas como ferramentas eficazes para a previsão de séries temporais no setor elétrico. Os resultados obtidos podem incentivar a implantação dessas técnicas para extrair conhecimento sobre o comportamento do sistema de energia brasileiro. Isso é particularmente relevante, dado que o preço da energia frequentemente exhibe sazonalidade, alta volatilidade e picos, e a geração de energia eólica é amplamente influenciada pelas condições climáticas. Para modelar a previsão dessas duas séries temporais, utilizamos o banco de dados sobre o PLD, focando especialmente no preço médio da energia do Sistema Nacional Brasileiro. As variáveis mais relevantes estão relacionadas às condições hidrológicas, carga elétrica e preço dos combustíveis das unidades térmicas. Para a coleta das variáveis relacionadas à energia eólica, foram considerados dois locais distintos na região nordeste do Brasil: Macau e Petrolina. Para o estudo de previsão, utilizamos uma Rede Neural Perceptron Multicamadas (MLP), uma Long Short Term Memory (LSTM), o Auto-Regressive Integrado de Média Móveis (ARIMA) e a Máquina de Suporte de Vetores (SVM) para determinar as linhas bases nos resultados da previsão. Para aprimorar os resultados destes regressores, utilizamos duas abordagens distintas de previsão. Uma das abordagens consistiu na combinação das técnicas de Redes Neurais Artificiais Profundas, baseada na Meta-Heurística do Algoritmo Genético Canônico (AG), para ajustar os hiperparâmetros dos regressores MLP e LSTM. Já a segunda estratégia focou em comitês de máquinas, os quais incluíam MLP, Árvore de Decisão, Regressão Linear e SVM em um comitê, e MLP, LSTM, SVM e ARIMA em outro. Essas abordagens consideraram dois tipos de votação, voting average (VO) e voting weighted average (VOWA), para avaliar o impacto no desempenho do comitê de máquinas.

***Palavras-chave:*** Preço de Liquidação das Diferenças. Velocidade do Vento. Sistema Elétrico Brasileiro. Aprendizagem de Máquina. Comitê de Máquinas. Previsão

## ABSTRACT

*The overview of electric energy in Brazil is influenced by a variety of complex factors and nonlinear relationships, making forecasting challenging. With the increasing demand for energy and growing environmental concerns, it is crucial to seek solutions based on clean and renewable energy practices, aiming to make the energy market more sustainable. These practices aim to reduce waste and optimize the efficiency of processes involved in the operation of electricity distribution and generation technologies. A promising approach to enable sustainable energy is the application of forecasting techniques for various variables in the energy market. This thesis proposes an empirical analysis of the use of regressors to make predictions in the databases of the Price of Settlement Differences (PLD) in the Brazilian market and wind speed in wind turbines in Northeast Brazil, through principal component analysis. We aim to provide significant information about machine learning techniques that can be employed as effective tools for time series prediction in the electric sector. The results obtained may encourage the implementation of these techniques to extract knowledge about the behavior of the Brazilian energy system. This is particularly relevant, given that energy prices often exhibit seasonality, high volatility, and peaks, and wind power generation is widely influenced by weather conditions. To model the prediction of these two time series, we use the database on the Price of Settlement Differences (PLD), focusing especially on the average energy price of the Brazilian National System. The most relevant variables are related to hydrological conditions, electrical load, and fuel prices for thermal units. For collecting variables related to wind energy, two distinct locations in the Northeast region of Brazil were considered: Macau and Petrolina. For the prediction study, we use a Multilayer Perceptron Neural Network (MLP), a Long Short Term Memory (LSTM), Auto-Regressive Integrated Moving Average (ARIMA), and Support Vector Machine (SVM) to determine baseline results in prediction. To enhance the results of these regressors, we employ two different prediction approaches. One approach involves combining deep artificial neural network techniques based on the Canonical Genetic Algorithm (AG) meta-heuristic to adjust the hyperparameters of MLP and LSTM regressors. The second strategy focuses on machine committees, which include MLP, decision tree, linear regression, and SVM in one committee, and MLP, LSTM, SVM, and ARIMA in another. These approaches consider two types of voting, voting average (VO) and voting weighted average (VOWA), to assess the impact on the performance of the machine committee.*

**Keywords:** Price of Liquidation Differences. Wind Speed. Brazilian Electric System. Machine Learning. Machine Committee. Forecast

## LISTA DE ILUSTRAÇÃO

Figura 1- Matriz de Produção de Energia Elétrica no Brasil.....	15
Figura 2- Exemplo de uma Arquitetura <i>Deep Learning</i> .....	19
Figura 3- Exemplo de um Ensemble .....	19
Figura 4- A Hierarquia do Aprendizado.....	31
Figura 5- Topologia Genérica do Perceptron Multicamada .....	34
Figura 6- Rede Neural Recorrente.....	36
Figura 7- <i>Long Short Term Memory</i> (LSTM) .....	37
Figura 8- Fluxograma de um AG Clássico.....	39
Figura 9- Algoritmo Básico do Método de Seleção por Torneio. ....	41
Figura 10- Possíveis Hiperplanos de Separação para uma Data Set Bidimensional.....	42
Figura 11- SVM não Linear.....	43
Figura 12- Transformação não Linear entre Espaço Inicial (à esquerda) e Espaço Depois da Transformação (à direita) .....	43
Figura 13- Comportamento do SVM utilizando diferentes valores para C.....	44
Figura 14- Comportamento do SVM utilizando diferentes valores para Gamma.....	45
Figura 15- Escopo Geral.....	49
Figura 16- Modelo Preditor de Série Temporal Combinando Algoritmo Genético e Rede Neural Profunda LSTM (exemplo de predição do preço de eletricidade PLD).....	50
Figura 17- Fluxo de um Algoritmo Genético .....	52
Figura 18- Indivíduo da População .....	53
Figura 19- Mecanismo de Seleção.....	55
Figura 20- Cruzamento com Dois Pontos de Corte.....	56
Figura 21- Mutação em dois Genes de um Indivíduo .....	56

Figura 22- Proposta de Comitê de Máquinas: “Ensemble 01” .....	59
Figura 23- Proposta de Comitê de Máquinas: “Ensemble 02” .....	60
Figura 24- Matriz de <i>Scatter</i> , plot para 6 Variáveis Dados PLD-Região Norte.....	65
Figura 25- Matriz de <i>Scatter</i> , plot para 5 Variáveis Dados Velocidade do Vento- Macau.....	66
Figura 26- Separação de Treino e Teste da Base de Dados do PLD – Norte .....	69
Figura 27- Separação de Treino e Teste da Base de Dados da Velocidade do Vento – Macau .....	69
Figura 28.a- PLD: Convergência da Rede LSTM para a Região Sudeste.....	76
Figura 28.b- PLD: Previsão da Rede LSTM para Região Sudeste.....	77
Figura 29- Previsão do ARIMA com 12 Passos à Frente - Velocidade do Vento .....	82
Figura 30- Previsão do ARIMA com 12 Passos à Frente – PLD .....	82
Figura 31 – Atrasos de um e dois Períodos da Variável “Velocidade” – Velocidade do Vento.....	84
Figura 32- Comparação dos Resultados da Previsão Direta para o Preço Médio do PLD com os Métodos <i>Bayes Search</i> e <i>Random Search</i> .....	85
Figura 33- Comparação dos Resultados da Previsão para Velocidade do Vento com os Métodos <i>Bayes Search</i> e <i>Random Search</i> .....	86
Figura 34- Configuração ótimas do AG + LSTM para a Previsão do PLD da Região Norte .	87
Figura 35- Gráfico do MSE para a Predição do PLD.....	89
Figura 36- Configurações ótimas do AG + LSTM para a Previsão da Velocidade do Vento - Parque Eólico da Cidade de Macau.....	90
Figura 37- Gráfico do MSE para Predição da Velocidade do Vento .....	91
Figura 38- Gráfico do MSE para Predição dos Melhores Resultados do PLD.....	92
Figura 39- Gráfico do MSE para Predição dos Melhores Resultados da Velocidade do Vento .....	92
Figura 40- Gráfico do MSE para Prever os Melhores Resultados de PLD .....	98

Figura 41- Gráfico do MSE para Prever os Melhores Resultados de Velocidade do Vento... 99

## LISTA DE TABELAS

Tabela 1- Resultado da Pesquisa Bibliográfica .....	22
Tabela 2- Parâmetros para Simulação com AG.....	57
Tabela 3- Hiperparâmetros dos Membros dos Ensembles .....	61
Tabela 4- Medições de Posição e Dispersão na Região Norte .....	67
Tabela 5- Medições de Posição e Dispersão de Macau.....	67
Tabela 6- Aplicação de Hiperparâmetros .....	71
Tabela 7- Configurações das RNA.....	75
Tabela 8- Resultados para o PLD - Medida de Erro MSE .....	76
Tabela 9- Resultados para Velocidade do Vento - Medida de Erro MSE.....	78
Tabela 10- Resultados do Teste Dickey-Fuller .....	78
Tabela 11- Estimação dos Parâmetros.....	80
Tabela 12- Erros de Previsão para Velocidade do Vento - Macau - ARIMA (23,0,25) .....	81
Tabela 13- Erros de Previsão para PLD - Região Norte - ARIMA (16,0,24) .....	81
Tabela 14- Resultados da Otimização de Hiperparâmetros PLD – Região Norte - SVM .....	84
Tabela 15- Resultados das Previsões <i>Random Search</i> – PLD - Região Norte - SVM.....	85
Tabela 16- Resultados da Otimização de Hiperparâmetros Velocidade do Vento – Macau - SVM .....	85
Tabela 17- Resultados das Previsões <i>Random Search</i> - Velocidade do Vento – Macau - SVM .....	86
Tabela 18- Resultados da Combinação AG + Rede Neural para o PLD.....	87
Tabela 19- MSE para até 12 Passos à Frente de Predições - Região Norte .....	88
Tabela 20- Resultado da Combinação AG + Rede Neural para Velocidade do Vento.....	89

Tabela 21- Resultado da Combinação AG + Rede Neural para Velocidade do Vento – Macau .....	90
Tabela 22- Pesos dos Membros Ensemble - VOWA .....	94
Tabela 23- Resultados Comparativos dos Membros “Ensemble 01” com VOA e “Ensemble 01” com VOWA – Região Norte .....	94
Tabela 24- Resultados Comparativos dos Membros “Ensemble 02” com VOA e “Ensemble 02” com VOWA – Região Norte .....	95
Tabela 25- Resultados do “Ensemble 01” com a substituição do SVM pelo ARIMA – Região Norte .....	95
Tabela 26- Resultados do “Ensemble 02” com a substituição do SVM pelo ARIMA – Região Norte .....	96
Tabela 27- Resultados Comparativos dos Membros “Ensemble 01” com VOA e “Ensemble 01” com VOWA - Macau .....	96
Tabela 28- Resultados Comparativos dos Membros “Ensemble 02” com VOA e “Ensemble 02” com VOWA - Macau .....	96
Tabela 29- Resultados do “Ensemble 01” com a Substituição do SVM pelo ARIMA - Macau .....	97
Tabela 30- Resultados do “Ensemble 02” com a Substituição do SVM pela Regressão Linear - Macau .....	97
Tabela 31- Comparação do Modelo de Previsão Proposto com outros Modelos Publicados para Velocidade do Vento.....	100
Tabela 32- Comparação do Modelo de Previsão Proposto com outros Modelos Publicados para o PLD.....	100

## Sumário

<b>1 INTRODUÇÃO</b> .....	13
<b>1.1 Motivação</b> .....	13
<b>1.2 Objetivos</b> .....	20
1.2.1 Objetivo Geral .....	20
1.2.2 Objetivos Específicos .....	20
<b>1.3 Estrutura da Tese</b> .....	211
<b>2 ESTUDO RELACIONADO E CONTRIBUIÇÕES</b> .....	222
<b>2.1 Introdução</b> .....	222
<b>2.2 Contribuição</b> .....	28
<b>2.3 Considerações Finais</b> .....	30
<b>3 FUNDAMENTAÇÃO TEÓRICA</b> .....	30
<b>3.1 Aprendizagem de Máquina</b> .....	30
<b>3.2 Séries Temporais</b> .....	31
3.2.1 Métricas de Avaliação para Séries Temporais .....	32
3.2.1.1 Tipos de Métricas .....	32
<b>3.3 Algoritmos de Aprendizagem de Máquina</b> .....	33
3.3.1 Perceptron de Múltipla Camadas.....	33
3.3.2 <i>Long Short Term Memory</i> .....	35
3.3.3 Algoritmo Genético .....	39
3.3.3.1 Estrutura Padrão de um AG .....	39
3.3.3.2 Seleção e Elitismo .....	40
3.3.4 Máquinas de Vetor de Suporte .....	42
3.3.4.1 Máquinas de Vetor de Suporte Não Linear .....	42
3.3.4.2 Hiperparâmetros do SVM .....	44
3.3.5 Modelo Auto-Regressivo Integrado de Médias Móveis .....	45
<b>4 MODELO PREDITOR PARA USO NO SISTEMA ELÉTRICO BRASILEIRO</b> .....	49
<b>4.1 Redes Neurais Artificiais Profundas com Hiperparâmetros Otimizados pelo Algoritmo Genético</b> .....	499
4.1.1 Fase I: Iteração do Algoritmo Genético .....	51

___ 4.1.1.1 Codificação do Problema .....	52
___ 4.1.1.2 População .....	54
___ 4.1.1.3 Avaliação da População .....	54
___ 4.1.1.4 Seleção .....	54
___ 4.1.1.5 Elitismo .....	555
___ 4.1.1.6 Cruzamento .....	555
___ 4.1.1.7 Mutação.....	56
4.1.2 Fase II: Cálculo da Função Fitness (Erro de Predição das Redes Neurais Artificiais Profundas) .....	577
<b>4.2 Comitê de Máquinas.....</b>	<b>58</b>
<b>4.3 Aspectos Gerais e Relevantes dos Modelos Propostos</b> .....732	
4.3.1 Base de Dados.....	732
4.3.2 Pré-Processamento.....	738
4.3.3 Técnicas e Métodos para Otimizar Parâmetros.....	70
4.3.4 Métricas de avaliação.....	71
4.3.5 Custo de Treinamento.....	72
<b>4.4 Considerações Finais.....</b>	<b>73</b>
<b>5 RESULTADOS E DISCUSSÕES .....</b>	<b>74</b>
<b>5.1 Resultados Baselines.....</b>	<b>74</b>
<b>5.1.1 RNA (MLP e LSTM).....</b>	<b>75</b>
___ 5.1.1.1 Resultados <i>Baselines</i> para o PLD.....	75
___ 5.1.1.2 Resultados <i>Baselines</i> para Velocidade do Vento .....	78
<b>5.1.2 ARIMA .....</b>	<b>78</b>
<b>5.1.3 SVM .....</b>	<b>82</b>
<b>5.2 Resultados das combinações AG + LSTM e AG + MLP .....</b>	<b>87</b>
5.2.1 Resultados da Combinação para o PLD .....	87
5.2.2 Resultados da Combinação para Velocidade do Vento.....	89
<b>5.3 Resultados dos Comitê de Máquinas .....</b>	<b>93</b>
5.3.1 Resultados Comitê de Máquinas para PLD – Região Norte .....	94

5.3.2 Resultados Comitê de Máquinas para Velocidade do Vento – Região Macau ...	96
<b>5.4 Comparação dos Resultados das Combinações AG + LSTM e AG + MLP com Comitê de Máquinas.....</b>	<b>97</b>
<b>5.5 Considerações Finais .....</b>	<b>100</b>
<b>6 Conclusão.....</b>	<b>101</b>
6.1 Implicações Práticas.....	102
6.2 Pontos Fortes.....	103
6.3 Fraquezas.....	103
6.4 Trabalhos Futuros.....	104
<b>REFERÊNCIAS.....</b>	<b>106</b>
<b>APÊNDICE A: Resultados das Métricas de Avaliação para 1 Passo à Frente dos Modelos.....</b>	<b>115</b>

# 1 INTRODUÇÃO

## 1.1 Motivação

Manter a integridade ambiental não apenas atende a uma exigência do Ministério do Meio Ambiente (MMA) brasileiro, mas também é uma preocupação compartilhada por todos os indivíduos. Isso se deve ao fato de que, juntamente com os avanços, a modernização e a melhoria da qualidade de vida da sociedade, surgem também problemas ambientais significativos. A fabricação de equipamentos, o descarte inadequado de periféricos obsoletos, o alto consumo de energia e o aumento do efeito estufa são exemplos claros desse impacto negativo.

No horizonte do futuro terrestre, há uma sombra crescente que ameaça a sustentabilidade de toda forma de vida: as emissões de gases de efeito estufa. Este não é apenas um dilema ambiental, mas sim um chamado urgente à ação para garantir a sobrevivência de nosso planeta e suas criaturas.

Os dados revelam uma realidade preocupante: as emissões de gases de efeito estufa estão alcançando níveis alarmantes. Esses gases, provenientes de atividades humanas como a queima de combustíveis fósseis e desmatamento, estão exacerbando o aquecimento global. Como resultado, testemunhamos mudanças climáticas drásticas que comprometem não apenas os ecossistemas, mas também nossos próprios modos de vida (SEEG, 2021; MCTI, 2022; CNN BRASIL, 2023).

O aquecimento da Terra não é apenas uma questão de desconforto climático; é uma ameaça tangível à vida como a conhecemos. Espécies enfrentam desafios crescentes para sua sobrevivência, com habitats se tornando inóspitos e condições de reprodução se deteriorando. Animais marinhos sofrem com a acidificação dos oceanos, enquanto criaturas terrestres lutam para se adaptar a novos padrões climáticos imprevisíveis (G1; VEJA, 2024).

A necessidade de agir é mais premente do que nunca. Se os países não reduzirem suas emissões de gases de efeito estufa, o futuro do planeta estará em sério risco. Alcançar os objetivos climáticos torna-se uma tarefa cada vez mais difícil, e as consequências de nossa inação são profundas e duradouras (DASGUPTA, 2024; OTOSAKA, 2023).

Olhando para os oceanos, vemos um lembrete contundente do que está em jogo. As temperaturas recordes do oceano do ano passado persistem em 2024, sinalizando uma escalada contínua do aquecimento global. As correntes marinhas estão mudando, os recifes de coral estão morrendo e os ecossistemas marinhos estão sob pressão sem precedentes (ONU, 2023).

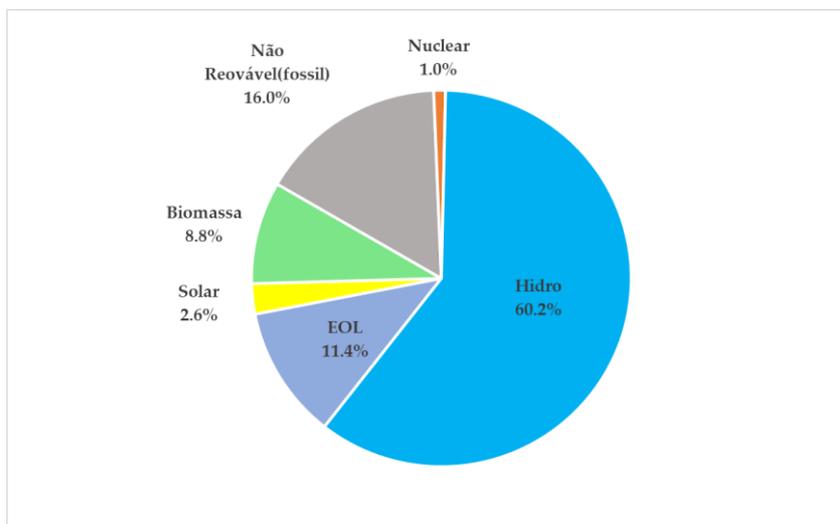
No entanto, há esperança no ar, mesmo que frágil. À medida que mais pessoas se conscientizam sobre os impactos das emissões de gases de efeito estufa, cresce também o impulso por mudanças significativas. A transição para fontes de energia renovável, a adoção de práticas agrícolas sustentáveis e o investimento em tecnologias de captura de carbono oferecem caminhos para mitigar os piores efeitos das mudanças climáticas.

Diante da realidade dos diversos impactos que nosso planeta está enfrentando, tais como o aumento do efeito estufa, as mudanças climáticas, a poluição do ar, a poluição da água, a desflorestação e outros, é imperativo adotar intervenções harmoniosas e sustentáveis. Através dessas medidas, é possível evitar ou reduzir desperdícios de recursos, tanto de insumos quanto de equipamentos, e mitigar a poluição gerada por fontes contaminadoras, como o carvão. Além disso, tais ações visam diminuir as emissões de gases de efeito estufa, combater o desmatamento e promover o uso consciente dos recursos naturais.

Nesse contexto, é crucial considerar não apenas a projeção do consumo, mas também a possibilidade de reduzir o consumo de energia elétrica e buscar a eficiência energética. Essa abordagem é essencial, especialmente quando se trata de fontes não renováveis, que são esgotadas durante o processo de geração de energia ao longo do tempo. Além dos benefícios financeiros advindos da redução do consumo de energia, é fundamental ressaltar os impactos positivos para o meio ambiente. A preservação ambiental é uma preocupação crescente, e a economia de energia desempenha um papel significativo nesse aspecto. Ao utilizar os recursos energéticos de maneira mais eficiente, é possível reduzir a demanda por fontes não renováveis, mitigar a emissão de gases de efeito estufa e minimizar a degradação ambiental associados à geração de energia.

De acordo com o Ministério de Minas e Energia (MME), a oferta interna de energia elétrica no Brasil é originária das seguintes fontes (ver Figura 1):

Figura 1 - Matriz de Produção de Energia Elétrica no Brasil



Fonte: Balanço Energético Nacional - BEN (2022, p. 15)

Conforme observado, a geração hidráulica representa a principal fonte de energia elétrica no Brasil. No entanto, as contribuições das outras fontes são significativas, especialmente durante os períodos de escassez de chuvas, quando há um alto consumo de energia.

Atualmente, a matriz energética do país é composta por diversas fontes de energia elétrica, incluindo hidroelétricas e termoelétricas, com a energia eólica emergindo como uma das formas mais promissoras de geração. Isso é especialmente verdadeiro na região Nordeste, onde existem 258 parques eólicos na Bahia, 225 no Rio Grande do Norte, 105 no Piauí e 98 no Ceará (INFOVENTO, 2023).

Diante desses fatos, entre as opções de energia renovável, a geração eólica é uma das fontes que mais crescem no mundo. Nos cálculos realizados no Mercado Livre de Energia para dimensionar o consumo e os gastos com insumos, um dos fatores mais importantes é o Preço de Liquidação de Diferenças (PLD).

Para estimar o PLD, são avaliadas as diferenças entre o volume de energia gerada ou contratada e o total consumido pelo mercado. Em outras palavras, o PLD harmoniza os custos entre a demanda e a oferta do setor energético do país.

A Câmara de Comercialização de Energia Elétrica (CCEE) é a instituição responsável pela gestão do mercado de energia elétrica no Brasil e utiliza modelos matemáticos para

calcular o PLD. Esse cálculo é uma métrica que baliza o valor da energia no Mercado de Curto Prazo (MCP). Atualmente, CCEE utiliza o modelo computacional conhecido como NEWAVE (*New Agent for Voltage and Energy*), desenvolvido pelo Operador Nacional do Sistema Elétrico (ONS), para estimar o PLD. O NEWAVE é um modelo de otimização energética de longo prazo que considera diversas variáveis e restrições do sistema elétrico, como a disponibilidade de recursos hídricos, a demanda por energia, as características dos diferentes tipos de usinas geradoras, entre outros. Com base nessas informações, o NEWAVE calcula o PLD, que é o preço utilizado nas transações de energia no mercado elétrico brasileiro, refletindo as condições de oferta e demanda de energia em todo o país.

Além disso, a CCEE também está implementando o DESSEM (Despacho Hidrotérmico de Curtíssimo Prazo), um modelo computacional mais avançado e detalhado que visa aprimorar a previsão e o despacho da geração de energia elétrica. O DESSEM considera aspectos operativos de curto prazo, como previsão de aflúências, custos marginais de operação das usinas, limitações de transmissão, entre outros fatores, para determinar a operação ótima do sistema elétrico em intervalos de tempo mais curtos, como horas ou dias. Essa tecnologia proporciona uma visão mais precisa e dinâmica das condições do sistema elétrico, contribuindo para uma gestão mais eficiente e segura do fornecimento de energia elétrica no Brasil (ANELL; MME; CCEE, 2023).

Hoje, à medida que nos aprofundamos no mundo da modelagem computacional, surgem abordagens inovadoras. O DESSEM, uma modelagem computacional avançada, é um exemplo notável disso. Mas o que realmente impulsiona sua eficácia são os algoritmos de aprendizado de máquina. Esses algoritmos permitem que o DESSEM aprenda com dados históricos e tome decisões cada vez mais precisas e ágeis. Eles podem identificar padrões sutis nos dados, prever tendências futuras e otimizar o uso dos recursos disponíveis. Essa capacidade de adaptação e aprimoramento contínuo são essenciais em um mercado tão dinâmico quanto o de energia.

Além disso, o aprendizado de máquina também desempenha um papel fundamental no desenvolvimento de estratégias de mitigação de riscos e na tomada de decisões baseadas em dados. Ele pode analisar uma ampla gama de variáveis, desde condições climáticas até flutuações de preços, para ajudar os operadores a antecipar e responder proativamente às mudanças no mercado.

Em resumo, o Aprendizado de Máquinas (AM) é uma abordagem que possibilita que as máquinas aprendam a executar tarefas específicas ao identificar padrões e correlações nos dados por meio de algoritmos. Entre as várias aplicações do AM, destaca-se a previsão de séries temporais como uma área em que tem sido altamente eficaz (LUDERMIR, 2021; VALENCIA-ARIAS *et al.*, 2024).

A previsão de séries temporais é uma tarefa importante em várias áreas, como finanças, economia, engenharia, saúde, entre outras. Ela envolve a análise de dados que mudam ao longo do tempo, com o objetivo de identificar tendências, sazonalidades e padrões que possam ser utilizados para fazer previsões futuras (KRAMAR e ALCHAKOV, 2023).

O AM oferece uma abordagem alternativa para a previsão de séries temporais, na qual os algoritmos são treinados em dados históricos para fazer previsões futuras. Existem várias técnicas de Aprendizado de Máquinas para a previsão de séries temporais, incluindo Redes Neurais Artificiais (RNA), Máquina de Vetor de Suporte (SVM), Árvores de Decisão, Florestas Aleatórias e Comitê de Máquinas - Ensemble (LIM e ZOHREN, 2021; FOROOTAN, 2022).

As RNA, em particular, têm se mostrado altamente eficazes na previsão de séries temporais. Elas possuem a capacidade de capturar padrões complexos e não lineares nos dados, sendo possível treiná-las em grandes conjuntos de dados para alcançar uma precisão muito alta nas previsões (CAO *et al.*, 2019; ALSUBAIE *et al.*, 2024).

O uso de Comitês de Máquinas (ensemble) na previsão de séries temporais é uma estratégia poderosa para aprimorar a precisão e a robustez das previsões. Ao combinar múltiplos modelos, é possível capturar uma variedade de padrões nos dados (CARNEIRO *et al.*, 2022)

Em resumo, AM é uma técnica promissora para a previsão de séries temporais, oferecendo uma abordagem alternativa aos métodos tradicionais baseados em modelos estatísticos que tentam explicar a dinâmica da série temporal. No entanto, esses modelos podem ser limitados em sua capacidade de identificar padrões e outros.

Na oportunidade, as técnicas de AM têm sido aplicadas com sucesso em uma ampla variedade de campos. No entanto, para obter um bom desempenho desses modelos, é essencial ajustar os hiperparâmetros corretamente.

Os hiperparâmetros são parâmetros que não são aprendidos pelo modelo, mas que precisam ser ajustados pelo programador. Na área de RNA, alguns exemplos comuns incluem

o tamanho do batch, o número de épocas, a taxa de aprendizado, o número de camadas ocultas e o número de neurônios em cada camada oculta. Esses hiperparâmetros desempenham um papel importante no desempenho do modelo, afetando sua precisão e tempo de treinamento (ELSKEN *et al.*, 2019).

Uma sintonia adequada dos hiperparâmetros é crucial para encontrar a combinação ideal de valores que resultem no melhor desempenho possível do modelo. Por outro lado, uma configuração inadequada dos hiperparâmetros pode levar a um modelo com desempenho insatisfatório e uma experiência negativa para o usuário. Um exemplo disso é a taxa de aprendizado. Se essa taxa for muito alta, o modelo pode não convergir durante o treinamento. Isso significa que o algoritmo de otimização não consegue encontrar os melhores valores para os pesos da rede, resultando em um processo de treinamento instável ou até mesmo sem progresso. Por outro lado, se a taxa de aprendizado for muito baixa, o modelo pode levar muito tempo para convergir, prolongando excessivamente o tempo de treinamento

Existem várias técnicas de sintonia de hiperparâmetros disponíveis, como busca em grade, busca aleatória, busca bayesiana e otimização por enxame de partículas. Essas técnicas são usadas para encontrar a combinação ideal de hiperparâmetros que leva ao melhor desempenho do modelo (HERNANDEZ *et al.*, 2022).

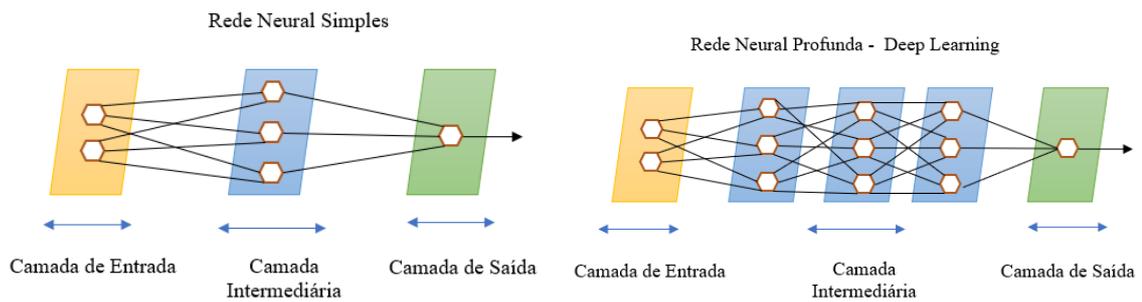
A sintonia adequada dos hiperparâmetros é fundamental para obter um bom desempenho do modelo de AM. O ajuste correto dos hiperparâmetros pode melhorar significativamente a precisão do modelo e reduzir o tempo de treinamento.

Por este motivo, grandes esforços têm sido empreendidos para elaboração de técnicas que permitam o ajuste fino dos parâmetros do modelo.

Além disso, algumas citações rotineiramente em revistas como *Science*, *Nature*, *Nature Methods* e *Forbes* apontam que a utilização de técnicas de aprendizagem profunda (*Deep Learning*), para modelar e controlar processos complexos é uma das metodologias mais utilizadas atualmente, conseqüentemente, o interesse em aprendizagem profunda tem sido intenso, com cobertura constante no âmbito acadêmico (DATA SCIENCE ACADEMY, 2018; HASSANIEN, 2020). O que provavelmente tem relação com suas potencialidades e possibilidades de aplicação em problemas do mundo real, que se estendem às diversas áreas de conhecimento, tais como engenharias, matemática, economia, finanças, estatística e física (CARVALHO *et al.*, 2011). Entretanto, por serem altamente customizáveis, a escolha da

arquitetura de uma *Deep Learning* (ver Figura 2) para um dado problema acaba se tornando o principal alvo, visto que uma arquitetura de RNA influencia diretamente na resolução do problema.

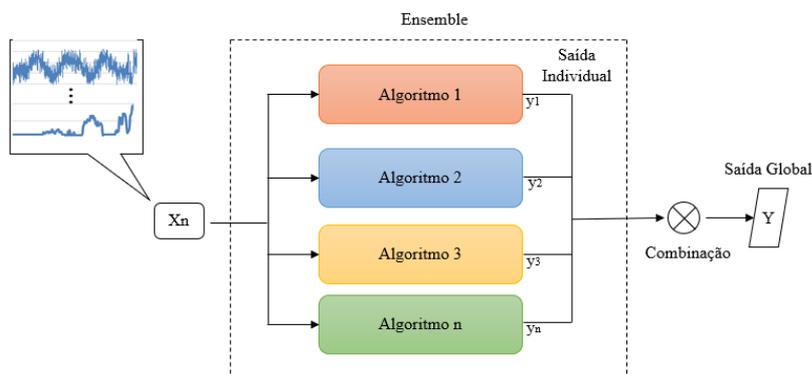
Figura 2 - Exemplo de uma Arquitetura *Deep Learning*



Fonte: Autor (2024)

Destaca-se também a ideia por trás dos ensembles que representam uma abordagem poderosa e cada vez mais popular no campo da AM. Em vez de confiar em um único modelo para fazer previsões ou classificações, os ensembles combinam as previsões de vários modelos individuais para obter uma decisão final. (SAGI, 2018; ZAO, 2024) (Ver Figura 3).

Figura 3 - Exemplo de um Ensemble



Fonte: Autor (2024)

Assim, surge o interesse na utilização de técnicas eficientes de AM combinadas com o emprego de meta-heurísticas para ajustar seus hiperparâmetros e, adicionalmente, considerar a diversidade e complementaridade dos modelos individuais dentro do ensemble.

Neste contexto, é apresentada uma análise comparativa entre o comitê de máquinas e a combinação de RNA e Algoritmo Genético (AG) na previsão de parâmetros elétricos, com foco no PLD e na velocidade do vento para energia eólica. Para alcançar os objetivos deste estudo, considerou-se as seguintes abordagens e métodos: modelo de árvore de decisão, regressão linear ARIMA e SVM, métodos de busca aleatória e de busca bayesiana para otimizar os hiperparâmetros do SVM e a função auto-ARIMA com o algoritmo de busca em grade para otimizar os parâmetros ARIMA. Em relação às abordagens que envolvem RNA, incluiu-se *Long Short Term Memory* (LSTM) e Perceptron Multicamadas (MLP) com hiperparâmetros otimizados por meio de AG. É importante notar que a aplicação dessas diversas classes de modelos no mesmo estudo comparativo, com o objetivo de prever o preço da eletricidade e a velocidade do vento no Brasil, não é encontrada na literatura existente. Embora existam estudos que utilizem modelos de previsão nos dados analisados, este trabalho contribui significativamente para o campo de pesquisa, abrindo espaço para discussões e aplicações futuras.

## 1.2 Objetivos

### 1.2.1 Objetivo Geral

Esta tese de doutorado tem como objetivo realizar uma análise empírica do uso de alguns regressores para prever séries temporais do sistema de energia elétrico brasileiro.

### 1.2.2 Objetivos Específicos

Aplicar seis tipos de Preditores para duas Séries Temporais, a de PLD e a de Velocidade do Vento;

- a) Empregar a Meta-Heurística AG para sintonizar Hiperparâmetros dos regressores MLP e LSTM;
- b) Utilizar a diversidade do comitê de máquinas para aprimorar o desempenho dos preditores;
- c) Usar testes Paramétrico e Não Paramétrico para caracterizar estatisticamente o desempenho dos vários preditores.

### 1.3 Estrutura da Tese

Esta tese está dividida em seis capítulos, conforme descrito a seguir:

- a) Capítulo 1: Apresenta a motivação, contribuição e relevância do tema abordado na pesquisa, além de definir os objetivos geral e específicos do estudo;
- b) Capítulo 2: Neste capítulo, é realizado um levantamento do estado da arte, abordando um panorama geral das principais pesquisas e trabalhos relevantes que se enquadram no contexto do problema em questão;
- c) Capítulo 3: são apresentados os elementos básicos do aprendizado de máquina, incluindo uma explanação sobre os algoritmos mais relevantes e suas características predominantes;
- d) Capítulo 4: é apresentada a modelagem para a predição do sistema de energia brasileiro, destacando as técnicas bioinspiradas utilizadas no processo;
- e) Capítulo 5: são apresentados os resultados obtidos na validação e comparação dos modelos preditivos desenvolvidos, fornecendo uma análise detalhada de sua eficácia e desempenho em relação aos objetivos propostos;
- f) Capítulo 6: Este capítulo apresenta as conclusões obtidas a partir da pesquisa realizada, oferecendo uma compreensão sobre os resultados alcançados e destacando suas implicações no contexto do problema estudado. Além disso, são sugeridos possíveis trabalhos futuros que possam expandir e aprimorar as contribuições desta tese.

## 2 ESTUDO RELACIONADO E CONTRIBUIÇÕES

### 2.1 Introdução

Atualmente, a busca por melhorias é inerente a diversos setores, incluindo indústria, transporte, educação e energia, especialmente no contexto do desenvolvimento sustentável. Diante disso, a demanda por inovações tecnológicas e pesquisas que visam harmonizar ideias e estratégias tem sido crescente. Vários estudos científicos têm sido conduzidos e publicados na literatura acadêmica (VICARI, 2021; GONTIJO, 2020), com o intuito de prever e fornecer descobertas relevantes no contexto da energia elétrica. Este capítulo tem como objetivo oferecer uma visão abrangente das principais pesquisas realizadas tanto no âmbito nacional quanto internacional, que buscam desenvolver modelos visando aprimorar e alcançar resultados satisfatórios na previsão da energia elétrica. A pesquisa foi realizada a partir de março de 2019, com o apoio do Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - <https://www.periodicos.capes.gov.br>, que é um dos maiores acervos científicos virtuais do país. Este portal oferece acesso a uma ampla variedade de conteúdos, incluindo aqueles produzidos nacionalmente e outros disponibilizados por editoras internacionais, contribuindo significativamente para o avanço da pesquisa e do ensino no Brasil (Ver Tabela 1).

Tabela 1 - Resultado da Pesquisa Bibliográfica

Pesquisa Bibliográfica	Qtd. Artigos
<i>Machine Learning Time Series</i>	68.557
<i>Machine learning to predict the price of Brazilian electricity</i>	136
<i>Recurrent neural networks for forecasting Brazilian electricity</i>	72
<i>Deep neural networks for forecasting the price of electricity in Brazil</i>	113
<i>Long Short Term Memory Brazilian Electricity System</i>	78
<i>Incorporating a combination of long short-term memory (LSTM) and genetic algorithm techniques for optimizing the Brazilian electricity system</i>	79
<i>hybrid model for the Brazilian electrical system</i>	3.316

<i>hybrid models to predict the price of Brazilian electricity</i>	208
<i>Assemble to predict for the Brazilian electrical system</i>	467
<i>Assemble to predict the price of Brazilian electricity</i>	-

Após execução da pesquisa citada acima foi observado as seguintes características. Quando se trata de previsão de preço da eletricidade, pesquisas anteriores destacaram a utilização de métodos de AM, com uma prevalência significativa de *Deep Learning* para previsões (HONG, 2020). Um exemplo disso é o trabalho de Ozcanli (2020), que abordou métodos e aplicações de *Deep Learning* para sistemas de energia elétrica em uma revisão sistemática. Da mesma forma, Abedinia (2020) observou uma tendência na literatura internacional de combinar modelos com o objetivo de obter previsões mais precisas. Uma dessas estratégias envolve a previsão de carga de curto prazo e precificação de eletricidade em sistemas elétricos ou de energia isolados, usando uma abordagem que combina redes neurais e um algoritmo de otimização baseado em gravidade (HEYDARI *et al.*, 2020).

Em linha com pesquisas anteriores, Chen (2020) abordou a combinação de AG e modelos baseados em AM na avaliação da vulnerabilidade ou propensão a deslizamentos de terra. Luo (2020) apresentou uma arquitetura de RNA utilizando um AG para estabelecer uma estrutura *feedforward* profunda para o consumo diário de eletricidade para o próximo dia e semana em um edifício de um campus no Reino Unido. Alencar (2018) sugeriu um modelo combinado integrando séries temporais e RNA para a geração de previsão de energia eólica. Além disso, a metodologia de *Deep Learning*, baseada em RNA recorrentes LSTM, combinada com o AG canônico, conforme mencionado por Junior (2023), foi aplicada para auxiliar na análise de sinais derivados de exames de eletroencefalograma (EEG). No entanto, é crucial observar que, até o momento, essa metodologia não foi explorada para previsão de eletricidade dentro do contexto brasileiro. Nesta tese a aplicação de AG difere da literatura na forma como o indivíduo/solução é construído e interpretado, bem como no procedimento aplicado à função de aptidão.

No que diz respeito ao uso do AG em combinação com RNA MLP e LSTM para a tarefa de previsão no setor elétrico, a literatura mostra que o AG pode ser usado para ajustar os hiperparâmetros desses dois modelos neurais ou pode ser usado para encontrar os valores dos

pesos e dos vieses de ambos os modelos. No artigo Li *et al.* (2020), o AG é utilizado para otimizar a taxa de aprendizado e o número de iterações de uma LSTM na tarefa de prever a carga de curto prazo com a LSTM-AG, aumentando a precisão da previsão em 63% quando comparada à previsão de uma LSTM padrão. Os autores não mencionam qual banco de dados é utilizado. Ainda sobre previsão de carga, a referência Zulfiqar (2022) utiliza o AG para ajustar os hiperparâmetros, como tamanho dos lotes, quantidade de neurônios, função de ativação e duração das épocas de uma LSTM com um extrator de características que define as entradas da rede neural. O modelo AG-LSTM-extractor apresenta uma medida de erro MAPE de 0.6710, que é menor do que o dos modelos AG-LSTM, LSTM-NSGA-II, LSTM-PSO e LSTM. Neste artigo, o banco de dados de Nova Inglaterra é utilizado. A referência Bendali *et al.* (2020) utiliza o AG para encontrar os melhores hiperparâmetros, o número de tamanhos de janela e o número de unidades (neurônios) de uma LSTM para prever a irradiância solar de curto prazo. O banco de dados utilizado é a irradiância global horizontal (GHI) na localização de Fes (33,3° N, -5,0° E, altitude = 579 m), separado por estações. Os resultados do modelo de previsão AG-LSTM são comparados com os modelos LSTM, Gate Recurrent Unit (GRU) e Recurrent Neural Network (RNN), bem como com as versões GRU + AG e RNN + AG. Os resultados indicam que os modelos aprimorados utilizando o AG superam suas respectivas metodologias padrão. No artigo Al Mamun *et al.* (2019), os autores empregam AG para otimizar o tamanho da janela de tempo para a tarefa de previsão de séries temporais, utilizadas como dados de entrada para a RNN. Eles também otimizam a quantidade de unidades LSTM nas camadas na seção oculta e o tamanho dos lotes para prever a carga do sistema elétrico australiano, com dados coletados fornecidos pelo operador do mercado de energia australiano. Os resultados da simulação destacam que a estratégia híbrida recomendada envolvendo AG e modelo de rede LSTM apresenta desempenho superior em comparação com outros modelos convencionais, como MLP, SVM e o modelo LSTM tradicional, produzindo os menores valores de RMSE e MAE de 87.304 e 118.007, respectivamente. O modelo sugerido demonstra uma redução de erro de 5,89% e 8,19% em comparação com a abordagem LSTM nos valores de RMSE e MAE. A referência Huang *et al.* (2023) utiliza o AG para selecionar o algoritmo de otimização (por exemplo, Adam, Rmsprop e Sgdm), uma taxa de aprendizado inicial  $\delta$ , e a quantidade de camadas ocultas, bem como a quantidade de neurônios dentro de cada camada oculta de uma LSTM para prever a velocidade do vento em uma cidade na China (instalações de energia eólica situadas na Mongólia Interior) e em outra cidade na Espanha (instalações de energia eólica situadas em Sotavento, Galícia). Um ensemble LSTM + AG também é utilizado para aprimorar

a previsão. O LSTM + AG proposto superou os cinco modelos de previsão restantes comparados (WNN - rede neural de ondaleta, BPNN - redes neurais de retropropagação, DBN - rede de crença profunda, ELM - máquina de aprendizado extremo e LSTM padrão), com o melhor valor de MAE de 0,53989, RMSE de 0,74116 e MAPE de 8,49605. O LSTM padrão foi o segundo melhor modelo, e o BPNN foi identificado como o modelo menos eficiente. No artigo Shahid *et al.* (2021), os pesquisadores utilizam o AG para ajustar as dimensões da janela de séries temporais e a quantidade de neurônios nas camadas LSTM, buscando prever a velocidade do vento em sete parques eólicos na Europa. O modelo LSTM + AG, em média, melhora as previsões de energia eólica em 6% a 30% em comparação com técnicas existentes, como SVR com kernel polinomial, SVR com kernel RBF, SVR com kernel linear e LSTM padrão. No estudo Ul Islam *et al.* (2014), é apresentada uma abordagem que utiliza uma MLP otimizada por um AG para estimar a carga do sistema elétrico australiano. O AG ajusta o número de camadas e neurônios na MLP. O modelo otimizado pelo AG possui uma camada oculta com dez neurônios (dado que a MLP tem seis variáveis de entrada), alcançando uma previsão com um Erro Quadrático Médio (MSE) de  $4.226 \times 10^{-7}$ . No artigo Izidio (2021), um AG é utilizado para configurar os parâmetros da MLP, incluindo a função de ativação, algoritmo de ajuste de peso e viés, e a quantidade de neurônios na camada oculta, com o objetivo de prever o consumo de energia para medidores inteligentes. O sistema implementado inicialmente realiza uma previsão linear do consumo, seguida por uma previsão não linear do resíduo desta previsão linear. Finalmente, uma MLP combina ambas as previsões. O framework híbrido sugerido é avaliado em comparação com metodologias anteriores, abrangendo métodos individuais, de conjunto e híbridos. Entre os modelos individuais, destacam-se SARIMA (Média Móvel Integrada Autoregressiva Sazonal), MLP, SVR, LSTM, LR (Regressão Linear) e árvores de classificação e regressão. Os resultados apresentados indicam que os modelos híbridos, incluindo o proposto no artigo, tendem a fornecer melhores resultados no que diz respeito ao consumo de energia. Esses resultados foram seguidos por técnicas de ensemble (como *bagging*) e modelos individuais. O sistema híbrido proposto foi avaliado utilizando informações coletadas por meio da infraestrutura de *smart grid* implementada em uma estrutura residencial. Este edifício de três andares está situado no distrito de Xindian, na cidade de Nova Taipé, Taiwan. É digno de nota que os estudos referenciados focam exclusivamente em um tipo de dado (carga/consumo de energia, irradiância solar ou velocidade do vento), e nenhum deles utiliza dados do Brasil. Em (HUANG *et al.*, 2023; ELSKEN *et al.*, 2019), o mesmo tipo de dado (velocidade do vento) é utilizado, embora proveniente de bancos de dados diferentes.

No artigo Shahid (2021), ao explorar o uso de ensembles, foi introduzida uma abordagem para previsão de velocidade do vento fundamentada na utilização da decomposição em modo empírico (EMD) em conjunto com o algoritmo genético e rede neural de retropropagação (AG-BPNN). O AG é empregado para calcular os valores de peso iniciais do BPNN. A decomposição em modo empírico coletivo (EEMD) representa uma evolução do EMD, demonstrando eficácia na resolução do problema de mistura de modos, decompondo e transformando os dados originais em sinais com frequências variáveis que apresentam características mais próximas da estacionariedade. Cada sinal derivado atua como entrada para o modelo AG-BPNN. A informação final de previsão da velocidade do vento é derivada pela combinação das previsões individuais de cada sinal decomposto. Uma análise realizada em uma instalação eólica na Mongólia Interior, na China, destaca que a abordagem híbrida sugerida supera significativamente a abordagem de previsão tradicional AG-BPNN, utilizando EMD em conjunto com o método da Rede Neural Wavelet com o método da Rede Neural Wavelet em termos de precisão. Um modelo chamado ISt-LSTM-informer, que utiliza um algoritmo de conjunto aprimorado chamado *stacking*, é proposto no artigo (WANG *et al.*, 2016) para prever a produção de energia fotovoltaica (PV). Utilizando dados históricos do centro solar DKA (<https://ulurusolar.com.au>), uma plataforma de acesso público para sistemas de energia fotovoltaica (PV) em Uluru, Austrália, o modelo emprega o algoritmo de regressão linear múltipla (LR) como um meta-modelo para incorporar meta-características, fornecendo previsões precisas de curto e médio prazo para a energia fotovoltaica (PVPF). Quando comparado a quatro outros métodos (LSTM padrão e informer, LSTM bidirecional, BiLSTM e Autoformer), o ISt-LSTM-Informer demonstra eficiência superior em todos os quatro indicadores de desempenho: MAE, MAPE, RMSE e R2. No artigo Cao *et al.* (2023), os autores propõem um método de previsão que combina vários modelos usando algoritmo genético-memória de curto prazo longo (AG-LSTM) complementar, decomposição em modo empírico de conjunto (CEEMD), otimização de enxame de partículas-máquina de vetor de suporte (PSO-SVM) e fusão de bases radiais-autoencoder (RBF-AE). Este método é empregado para previsão de carga no sistema energético integrado regional (RIES) (Estação de Tratamento de Água de Jieyuan, a Estação de Tratamento de Água de Lingzhuang, a Estação de Tratamento de Água de Xinkaihe e a Estação de Tratamento de Água de Jinbin - em Tianjin). A sequência de carga é decomposta em componentes de funções intrínsecas de modo de frequência (IMF) usando CEEMD. Os componentes da IMF são categorizados de acordo com sua taxa de pontos de cruzamento zero e entropia de amostragem (SE), levando à formação de três categorias

distintas: componentes de alta, média e baixa frequência. Após isso, o componente de carga de alta frequência, caracterizado por uma aleatoriedade pronunciada, é previsto por meio do uso de AG-LSTM. O componente de carga com frequência média, caracterizado por uma redução na aleatoriedade, é previsto por meio da utilização de RBF-AE. O componente de carga de baixa frequência, estável e cíclico, é previsto usando PSO-SVM. Os resultados previstos por esses três modelos são combinados para derivar o valor previsto final. Esses resultados são então comparados com os modelos padrão BPNN, SVM, RBF e LSTM. O conjunto de dados de carga multivariada utilizado neste estudo foi coletado no campus da Arizona State University, em Tempe, nos Estados Unidos (<http://cm.asu.edu/>, acessado em 5 de janeiro de 2024), enquanto os dados climáticos foram adquiridos no Laboratório Nacional de Energia Renovável (<http://maps.nrel.gov/nsrdb-viewer/>, acessado em 5 de janeiro de 2024). Os resultados validam a eficácia do modelo preditivo na manipulação de sequências não estacionárias de consumo de eletricidade e demonstram uma precisão máxima na previsão. No artigo Shi (2024), é proposta uma estrutura de conjunto de redes neurais (eNN), utilizando LSTM, SVM, BPNN e ELM para prever a geração de energia eólica e solar na China. Três algoritmos foram empregados para melhorar a precisão das previsões. Inicialmente, as séries temporais voláteis de energia eólica e solar passam por uma decomposição em sub-sequências suavizadas usando a decomposição de modo variacional (VMD) para mitigar os efeitos indesejáveis causados pela volatilidade. Posteriormente, modelos básicos (ELM, SVM, LSTM e BPNN), otimizados pelo algoritmo de exploração Sparrow com aprendizado baseado em elite (EOS-SA), são utilizados para prever a geração de energia eólica e solar com base nessas sub-sequências decompostas. Por fim, os resultados previstos pelo eNN são reconstruídos ponderando as previsões entre os quatro modelos. O peso de cada modelo é determinado pelo método dos mínimos quadrados. Todos os dados utilizados nesta investigação foram obtidos a partir dos dados estatísticos fornecidos pelo Instituto Nacional de Estatísticas da China (<https://data.stats.gov.cn>, acessado em 7 de janeiro de 2024). Os resultados indicaram que o eNN proposto alcançou as métricas mais baixas de RMSE, MAE, MSE e MAPE para a previsão tanto da energia eólica quanto solar, quando comparado aos modelos EOSSA-LSTM, EOSSA-ELM, EOSSA-SVM, EOSSA-BP, bem como aos modelos convencionais ELM, SVM, BPNN e LSTM. No estudo Wu (2021), é apresentada uma abordagem multiestágios inovadora para antecipar os preços da eletricidade em ambientes comerciais e residenciais no Brasil para vários horizontes temporais (um, dois e três meses à frente). A metodologia proposta combina uma etapa inicial de pré-processamento, empregando a decomposição em modo empírico de

conjunto (CEEMD) em conjunto com o algoritmo de otimização inspirado no comportamento do coiote (COA), com o objetivo de definir os hiperparâmetros do CEEMD. Essa abordagem aborda não linearidades nos dados de séries temporais, melhorando o desempenho do modelo. Posteriormente, quatro modelos de aprendizado de máquina, a saber, ELM, processo gaussiano, GBM e a máquina de vetores relevantes, são aplicados para treinar e fazer previsões dos componentes do CEEMD. Por fim, na fase conclusiva, os resultados obtidos das etapas anteriores são integrados diretamente para a formação de um aprendizado de ensemble diversificado, composto por componentes, com o objetivo de fornecer as previsões finais. Esses dados foram obtidos do site do Instituto de Pesquisa Econômica Aplicada (IPEA), acessível em <http://www.ipeadata.gov.br/Default.aspx>, acessado em 9 de janeiro de 2024. Através de comparações desenvolvidas (com ELM padrão, processo gaussiano, GBM e RVM, e com modelos de conjunto homogêneos em várias etapas COACEEMD-GBM, COA-CEEMD-GP, COA-CEEMD-RVM e COA-CEEMD-ELM), os resultados indicaram que a combinação de COA-CEEMD com uma abordagem de aprendizado de conjunto diversificado pode gerar previsões precisas. Esses artigos trabalham apenas com um tipo de dado (preço, carga, energia solar ou velocidade do vento), exceto em Ribeiro *et al.* (2020) onde os dados de velocidade do vento e energia solar são usados juntos. Apenas Wu (2021) utilizou dados brasileiros (preços de eletricidade comerciais e residenciais).

## 2.2 Contribuição

As contribuições desta tese são: (1) usar o AG para ajustar os hiperparâmetros do MLP e LSTM (um sistema híbrido mais complexo que os modelos MLP e LSTM) para prever diferentes tipos de dados (Preços e Velocidade do Vento). O processo de aprendizagem AG + DNNs é composto por duas fases distintas. Na Fase I, que envolve a interação do AG, busca-se criar ou encontrar o modelo neural com as condições mais favoráveis de capacidade de generalização para a predição das séries temporais estudadas. Já na Fase II, denominada como cálculo da Função *Fitness*, o AG fornece à DNN os valores decodificados do cromossomo binário para os hiperparâmetros da rede, permitindo assim a execução do processo de aprendizagem. O resultado do aprendizado, após algumas interações, é retornado para o AG por meio do valor da função *Fitness*. Esse processo, composto pela Fase I seguida da Fase II, é repetido até que o AG atinja a convergência, encontrando assim uma configuração ótima para os hiperparâmetros da LSTM; (2) construção do cromossomo para definir os hiperparâmetros e arquitetura da RNA. Esta contribuição é de suma importância na forma como o

indivíduo/solução é concebido e interpretado, assim como no procedimento aplicado à função de aptidão; (3) comparar os resultados dos sistemas híbridos AG+MLP e AG+LSTM com Ensembles heterogêneos. Um Ensemble funciona através da integração dos resultados de vários modelos de aprendizado, conhecidos como membros do comitê. Esses membros consistem em diferentes algoritmos de aprendizado que utilizam o método de votação classificada. Cada membro do comitê emite sua própria previsão ou classificação para uma entrada específica. Em seguida, uma decisão final é tomada com base na média simples e ponderada das previsões dos membros individuais. Essa estratégia geralmente resulta em previsões mais precisas e robustas do que as fornecidas por qualquer membro individual, pois aproveita a diversidade de abordagens e compensa possíveis fraquezas de cada modelo. (4) Aplicar esses modelos ao sistema elétrico brasileiro representa uma estratégia inovadora. A integração de regressores com meta-heurísticas possibilita que o algoritmo aproveite o melhor dos dois mundos, permitindo encontrar soluções mais precisas e confiáveis. As meta-heurísticas são algoritmos projetados para explorar soluções ótimas em problemas complexos, enquanto os regressores são modelos matemáticos destinados a estabelecer relações entre um conjunto de variáveis e a variável de interesse. Essa abordagem combinada tem o potencial de aprimorar significativamente a análise e o planejamento do sistema elétrico brasileiro, oferecendo percepções valiosas para otimização e tomada de decisões.

### **2.3 Considerações Finais**

Neste capítulo, foi apresentada uma seleção de aplicações descritas por diversos autores, destacando pesquisas significativas e relevantes na literatura relacionadas à previsão. Em particular, enfatizou-se o uso de uma rede neural recorrente especial chamada *Long Short-Term Memory* (LSTM) e comitês de máquinas. A principal contribuição deste trabalho é o desenvolvimento de um modelo de previsão e sua aplicação no contexto brasileiro, incorporando uma combinação de técnicas de IA que geralmente não são associadas entre si. No próximo capítulo, serão apresentados os conceitos básicos da metodologia proposta, juntamente com a descrição de algumas técnicas de AM.

### 3 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica deste trabalho, destaca pontos principais empregados na elaboração da pesquisa como conceitos e tipos de aprendizagem de máquina, os algoritmos de aprendizagem de máquina especificamente MLP, LSTM, SVM, ARIMA e AG, características de uma série temporal, métricas de avaliação e outros.

#### 3.1 Aprendizagem de Máquina

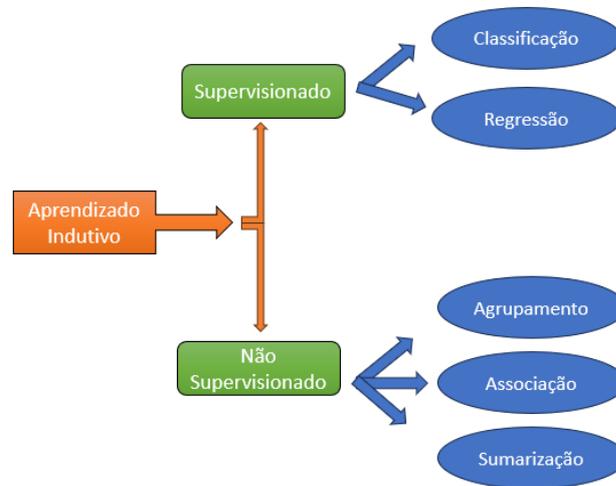
Aprendizagem de Máquina (AM) é um tema em constante ascensão e acelera o crescimento de soluções para problemas reais e não-lineares nos últimos anos, por sua capacidade de poder computar qualquer função. Por conta disso, esse campo vem sendo explorado e aplicado em áreas que necessitam de otimização como setores médicos, reconhecimento de imagens, previsão da energia elétrica, dentre outras (Arif *et al.*, 2020)

Segundo Oliveira (2018), AM é um campo da inteligência artificial que realiza tarefas específicas através da capacidade de aprender ou “treinar” a partir de quantidades de dados disponibilizado para o algoritmo, permitindo que ele se ajuste e aprimore o seu desempenho com as iterações do aprendizado.

Em geral, AM pode ser classificada em dois tipos: supervisionadas e não supervisionadas. A supervisionada de forma geral é utilizada quando as classes alvos da construção de um modelo são conhecidas ou fornecidas para prever. A técnica não supervisionada está relacionada a situações, com o objetivo de desenvolver um modelo, em que os rótulos alvos não são conhecidos e o conjunto de dados sofre agrupamentos com características similares (CONTE, 2015).

Segundo Carvalho *et al.* (2011) várias formas de aprendizado podem ser utilizadas nas arquiteturas RNA, entre elas os “aprendizados de regressão” que ocorre quando o resultado é numérico. Para exemplificar, imagine o caso onde se deseja calcular o preço de uma piscina tendo como base características como profundidade, iluminação e outros. Com base em registros de piscinas similares, o algoritmo aprende a estimar novos projetos de construção. Na Figura 4, é mostrado o fluxo do aprendizado indutivo, base da AM.

Figura 4 - A Hierarquia do Aprendizado



Fonte: Autor (2023)

A essência deste trabalho é explorar o aprendizado supervisionado, com destaque nos algoritmos de regressão. Esses algoritmos são particularmente úteis quando se trata de lidar com problemas em que há uma relação não-linear entre as entradas e as saídas, sendo a variável a prever um atributo numérico (CARVALHO *et al.*, 2011).

### 3.2 Séries Temporais

Uma série temporal é uma sequência de elementos observada em um período de tempo. Normalmente, este conjunto de observações é realizado com o objetivo de explorar ou compreender o comportamento de uma variável alvo, isto é, verificar alguma tendência de crescimento, tomar decisões a partir da compreensão geradas etc. (BROWNLEE, 2020). Um tipo de estudo importante em séries temporais é a previsão de valores futuros (HAHN, 2023).

As series temporais estão presentes em diversas áreas especificadamente na economia, climatologia, marketing e outras. Apresentam propriedades importantes para análise de dados como tendência da série, sazonalidade, monitoramento de ciclos, erro aleatório e estacionariedade (CHEN, 2020).

Uma série temporal pode ser representada por um modelo matemático que possua premissas de apresentar a estacionariedade, exemplo ARIMA (GOYAL, 2020), ou AM com uma RNA (LIU, 2022).

### 3.2.1 Métricas de Avaliação para Séries Temporais

Avaliar um conjunto de dados de um determinado evento, de maneira sequencial, durante um período é fundamental na identificação de padrões e, assim, torna-se fundamental na criação de modelos de previsão que têm como alimentação uma série temporal. Além de verificar se o modelo preditivo desenvolvido está ajustado à série temporal é preciso medir o desempenho do modelo.

#### 3.2.1.1 Tipos de Métricas

- Erro Médio Absoluto – MAE

O erro médio absoluto, do inglês *Mean Absolute Error* – MAE, é calculado a partir da média dos erros absolutos. Cada erro, pode ser entendido como a diferença entre  $Y$  e  $\hat{Y}$  e assim temos (ver Equação (1)):

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad 1$$

O MAE é frequentemente utilizado em séries temporais por medir apenas a distância do valor real para o predito e é menos influenciado por valores extremos (*outliers*)

- Erro Quadrático Médio - MSE

O erro quadrático médio – MSE, do inglês *Mean Squared Error*, é muito utilizado para verificar a acurácia de modelos, além de adicionar um maior peso aos erros significativos, isto é, durante o processo de cálculo, cada erro é elevado ao quadrado individualmente e após é calculado a média desses erros quadráticos, isto é (ver Equação (2)),

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad 2$$

Essa métrica é muito sensível a *outliers* (valores discrepantes) a cargo do expoente ao quadrado que o erro assume.

- Raiz Quadrada do Erro Médio - RMSE

A raiz quadrada do erro médio, do inglês *Root Mean Squared Error*, é a medida da diferença quadrática entre cada ponto previsto e o ponto real. Geralmente é um valor grande

porque elevamos ao quadrado cada diferença para remover os valores negativos antes de calcular a média (ver Equação (3)).

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad 3$$

O RMSE apresenta valores do erro nas mesmas dimensões da variável analisada.

### 3.3 Algoritmos de Aprendizagem de Máquina

Quando se fala de Aprendizagem de Máquina ou *Machine Learning* no inglês, existem diversas técnicas onde se consegue aprender através de treinamento, inspiradas no funcionamento do cérebro, e prever situações com a experiência adquirida, inspiradas pela biologia evolutiva etc. Entretanto cada uma das técnicas tem seus pontos positivos e negativos para solucionar problemas em um campo do conhecimento humano. Nesta Seção são apresentados alguns dos algoritmos e os principais conceitos necessários à compreensão do trabalho desenvolvido.

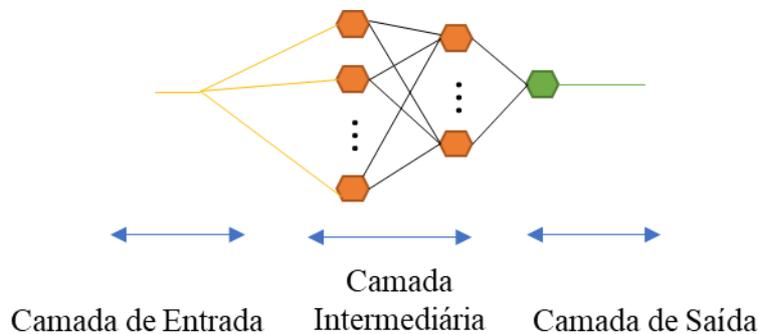
#### 3.3.1 Perceptron de Múltipla Camadas

É importante destacar que as redes Perceptron são uma extensão do trabalho de McCulloch-Pitts proposto por Rosenblatt (1962), sua arquitetura consiste em uma camada de entrada e uma de saída e este é considerado o primeiro modelo de redes neurais capaz de aprender somente sobre problemas lineares, isto é, convergir para um conjunto de parâmetros que determine uma reta que separe em diferentes classes as entradas apresentadas.

O Perceptron de Múltiplas Camadas é constituído de uma ou mais camadas intermediárias, permitem aprender tarefas mais complexas (não-lineares) e de acordo com (DATA SCIENCE ACADEMY, 2018; ARAÚJO, 2022) é o exemplo simples da aprendizagem profunda (*Deep Learning*) aplicado na esfera como reconhecimento de séries temporais e previsão de padrões.

As *Perceptron Multicamadas* (MLPs), representação visual na Figura 5, são redes de grande relevância para o cenário de AM, pois são a base para diversas aplicações hoje existentes.

Figura 5 - Topologia Genérica do Perceptron Multicamadas



Fonte: Autor (2024)

As redes neurais do tipo MLP são formadas de três partes: a parte 1 é conhecida como camada de entrada, a qual é composta por um conjunto de unidades sensoriais; a parte 2 consiste em uma ou mais camadas intermediárias com um número variado de neurônios; e a terceira parte também é constituída por neurônios e denominada de camada de saída.

Segundo Haykin (2008), nesse tipo de rede neural artificial, o sinal de entrada avança de camada em camada. A saída gerada pela rede é comparada a um alvo chamado saída desejada. A diferença entre a saída obtida da rede e a saída desejada é denominado erro. O sinal do erro é retro propagado camada por camada, a partir da camada de saída para atualizar os pesos sinápticos. Por conseguinte é definir alguns dos principais hiperparâmetros da rede MLP, mas existem outros que podem ser ajustados dependendo do problema em questão:

- Número de camadas: define quantas camadas de neurônios a rede terá. Geralmente, redes MLP possuem pelo menos uma camada oculta, além da camada de entrada e da camada de saída;
- Número de neurônios por camada: determina quantos neurônios serão incluídos em cada camada da rede. Esse número pode variar dependendo da complexidade do problema e da quantidade de dados disponíveis para treinamento;
- Taxa de aprendizado (*learning rate*): define o tamanho do passo que a rede dará durante o treinamento para ajustar os pesos sinápticos. Uma taxa de aprendizado alta pode fazer com que a rede oscile e nunca chegue a convergir, enquanto uma taxa de aprendizado baixa pode tornar o processo de treinamento lento demais;

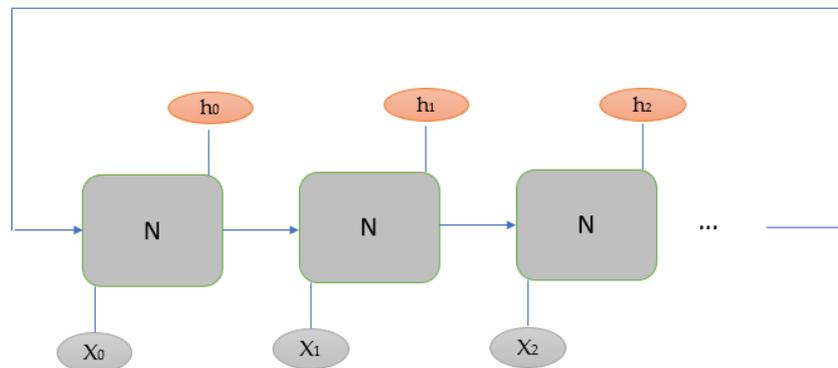
- Número de épocas: determina quantas vezes a rede percorrerá todo o conjunto de dados de treinamento durante o treinamento. Esse valor também pode variar dependendo da complexidade do problema e da quantidade de dados disponíveis;
- Função de ativação: define a forma como a saída de cada neurônio será transformada para ser passada para a próxima camada. Algumas das funções de ativação mais comuns em redes MLP incluem a *sigmóide*, a tangente hiperbólica e a *ReLU* (*Rectified Linear Unit*);
- Regularização: inclui técnicas como a L1 e L2 *regularization*, que ajudam a evitar *overfitting* durante o treinamento, ou seja, a rede ficar muito ajustada aos dados de treinamento e não conseguir generalizar bem para novos dados de teste.

### 3.3.2 Long Short Term Memory

Nas redes neurais de camadas múltiplas o aprendizado está relacionado com os ajustes de parâmetros para otimizar o erro. As mesmas são classificadas de redes neurais *feedforward* quando os dados sempre são propagados para frente, nunca são enviados de volta. No entanto, modelos de redes neurais artificiais em que loops de feedback são possíveis, isto é, as informações também são enviadas de volta, são denominados de redes neurais recorrentes.

Uma rede neural recorrente (RNR) é uma classe de RNA que possui loops, ou seja, arquiteturas que permitem a passagem da informação por meio de múltiplas cópias da mesma rede, permitindo que as informações persistam. Quando ‘desdobramos’ o loop de uma RNR, é possível visualizar a rede como uma série de cópias em sequência, onde cada iteração passa sua saída para a próxima. A Figura 6 apresenta essa visão desdobrada da RNR. Essa capacidade de manter uma memória interna é o que torna as RNRs especialmente úteis em tarefas que envolvem sequências, como análise de texto, reconhecimento de fala e previsão.

Figura 6 - Rede Neural Recorrente



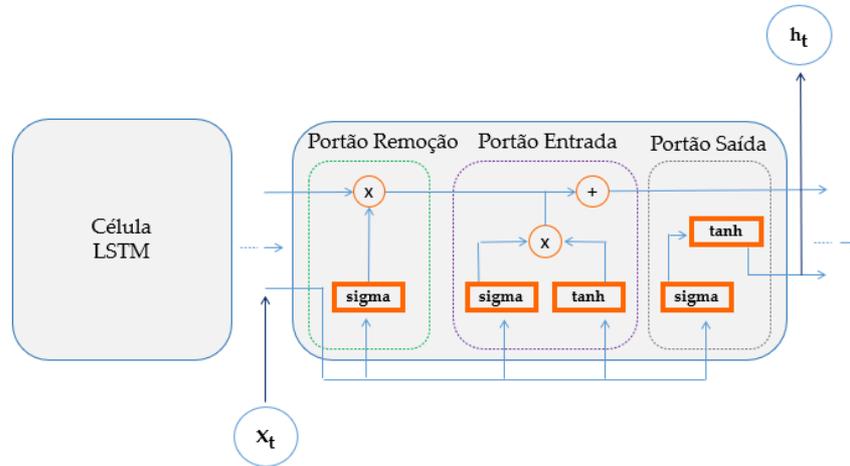
Fonte: Autor (2024)

As redes neurais recorrentes apresentam variações para superar e solucionar problemas que as redes neurais artificiais padrões não podem alcançar.

Segundo DATA SCIENCE ACADEMY (2018) e Venkatachalam *et al.* (2023) o *Long Short Term Memory* (LSTM) é um tipo de rede neural recorrente especial com propriedades de memorizar informações por um longo período de tempo por conta de sua estrutura de células, o que permite superar as Redes Neurais Tradicionais *feedforward* em tarefas onde a saída atual depende de saídas anteriores.

A arquitetura LSTM consiste em células de memória que contêm um estado de célula e um estado de porta. A célula de memória é responsável por manter uma memória de longo prazo, enquanto as portas controlam o fluxo de informações dentro e fora da célula. As portas são compostas por camadas densas, cada uma com ativações sigmóides que variam entre 0 e 1, permitindo que a rede decida quais informações devem ser esquecidas e quais devem ser lembradas. A Figura 7 ilustra a base da estrutura LSTM.

Figura 7 - Long Short-Term Memory (LSTM)



Fonte: Autor (2024)

a) **Portão de Remoção:** os dados processados que não são mais úteis no estado da célula são retirados com a porta de remoção. Duas entradas:  $x_t$  (entrada no momento específico) e  $h_{t-1}$  (saída de célula anterior) são alimentados ao portão e multiplicados por matrizes de peso, seguidas pela adição do bias. O resultado da operação é encaminhado para uma função de ativação cuja saída é binária. Se para um determinado estado de célula a saída for 0, a informação é descartada e para saída 1, a informação é armazenada para fins de processamento futuro;

b) **Portão de Entrada:** Adição dos dados operacionais indispensáveis ao estado da célula são feitos pelo portão de entrada. Primeiro, a informação é ajustada utilizando a função Sigmóide que filtra os valores a serem lembrados de forma similar ao portão de remoção usando as entradas  $h_{t-1}$  e  $x_t$ . Então, um vetor é criado usando a função tanh com saída de -1 a +1, que contém todos os valores possíveis de  $h_{t-1}$  e  $x_t$ . Os valores do vetor e os valores ajustados são multiplicados para obter as informações úteis;

c) **Portão de Saída:** A operação de remover informações úteis do estado da célula atual para ser exibida como uma saída é feita pelo portão de saída. Primeiro, um vetor é gerado aplicando a função tanh na célula. Então, a informação é regulada usando a função sigmóide que filtra os valores a serem lembrados usando as entradas  $h_{t-1}$  e  $x_t$ . Os

valores do vetor e os valores regulados são multiplicados para serem enviados como uma saída e entrada para a próxima célula.

A LSTM foi projetada para lidar com o problema de desvanecimento do gradiente em redes recorrentes convencionais. A LSTM é capaz de aprender dependências temporais de longo prazo, mantendo a capacidade de lidar com sequências de entrada de comprimento variável. Os hiperparâmetros da LSTM incluem:

- Número de unidades: o número de células de memória na camada da LSTM. Um número maior de unidades pode aumentar a capacidade da rede, mas também pode tornar o treinamento mais lento e aumentar o risco de *overfitting*;
- Número de camadas: a quantidade de camadas de células LSTM empilhadas na rede. Adicionar camadas pode melhorar a capacidade da rede para aprender representações mais complexas, mas também aumenta a complexidade da rede e aumenta a possibilidade de *overfitting*;
- Taxa de aprendizado: a taxa de ajuste dos pesos da rede durante o treinamento. Uma taxa de aprendizado alta pode levar a treinamento mais rápido, mas também pode levar a instabilidade e divergência no treinamento. Por outro lado, uma taxa de aprendizado muito baixa pode resultar em treinamento muito lento ou preso em mínimos locais;
- Função de ativação: a função que é aplicada aos resultados das portas. A função mais comum é a função *sigmóide*, mas outras funções, como a função tangente hiperbólica ou a função *ReLU*, também podem ser usadas;
- Regularização: métodos usados para prevenir o *overfitting*, como a regularização L1 ou L2, *dropout* ou *early stopping*;
- Tamanho do batch: o número de amostras de entrada que são processadas em paralelo durante o treinamento. Um tamanho de lote maior pode acelerar o treinamento, mas também pode exigir mais memória e processamento.

Esses são alguns dos hiperparâmetros mais comuns usados na arquitetura LSTM, e sua seleção pode influenciar significativamente o desempenho da rede. É importante ajustar esses parâmetros durante o treinamento para encontrar a combinação ideal que produz o melhor desempenho na tarefa em questão.

### 3.3.3 Algoritmo Genético

Segundo Kochenderfer and Wheeler (2019) o campo da Computação Evolutiva (CE) é formado por técnicas computacionais construídas sobre metáforas da natureza. O Algoritmo Genético (AG) é uma das primeiras metáforas, proposto por John Holland, na década de 1950, que explora os relacionamentos entre paradigmas e metáforas da Teoria da Evolução da Espécies de Charles Darwin como seleção natural, com as convicções da genética, como cruzamento e mutação. Os AG são algoritmos heurísticos de busca para achar soluções aproximadas em problemas de otimização, também aplicados em AM, que utiliza um processo de selecionar as soluções potencias para o problema, em um grupo de soluções geralmente geradas aleatoriamente, e efetua sobre elas procedimentos individuais ou em conjunto, visando buscar soluções ótimas.

#### 3.3.3.1 Estrutura Padrão de um AG

De maneira geral, um AG padrão pode ser descrito através do fluxograma apresentado na Figura 8.

Figura 8 - Fluxograma de um AG Clássico



Fonte: Autor (2023)

No AG padrão, como apresentado por Kochenderfer and Wheeler (2019), o algoritmo inicia criando um conjunto de indivíduos ou soluções aleatórias traduzidas em *strings* de binários, conhecido como população. Cada indivíduo da população tem um *string* específico

representando um cromossomo, descrevendo uma possível solução para o problema. Para cada cromossomo, é determinado um índice de aptidão que indica a ‘qualidade’ dessa possível solução para o problema enfrentado. A estrutura da codificação dos cromossomos varia de problema para problema.

A etapa de seleção e os operadores genéticos de cruzamento e mutação são os encarregados de gerar os novos indivíduos para compor a nova população através de um processo chamado de gerações.

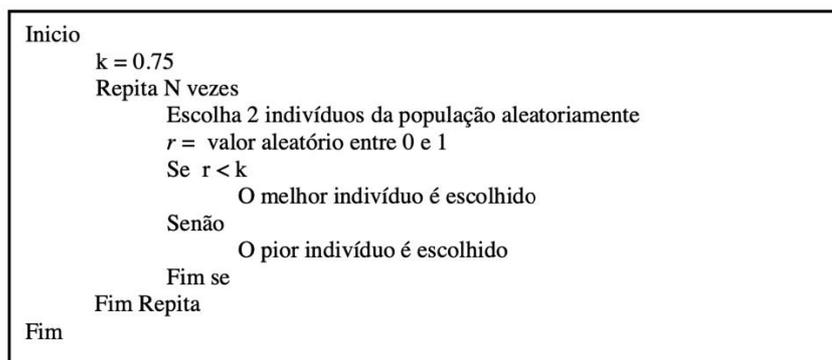
As iterações de seleção visam escolher soluções, conforme seu valor de aptidão, para misturá-las através do cruzamento, assim gerando novos indivíduos. Posteriormente estes novos indivíduos podem ser modificados pelo operador de mutação. Após várias iterações sucessivas, o AG converge os indivíduos da população para uma região do espaço de busca indicando o fim da evolução.

#### 3.2.3.2 Seleção e Elitismo

O propósito da etapa de seleção do AG é separar os membros da população que devem reproduzir. Em problemas de otimização, esta escolha deve ser realizada de tal maneira que os indivíduos de maior probabilidade para criar descendentes sejam os mais adaptados ao meio, ou seja os de maiores valores de aptidão. Existem vários mecanismos de seleção entre eles o torneio.

Neste método de seleção, sorteiam-se  $p$  indivíduos ao acaso da população para formar uma subpopulação temporária. Deste grupo, o indivíduo selecionado dependerá de uma probabilidade  $k$  definida previamente. Este procedimento é repetido para cada indivíduo a ser selecionado. Um exemplo básico da implementação deste algoritmo é mostrado na Figura 9, onde  $p=2$ .

Figura 9 - Algoritmo Básico do Método de Seleção por Torneio



Fonte: Mitchell (1997)

O elitismo é uma técnica aplicada para maximizar a convergência dos AGs, em outras palavras, esta técnica mantém os indivíduos mais aptos nas gerações seguintes. O elitismo consiste basicamente em realizar o processo em duas etapas:

- a) Seleciona-se uma elite de  $r$  membros entre os melhores da população inicial, os quais são incorporados diretamente na população final;
- b) O restante da população final é obtido a partir dos  $(n - r)$  elementos restantes da população inicial de tamanho  $n$ .

Além dos operadores genéticos e como o indivíduo é codificado, existem outros parâmetros que influem no comportamento dos AGs, entre eles são:

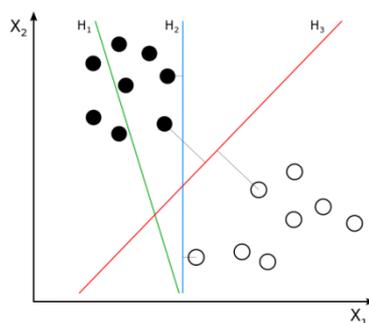
- a) Tamanho da população: Estabelece a quantidade de indivíduos que influencia no desempenho global e a eficiência do AG. Com uma população reduzida o desempenho pode diminuir, pois será explorado poucas soluções. Uma grande população fornece uma cobertura representativa do domínio do problema. Contudo, com um tamanho da população elevado, é necessário mais poder computacional;
- b) Taxa de Cruzamento: Define a probabilidade com que um cruzamento acontecerá. Quanto maior for essa taxa, mais rapidamente novas estruturas serão adicionadas na população. No entanto, se for muito elevada, grande parte da população será substituída, e pode ocorrer perda de estruturas de alta aptidão. Com um valor baixo, o algoritmo pode torna-se lento;

c) Taxa de Mutação: Define a probabilidade de ocorrência de uma mutação. Uma taxa reduzida de mutação previne a convergência prematura para um máximo local, viabilizando o algoritmo explorar melhor as soluções. Uma taxa de mutação elevado faz com que o processo de busca se torne casual.

### 3.3.4 Máquinas de Vetor de Suporte

É um algoritmo de aprendizagem de máquina utilizado para classificação ou regressão. A Máquina de Vetor de Suporte (SVM) objetiva encontrar uma reta de separação ou fronteira (também chamada de hiperplano) para um dado conjunto de dados. Os pontos de dados que estão mais próximos do hiperplano são chamados de vetores de suporte e determinam a localização do hiperplano (CERVANTES, 2020). Pode-se exemplificar essa ideia na Figura 10.

Figura 10 – Possíveis hiperplanos de Separação para uma *Data set* Bidimensional



Fonte: Wikimedia Commons (2012)

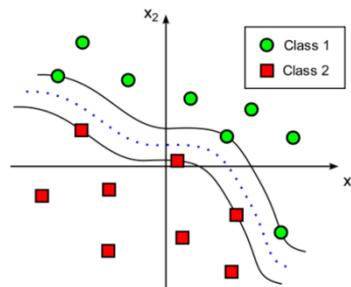
No caso da Figura 10, existem duas classes de observações, bola branca e preta, e três posições possíveis para o hiperplano. Portanto, os dados precisam ser totalmente separáveis. O classificador busca determinar essa posição a partir da maximização da margem entre os vetores de suporte (Fronteira) e elementos do grupo mais próximos.

A reta ótima ( $H_3$ ) é a mais distante entre os dois grupos, analisando os dois pontos, um de cada grupo, mais próximo a reta (indicados pelas linhas cinzas). Por conseguinte, o modelo conseguirá prever o novo dado a qual classe pertence verificando de qual lado da reta se encontra.

### 3.3.4.1 Máquinas de Vetor de Suporte Não Linear.

A SVM não limita seu uso em dados cujas as classes são linearmente separáveis, ou melhor, existem grupos que não podem ser separados somente por uma reta (hiperplanos), ver Figura 11.

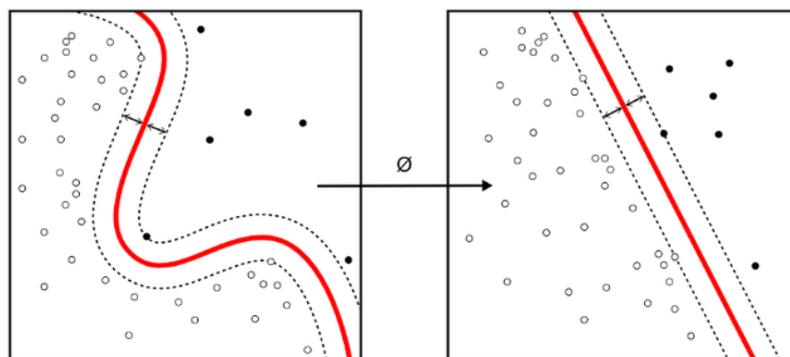
Figura 11 – SVM não Linear



Fonte: Ruben Ruiz-Gonçalves *et al.* (2014)

De acordo com a Figura 11, fica evidente a impossibilidade de separar as classes com uma linha reta. Dessa forma, técnicas de SVM não linear foram desenvolvidas para resolver problemas de natureza mais complexas. Como exemplo a técnica *Kernel Trick* que aplica uma transformação não-linear do espaço para depois separar os grupos com um SVM Linear (GORDIM *et al.*, 2021). Através de uma função, mapeia o conjunto de dados para um espaço de maior dimensão em que se torna possível obter um hiperplano que separe as classes corretamente (ver Figura 12).

Figura 12 – Transformação não Linear entre Espaço Inicial (à esquerda) e Espaço Depois da Transformação (à direita)



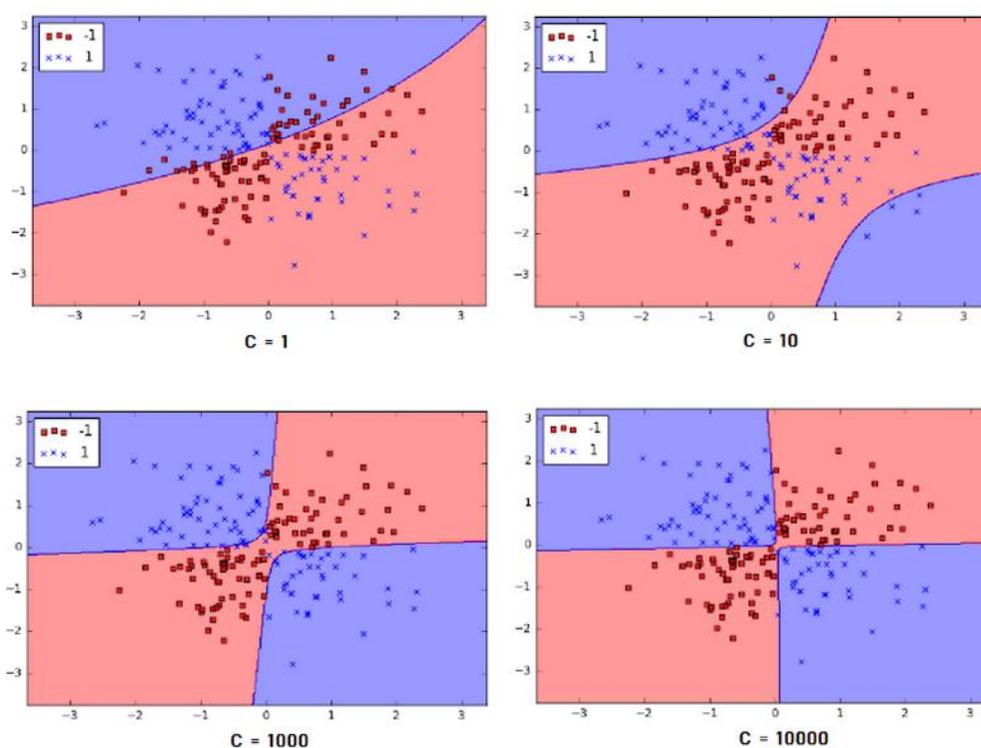
Fonte: Wikimedia Commons (2016)

Em alguns casos, introduzir uma reta que separe 100% grupos de um conjunto de dados é impossível, mesmo após aplicação de técnicas não lineares. Nesses casos os hiperparâmetros são otimizados para os ajustes finos dos modelos para buscar uma separação com o menor erro possível, para alterar a velocidade de treinamento, para evitar o chamado *Overfitting* e outros.

### 3.3.4.2 Hiperparâmetros do SVM

Os principais hiperparâmetros são denominados C e Gamma: o parâmetro C assume o compromisso por determinar um custo para cada registro, durante o treinamento, classificado de forma errada. Para tanto, um elevado valor C condiciona o algoritmo a treinar o modelo visando a separação completa entre as classes e possibilitar causar *Overfitting* e demandar muito tempo de treinamento. Por outro lado, valores de C muitos baixos flexibilizam o processo de treinamento e consente fronteiras de decisão com erros, porém pode ocasionar a um *underfitting*. A Figura 13 exemplifica o comportamento do SVM utilizando diferentes valores para C.

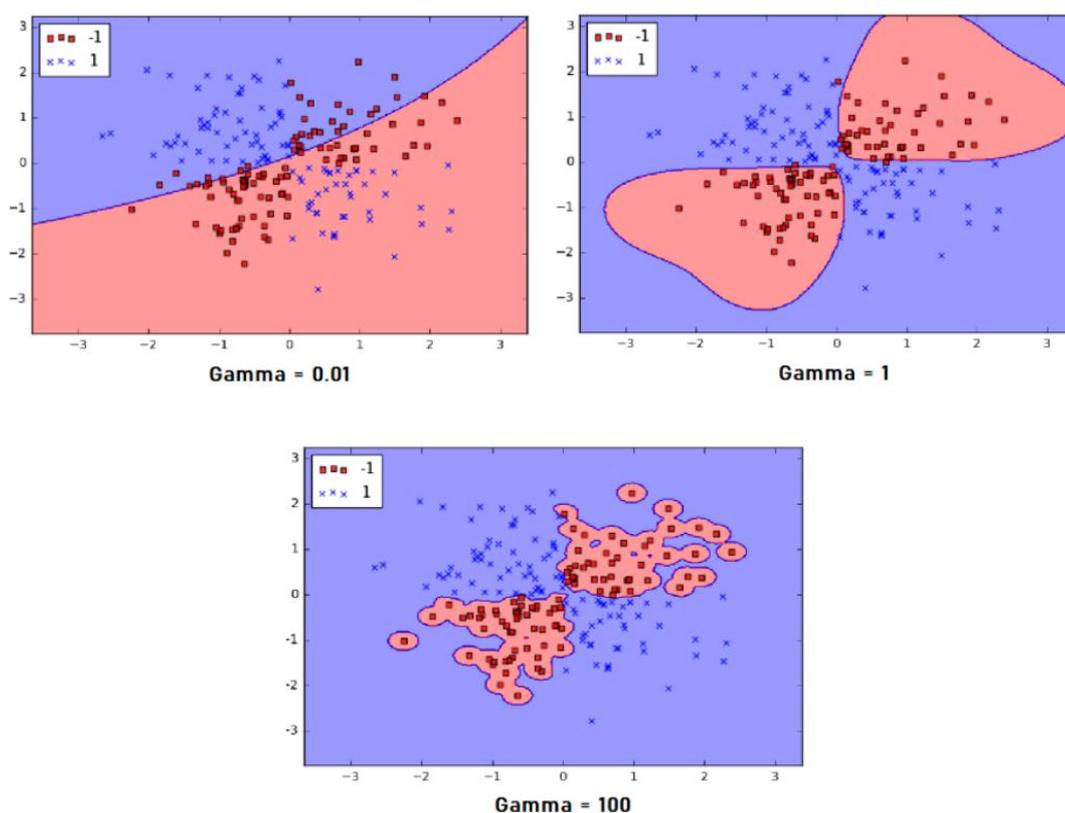
Figura 13 – Comportamento do SVM utilizando diferentes valores para C



Fonte: Chris Albon (2017)

O parâmetro Gamma é adotado para promover maior eficácia e atuação no distanciamento dos registros selecionados para o cálculo da fronteira de decisão. O parâmetro Gamma utiliza-se da aplicação de pesos para a distância entre registros, atribuindo maior ou menor relevância à amostras próximas ou distantes da fronteira de decisão. A Figura 14 apresenta o comportamento do SVM utilizando diferentes valores para o parâmetro Gamma.

Figura 14 – Comportamento do SVM utilizando diferentes valores para Gamma



Fonte: Chris Albon (2017)

O Parâmetro Gamma é introduzido em modelos SVM não linear, através das funções de *Kernel*, especificadamente o *Radial Basis Function* (RBF), polinomial e Sigmoide (HORAK *et al.*, 2020).

### 3.3.5 Modelo Auto-Regressivo Integrado de Médias Móveis

O Modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA) se baseia na metodologia tradicional da estatística proposto por George Box e Gwilym Jenkins, o método

Box-Jenkins, onde o modelo é direcionado quando o processo estocástico é não estacionário (Box e Jenkins, 1976).

PRISCILLA, 2022 divide o modelo ARIMA em três argumentos,  $p$ ,  $d$  e  $q$ , onde  $p$  é o atraso autoregressivo,  $d$  ordem de diferenciação e  $q$  é o parâmetro de média móvel a saber:

Diferenciação:

Os registros de uma base de dados geralmente possuem um grande volume não estacionários, sendo essencial a necessidade de uma transformação nas séries temporais (PAPASTEFANOPOULOS *et al.*, 2020) para gerar séries intermediárias que levem a uma série estacionária. Dependendo da base de dados, se faz necessário várias transformações (diferenciação), obtendo o valor de  $d=n$  ( $n$  é a quantidade de diferenciação).

Função de AutoCorrelação:

É a parte  $q$  do modelo ARIMA. É uma outra maneira de analisar a estacionariedade de uma série, que é uma medida do relacionamento linear entre os valores dentro de uma mesma série, isto é, a função de AutoCorrelação é uma forma a demonstrar o quanto um ponto influenciou o outro a cada variação do tempo da observação.

Função de AutoCorrelação Parcial:

A Função de AutoCorrelação Parcial está orientada a classificação do valor  $p$ , denominado atraso da autoregressão, também encontrada da mesma maneira da análise da Função de Autocorrelação.

Diante do contexto, para criação do ARIMA pelo método Box-Jenkins, são seguidas 3 etapas repetidamente até encontrar o modelo ideal:

- Identificação, é realizado testes estatísticos para compreensão das características das séries (*Dickey-Fuller Test*);
- Estimação, por meio da análise da Auto Correlação e Auto Correlação Parcial, é determinado os pontos utilizados para o modelo ARIMA( $p,d,q$ );
- Diagnóstico, etapa onde se valida os resultados por meio dos resíduos gerados e verificação da validade dos pressupostos;

O Teste *Dickey-Fuller* é uma avaliação da hipótese nula que verifica se uma série temporal possui raiz unitária. Caso a série possua raiz unitária, é considerada não estacionária. Por outro lado, se não houver raiz unitária, é considerada estacionária (FRACARO, 2018). Esse teste é um dos principais fatores na determinação da necessidade de aplicar diferenciação para transformar uma série não estacionária em estacionária. A diferenciação indica o valor do integrador  $d$  do modelo ARIMA.

Para interpretar a saída com *p-value* do Teste *Dickey-Fuller* (Ferreira e Patino, 2015), se a hipótese nula for rejeitada para  $p\text{-value} < 0,05$ , a série é considerada estacionária, ou seja, não possui raiz unitária. Caso o *p-value* seja  $> 0,05$ , a série é considerada não estacionária e requer uma transformação matemática para aplicação do modelo ARIMA.

Na etapa de estimação é realizado o ajuste do modelo ARIMA aos dados de entrada. Isso envolve estimar os parâmetros do modelo ARIMA para encontrar os valores que melhor se ajustam aos dados históricos. O método de estimação mais comum é o método de Máxima Verossimilhança (MLE, do inglês *Maximum Likelihood Estimation*) (AGRESTI *et al.*, 2021), que procura maximizar a probabilidade de observar os dados de entrada com base nos parâmetros do modelo.

O Critério de Informação de *Akaike* (AIC) é uma métrica usada em estatística e aprendizado de máquina para avaliar a qualidade do ajuste de um modelo estatístico aos dados observados. Foi desenvolvido por Hirotugu Akaike em 1974 (EDI *et al.*, 2023).

O AIC compara modelos diferentes com base em sua capacidade de ajustar os dados e, ao mesmo tempo, evitar o ajuste excessivo. Ele leva em consideração tanto o ajuste do modelo aos dados (por exemplo, quão bem o modelo descreve os dados observados) quanto a complexidade do modelo (por exemplo, quantos parâmetros ele usa).

O AIC é calculado usando a fórmula  $AIC = 2k - 2\ln(L)$ , onde  $k$  é o número de parâmetros no modelo e  $L$  é a máxima verossimilhança do modelo (ou seja, quão prováveis são os dados observados, dado o modelo).

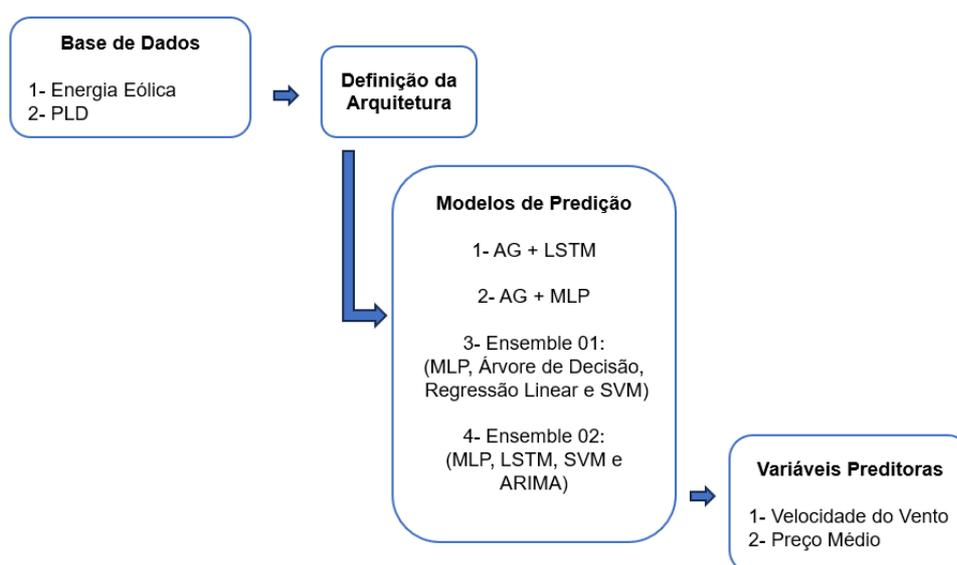
Quanto menor o valor do AIC, melhor é o ajuste do modelo aos dados. Portanto, o AIC pode ser usado para comparar modelos diferentes e selecionar aquele que melhor se ajusta aos dados observados, enquanto minimiza a complexidade do modelo.

Na etapa de diagnóstico é feita uma análise dos resíduos do modelo ARIMA para verificar se eles são aleatórios e apresentam uma distribuição normal. Se os resíduos não apresentarem essas características, isso pode indicar que o modelo ARIMA não é adequado para os dados de entrada ou que pode haver outros fatores que não foram considerados. Alguns métodos utilizados na etapa de diagnóstico incluem análise gráfica dos resíduos, teste de normalidade dos resíduos e autocorrelação dos resíduos.

## 4 MODELO PREDITOR PARA USO NO SISTEMA ELÉTRICO BRASILEIRO

Para este trabalho, destacou-se um modelo híbrido que combina técnicas de Aprendizado Supervisionado com RNA, incluindo LSTM e MLP. No modelo proposto, o AG é responsável por selecionar e ajustar os hiperparâmetros desses algoritmos, visando aumentar a eficiência do modelo e seu desempenho. Além disso, foram explorados Comitês de Máquinas, os quais incluem diferentes algoritmos como MLP, Árvore de Decisão, Regressão Linear e SVM em um ensemble, e MLP, LSTM, SVM e ARIMA em outro. A Figura 15 oferece uma representação visual do objetivo deste estudo.

Figura 15: Escopo Geral



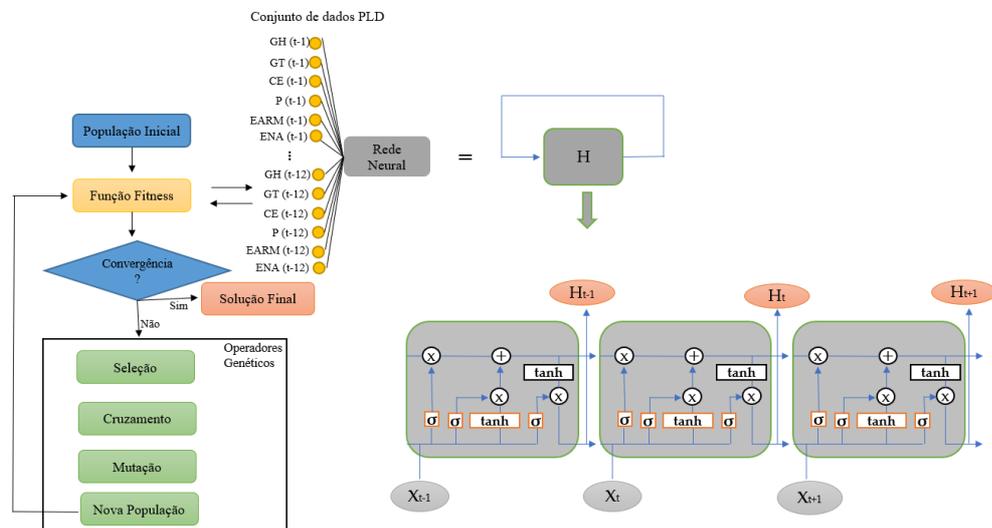
### 4.1 Redes Neurais Artificiais com Hiperparâmetros Otimizados pelo Algoritmo Genético

Sistemas Híbridos Inteligentes são modelos que resultam da combinação de duas, ou mais técnicas diferentes, visando unir vantagens para alcançar resultados bem mais satisfatórios ou para suprir deficiências individuais. Neste sentido, para alcançar o objetivo dessa tese, foi proposto a combinação de duas técnicas de inteligência computacional como estratégia, o AG e RNA, especificadamente LSTM e MLP. Neste contexto, para criação da arquitetura do modelo preditor, o AG teve como objetivo encontrar o número ótimo de camadas de neurônios, também o número ótimo de neurônios ou células em cada camada, além do valor ótimo da regularização (*Dropout*), etc da LSTM e MLP. O AG foi implementado para criar o modelo

preditor ótimo com a LSTM e MLP para algumas variáveis importantes do sistema elétrico do Brasil.

A pesquisa foi guiada basicamente por duas fases. A fase I (iteração do AG), utiliza o AG para criar/encontrar o modelo neural com condições mais favorável de capacidade de generalização para predição das séries temporais estudadas. Na fase II (cálculo da Função *Fitness*), o AG fornece à Rede Neural Artificial Profunda, LSTM e MLP, os valores decodificados do cromossomo binário para os hiperparâmetros da rede de forma que seja realizado o processo de aprendizagem na tarefa de predição de uma série temporal. O resultado do processo de aprendizado, após várias iterações até o total de 140, ou até a "Parada Antecipada" (ver subseção 5.1.1.1), é devolvido para o AG no valor da Função *Fitness*. Esse processo, fase I + fase II, perdura até a convergência do AG, que encontra uma configuração ótima para o hiperparâmetros da LSTM e MLP. A Figura 16 destaca as técnicas empregadas no modelo.

Figura 16 – Modelo Predictor de Série Temporal Combinando Algoritmo Genético e Rede Neural Profunda LSTM (exemplo de predição do preço de eletricidade PLD).



Fonte: Autor (2024)

A fase I inicia-se com os indivíduos (genótipos) aleatórios, e com todos os valores dos parâmetros descritos abaixo (seção 4.1). A fase II se dá quando a RNA atinge a quantidade máxima de épocas estabelecidas para essa fase de sintonia de hiperparâmetros. Após o treinamento da rede neural, faz-se necessário retornar à fase I o erro de predição que representa

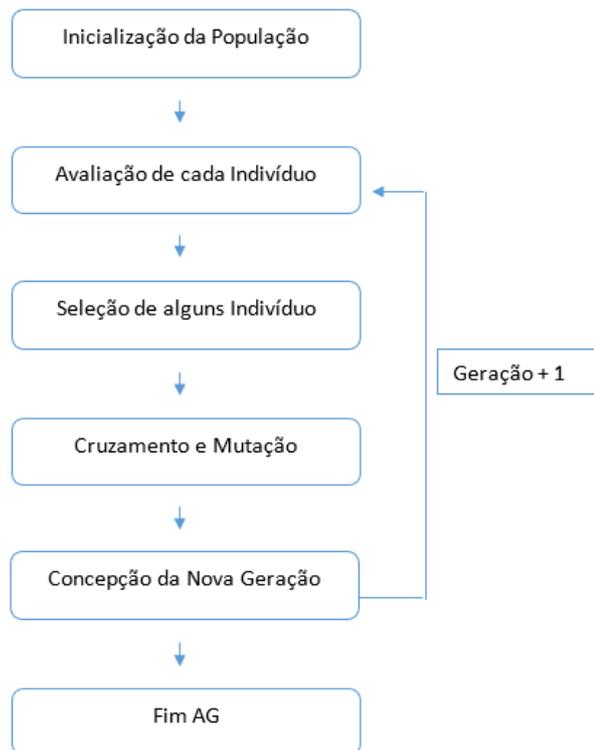
a função de avaliação do indivíduo. É importante observar que o número máximo de épocas para esta fase de treinamento da rede neural é pequeno, nesse momento de sintonia dos hiperparâmetros da LSTM. Após o AG ter realizado a sintonia dos hiperparâmetros a LSTM é treinada em um cenário mais exigente, onde o número máximo de épocas é maior, o valor mínimo de erro de predição é menor, etc.

No modelo da série temporal AG + LSTM , a previsão de valores futuros (1, 3, 8 e 12 passos à frente) de algumas variáveis importantes do sistema elétrico brasileiro é estimado. No caso do PLD são preditos média semanal e para velocidade do vento o valor diário. Os resultados obtidos para os dois estudos são realizados com 12 medidas atrasadas das variáveis de entrada e faz a predição  $k$  passos à frente da variável de saída. A definição das 12 medidas atrasadas para prever pode ser justificada pelas informações históricas (As medidas atrasadas podem conter informações importantes sobre o comportamento passado da variável, que são úteis para prever seu comportamento futuro) e melhoria da precisão (Em muitos casos, incluir mais medidas atrasadas pode aumentar a precisão do modelo de previsão, especialmente se houver uma relação significativa entre os valores passados e futuros da variável). Além disso, vale ressaltar que as 12 medidas atrasadas utilizadas para prever os passos seguintes também foram pela estratégia heurística de máximo de 12 medidas de horizonte de previsão. Foram realizados testes com menos de 12 medidas atrasadas na entrada do preditor, mas os resultados não foram satisfatórios.

#### 4.1.1 Fase I: Iteração do Algoritmo Genético

No AG, inicialmente tem-se a geração aleatória de uma população formada por indivíduos (possíveis soluções do espaço de busca). Estes são avaliados por uma função *fitness* e selecionados e, em seguida combinados entre si e sofrem mutação para obter uma nova população de indivíduos, uma nova geração, que espera-se sejam ‘melhores’ do que aqueles da população anterior. Os procedimentos computacionais do AG estão na Figura 17.

Figura 17 - Fluxo de um Algoritmo Genético



Fonte: Autor (2023)

Dessa forma, a primeira fase é caracterizada por encontrar a melhor configuração do modelo preditor de séries temporais do setor de energia do Brasil.

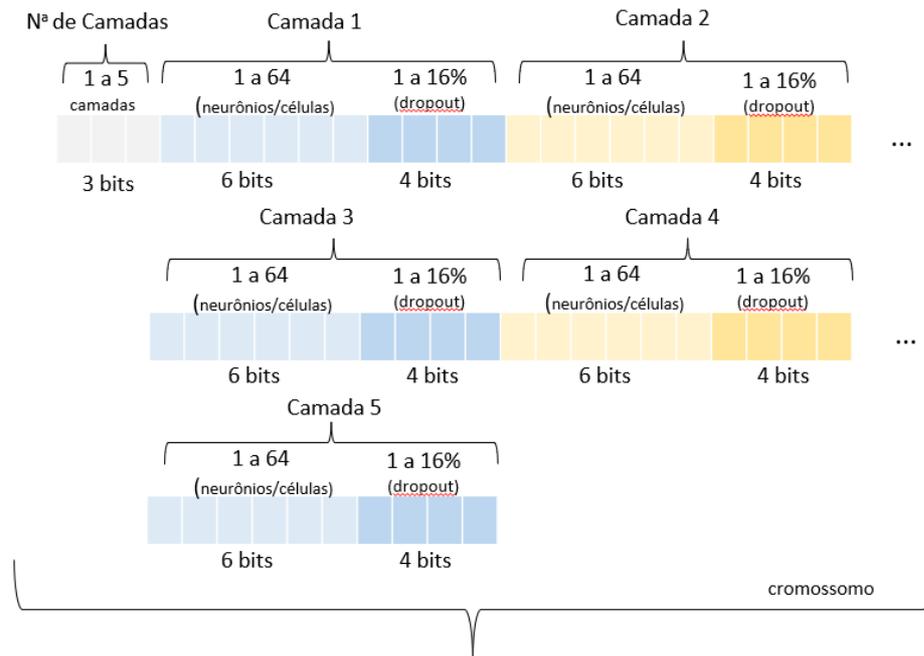
O AG utilizado nesse trabalho é o canônico com as características de Elitismo e seleção por Torneio (YANG, 2021). A seguir são apresentados os parâmetros do AG utilizado neste trabalho:

#### 4.1.1.1 Codificação do Problema

Inicialmente, define-se o genoma (representação ou descrição da solução/indivíduo que satisfaz a condição do problema) e suas condições que caracterizam uma solução factível. Esta solução é formada por variáveis do problema enfrentado, formando um vetor de números denominado de cromossomo. Logo, cada indivíduo tem uma cadeia de genes com seus

respectivos alelos (valor das variáveis) que codifica o seu cromossomo. Neste contexto, o problema em questão pode ser codificado conforme a Figura 18.

Figura 18 - Indivíduo da População



Fonte: Autor (2023)

Na Figura 18, os indivíduos são representados genotípicamente por vetores binários totalizando 53 bits, onde o primeiro grupo de três genes determina o número de camadas do modelo neural artificial LSTM e MLP e a cada grupo de 10 bits seguintes a estrutura da camada, sendo os primeiros 6 bits para indicar a quantidade de células, variando de 1 à 64, e os 4 bits restantes para definir o *Dropout*, balanceando de 1 à 16%.

O *Dropout* é uma técnica que possibilita modificar a rede eliminando aleatoriamente e temporariamente alguns dos neurônios das camadas intermediárias. Esse procedimento é como se estivéssemos treinando um grande número de redes neurais diferentes e, assim, as redes se adaptarão de diferentes maneiras à redução do *overfitting* (DATA SCIENCE ACADEMY, 2018; ILIADIS *et al.*, 2023).

#### 4.1.1.2 População

Antes de mais nada, uma população é um conjunto de indivíduos. Cada indivíduo é representado pelos seus cromossomos e em cada interação do AG, uma nova população será gerada. Por fim, na solução em questão, o AG inicia com um número pré-definido de soluções iniciais, 80 indivíduos, com os seus valores de genes gerados de forma aleatória. A quantidade estabelecida de indivíduos foi definida com base na estratégia de equilíbrio entre diversificação (exploração) e intensificação. Isso significa que pode melhorar a capacidade do algoritmo de encontrar soluções de alta qualidade de forma eficiente.

#### 4.1.1.3 Avaliação da População

Em termos de cálculos da função *Fitness* (Fase II), que mede o quão adaptado está o indivíduo, a população é avaliada segundo o aprendizado da RNA profunda, isto é o menor erro de previsão.

#### 4.1.1.4 Seleção

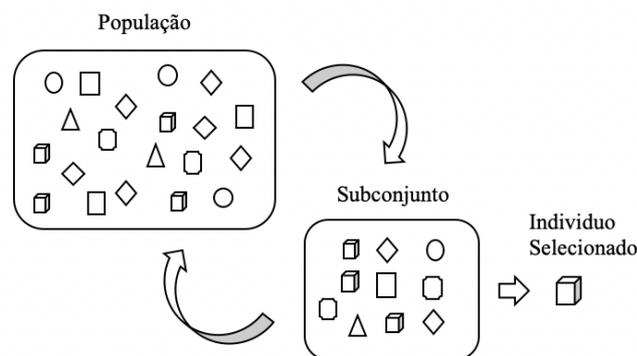
Uma das técnicas utilizadas em AG é a seleção aleatória, que consiste em selecionar aleatoriamente um conjunto de cromossomos da população atual (grupo de soluções). Esse processo de seleção aleatória é uma estratégia utilizada para explorar a diversidade genética da população, permitindo que cromossomos (pais da próxima geração) que possuem características únicas e vantajosas em relação aos demais, possam ser mantidos e contribuir para a geração de novas soluções.

Porém, é importante ressaltar que a seleção aleatória não garante a escolha dos melhores cromossomos da população, podendo haver indivíduos com características vantajosas que não são selecionados. Por isso, essa técnica geralmente é combinada com outras estratégias de seleção, como a seleção por torneio ou a seleção proporcional à aptidão, como a roleta, o ranking ou o método da amostragem estocástica universal.

Na seleção aleatória por torneio, é selecionado um conjunto de indivíduos da população aleatoriamente, sem considerar seus valores de aptidão. Em seguida, os indivíduos são avaliados e é selecionado o melhor indivíduo do conjunto como o vencedor do torneio (NIELSEN, 2021).

O tamanho do conjunto de indivíduos selecionados para o torneio é menor do que o tamanho da população. Isso permite que os indivíduos menos aptos tenham uma chance de serem selecionados e avaliados em relação aos mais aptos, aumentando a diversidade genética na população. O processo de seleção por torneio pode ser repetido várias vezes para selecionar vários indivíduos conforme pode ser observado na Figura 19.

Figura 19 - Mecanismo de Seleção



Fonte: Autor (2023)

#### 4.1.1.5 Elitismo

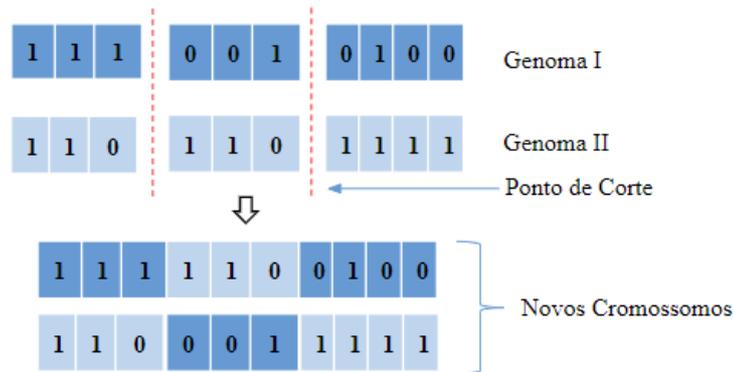
A fim de prevenir a perda de informações importantes presentes nas melhores soluções já encontradas e consequentemente aumentar o desempenho do AG foi implementada a técnica Elitista que consiste, neste trabalho, em escolher os “n” melhores indivíduos da população anterior (com  $n=1$ ) para estar na população atual. Isso garante que o melhor indivíduo da geração anterior seja preservado na geração atual.

#### 4.1.1.6 Cruzamento

Cruzamento é um dos operadores genéticos da técnica de AG aplicado sobre dois indivíduos, ou mais, com o objetivo de criar novos indivíduos. Não obrigatoriamente todos os participantes selecionados devem realizar o cruzamento. Assim, a taxa de *crossover* é uma variável aplicada para gerenciar a frequência com que acontecerá o cruzamento. Nesta etapa são eleitos ocasionalmente dois indivíduos para compor o cruzamento. Uma taxa de cruzamento de 75% sugere que o cruzamento ocorrerá consistentemente, enquanto uma taxa de 25% indica que o cruzamento nunca ocorrerá. Essas porcentagens são amplamente empregadas na maioria

das publicações sobre AG (KOTYRBA, 2022). Em seguida são designados arbitrariamente dois pontos de corte, e então seus genes são cruzados para obter dois novos indivíduos que vão compor a nova população. Na Figura 20 é possível observar a aplicação deste operador a um par de cromossomo.

Figura 20 - Cruzamento com Dois Pontos de Corte

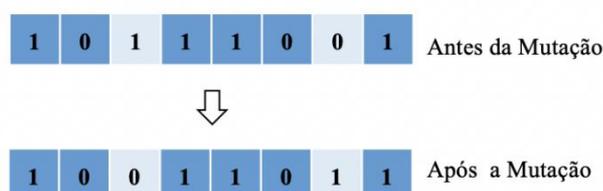


Fonte: Autor (2023)

#### 4.1.1.7 Mutação

A fim de introduzir e explorar novos indivíduos, além de possibilitar que a população recupere informação genética perdida, que não pode ser recuperada apenas por meio do cruzamento (Sivanandam, 2008), o operador de mutação, para o problema em questão, corresponde a verificação (probabilidade de mutação) e alteração através do processo de simples troca nos genes de cada indivíduo. Portanto, um gene selecionado aleatoriamente é modificado de acordo com uma taxa de mutação de 1%, como ilustrado na Figura 21. Este valor é amplamente utilizado na maioria das publicações sobre AG, conforme destacado no Zolpakar *et al.* (2021).

Figura 21 - Mutação em dois Genes de um Indivíduo



Fonte: Autor (2023)

#### 4.1.2 Fase II: Cálculo da Função *Fitness* (Erro de predição das Redes Neurais Artificiais Profundas)

Para utilizar a combinação de AG com LSTM ou MLP, é necessário que o AG tenha acesso à informação de *fitness* de cada indivíduo. Como cada indivíduo é, na verdade, um conjunto de hiperparâmetros que define um modelo preditor de série temporal, é preciso treinar o modelo com esses hiperparâmetros para avaliar a sua qualidade.

Dessa forma, a utilização do AG nessa combinação demanda que um modelo neural preditor seja treinado rapidamente para avaliar cada conjunto de hiperparâmetros gerado pelo AG. Esse processo de treinamento é fundamental para medir a qualidade de cada indivíduo e selecionar aqueles que têm maior aptidão para a solução do problema.

Portanto, a combinação AG+LSTM ou AG+MLP requer a interação entre o AG e o modelo neural preditor, de forma a otimizar o processo de busca dos melhores hiperparâmetros para o modelo. É fundamental ressaltar que, após a sintonia dos hiperparâmetros do modelo preditor, é imprescindível avaliá-lo para verificar a sua capacidade de generalização e previsão. Essa etapa é crucial para garantir que o modelo seja capaz de fornecer resultados precisos e confiáveis em situações reais.

Nesse processo, a avaliação do modelo preditor, são utilizadas métricas de desempenho que comparam as previsões do modelo com os dados observados. Desse modo, para calcular o erro da função de avaliação – *Fitness*, usa-se o Erro Médio Quadrático –MSE.

A Tabela 2 resume os parâmetros utilizados para a simulação do AG empregado neste trabalho.

Tabela 2 - Parâmetros para Simulação com AG

População	80 indivíduos
Elitismo	São escolhidos os $n$ melhores indivíduos da população anterior ( $n=1$ )

Probabilidade de cruzamento	75% (com dois pontos de corte)
Mutação	Os genes dos indivíduos podem mudar com a probabilidade de 1%
Seleção	Seleção aleatória por Torneio

## 4.2 Comitê de Máquinas

Os Comitês de Máquinas (ensembles) têm se destacado como uma abordagem eficaz para aprimorar o desempenho dos sistemas de AM (CARNEIRO, 2022). Em vez de depender de um único modelo, um comitê consiste em vários modelos de aprendizado, conhecidos como membros, que colaboram para tomar decisões mais precisas e robustas.

A ideia por trás do ensemble é explorar a diversidade e complementaridade dos modelos individuais. Cada membro do comitê pode ser treinado de maneira independente, utilizando diferentes algoritmos de aprendizado, conjuntos de dados ou configurações de hiperparâmetros. Ao combinar as previsões de cada membro, o comitê pode capturar diferentes perspectivas e reduzir o impacto de possíveis erros individuais, resultando em previsões mais confiáveis e precisas.

Existem diversas técnicas para combinar as previsões dos membros do comitê, como votação, *boosting*, *bagging*, entre outras (WANG *et al.*, 2016). Essas estratégias de agregação permitem aproveitar as fortalezas de cada modelo, enquanto mitigam suas fraquezas.

Dado o contexto de diversidade e complementaridade dos modelos individuais presentes no ensemble, optou-se por adotar o método de votação para combinar as previsões (SAGI, 2018). Nesse método, cada membro emite uma previsão para uma amostra específica, e a previsão final é determinada pela média das previsões individuais.

A votação é uma técnica simples e robusta que permite aproveitar as diversas perspectivas dos membros do comitê, enquanto mitiga possíveis erros individuais. Ao contar

com a contribuição de cada modelo, é possível obter resultados mais confiáveis e precisos em uma variedade de tarefas de aprendizado. Destaca-se que a escolha do método de votação para este trabalho foi embasada em estudos anteriores e na natureza do problema em questão. Essa estratégia se mostrou adequada para alcançar um equilíbrio entre acurácia e simplicidade no processo de tomada de decisão (SOLANO; ALYMANI *et al.*, 2023).

Os membros dos dois ensembles serão compostos por quatro modelos distintos: ensemble 01 (árvore de decisão, MLP, regressão linear e SVM) e ensemble 02 (MLP, LSTM, SVM e ARIMA). É crucial destacar que a constituição dos modelos nos ensembles foi planejada para garantir que cada componente generalize de maneiras diferentes. Além disso, na formação dos modelos, é considerada a inclusão daqueles que demonstraram resultados favoráveis em aplicações específicas. Por exemplo, em Zu *et al.* (2023), observou-se uma melhoria na precisão da previsão da resistência à compressão uniaxial (UCS) com o uso de SVM. Domingues *et al.* (2022) apresentaram uma nova contribuição com a regressão linear, combinando a Análise Dimensional (AD) para resolver problemas de instabilidade e erro. Por fim, Ospina *et al.* (2023) com ARIMA ofereceram percepções sobre as implicações e adaptabilidade de métodos preditivos no contexto de uma pandemia global, entre outros. As Figuras 22 e 23 representam os constituintes e a abordagem para combinar os votos no processo de tomada de decisão.

Figura 22 - Proposta de Comitê de máquinas: “Ensemble 01”.

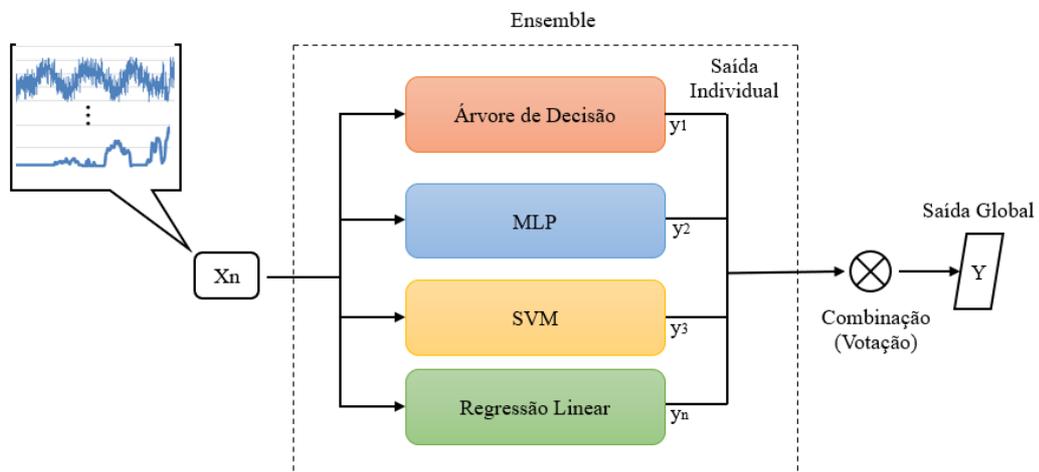
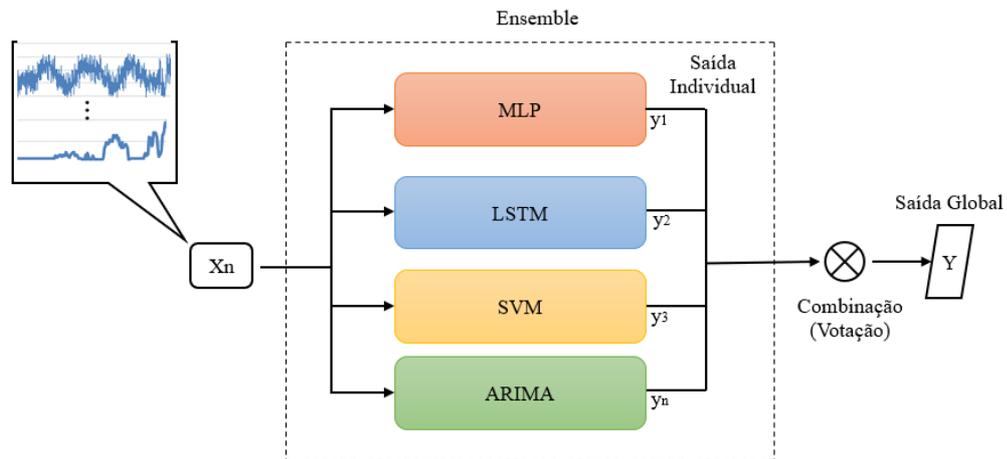


Figura 23 - Proposta de Comitê de Máquinas: “Ensemble 02”.



Essa combinação de modelos foi selecionada para explorar a diversidade de técnicas de AM e suas habilidades para lidar com diferentes aspectos do problema em questão (CHRISTOPH *et al.*, 2019).

A árvore de decisão é conhecida por sua capacidade de criar regras claras e interpretações intuitivas, tornando-a útil para tomar decisões relacionadas a atribuições de regressão e classificação.

Por outro lado, as RNAs são capazes de lidar com problemas mais complexos, aproveitando a capacidade de aproximação universal das redes neurais (VILLANUEVA, 2006; ZHOU, *et al.*, 2020).

A regressão linear é uma técnica tradicional que fornece uma abordagem simples e interpretável para problemas de regressão, enquanto a SVM é eficaz na separação de classes em problemas de classificação (LV, 2023).

O modelo ARIMA é uma técnica estatística usada em uma variedade de cenários onde se deseja prever valores futuros analisando padrões históricos dentro de dados de séries temporais (WANG, 2018).

Ao combinar esses modelos, o objetivo foi capturar diferentes perspectivas e características do conjunto de dados, aproveitando as forças individuais de cada um. Considerou-se que, com essa diversidade, essa combinação provavelmente forneceria resultados mais precisos e confiáveis para alcançar os objetivos estabelecidos.

A definição dos parâmetros para os melhores modelos dos ensembles de previsão está resumida na Tabela 3.

Tabela 3. Hiperparâmetros dos Membros dos Ensembles

<b>Membros "Ensemble 01"</b>	<b>Parâmetros</b>
Árvore de Decisão	Medição de cada nó da árvore: mean squared error – MSE; Node splitting: "best" (melhor divisão possível); Tree depth: 5;
Regressão Linear	Método fit; fit_intercept: True; copy_X: True; n_jobs: None; Positive bool: False.
SVM	Método random search (Base de dados - PLD): C: 70.645; Gamma: $4.64 \times 10^{-6}$ ; Epsilon: 0.230; Método Bayes search (Base de dados - Macau): C: 5.202; Gamma: auto Epsilon: 0.061.
MLP	2 Camadas escondida de 64 neurônios cada; Função de ativação não-linear em todas as camadas escondidas: "relu"; Definição épocas para o treinamento: 50; Função de ativação na camada de saída: Linear; Atualização dos pesos: 'adam'.
<b>Membros "Ensemble 02"</b>	<b>Parâmetros</b>
MLP e LSTM	Definição épocas para o treinamento: 100; não linear em todas as camadas escondidas: "relu"; Função de ativação na camada de saída: Linear; Atualização dos pesos: 'RMSProp optimizer'.
SVM	Método Random search (Base de dados - PLD): C: 70.645; Gamma: $4.64 \times 10^{-6}$ ; Epsilon: 0.230;  Método Bayes search (Base de dados - Macau): C: 5.202; Gamma: auto Epsilon: 0.061.
ARIMA	Teste de estacionariedade: Dickey–Fuller;

---

Técnica Grid-Search;  
 Método tradicional da estatística Box-Jenkins proposta por George Box e Gwilym Jenkins, que consiste na utilização de modelos Auto-Regressivos Integrados de Médias Móveis (ARIMA(p,d,q))

---

Sobre a atualização dos pesos das redes neurais, o algoritmo de treinamento utilizou a taxa de aprendizagem adaptativa com o *RMSProp* (FANGYU *et al.*, 2019).

Esse otimizador foi definido conforme a Equação (4):

$$E[g^2]_t = 0.9 \cdot E[g^2]_{t-1} + 0.1 \cdot g_t^2, \quad 4$$

Onde  $g$  é o gradiente da função de custo,  $g_t$  é o gradiente da função de custo no tempo  $t$ , e  $E[g^2]_x$ , é a média móvel dos quadrados dos gradientes no tempo  $x$ . A atualização dos vetores de parâmetro é dada pela Equação (5):

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t, \quad 5$$

Onde  $\eta$  é a taxa de aprendizagem,  $g_t$  é o gradiente da função de custo no tempo  $t$ ,  $E[g^2]_t$  é a média móvel dos quadrados dos gradientes no tempo  $t$  e  $\epsilon$  é o termo que serve para evitar a divisão por zero.

### 4.3 Aspectos Gerais e Relevantes dos Modelos Propostos

#### 4.3.1 Base de Dados

Neste trabalho foram utilizados dois bancos de dados reais: a base de dados do Preço de Liquidação das Diferenças - PLD, do Sistema Nacional Brasileiro, com registros de preços da eletricidade das regiões do Norte, Nordeste, Sul e Sudeste no período de 04 agosto de 2001 a 16 de dezembro 2009, fornecendo um conjunto de 440 amostras semanais para cada Região (HOMEPAGE CCEE, 2020); e a base de dados de Energia eólica, com informações de dois locais diferentes na região nordeste do Brasil: Macau (latitude 5 ° 9'3.726 "no sul, longitude 36 ° 34 '23.3112" no oeste), e Petrolina (latitude 09 ° 04 '08 "no sul, longitude 40 ° 19' 11" no oeste) no período 01 De junho de 2016 a 31 de maio de 2017 (HOMEPAGE SONDA, 2020), totalizando 4900 amostras cada.

Para a previsão do PLD utilizou-se as seguintes variáveis exógenas como entrada do modelo preditor: Geração Hidráulica (GH) e térmica (GT), Carga de Energia (CE), Preço do PLD (P) e Reservatórios EARM e ENA, com base na técnica de ranqueamento de variáveis explanatória que avalia o poder preditivo de uma variável individualmente a partir de uma função critério. Maiores detalhes sobre a seleção das variáveis são encontrados em (FILHO *et al.*, 2014) que descreve e apresenta a seleção das variáveis explanatórias a partir do teste F ANOVA (Análise de Variância).

Para a velocidade do vento de Aerogeradores do Nordeste do Brasil as variáveis exógenas na entrada do preditor foram: Velocidade (V), Temperatura do Ar (TA), Umidade do Ar (UA), Pressão Atmosférica (PA) e Direção do Vento (DV). Neste trabalho as séries históricas meteorológicas utilizadas foram inspiradas em (ALENCAR, 2018) onde foram obtidas no banco de dados disponibilizado pelo sistema de organização nacional de dados ambientais (SONDA) e o processo de validação dos dados obtidos pela estação SONDA baseia-se na estratégia de controle de qualidade de dados adotada pela *Baseline Surface Radiation Network* (BSRN) e os dados complementares foram disponibilizados diretamente com técnicos da estação.

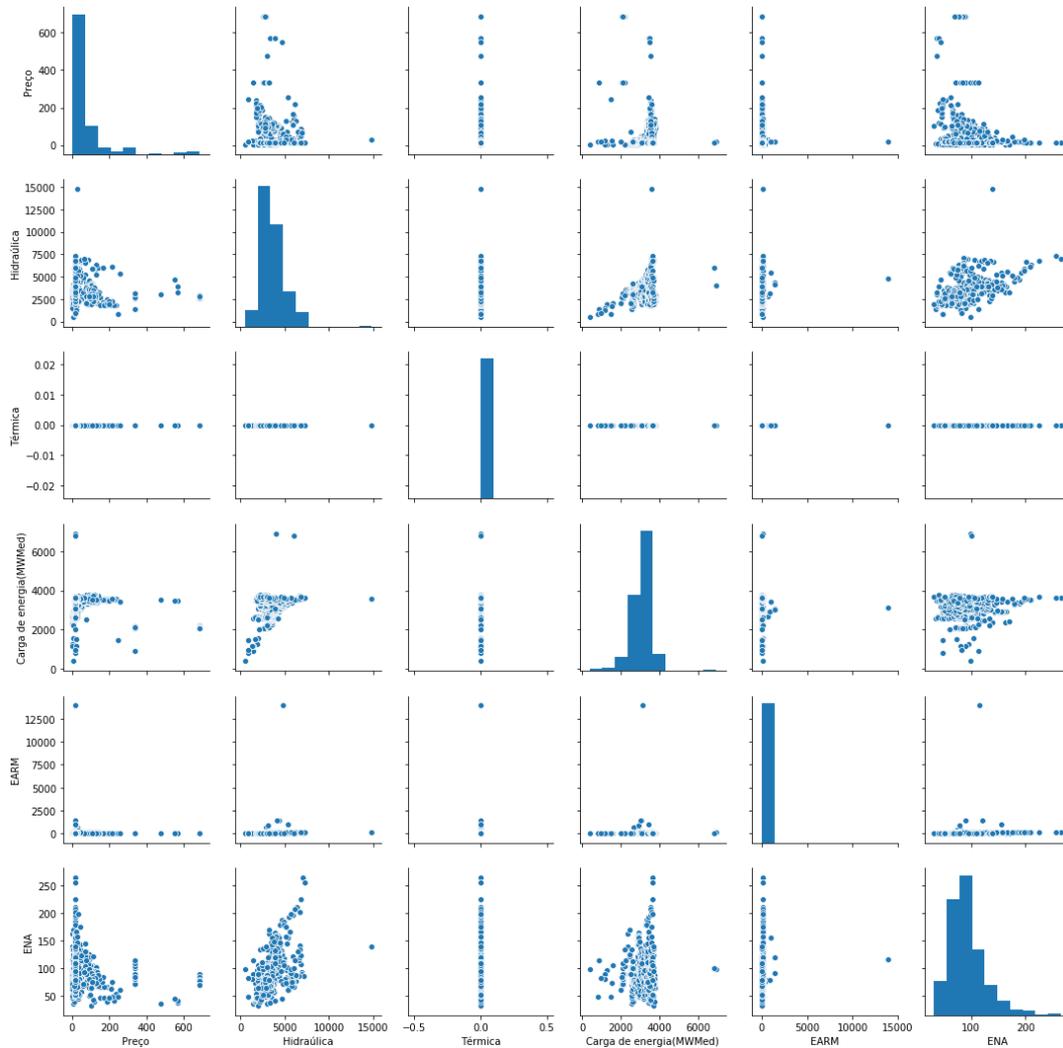
Na oportunidade, as variáveis de entrada, também conhecida como variáveis exógenas, desempenham um papel crucial na análise e no funcionamento dos sistemas energéticos. Para garantir a precisão e a confiabilidade das análises, essas variáveis são frequentemente extraídas de fontes confiáveis e autorizadas. No contexto do setor elétrico brasileiro, o Operador Nacional do Sistema (ONS), responsável por coordenar e supervisionar a operação das instalações de geração e transmissão de energia elétrica no país, disponibiliza uma série de informações pertinentes, incluindo dados sobre GH, GT, CE, EARM e ENA. Enquanto os valores de preços são geralmente obtidos da Câmara de Comercialização de Energia Elétrica (CCEE) órgão responsável por viabilizar e operacionalizar a comercialização de energia elétrica no mercado brasileiro, as informações relacionadas às variáveis operacionais e climáticas cruciais para a gestão do sistema elétrico são fornecidas pelo ONS. Ao integrar esses dados de diferentes fontes, os analistas e operadores do setor elétrico podem realizar avaliações abrangentes, tomar decisões informadas e garantir a estabilidade e a eficiência do sistema elétrico nacional. Portanto, a colaboração entre instituições como o ONS e a CCEE desempenha um papel fundamental na sustentabilidade e no desenvolvimento contínuo do setor energético brasileiro.

De modo a facilitar e possibilitar uma análise do comportamento dos dados e, assim, identificar algum padrão ou tendência, são representados graficamente apenas dados de PLD para uma região e dados de Velocidade do vento para um parque eólico. Contudo, foram realizadas a mesma análise para os demais dados.

As Figuras 24 e 25 ilustram a matriz de Scatter que é uma ferramenta gráfica que permite visualizar a relação entre duas ou mais variáveis, mostrando como elas se relacionam. Essa ferramenta é particularmente útil para a análise de dados em que há várias variáveis que podem afetar o resultado final.

A matriz é composta por uma combinação de gráficos de dispersão que relacionam as variáveis duas a duas. Cada ponto na matriz representa uma observação para as variáveis em questão. A disposição diagonal é ocupada por gráficos de histograma para cada uma das variáveis, mostrando a distribuição de frequência dos dados para essa variável em particular. Isso permite ter uma visão geral da distribuição dos dados e identificar possíveis anomalias ou desvios em relação a uma distribuição normal. Além disso, também pode ajudar a identificar *outliers* ou valores extremos que possam estar influenciando a análise de correlação entre as variáveis. Em resumo, a disposição diagonal é uma forma de visualização das distribuições individuais de cada variável na matriz de scatter.

Figura 24: Matriz de Scatter, plot para 6 Variáveis Dados PLD- Região Norte

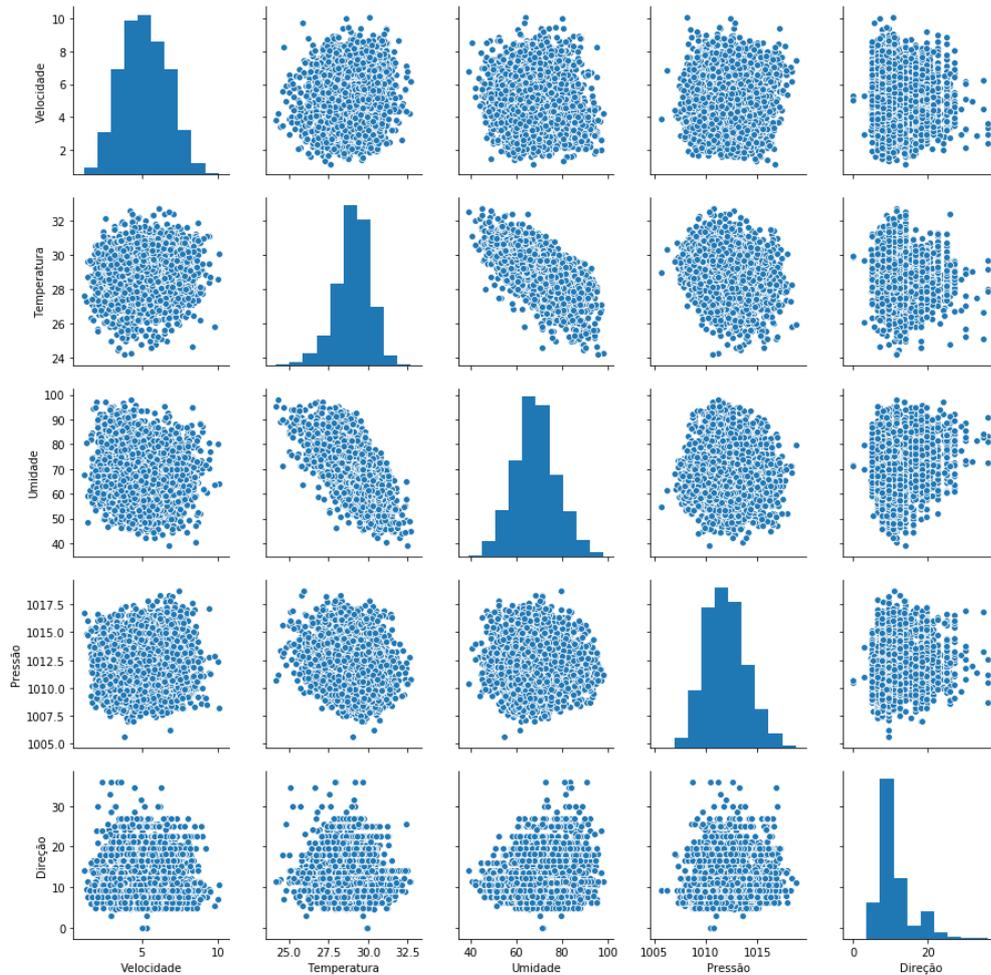


Na Figura 24, apresenta-se a matriz de scatter com o plot para 6 variáveis de dados do PLD na região Norte: Preço, Hidráulica, Térmica, Carga de Energia, EARM e ENA.

Ao analisar a posição dos pontos na matriz de scatter, é possível identificar possíveis relações entre as variáveis, por exemplo se há uma correlação positiva ou negativa entre elas. Por exemplo, pode-se observar que há uma forte correlação positiva entre a geração hidrelétrica e a geração térmica na região norte, o que pode indicar que quando uma dessas fontes de energia aumenta, a outra também tende a aumentar. Além disso, é possível verificar que há uma correlação negativa entre o valor da carga de energia e o valor da energia no mercado spot, o

que pode sugerir que quando a carga de energia aumenta, o valor da energia no mercado spot tende a diminuir.

Figura 25: Matriz de Scatter, plot para 5 Variáveis Dados Velocidade do Vento- Macau



Na Figura 25, tem-se a matriz de scatter para 6 variáveis relacionadas à velocidade do vento em Macau: Velocidade, Temperatura, Umidade, Pressão e Direção. Cada variável é plotada em um eixo separado, e os pontos no gráfico representam os valores observados para cada variável.

Ao analisar a posição dos pontos na matriz de scatter, é possível identificar possíveis relações entre as variáveis. Por exemplo, podemos observar que a velocidade do vento tem uma correlação negativa com a pressão, o que sugere que quando a pressão atmosférica aumenta, a velocidade do vento tende a diminuir. Por outro lado, pode-se observar uma correlação positiva

entre a velocidade do vento e a temperatura, o que indica que quando a temperatura aumenta, a velocidade do vento também tende a aumentar.

Além disso, é importante ressaltar que a direção do vento também pode influenciar a velocidade. Na Figura 25, é possível observar que existem pontos concentrados em certas áreas do gráfico, o que pode indicar a presença de padrões de vento específicos na região analisada.

As Tabelas 4 e 5 apresentam as análises estatísticas relativas às cinco e seis séries históricas, respectivamente. Estas análises incluem medidas de tendência central e dispersão, proporcionando uma visão abrangente dos dados.

Tabela 4 - Medições de Posição e Dispersão na Região Norte.

Medidas	PLD	Hidraulica	Térmica	Carga	EARM	ENA
Média	66,66	3.636,02	0,00	3.091,66	106,44	93,00
Mínimo	4,00	539,00	0,00	376,43	8,73	31,86
Máximo	684,00	14.795,00	0,00	6.922,29	13.948,34	264,43
Amplitude	680,00	14.256,00	0,00	6.545,86	13.939,61	232,57
variância	133,53	18.106,00	0	345.297,7519	446.677,2	1.130,133
Desvio Padrão	115,68	1347,11	0	588,2847683	669,0948	33,65546
mediana	18,59	3382,14	0,00	3.165,43	67,20	87,29
Moda	18,59	3876	0	2.622,14	73,81	97

Tabela 5 - Medições de Posição e Dispersão de Macau.

Medidas	Velocidade	Temperatura	Umidade	Pressão	Direção
Média	5,17	29,04	68,42	1011,94	11,07
Mínimo	1,15	24,20	39,00	1005,63	0,00
Máximo	10,07	32,70	98,00	1018,73	36,00
Amplitude	8,92	8,50	59,00	13,10	36,00
variância	2,25051	1,21818	85,6184	3,7769	19,6294
Desvio Padrão	1,50032	1,10383	9,25397	1,94362	4,43096
mediana	5,10	29,13	68,00	1011,87	9,33
Moda	5,3	29	71	1012,3	9,33

### 4.3.2 Pré-processamento

As variáveis de entradas utilizadas no treinamento e teste do preditor, para que todos os modelos viessem a performar melhor, receberam o pré-processamento da normalização (PUHEIM, 2014), colocando os dados na escala comum de 0 a 1, sem distorcer as diferenças nos intervalos de valores.

O método usado para normalização das bases de dados foi o seguinte (ver Equação (6)):

$$V_n = \frac{(V - V_{\min})}{(V_{\max} - V_{\min})} \quad 6$$

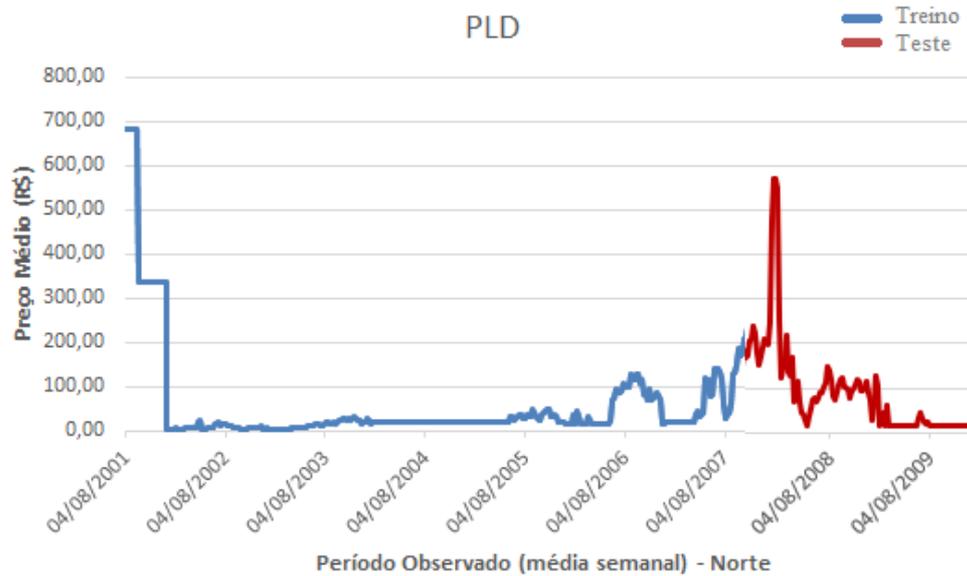
Onde:

- a)  $V_n$  = Valor normalizado;
- b)  $V$  = Valor não normalizado;
- c)  $V_{\min}$  = Valor mínimo apresentado pela variável;
- d)  $V_{\max}$  = Valor máximo apresentado pela variável.

Vale ressaltar que em função da preponderância das condições hidrológicas, na demanda de energia, nos preços de combustível, no custo de déficit e na disponibilidade de equipamentos de geração e transmissão no parque de geração brasileiro as variáveis usadas neste trabalho têm sido mais comumente examinadas em estudos de previsão (DUMITRU, 2017).

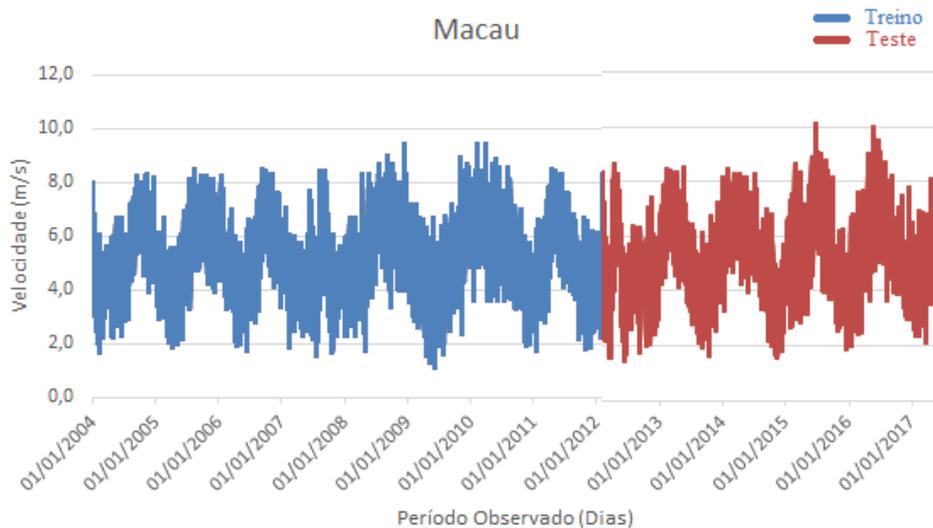
Para a Rede Neural Profunda (DNN) e o AG, um método de amostragem de validação (*Holdout*) foi aplicada para obter estimativas de desempenho preditivos mais confiáveis definindo subconjunto de treinamento e testes (CARVALHO, 2011). Cada base de dados teve suas amostras divididas em 75% para treinamento empregado para indução e ajustes de modelos e 25% para testes que simulam a predição de objetos novos ao preditor aos quais não foram vistos na fase de treinamento. Esses subconjuntos são disjuntos para assegurar que as medidas de desempenho fossem obtidas a partir de uma amostra diferente daquele usado no aprendizado. Nas Figuras 26 e 27, pode-se visualizar a definição das séries temporais de dados do PLD no Norte Brasileiro e velocidade do vento na cidade de Macau.

Figura 26 Separação de Treino e Teste da Base de Dados do PLD - Norte



A Figura 26 apresenta a evolução do Preço Médio do PLD ao longo de um período observado, dividido em duas partes: a parte vermelha, que representa os dados de teste, e a parte azul, que representa os dados de treino.

Figura 27 Separação de Treino e Teste da Base de Dados da Velocidade do vento - Macau



A Figura 27, a parte vermelha do gráfico representa os dados de teste, enquanto a parte azul representa os dados de treinamento. No eixo x, tem-se o período observado, e no eixo y a velocidade do vento registrada na cidade de Macau. É importante lembrar que o uso de dados

de treinamento e teste é uma prática comum em modelagem preditiva para avaliar a capacidade do modelo de generalizar para novos dados.

Na ocasião, foram aplicadas 10 partições para validação cruzada, reservando 20% dos dados de treinamento para a validação. Dessa forma, foram realizados 10 treinamentos distintos, em que cada um utilizava 80% dos dados para treinamento e 20% para validação.

O processo de validação cruzada envolve dividir o conjunto de dados disponível em dois ou mais conjuntos complementares: um conjunto de treinamento e um conjunto de validação. O modelo é treinado no conjunto de treinamento e, em seguida, avaliado no conjunto de validação. Esse processo é repetido várias vezes, com diferentes divisões dos dados em conjuntos de treinamento e validação.

Existem várias técnicas de validação cruzada, sendo a mais comum a validação cruzada *k-fold*. Nessa abordagem, o conjunto de dados é dividido em *k* partições (ou *folds*) de tamanho aproximadamente igual. O modelo é treinado *k* vezes, cada vez usando *k-1* partições como conjunto de treinamento e uma partição como conjunto de validação. O desempenho do modelo é então calculado como a média dos resultados obtidos em cada uma das *k* iterações.

Para simular os membros do ensemble, os dados de PLD e velocidade do vento passaram por pré-processamento (limpeza de dados, tratamento de valores faltantes, normalização de dados, codificação de variáveis categóricas, etc.). Durante este pré-processamento foi utilizada uma técnica de balanceamento de dados, que visou distribuir uniformemente o trabalho ou carga entre os recursos disponíveis (PUHEIM, 2014). Este procedimento garantiu que os dados sejam equilibrados e representativos para análise posterior. Após, os dados foram divididos em conjuntos de dados de treinamento e avaliação. Foi escolhida uma proporção de 75% para o conjunto de dados para fins de treinamento e 25% para o conjunto de testes. Esta divisão foi de grande importância, pois permitiu avaliar a eficácia do modelo em um cenário de teste independente e avaliar sua capacidade de generalização para novos dados.

#### 4.3.3 Técnicas e Métodos para Otimizar Parâmetros

Para a criação dos membros ARIMA e SVM, foi adotada uma estratégia diferente daquela utilizada pelo AG (LEE *et al.*, 2021). Em vez de usar o AG para ajustar os parâmetros desses modelos, métodos de busca aleatória e busca bayesiana foram empregados para SVM, e a função auto-ARIMA da biblioteca pm-ARIMA em Python, junto com o algoritmo de busca

em grade, foi usado para ARIMA (GÉRON, 2020). A Tabela 6 resume a otimização do hiperparâmetro.

Tabela 6. Aplicação de Hiperparâmetros

Otimização de hiperparâmetros	ARIMA	SVM	LSTM	MLP	Árvore de Decisão	Regressão Linear	Características
Auto ARIMA	✓	x	x	x	x	x	Determina os valores de P, D e Q; utiliza o critério de informação de Akaike (AIC) para escolher o melhor modelo.
Grid Search	✓	x	x	x	x	x	Maior foco nos erros; teste em todas as combinações possíveis; identifica o modelo ideal com base em erros.
Random Search	x	✓	x	x	x	x	Otimização C, Epsilon e Gama; teste em combinações diferentes, mas de forma aleatória; o objetivo é minimizar o tempo de execução.
Baves Search	x	✓	x	x	x	x	Otimização C, Epsilon e Gama; agiliza a busca, pois reaproveita informações em pontos de interações passadas;
Método fit	x	x	x	x	x	✓	Encontrar os melhores parâmetros que minimizem o erro entre as previsões do modelo e os valores reais no conjunto de treinamento.
Algoritmo Genético	x	x	✓	✓	x	x	Canônico com características de elitismo e seleção por torneio.
Ajustado manualmente	x	x	✓	✓	✓	x	Número de camadas, número de neurônios/células, atualização de pesos; tempos de treinamento, função de ativação da camada oculta e de saída, métricas de avaliação, profundidade da árvore e divisão de nós.

O símbolo “✓” denota a utilização do hiperparâmetro, enquanto “x” significa a sua não utilização.

#### 4.3.4 Métrica de avaliação

Para validar os modelos desenvolvidos foi utilizada a medida de desempenho estatístico denominada erro quadrático médio (MSE). Esta medida proporciona uma compreensão do comportamento do modelo em relação aos dados analisados, permitindo uma avaliação de quão bem ele pode fazer previsões precisas. Embora existam outras medidas de desempenho

disponíveis, o MSE foi considerado o mais adequado para este estudo devido às suas características benéficas para problemas de regressão, bem como às informações obtidas em (CARVALHO *et al.* 2011; BRADSHAW, 2018).

#### 4.3.5 Custo de Treinamento

Todos os experimentos foram conduzidos em um único sistema computacional equipado com processador Intel Core i5, 8 GB de RAM e armazenamento SSD de 128 GB. A escolha da linguagem de programação Python baseou-se em vários critérios, abrangendo a sua prevalência no campo da ciência de dados e aprendizagem automática, bem como a disponibilidade de um extenso conjunto de bibliotecas e ferramentas especializadas, simplificando a implementação e análise dos modelos. Essas bibliotecas incluem o Pandas, *Scikit-Learn*, *Matplotlib*, *Pmdarima.arima* e *Keras*.

O Pandas é uma biblioteca para a manipulação e análise de dados em Python, fornecendo estruturas de dados flexíveis e eficientes para trabalhar com dados tabulares. Ele permite a limpeza, transformação e manipulação de dados, além de oferecer recursos para combinar e juntar conjuntos de dados (PANDAS DOCUMENTATION, 2023).

O *Scikit-Learn* é uma das bibliotecas mais usadas para aprendizado de máquina em Python. Ele oferece uma ampla gama de algoritmos de AM supervisionado e não supervisionado, além de ferramentas para pré-processamento de dados, validação cruzada, seleção de modelos e muito mais.

*Scikit-Optimize (skopt)* é uma biblioteca de otimização de hiperparâmetros para Python que fornece ferramentas eficazes para buscar os melhores hiperparâmetros para um modelo de AM. O *skopt* utiliza algoritmos de otimização Bayesianos para buscar a combinação mais adequada de hiperparâmetros para um modelo (SCIKIT-LEARN DOCUMENTATION, 2023).

O *Matplotlib* é uma biblioteca de visualização de dados em Python, que permite a criação de gráficos e visualizações de alta qualidade. Ele oferece uma ampla gama de estilos e opções de personalização para criar visualizações de dados que atendam às necessidades do usuário (MATPLOTLIB DOCUMENTATION, 2023).

A biblioteca *Pmdarima* implementa uma interface de usuário amigável para ajustar e prever modelos ARIMA, com recursos para lidar com diferentes tipos de dados, como séries temporais sazonais e exógenas. A biblioteca também possui ferramentas para diagnóstico de

modelo, como visualização de resíduos e testes estatísticos (PMDARIMA DOCUMENTATION, 2023).

Por fim, o *Keras* é uma biblioteca de AM de alto nível, que oferece suporte à criação de redes neurais profundas (*Deep Learning*) e outras arquiteturas de modelo de Aprendizado de Máquina. Ele é compatível com várias outras bibliotecas populares de AM em Python, como *TensorFlow* e *Theano*, permitindo que os usuários escolham a melhor ferramenta para o seu trabalho. Com sua facilidade de uso e documentação abrangente, *Keras* é amplamente utilizado por pesquisadores e profissionais de todo o mundo para a construção de modelos de AM em Python. (KERAS DOCUMENTATION, 2023).

O tempo médio de treinamento dos modelos foi de aproximadamente dois dias para algoritmos individuais, cinco dias para comitê de máquinas e nove dias para a hibridização entre aprendizado profundo, aprimorado por meio de hiperparâmetros otimizados via AG canônico.

#### **4.4 Considerações Finais**

No presente capítulo, estabeleceu uma base fundamental para a compreensão do processo de criação do modelo "Redes Neurais Artificiais com Hiperparâmetros Otimizados pelo Algoritmo Genético" e "Comitê de Máquinas" para previsão de algumas variáveis do sistema elétrico brasileiro. Foram discutidas as definições dos parâmetros que influenciam no comportamento das técnicas de IA, bem como aspectos gerais e relevantes dos modelos propostos, tais como base de dados, pré-processamento, técnicas e métodos para otimização de parâmetros, métricas de avaliação e custo de treinamento. A seguir, serão apresentados os resultados e as discussões decorrentes da utilização dos modelos preditivos.

## 5 RESULTADOS E DISCUSSÕES

Neste capítulo, após a separação das bases de dados e os ajustes dos sinais de entrada, são apresentados os resultados provenientes dos seguintes preditores:

- a) *Perceptron Multicamadas* (MLP);
- b) *Long Short Term Memory* (LSTM);
- c) *Autoregressive Integrated Moving Average* (ARIMA);
- d) *Support Vector Machine* (SVM);
- e) Combinação do Algoritmo Genético e MLP;
- f) Combinação do Algoritmo Genético e LSTM;
- g) Comitê de Máquinas (“ensemble 01”): MLP, Árvore de Decisão, Regressão Linear e SVM);
- h) Comitê de Máquinas (“ensemble 02”):MLP, LSTM, SVM e ARIMA;

### 5.1 Resultados Baselines

Como ponto de referência de modelos estabelecidos e aprovados, as *baselines* são consideradas uma medida importante ou fundamental na avaliação de novas abordagens de AM. Elas fornecem uma referência objetiva para avaliar o desempenho em relação aos modelos existentes, mostrando o rendimento mínimo que uma nova abordagem deve superar para ser considerada uma melhoria.

No contexto do modelo preditivo em questão, foram considerados como baselines os métodos ARIMA, MLP, LSTM, SVM e a combinação do AG+MLP. Além disso, foi avaliada a combinação AG+LSTM com Ensemble como uma nova abordagem no cenário da previsão de algumas variáveis do setor elétrico do Brasil. Ao comparar os resultados obtidos pela nova abordagem com os resultados das *baselines* estabelecidas, é possível avaliar a eficácia da nova abordagem e identificar suas vantagens e desvantagens em relação aos métodos existentes.

Para a criação das baselines ARIMA e SVM ao invés de utilizar o AG para otimizar os parâmetros desses modelos, foram utilizados métodos de busca aleatória (*Random Search*) e busca bayesiana (*Bayesian Search*) para o SVM, e a função auto-arima da biblioteca pm-darima do Python juntamente com o algoritmo de busca em grade (*Grid-search*) para o ARIMA.

Para as baselines, foram apresentados resultados de previsão para 1, 3, 8 e 12 passos à frente. No entanto, cabe ressaltar que as análises foram concentradas nos modelos AG + MLP e AG + LSTM, deixando de lado a apresentação isolada de MLP e LSTM para esses diferentes horizontes de previsão.

### 5.1.1 RNA (MLP e LSTM)

Na tabela 7, tem-se a definição da estrutura e configuração das RNAs que basicamente é segmentada em definição da arquitetura da rede apropriada à aplicação, determinação da topologia da rede a ser utilizada, determinação de parâmetros do algoritmo de treinamento e funções de ativação.

Tabela 7 - Configurações das RNAs

<b>Configuração</b>	<b>MLP</b>	<b>LSTM</b>
Atualização dos pesos	Otimizador RMSProp	Otimizador RMSProp
A primeira camada escondida	30 neurônios e método <i>Dropout</i> com 5%;	30 neurônios e método <i>Dropout</i> com 5%;
A segunda camada escondida	25 neurônios e método <i>Dropout</i> com 0%	25 neurônios e método <i>Dropout</i> com 0%
A terceira camada escondida com	25 neurônios e método <i>Dropout</i> com 5%	25 neurônios e método <i>Dropout</i> com 5%
A camada de saída	1 neurônio	1 neurônio
Definição manual - épocas para o treinamento	100	100
Função de ativação não-linear em todas as camadas escondidas	“relu”	“relu”
Função de ativação na camada de saída	Linear	Linear

#### 5.1.1.1 Resultados *Baselines* para o PLD

O valor do PLD (Preço de Liquidação das Diferenças) é um fator crítico no mercado de energia elétrica brasileiro, pois representa o preço da energia no mercado spot. Os modelos de previsão de PLD utilizam uma variedade de sinais de entrada exógenas para fazer previsões precisas. Dois desses sinais de entrada comuns são a Geração Hidráulica (GH) e térmica (GT), que representam a quantidade de energia gerada por meio de usinas hidrelétricas e termelétricas,

respectivamente. A Carga de Energia (CE) é outro sinal de entrada importante, pois representa a demanda de energia elétrica em um determinado momento.

Os Reservatórios EARM (Energia Armazenada) e ENA (Energia Natural Afluente) também são sinais de entrada relevantes, pois fornecem informações sobre a quantidade de água armazenada nos reservatórios e a quantidade de água que está fluindo naturalmente para esses reservatórios.

A Tabela 8 reúne os resultados da diferença entre a previsão das RNAs *baselines* e o valor desejado (erro mínimo).

Tabela 8- Resultados para o PLD - Medida de Erro MSE

Rede Neural Artificial	Erro	Nordeste	Norte	Sudeste	Sul
<b>MLP</b>	<b>MSE</b>	0,00522	0,00242	0,00278	0,00271
<b>LSTM</b>		0,00725	0,00268	0,00190	0,00847

Como observado na Tabela 8, os resultados são satisfatórios e com base nos valores do MSE do PLD. Por região, quem apresentou melhores resultados foi a rede MLP exceto na região Sudeste.

Nas Figuras 28.a e 28.b, é evidente a convergência da RNA para o conjunto de treinamento na Figura 28.a, e o desempenho da RNA treinada para o conjunto de validação na Figura 28.b. Esses resultados correspondem ao modelo que alcançou a melhor métrica de avaliação (menor MSE) entre todas as regiões, conforme apresentado na Tabela 8.

Figura 28.a. PLD: Convergência da Rede LSTM para a Região Sudeste (eixo x=Ciclo da rede neural; eixo y= MSE; linha tracejada vertical amarela= indica em qual ciclo foi encontrado o melhor modelo de rede neural da validação)

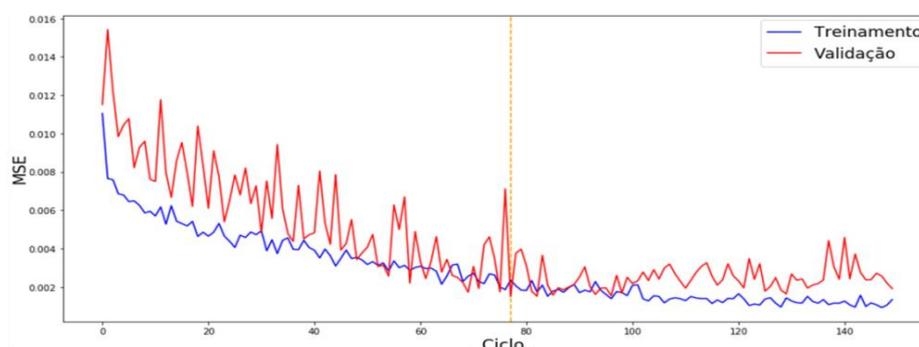
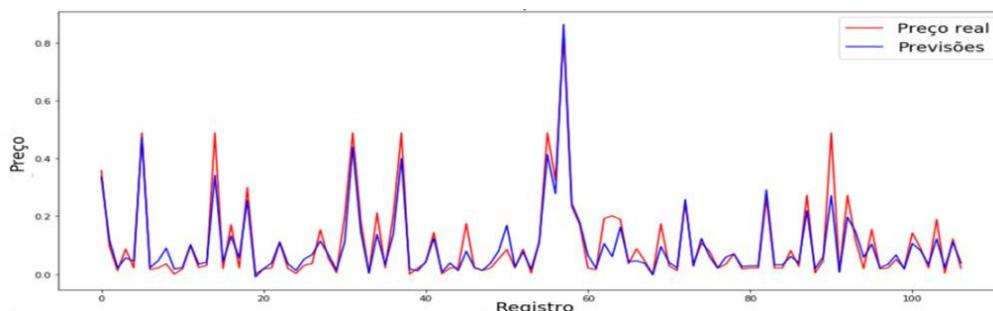


Figura 28.b. PLD: Previsão da rede LSTM para Região Sudeste (eixo x= Registro; eixo y= Preço)



Na Figura 28.a, é possível observar que o erro entre os dados reais (azul) e as previsões (vermelho) apresenta valores baixos. Para alcançar um bom desempenho no treinamento, utilizou-se o hiperparâmetro *Early Stopping* ("Parada Antecipada") com um valor de "min\_delta" de  $1e-10$ . Esse hiperparâmetro permite interromper o treinamento quando a taxa de erro na classificação nos dados de treinamento se mantém continuamente acima do valor estabelecido por um determinado período (15 ciclos).

A linha tracejada vertical em amarelo indica em qual ciclo foi encontrado o melhor modelo de rede neural da validação. A partir desse ponto, o modelo passa a ter um maior erro de generalização. Essa técnica permite evitar o *overfitting*, que é quando o modelo se ajusta demais aos dados de treinamento e perde a capacidade de generalizar para novos dados. Dessa forma, pode-se garantir que o modelo terá uma boa performance em novos dados e situações.

A Figura 28.b ilustra o comportamento da previsão do melhor modelo encontrado durante a convergência da LSTM para o conjunto de dados destinado à validação, o qual nunca foi visto pelo modelo anteriormente. No eixo x tem-se o registro de tempo e no eixo y temos o preço previsto. Pode-se observar que o modelo foi capaz de fazer previsões precisas, apresentando uma boa correspondência entre os dados reais (vermelho) e as previsões (azul). Essa previsão é extremamente importante para a tomada de decisão em muitos setores, principalmente no mercado financeiro, permitindo que os investidores se preparem para as flutuações futuras do preço.

### 5.1.1.2 Resultados *Baselines* para Velocidade do Vento

A Tabela 9 reúne os resultados de treino para os dois modelos de RNA.

Tabela 9- Resultados para Velocidade do Vento - Medida de Erro MSE

Rede Neural	Erro	Macau	Petrolina
MLP	MSE	0,01476	0,00861
LSTM		0,01844	0,00840

O estudo comparou o desempenho dessas técnicas na previsão da velocidade do vento em duas cidades diferentes, Macau e Petrolina. Os resultados mostraram que, para a cidade de Macau, a rede MLP foi a mais eficaz na previsão da velocidade do vento. Já para a cidade de Petrolina, a LSTM foi um pouco melhor na previsão da velocidade do vento.

### 5.1.2 ARIMA

O processo de construção do modelo ARIMA consiste em aplicar algumas ações essenciais para obtenção do modelo mais adequado para previsão da série temporal.

Neste sentido foi aplicado o teste de estacionariedade de *Dickey-Fuller*, função presente na biblioteca *Statsmodels* do Python, para determinar se a média e a variância são constantes e comprovar se as características dos dados condizem com os padrões de uma série estacionária (SALES *et al.*, 2018). Consoante os ajustes realizados nos dados, foi possível analisar o comportamento e posteriormente criar as previsões da base amostral. A Tabela 10 exhibe os resultados do teste de *Dickey-Fuller* para a região Norte e Macau. A seleção dessas áreas entre outras regiões como Sul, Sudeste, Nordeste e Petrolina foi feita com o objetivo de facilitar futuras comparações entre os resultados da linha de base e os modelos aprimorados. Além disso, essa escolha visa estabelecer uma relação clara de aprimoramento ao longo do tempo.

Tabela 10 – Resultados do Teste Dickey-Fuller

	Teste estatístico	P-Value	Valores Críticos		
			1%	5%	10%
PLD	-4,458580	0,000233	-3,445794	-2,868349	-2,570397
Vento	-5,147908	0,000011	-3,431694	-2,862134	-2,567086

Com base no teste de estacionariedade observado na Tabela 6, o *p-value* está abaixo dos 5% nas duas variáveis analisadas, e além disso, o teste estatístico é inferior aos valores críticos 1%, 5% e 10%, nesse caso, rejeitando a hipótese nula para série temporal não estacionária, visto que o teste assumiu a hipótese alternativa que indica a série como estacionária (SILVA, 2022). Mediante isso, dispensa-se a transformação dos dados para ajustar o modelo, o que foi sustentado com os resultados da otimização de hiperparâmetros para o *ARIMA*, haja vista, que as melhores configurações encontradas não trouxeram diferenciações no parâmetro “D”, responsável pela transformação da série temporal.

Considerou-se o método tradicional da estatística Box-Jenkins proposta por George Box e Gwilym Jenkins, que consiste na utilização de modelos Auto-Regressivos Integrados de Médias Móveis (*ARIMA(p,d,q)*) para prever as séries temporais. O modelo *ARIMA* requer a definição dos parâmetros “p” responsáveis pela ordem do modelo autorregressivo, “d” para a quantidade de diferenciações necessárias para tornar a série estacionária e “q” que concerne a parte da definição da ordem de médias móveis do modelo de séries temporais (BORSATO, 2019).

O processo de definição dos parâmetros para avaliar o modelo e facilitar na busca por configurações que apresentam a melhor performance para os dados apresentados ocorreram de forma automática por meio das técnicas de “estimação de parâmetros” como *auto-arima* (função da biblioteca *pm-darima* do Python) e algoritmo de pesquisa em grade (*Grid-search*), utilizado para testar combinações exaustivas de parâmetros frente uma métrica de erro.

O *auto-arima* busca encontrar o conjunto de parâmetros que melhor se ajusta a um modelo *ARIMA*, nesse caso, tenta determinar a ordem de diferenciações para o parâmetro “d” e testar os valores de “p” e “q” para determinar os cálculos de *Akaike Information Criterion* (AIC) em cada combinação de parâmetros (AWAN, 2020).

O algoritmo *Grid-search* consiste em definir uma quantidade de valores para os parâmetros e iterar a maior quantidade possível de combinações a fim de identificar aquela com o menor erro (KIRCHOFF, 2019).

A partir dos modelos encontrados na etapa de estimação dos parâmetros, o que acontece em dois momentos, a previsão dentro da amostra e a previsão fora da mostra a saber:

Dentro da amostra: avaliar a eficiência dos modelos gerados pelo auto-arima e o *grid-search*, logo, a combinação de parâmetros que obtiver o melhor desempenho aos dados, nesta etapa, será selecionado como o modelo ideal para criar as previsões fora da amostra.

Fora da amostra: Após modelo selecionado, a base de dados estará segmentada com 75% de dados para treinamento e 25% para teste a qual o modelo não conhece para validar o modelo.

Algumas técnicas foram utilizadas para “estender as previsões além da base amostral” como o algoritmo de validação Walk-Forward com as abordagens *Expanding Window* (EW) que possibilita adicionar novos valores a base de treinamento conforme o conjunto de teste seja incrementado e conseqüentemente expandido o horizonte de treinamento à cada valor predito. O *Sliding Window* (SW) funciona de forma similar ao EW, porém remove da janela de treinamento os valores mais antigos e por conseguinte trabalha com faixas de valores mais recentes (KIRCHOFF, 2019).

Diante do exposto, a Tabela 11 apresenta os erros de previsão, Erro Médio Quadrático (MSE), para avaliação dos modelos encontrados durante a estimação dos parâmetros.

Tabela 11 – Estimação dos Parâmetros

Dados	AUTO- ARIMA		GRID-SEARCH	
	Vento Macau	PLD Norte	Vento Macau	PLD Norte
Modelo	(10,0,5)	(1,0,1)	(23,0,25)	(16,0,24)
MSE	1,050	1.830,600	1,036	1.778,597

Conforme Tabela 11, os modelos ARIMA (23,0,25) para a velocidade do vento e o ARIMA (16,0,24) para o PLD obtiveram os menores erros e melhores ajustes na previsão dentro da amostra com a técnica *Grid-Search*.

Na seleção dos melhores modelos, foi criado previsões fora da amostra com 1, 3, 8 e 12 passos à frente em ambas as bases de dados, isto é, as previsões são extrapoladas com um salto de até 12 dias a frente para velocidade do vento e 12 semanas para o PLD.

As Tabelas 12 e 13 apresentam as eficiências dos modelos ao prever fora da amostra.

Tabela 12 – Erros de Previsão para Velocidade do Vento - Macau - ARIMA (23,0,25)

	Expanding Window (EW)	Sliding Window (SW)
Passos	MSE	
1	0,177	0,177
3	0,813	0,869
8	0,313	0,335
12	1,668	1,671

Tabela 13 – Erros de Previsão para o PLD - Região Norte - ARIMA (16,0,24)

	Expanding Window (EW)	Sliding Window (SW)
Passos	MSE	
1	80,398	80,398
3	33,797	102,629
8	740,740	2.971,923
12	2.972,642	2.384,905

Ressalta-se que os resultados apresentados nas Tabelas 12 e 13 tiveram as validações *Walk-Forward* nas abordagens EW e SW. Além da presença de *outliers* na base de dados sobre o PLD, existente entre os períodos de 2001 e 2002 e no início de 2008, sendo estes responsáveis por aumentar os erros obtidos durante a previsão. Diferente dos dados sobre a velocidade do vento que apresenta valores mais comportados.

As Figuras 29 e 30 apresentam o comportamento da previsão da velocidade do vento em Macau e do PLD da região Norte, ambas com 12 passos à frente.

Figura 29 – Previsão do ARIMA com 12 passos à frente – Velocidade do Vento

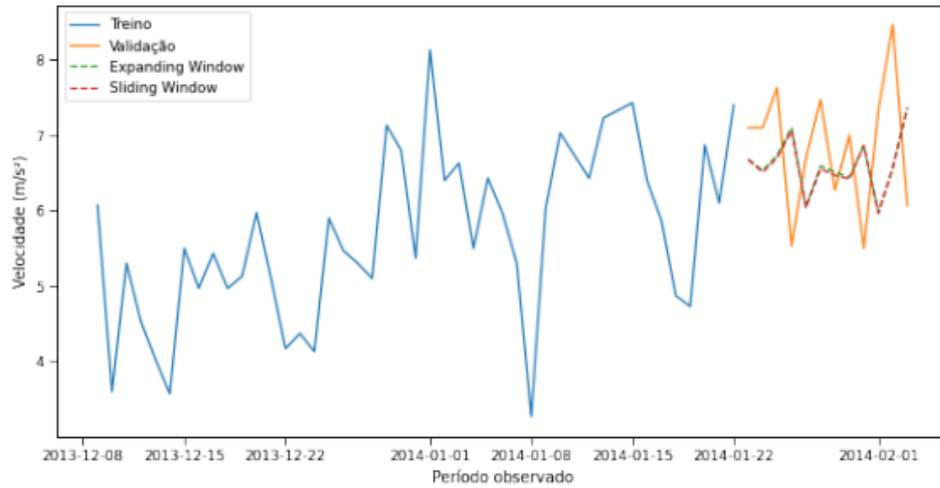
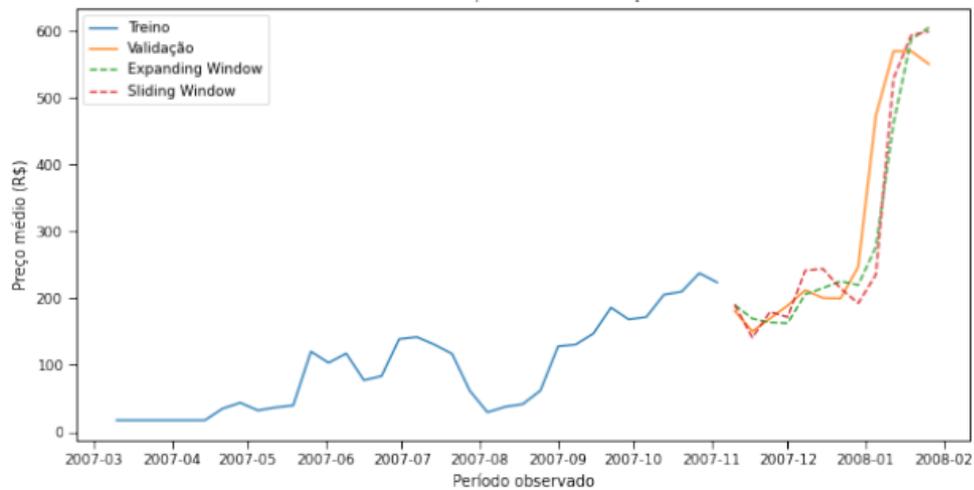


Figura 30 – Previsão do ARIMA com 12 passos à frente – PLD



Como observado nas Figuras 29 e 30, quanto ao desempenho tem-se pouca diferença nas duas abordagens, porém os resultados da previsão com aplicação do EW obteve-se os melhores resultados nos passos 3 e 8, e um erro um pouco elevado nos 12 passos à frente em uma análise comparativa com SW.

### 5.1.3 SVM

No que diz respeito a abordagem de Máquinas de Vetores de Suporte (SVM) foram aplicados alguns hiperparâmetros como o Epsilon, que controla a largura da via, além daqueles já utilizados para predição, como C, responsável por equilibrar a largura das vias com os limites de violação das margens, e o Kernel, que ajusta as previsões para resolver problemas não

lineares (GÉRON, 2019). Para tanto, para auxiliar no processo de definição dos parâmetros/configurações ideais foram utilizados os métodos *Random Search* e *Bayesian Search* das bibliotecas *sklearn* e *skopt*. Em conjunto, essas bibliotecas podem ser usadas para criar modelos de AM de alto desempenho com hiperparâmetros bem ajustados e otimizados. O *skopt* é especialmente útil quando a otimização de hiperparâmetros é um desafio, uma vez que pode lidar com espaços de busca de hiperparâmetros complexos e alta dimensionalidade.

Segundo (NEVES, 2020), o *Random Search* gera de forma aleatória os conjuntos de hiperparâmetros, sem a necessidade de testar todas as combinações para minimizar o tempo de execução do algoritmo. Por outro lado, o *Bayes Search* agiliza esse processo ao reutilizar as informações nos pontos de iterações passadas, ou seja, o método melhora a busca de hiperparâmetros a partir dos resultados de combinações nas execuções anteriores (ALVES *et al.*, 2022). Em ambos os métodos, os hiperparâmetros do modelo SVM otimizados foram, C, Epsilon e Gamma.

Destaca-se que a técnica de *Feature Engineering* foi utilizada para construir e desenvolver o modelo de AM. Essa técnica foi responsável por produzir novas variáveis de entrada com atrasos em relação à variável de saída em seu tempo atual. Esses atrasos foram definidos como  $X(t-1)$ ,  $X(t-2)$ ,  $X(t-3)$ ...  $X(t-n)$ , conforme recomendado por Brownlee (2017). Dessa forma, as variáveis remodeladas foram utilizadas como recursos de entrada  $X$  pelo modelo para aprender seus padrões e comportamentos, a fim de prever suas saídas  $Y$ . Para isso, os modelos foram ajustados usando tantos métodos diretos, que fazem uso de todas as entradas disponíveis, quanto métodos iterativos, que criam previsões em várias etapas de tempo, conforme sugerido por Lim (2021).

A Figura 31 apresenta uma amostra da base de dados de vento com o objetivo de demonstrar o comportamento da variável "velocidade" em 2004, juntamente com a aplicação do método *Feature Engineering* para os atrasos  $t1$  e  $t2$ .

Figura 31 - Atrasos de um e dois Períodos da Variável "Velocidade" – Velocidade do Vento

Dados Original		Atrasos		
Data	Velocidade	t-2	t-1	Alvo
01/01/2004	6,3			6,3
02/01/2004	4,6		6,3	4,6
03/01/2004	5,5	6,3	4,6	5,5
04/01/2004	8,0	4,6	5,5	8,0
05/01/2004	6,7	5,5	8,0	6,7
06/01/2004	5,8	8,0	6,7	5,8

No caso da Figura 31, as novas variáveis foram criadas a partir de atrasos de um e dois períodos da variável "velocidade". Essas novas variáveis são úteis para capturar tendências e padrões ao longo do tempo e podem ser utilizadas como entrada para modelos preditivos.

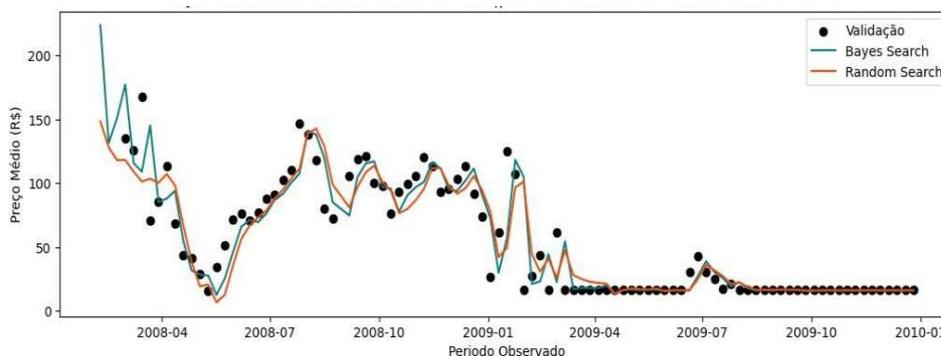
Os atrasos utilizados foram de 12 para o PLD e 88 para a velocidade do vento. Embora não havendo uma tabela com os atrasos separados, eles foram obtidos por meio de um algoritmo que testou diferentes números de atrasos. Os valores obtidos foram utilizados como base para os modelos de otimização de hiperparâmetros. Vale ressaltar que a quantidade ideal de atrasos pode variar dependendo do tamanho da amostra.

A Tabela 14 apresenta os resultados da otimização na base de dados do PLD e Figura 32 a previsão do PLD da região Norte.

Tabela 14. Resultados da Otimização de Hiperparâmetros PLD – Região Norte – SVM

Métodos	Hiperparâmetros			MSE
	C	Epsilon	Gamma	
Bayes Search (BS)	1000	1	9.70e-07	597.152
Random Search (RS)	70.645	0.230	4.64e-06	548.750

Figura 32. Comparação dos Resultados da Previsão Direta para o Preço Médio do PLD com os Métodos *Bayes Search* e *Random Search*.



Conforme a Tabela 14 os melhores ajustes foram encontrados com testes nas seguintes combinações de C, Epsilon e Gamma, com destaque para o método *Random Search*, que nesse caso, performou melhor na análise dos erros da previsão. Tendo em vista o resultado do método *Random Search*, a Tabela 15 apresenta os erros da previsão com 1, 3, 8 e 12 passos.

Tabela 15- Resultados das Previsões Random Search PLD - Região Norte - SVM

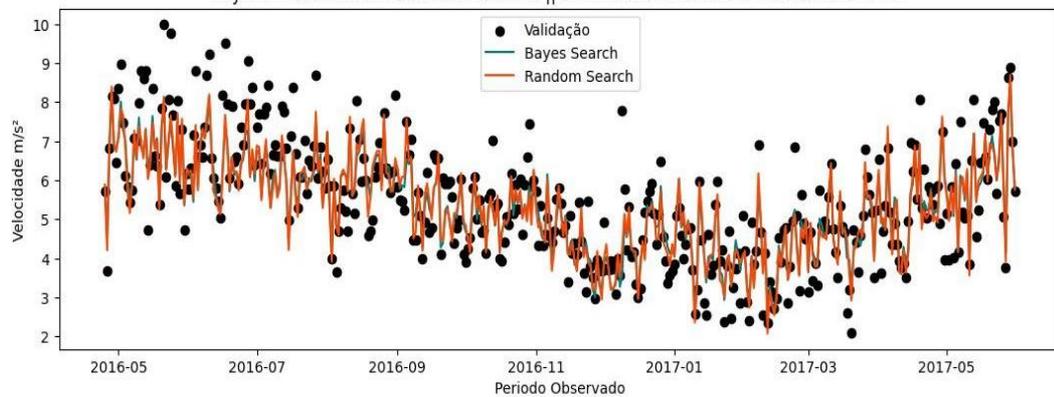
<b>SVM - Random Search - PLD</b>	
<b>Passos</b>	<b>MSE</b>
1	543.180
3	504.466
8	51473.100
12	21401.665

Referente os resultados da Velocidade do Vento - Macau (Ver Tabela 16 e Figura 33)

Tabela 16. Resultados da Otimização de Hiperparâmetros Velocidade do Vento - Macau - SVM

<b>Métodos</b>	<b>Hiperparâmetros</b>		<b>MSE</b>
	<b>C</b>	<b>Epsilon</b>	
<i>Bayes Search</i> (BS)	5.202	0.061	0.548
<i>Random Search</i> (RS)	31.623	1e-05	0.559

Figura 33. Comparação dos Resultados da Previsão para Velocidade do Vento com os Métodos *Bayes Search* e *Random Search*.



Na tabela 16, é possível analisar os erros de previsão com a métrica MSE para selecionar a configuração que melhor se ajustou aos dados, nesse caso *Bayes Search*. Além da observação da “ausência” de dois parâmetros do modelo SVM que não sofreram ajustes, sendo eles, Gamma que assume o valor padrão do modelo, que originalmente é definido como "scale" e o parâmetro Kernel, que possibilita a separação dos dados não linearmente separáveis, segundo (ACOSTA, 2021), assim, ele recebe como valor padrão *Radial Basis Function* (RBF), nesse caso, a otimização do parâmetro em ambos os métodos acrescia o tempo que os algoritmos ficam em execução, visto que ele requer um poder computacional maior que os demais parâmetros otimizados, como ressalta (GÉRON, 2019), e por esse motivo, optou-se por não o otimizar.

Na Tabela 17, são apresentados os resultados obtidos durante a previsão na base de dados da velocidade do vento.

Tabela 17. Resultados das Previsões *Bayes Search* - Velocidade do Vento – Macau - SVM

SVM – <i>Bayes Search</i> -Velocidade do Vento	
Passos	MSE
1	0.214
3	1.460
8	0372
12	0.369

## 5.2 Resultados das Combinações AG + LSTM e AG + MLP

### 5.2.1 Resultados da Combinação para o PLD

A seguir apresenta-se os resultados, ver Tabela 18, do desenvolvimento do modelo combinando AG + LSTM. Para fins de comparação com a abordagem proposta foi acrescentado resultados da combinação do AG + MLP.

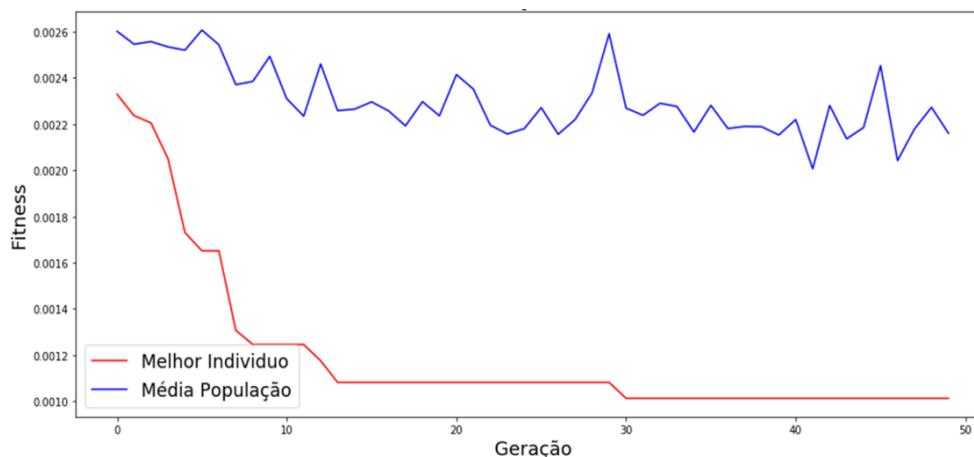
Tabela 18: Resultados da Combinação AG + Rede Neural para o PLD

Rede Neural + AG	Nordeste	Norte	Sudeste	Sul
MLP + AG	0.00267	0.00183	0.001024	0.00137
LSTM + AG	0.00260	0.001012	0.001015	0.00108

Ao se comparar os resultados da proposição do trabalho verifica-se que o modelo com a rede LSTM + AG foram satisfatórios com melhores resultados para o PLD nas quatro regiões.

O resultado da evolução do LSTM + AG para a região Norte é apresentado na Figura 34 (cor vermelha: curva do melhor indivíduo; cor azul: aptidão média de população por geração).

Figura 34 – Configurações ótimas do AG + LSTM Para a Previsão do PLD da Região Norte



Na Figura 34, pode-se observar que a aptidão da melhor configuração da LSTM melhora rapidamente (o erro de previsão cai rapidamente) e a partir da geração 10 se inicia a

convergência do AG para um ponto ótimo do espaço de busca das melhores configurações para a LSTM.

As principais características do melhor modelo ilustrado pela Figura 34 são resumidas na transcrição do genoma abaixo:

- Transcrição Genoma: [4, 46, 0.01, 57, 0.0, 2, 0.11, 8, 0.03, 0.00101].

A melhor solução foi obtida com arquitetura neural das quais 4 camadas escondidas, com 46, 57, 2 e 8 neurônios e com *Dropout* de 0.01, 0.0, 0.11 e 0.03 respectivamente, correspondendo a um erro previsão de 0.00101.

- **PLD: 12 passos à frente**

A Tabela 19 apresenta os resultados do modelo AG + LSTM quanto ao critério do erro médio quadrático- MSE para região Norte.

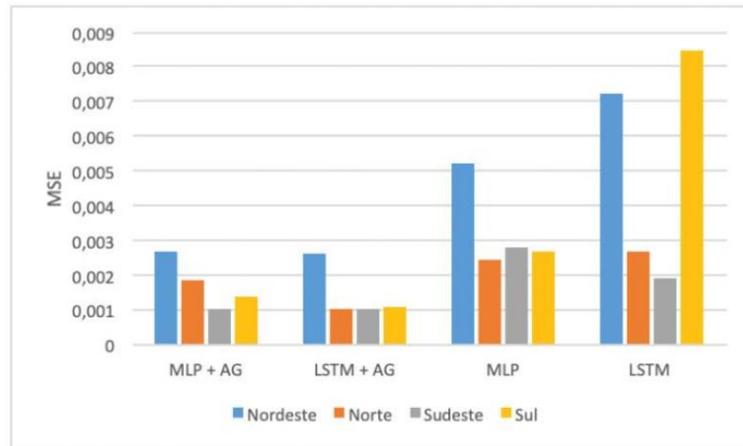
Tabela 19: MSE para até 12 Passos à Frente de Predições - Região Norte

Região Norte	1 passo	3 passos	8 passos	12 passos
<b>AG + LSTM</b>	0.00101	0.00217	0.00165	0.00174
<b>AG + MLP</b>	0.00183	0.00229	0.00160	0.00199

Os resultados mostram que o modelo híbrido proposto, AG+LSTM, apresenta menor erro para os 1, 3 e 12 passos. Além disso, o desempenho dos modelos comparados se deteriora conforme o horizonte temporal de previsão aumenta. No entanto, o horizonte de tempo de 12 semanas à frente é o mais apropriado para as práticas de gestão no mercado brasileiro.

Com base nos resultados obtidos nos experimentos. A Figura 35 mostra todos os modelos neurais com os seus erros mínimos mensurados referente ao PLD com um passo à frente.

Figura 35 Gráfico do MSE para a Predição do PLD



### 5.2.2 Resultados da Combinação para Velocidade do Vento

Na Tabela 20, pode-se observar os resultados do experimento para base de dados velocidade do vento.

Tabela 20: Resultado da Combinação AG + Rede Neural para Velocidade do Vento

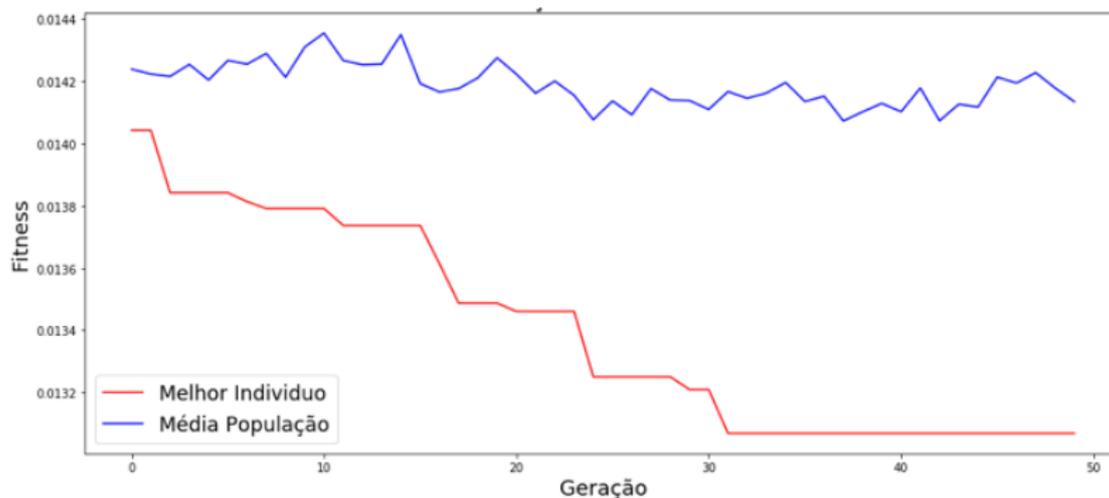
Rede Neural + AG	Macau	Petrolina
MLP + AG	0.01429	0.00808
LSTM + AG	0.01306	0.00815

Para velocidade do vento, a LSTM + AG também apresentou melhor resultado para cidade de Macau e tendo um MSE um pouco menor do que a MLP + AG.

A estratégia de utilizar o MLP + AG também apresentou resultados satisfatórios tendo valores de MSE um pouco menor do que a LSTM + AG em Petrolina.

O resultado da evolução do LSTM + AG para a cidade de Macau é apresentado na Figura 36.

Figura 36 – Configurações ótimas do AG + LSTM para a Previsão da Velocidade do Vento – Parque Eólico da Cidade de Macau



Pode-se observar na Figura 36 que desde as primeiras gerações a evolução das melhores configurações para a LSTM é pequena, com a convergência acontecendo um pouco depois da geração 30.

As principais características do melhor modelo ilustrada pela Figura 36 são resumidas na transcrição do genoma abaixo:

- Transcrição Genoma: [5, 5, 0.03, 45, 0.09, 36, 0.14, 63, 0.04, 2, 0.02, 0.01306]

A melhor solução foi obtida com arquitetura neural das quais 5 camadas escondidas, com 5, 45, 36, 63 e 2 neurônios e com Dropout de 0.03, 0.09, 0.14, 0.04 e 0.02 respectivamente, correspondendo a um erro previsão de 0.01306.

- **Velocidade do Vento: 12 passos à frente**

A Tabela 21 apresenta os resultados obtidos com o modelo AG + LSTM para a base de dados Macau.

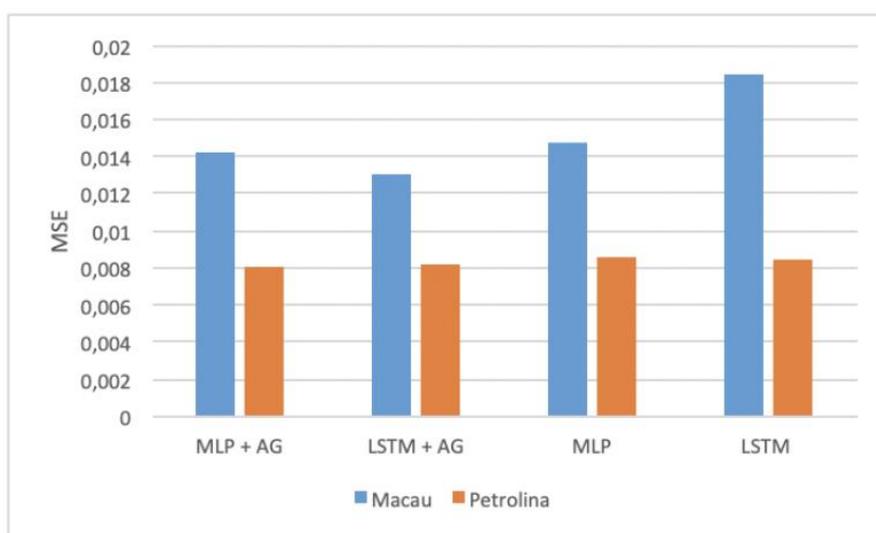
Tabela 21 - Resultados da combinação AG + Rede Neural para Velocidade do Vento - Macau

Cidade Macau	1 passo	3 passo	8 passos	12 passos
<b>AG + LSTM</b>	0.01306	0.02402	0.02209	0.02305
<b>AG + MLP</b>	0.01429	0.02606	0.02301	0.02520

O comportamento do modelo híbrido também foi testado em horizontes temporais superiores para cidade de Macau. Os resultados indicam que a técnica híbrida proposta é o melhor modelo em todos os horizontes de predição, porém, é notável que a precisão diminui a partir do terceiro passo à frente, embora permaneça razoável.

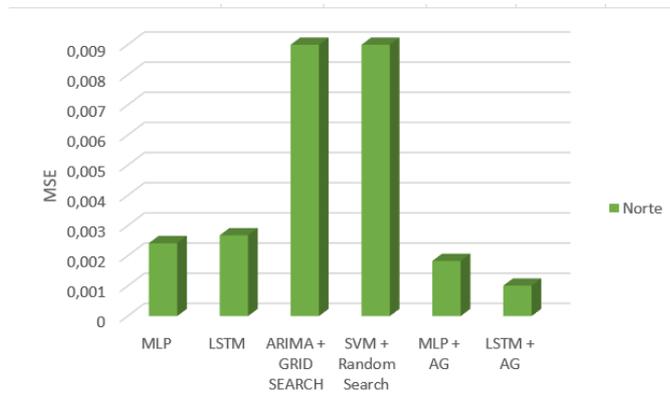
A Figura 37 apresenta os resultados obtidos (erro mínimo) dos modelos para Velocidade do Vento respectivamente para um passo à frente.

Figura 37 Gráfico do MSE para Predição da Velocidade do Vento



A seguir, são apresentadas as Figuras 38 e 39, que ilustram o cenário de todas as previsões realizadas pelos modelos para a região norte considerando um passo à frente. Além disso, é apresentada uma Figura para a velocidade do vento em Macau, visando a análise pontual do desempenho dos modelos.

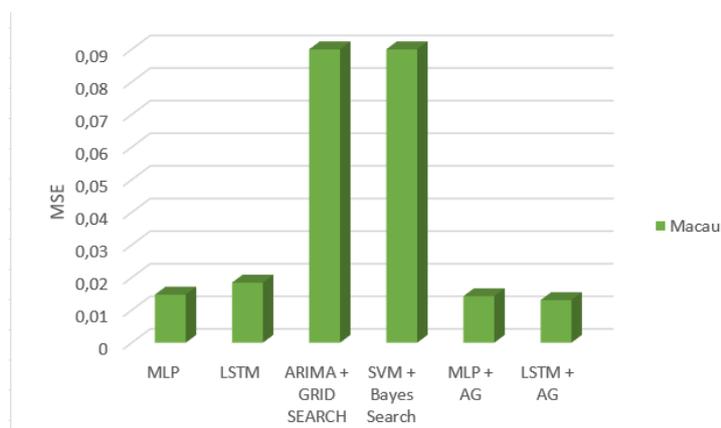
Figura 38 Gráfico do MSE para Predição dos Melhores Resultados do PLD



Os resultados obtidos nas análises revelaram que o modelo LSTM + AG obteve o melhor desempenho em relação aos demais modelos avaliados. O segundo melhor resultado foi do modelo MLP + AG, com uma diferença “considerável” em relação ao primeiro colocado. O modelo MLP apresentou um pequeno ganho sobre o LSTM, mas ainda assim ficou atrás do LSTM + AG e do MLP + AG.

Por outro lado, os resultados do modelo ARIMA + GRID SEARCH e SVM + RANDOM SEARCH foram significativamente mais altos, indicando uma menor capacidade de previsão em relação aos modelos combinados com o Algoritmo Genético

Figura 39 Gráfico do MSE para Predição dos Melhores Resultados da Velocidade do Vento



O modelo LSTM + AG obteve o melhor desempenho, ou seja, considerado o modelo com a melhor previsão para os dados da velocidade do vento. O segundo melhor resultado foi

do modelo MLP + AG, que apresentou uma pequena diferença em relação ao MLP. Surpreendentemente, o modelo LSTM obteve o quarto pior resultado dentre os modelos avaliados. Os resultados dos modelos *Grid Search* e *SVM + Bayes Search* apresentaram valores bem maiores e parecidos.

### 5.3 Resultados dos Comitê de Máquinas

Nesta seção, são apresentados os resultados comparativos entre "ensemble 01" e "ensemble 02" em relação à sua eficácia na previsão de séries temporais do PLD e velocidade do vento. Além disso, são abordados os resultados resultantes da substituição do componente SVM como uma estratégia de aprimoramento, destacando o impacto que a substituição de um componente do modelo tem sobre o resultado.

Foram analisados dois métodos de votação: votação média (VOA) e votação média ponderada (VOWA). No caso da VOA, os pesos são equivalentes e iguais a um, e o valor predito final é obtido calculando a média das previsões geradas pelos modelos de AM individuais (ver Equação (7)):

$$\hat{y} = \sum_{j=1}^m \hat{y}_j / m \quad 7$$

Onde  $m$  é o número de algoritmos de AM usados no ensemble,  $\hat{y}_j$  representa o valor previsto pelo algoritmo  $j$  e  $\hat{y}$  é o valor previsto final. No entanto, uma desvantagem do método de VOA é que todos os modelos do ensemble são considerados igualmente eficientes, sem levar em conta possíveis diferenças de desempenho entre eles.

O método VOWA, especifica um coeficiente de peso para cada membro do ensemble. Esse peso pode ser um número de ponto flutuante entre 0 e 1, onde a soma de todos os pesos é igual a 1, ou pode ser um número inteiro de 1, indicando o número de votos atribuídos ao respectivo membro do ensemble. O valor previsto final é obtido conforme mostrado na Equação (8), onde  $w_j$  representa o peso do algoritmo  $j$ .

$$\hat{y} = \frac{\sum_{j=1}^m (w_j \hat{y}_j)}{\sum_{j=1}^m w_j} \quad 8$$

Em relação aos resultados do ensemble com VOWA, a distribuição de pesos seguiu uma alocação de votos inteiros conforme indicado na Tabela 22. O modelo com melhor desempenho recebeu o maior número de votos, enquanto o modelo com o menor desempenho recebeu apenas um voto.

Tabela 22. Pesos dos Membros Ensemble - VOWA.

	<b>MLP</b>	<b>Árvore de Decisão</b>	<b>Regressão Linear</b>	<b>SVM</b>
“Ensemble 01”	4	3	2	1
	<b>MLP</b>	<b>LSTM</b>	<b>SVM</b>	<b>ARIMA</b>
“Ensemble 02”	4	3	2	1

A seleção de pesos como números inteiros e a votação no método VOWA são justificadas por sua interpretação mais intuitiva, pois refletem uma contagem direta, facilitando a compreensão dos resultados. Além da facilidade de implementação do método, tornando-o mais direto e menos complexo do que trabalhar com valores fracionários, também é mais fácil rastrear e entender como cada membro do grupo contribui para a decisão final. Expressar a força de uma preferência de forma discreta também é mais direto.

Considerando as informações fornecidas sobre os métodos de votação, os resultados obtidos usando as abordagens VOA e VOWA serão apresentados.

### 5.3.1 Resultados Comitê de Máquinas para PLD – Região Norte

Os resultados obtidos pelo conjunto são visíveis nas Tabelas 23 e 24.

Tabela 23. Resultados Comparativos dos Membros, “Ensemble 01” com VOA e “Ensemble 01” com VOWA – Região Norte.

<b>Métrica de Avaliação</b>	<b>Árvore de Decisão</b>	<b>MLP</b>	<b>SVM</b>	<b>Regressão Linear</b>	<b>“Ensemble 01” com VOA</b>	<b>“Ensemble 01” com VOWA</b>
MSE	0,00220	0,002941	0,74488	0,002353	0,18823	0,15082

Tabela 24. Resultados Comparativos dos Membros, “Ensemble 02” com VOA e “Ensemble 02” com VOWA – Região Norte.

<b>Métrica de Avaliação</b>	<b>MLP</b>	<b>LSTM</b>	<b>SVM</b>	<b>ARIMA (16, 0, 24)</b>	<b>“Ensemble 02” com VOA</b>	<b>“Ensemble 02” com VOWA</b>
MSE	0,00233	0,00252	0,78235	0,09720	0,22115	0,16823

Nas Tabelas 23 e 24, pode-se encontrar os resultados dos quatro componentes individualmente, bem como os resultados obtidos por meio da combinação usando os métodos de votação VOA e VOWA.

A votação ponderada, que atribui votos a cada modelo com base em seu desempenho, mostra uma melhoria nos resultados, como evidenciado na tabela acima.

É notável que no "ensemble 01", três dos membros demonstraram capacidade superior em aproximar a série PLD, enquanto o SVM exibiu o desempenho menos satisfatório.

Tanto "ensemble 01" quanto "ensemble 02" exibem resultados satisfatórios, mas não conseguem superar o desempenho individual dos membros, exceto pelo SVM.

É evidente que, na prática, o principal objetivo é obter um ganho de desempenho em comparação com o melhor classificador disponível quando considerado isoladamente. No entanto, essa melhoria pode não necessariamente resultar em taxas de precisão excepcionalmente altas e pode não ser capaz de superar o desempenho dos modelos individuais devido à "limitação da média".

Para explicar melhor, o desempenho do ensemble é, em última análise, uma combinação média ou ponderada das previsões dos modelos individuais. Se um dos modelos individuais, como o SVM, consistentemente gera previsões "imprecisas ou incorretas", esses resultados desfavoráveis eventualmente terão um impacto negativo no desempenho geral do ensemble.

Para resolver essa questão, o membro SVM no "ensemble 01" foi substituído pelo modelo ARIMA. No "ensemble 02", o SVM foi substituído pelo modelo de regressão linear (consulte as Tabelas 25 e 26).

Tabela 25. Resultados do “Ensemble 01” com a Substituição do SVM pelo ARIMA – Região Norte.

<b>Métrica de Avaliação</b>	<b>“Ensemble 01” com VOA</b>	<b>“Ensemble 01” com VOWA</b>
MSE	0,02617	0,01308

Tabela 26. Resultados do “Ensemble 02” com a Substituição do SVM pela Regressão Linear - Região Norte.

Métrica de Avaliação	“Ensemble 02” com VOA	“Ensemble 02” com VOWA
MSE	0,02610	0,01185

Existem inúmeros outros cenários nos quais a aplicação de ensembles pode ser altamente vantajosa. No entanto, abordar esses cenários diversos requer um entendimento profundo das condições que devem estar presentes para garantir o sucesso de uma abordagem de ensemble. Identificar essas condições às vezes pode não ser uma tarefa direta, então é uma sugestão prática aplicar ensembles e comparar o desempenho resultante com o obtido a partir de soluções individuais.

### 5.3.2 Resultados Comitê de Máquinas para Velocidade do Vento – Região Macau

Aqui estão os erros relacionados às previsões obtidas pelo ensemble. As Tabelas 27 e 28 contêm os detalhes desses resultados.

Tabela 27. Resultados Comparativos dos Membros “Ensemble 01” com VOA e “Ensemble 01” com VOWA—Macau.

Métrica de Avaliação	Árvore de Decisão	MLP	SVM	Regressão Linear	“Ensemble 01” com VOA	“Ensemble 01” com VOWA
MSE	0,11348	0,11731	0,85393	0,11235	0,30056	0,18988

Tabela 28. Resultados Comparativos dos Membros “Ensemble 02” com VOA e “Ensemble 02” com VOWA - Macau.

Métrica de Avaliação	MLP	LSTM	SVM	ARIMA (23, 0, 25)	“Ensemble 02” com VOA	“Ensemble 02” com VOWA
MSE	0,01348	0,01685	0,22247	0,16629	0,10561	0,07191

Os resultados dos membros indicam um desempenho satisfatório na previsão da velocidade do vento.

A inclusão do modelo SVM na votação final do ensemble pode introduzir viés nos resultados, com a VOA mostrando resultados satisfatórios, mas não superando o desempenho individual demonstrado.

Abaixo, as Tabelas 29 e 30 exibem resultados com a substituição do componente SVM, nomeadamente, no "ensemble 01", composto por árvores de decisão, MLP, ARIMA e regressão linear, e no "ensemble 02", composto por MLP, LSTM, regressão linear e ARIMA.

Tabela 29. Resultados do “Ensemble 01” com a Substituição do SVM pelo ARIMA—Macau.

Métrica de Avaliação	“Ensemble 01” com VOA	“Ensemble 01” com VOWA
MSE	0,12696	0,11797

Tabela 30. Resultados do “ensemble 02” com a substituição do SVM pela Regressão Linear – Macau

Métrica de Avaliação	“Ensemble 02” com VOA	“Ensemble 02” com VOWA
MSE	0,07640	0,04494

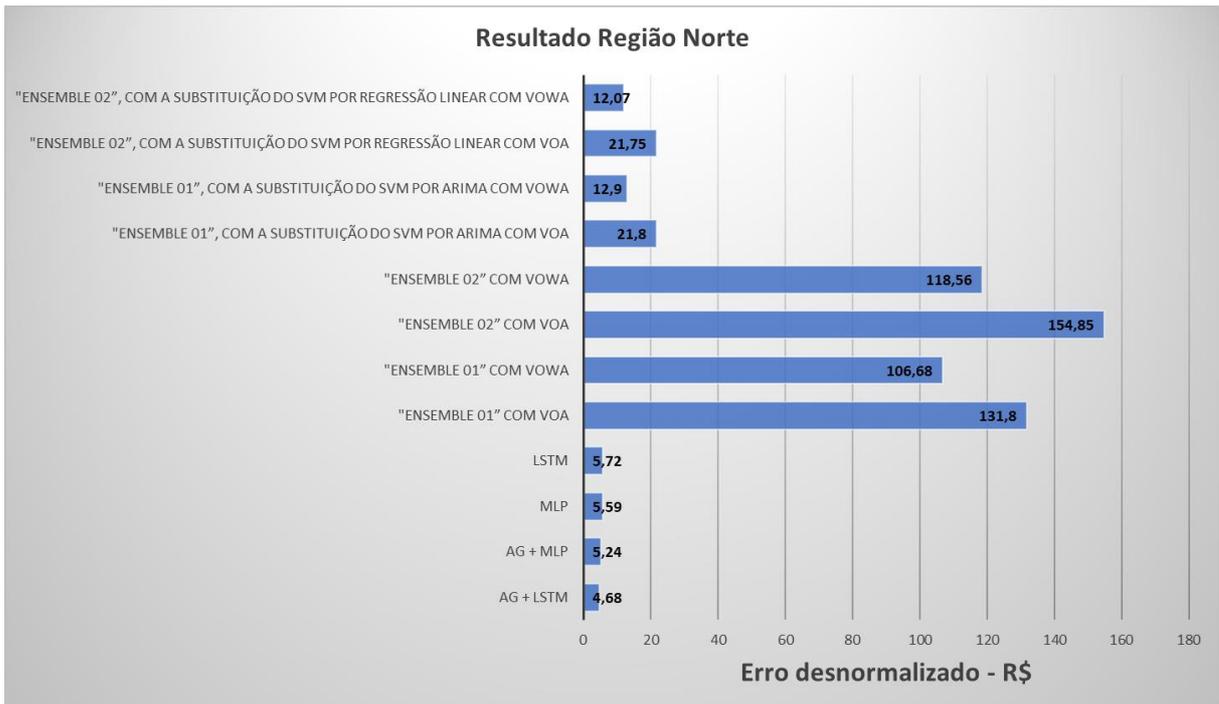
Com base nas tabelas acima, é evidente que formar o ensemble substituindo o membro SVM resultou em previsões mais satisfatórias, tanto com VOA quanto com VOWA. Além disso, os valores se aproximaram dos resultados individuais e, em alguns casos, superaram o desempenho do modelo ARIMA.

#### **5.4 Comparação dos Resultados das Combinações AG + LSTM e AG + MLP com Comitê de Máquinas.**

As Figuras 40 e 41 compilam os melhores resultados considerando a métrica de avaliação MSE para a previsão de um passo à frente dos modelos apresentados na pesquisa. Esses resultados são para os conjuntos de dados de PLD da região norte e velocidade do vento em Macau.

Os resultados dos membros individuais LSTM e MLP foram incluídos nas Figuras 40 e 41 como parte de uma linha de base com hiperparâmetros ajustados sem o uso do AG (parâmetros definidos na Tabela 3). Esta linha de base integra a técnica do AG para 'selecionar' e 'ajustar' os hiperparâmetros desses algoritmos, visando aprimorar a eficiência do modelo.

Figura 40. Gráfico MSE para Prever os Melhores Resultados de PLD

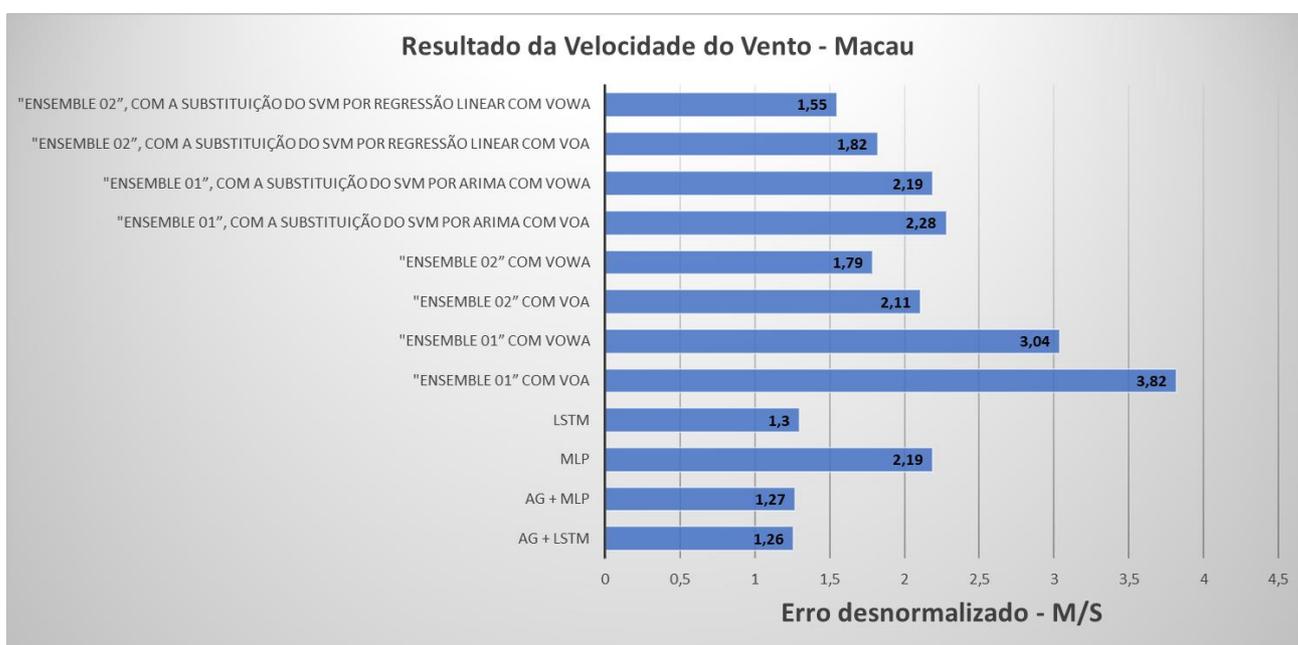


Para uma melhor compreensão, os valores foram "desnormalizados" utilizando um mínimo de R\$ 4,00 e um máximo de R\$ 684,00 como referência.

Os resultados obtidos na análise revelaram que o modelo AG + LSTM alcançou o melhor desempenho em comparação com os outros modelos avaliados (Figura 40). O segundo melhor resultado foi do modelo AG + MLP, com uma leve diferença em relação ao melhor desempenho.

Os ensembles que não utilizaram SVM demonstraram melhores desempenhos. Ao comparar diferentes métodos de votação, foi observado que o VOWA alcançou o melhor desempenho.

Figura 41. Gráfico MSE para Previsão dos Melhores Resultados de Velocidade do Vento.



Para uma melhor compreensão, os valores foram "desnormalizados" utilizando um mínimo de 1,15 m/s e um máximo de 10,07 m/s como referências.

Conforme evidenciado na Figura 41, o modelo AG + LSTM demonstrou um desempenho "superior", com uma margem de apenas 0,01 m/s em comparação com o segundo melhor modelo e uma diferença distinta de 0,29 m/s em relação ao "ensemble 02", que ficou em terceiro lugar, no qual o SVM foi substituído pelo modelo de regressão linear.

Dentro dos resultados dos ensembles, o "ensemble 02", que substituiu o SVM pelo modelo de regressão linear no VOWA, superou não apenas todos os outros ensembles, mas também os desempenhos individuais dos membros que compunham o "ensemble 02".

As Tabelas 31 e 32 fornecem uma comparação do modelo proposto, AG+LSTM, para previsão em relação a outros modelos publicados para velocidade do vento e PLD.

A avaliação comparativa com outros modelos é realizada utilizando o critério do erro quadrático médio (RMSE) para fornecer uma interpretação mais intuitiva dos resultados. Isso é particularmente relevante, considerando que as referências (CARNEIRO, 2022; CAMELO *et al.*, 2017; BELENTANI, 2023; SANTOS *et al.*, 2022) apresentam suas métricas em RMSE. Assim, a análise é realizada após extrair a raiz quadrada dos valores de erro quadrático médio (MSE), visando a uma compreensão mais direta e uma comparação consistente com as referências mencionadas.

Tabela 31. Comparação do Modelo de Previsão Proposto com outros Modelos Publicados para Velocidade do Vento.

<b>Model</b>	<b>RMSE</b>
AG + LSTM	0,113
MLP (CARNEIRO, 2022)	0,893
CFBP (CARNEIRO, 2022)	0,448
RBF (CARNEIRO, 2022)	0,548
SOM (CARNEIRO, 2022)	0,713
Ensemble (CARNEIRO, 2022)	0,188
ARIMA (CAMELO <i>et al.</i> , 2017)	0,72
HW (CAMELO <i>et al.</i> , 2017)	0,62
RNA (CAMELO <i>et al.</i> , 2017)	0,52
ARIMA + RNA (CAMELO <i>et al.</i> , 2017)	0,69
HW + RNA (CAMELO <i>et al.</i> , 2017)	0,25

É crucial enfatizar que este estudo está focado na previsão de um passo à frente, com a velocidade do vento como a variável preditiva. Para comparação com pesquisas similares, a referência (CARNEIRO, 2022) utiliza uma média de janela de treinamento de quatro dias para previsão de vento, enquanto (CAMELO *et al.*, 2017) é dedicado à previsão de séries temporais de média mensal.

Tabela 32. Comparação do Modelo de Previsão Proposto com outros Modelos Publicados para o PLD.

<b>Model</b>	<b>RMSE</b>
AG + LSTM	1,12
LSTM (BELENTANI, 2023)	32,25
Árvore de Decisão (SANTOS <i>et al.</i> , 2022)	82,41

No contexto da Tabela 32, dois modelos foram referenciados: LSTM e Árvore de Decisão. Os resultados mostram valores de erro de 1,2, 32,25 e 82,41, respectivamente, para esses modelos. Cada um desses métodos exibe características distintas em seu desempenho, conforme detalhadamente discutido por Belentani (2023) e Santos *et al.* (2022). A análise desses resultados oferece compreensões valiosos sobre a adequação de cada abordagem no contexto específico da aplicação em questão.

### 5.5 Considerações Finais

Neste capítulo, foram apresentados os resultados obtidos a partir dos algoritmos em sua forma individual, aprimorados pela combinação das redes neurais artificiais profundas com a

sintonia dos algoritmos genéticos. Além disso, discutiu-se os resultados dos comitês de máquinas considerando duas combinações de votos (WO, WPO). Também foram discutidas diferentes estratégias para otimização dos parâmetros dos modelos em questão, incluindo a descrição das configurações dos modelos e os resultados dos hiperparâmetros, com foco específico no cenário da energia elétrica brasileira, destacando as bases de dados do PLD e da velocidade do vento. Destaca-se também a comparação do modelo de previsão proposto com outros modelos publicados, considerando métricas como o RMSE.

Ao encerrar este capítulo, é imperativo ressaltar a importância vital de uma perspectiva ambiental no que diz respeito ao uso eficiente dos recursos energéticos e à expansão da oferta de energia proveniente de fontes renováveis. Estamos diante de um momento crucial na história, onde a conscientização sobre as consequências da degradação ambiental causada pela geração de energia convencional atingiu um ponto crítico.

Nossa pesquisa, desde sua concepção, foi motivada pelo desejo de mitigar as emissões de gases de efeito estufa e reduzir os impactos negativos associados à produção de energia. Reconhecemos que a transição para fontes renováveis não é apenas uma opção, mas uma necessidade urgente para preservar o nosso planeta para as gerações futuras.

Ao adotar abordagens que priorizam a eficiência energética e a utilização de fontes renováveis, podemos não apenas reduzir a poluição e a degradação ambiental, mas também promover um desenvolvimento sustentável que beneficie tanto o meio ambiente quanto as comunidades humanas.

Nossas considerações finais reforçam a convicção de que, ao integrar práticas e tecnologias que respeitam os limites do nosso ecossistema, estamos pavimentando o caminho para um futuro mais promissor e resiliente.

## 6 Conclusão

Esta tese propõe uma metodologia que emprega uma variedade de algoritmos existentes, incluindo adaptações de alguns, para prever séries temporais relacionadas ao sistema elétrico brasileiro. Notavelmente, a combinação de aprendizagem profunda com hiperparâmetros otimizados através do algoritmo genético canônico, utilizando uma representação individual, produz bons resultados quando comparado aos demais métodos estudados.

O objetivo foi aumentar a precisão dos resultados por meio da aplicação de técnicas avançadas de aprendizado de máquina. Dessa forma, buscou-se alcançar previsões mais precisas e confiáveis do preço das diferenças de liquidação, que é um fator de referência para os preços no mercado livre de energia no Brasil e para o planejamento do sistema de energia renovável (previsão de velocidade do vento).

Os resultados obtidos revelaram que o modelo AG + LSTM obteve o melhor desempenho de previsão em comparação com os outros modelos, com um erro MSE de 4,68 no conjunto de dados do PLD e um erro de 1,26 no conjunto de dados de Macau.

Em relação à implementação do modelo de votação, que combina as previsões de vários modelos em uma única previsão final, foi avaliado usando duas abordagens distintas: média de votação (VOA) e média ponderada de votação (VOWA). Os resultados obtidos indicam que a VOWA apresenta desempenho de previsão superior em comparação com a VOA. Isso destaca a importância de levar em consideração a contribuição relativa de cada modelo para a previsão final.

Quando modelos individuais (membros do comitê) apresentam resultados superiores ao do comitê, essa discrepância pode ser atribuída a vários fatores. A fusão de modelos individuais em um comitê pode não capturar totalmente as nuances dos dados. Nesse contexto, ajustar os hiperparâmetros para melhor refletir o desempenho de cada modelo torna-se crucial para aprimorar o resultado geral do comitê, entre outras considerações. É essencial enfatizar a busca contínua pela melhoria do desempenho do comitê, seja por meio de ajustes nos pesos, otimização do treinamento ou consideração de fatores específicos dos modelos individuais que contribuem para o sucesso.

A hibridização do aprendizado profundo, aprimorada por hiperparâmetros otimizados por meio do algoritmo genético canônico e a implementação de comitês de máquinas revelaram

implicações práticas de grande relevância, juntamente com suas forças e fraquezas para consideração.

### **6.1 Implicações Práticas**

**Maior Eficiência na Solução de Problemas Complexos:** A combinação dessas técnicas possibilita um manuseio mais eficiente de problemas complexos. Redes neurais podem oferecer característica de aprendizado e generalização, algoritmos genéticos contribuem para otimização e busca eficiente de soluções, enquanto comitês de máquinas agregam perspectivas diversas e robustez na tomada de decisões.

**Adaptação a Diferentes Contextos:** A combinação desses métodos pode ser ajustada para se adequar a diferentes domínios ou contextos de aplicação. Isso permite a criação de sistemas mais flexíveis e adaptáveis capazes de lidar com variações e mudanças nos dados ou ambientes.

### **6.2 Pontos Fortes**

**Sinergia Entre Técnicas Diversas:** A complementaridade entre redes neurais, algoritmos genéticos e comitês de máquinas permite explorar sinergias entre essas abordagens, aproveitando suas respectivas vantagens e minimizando limitações individuais.

**Maior Precisão e Generalização Aprimorada:** A combinação dessas técnicas pode resultar em modelos mais precisos com melhores características de generalização, especialmente ao lidar com conjuntos de dados complexos.

**Exploração de Soluções Ótimas e Diversificadas:** Algoritmos genéticos podem explorar eficientemente um espaço de soluções, enquanto comitês de máquinas podem fornecer uma variedade de opiniões, aprimorando a tomada de decisões. Isso pode levar à descoberta de soluções ótimas e diversificadas.

### **6.3 Fraquezas**

**Complexidade Computacional e Custos:** A combinação de diferentes técnicas pode aumentar a complexidade dos modelos e exigir recursos computacionais mais robustos, resultando em tempos de processamento mais longos e custos associados.

**Dificuldade na Interpretabilidade:** Às vezes, a combinação de várias técnicas pode tornar o modelo resultante mais desafiador de interpretar e explicar como se chegou a determinado resultado.

**Desafios na Sintonia de Parâmetros:** Otimizar e ajustar os parâmetros dessas técnicas combinadas pode ser desafiador e pode exigir expertise significativa e tempo para encontrar configurações ideais.

Além disso, conclui-se que o uso de modelos preditores no setor elétrico é um tópico de grande importância, conforme evidenciado pela abundância de publicações dedicadas a esse tema. Esta pesquisa contribui para a análise do uso de várias técnicas de aprendizado de máquina no contexto brasileiro. Há uma tendência crescente no uso de modelos híbridos com o objetivo de obter previsões mais precisas. Esta abordagem pode ser mais uma alternativa para impulsionar a busca por soluções que minimizem impactos globais, incorporando otimização em vários aspectos, como eficiência e rentabilidade.

#### **6.4 Trabalhos Futuros**

Para trabalhos futuros, sugerimos expandir a pesquisa explorando outras abordagens que possam complementar e aprimorar os resultados obtidos até o momento. Enquanto nossa investigação se concentrou principalmente na hibridação do AG + LSTM, na diversificação das configurações dos hiperparâmetros dos modelos e na exploração de diferentes algoritmos como MLP, LSTM, ARIMA, SVM, AG + MLP e AG + LSTM e o Comitê de Máquinas, reconhecemos que há cenários que não foram explorados.

Um desses cenários é a aplicação de abordagens de subamostragem do espaço de entrada e seleção de conjuntos de dados com alguma interseção, juntamente com a subamostragem dos atributos, a aplicação da análise de resíduos nas "variáveis de atraso" para prever um passo à frente. Além disso, um caminho interessante para pesquisa futura seria a comparação entre a arquitetura dos transformers no contexto do Deep Learning e as técnicas aqui apresentadas.

Portanto, recomendamos que futuros estudos explorem essas técnicas adicionais, investigando seu impacto na performance e na estabilidade dos modelos em diferentes contextos e domínios de aplicação. Ao fazê-lo, podemos continuar avançando no desenvolvimento de soluções mais eficazes e abrangentes para os desafios presentes em aprendizado de máquina e análise de dados.

Além disso, sugerimos que futuros estudos explorem não apenas os parâmetros e hiperparâmetros dos modelos, mas também incorporem outras variáveis relevantes utilizadas na tomada de decisões no setor elétrico. Isso inclui a investigação do uso de diferentes regressores para estimar, explorar novas formas de hibridização entre regressores e considerar a correlação entre os membros dos comitês de máquinas para melhorar a performance.

## REFERÊNCIAS

- ANELL. **Agência Nacional de Energia Elétrica**, 2023. Preço de Liquidação de Diferenças (PLD). Disponível em: <https://www.gov.br/aneel/>. Acesso em: 11/04/2023
- Abedinia, O., Lotfi, M., Bagheri, M., Sobhani, B., Shafie-khah, M. and Catalão, J., **Improved EMD-Based Complex Prediction Model for Wind Power Forecasting**. IEEE Transactions on Sustainable Energy, vol. 11, pp. 2790-2802, 2020. (DOI 10.1109/TSTE.2020.2976038).
- Alsubaie, Mohammed *et al.* **Alzheimer's Disease Detection Using Deep Learning on Neuroimaging: A Systematic Review**. Mach. Learn. Knowl. Extr., 6(1), 464-505, 2024 (<https://doi.org/10.3390/make6010024>)
- Alencar, D. **Modelo Híbrido Baseado em Séries Temporais e Redes Neurais para Previsão da Geração de Energia Eólica**. In: Programa em Pós-Graduação em Engenharia Elétrica, pp. 20-24. Universidade Federal do Pará, Belém (2018).
- Araújo, Dhiego. **Utilização de um Perceptron Múltiplas Camadas na Aproximação de Funções Contínuas**, Applied Computer Engineering 2, Atena, 2022 (DOI:10.22533/at.ed.4482216038)
- Al Mamun, A.; Hoq, M.; Hossain, E.; Bayindir, R. **A Hybrid Deep Learning Model with Evolutionary Algorithm for Short-Term Load Forecasting**. In Proceedings of the 8th International Conference on Renewable Energy Research and Applications (ICRERA), Brasov, Romania, 3–6 November 2019. (<https://doi.org/10.1109/ICRERA47325.2019.8996550>)
- Arif Muhammad *et al.* **Classification of Anxiety Disorders using Machine Learning Methods: A Literature Review**. Insights of Biomedical Research, vol. 4, 2020 (ISSN: 2642-4576).
- Alymani, M.; Mengash, H.A.; Aljebreen, M.; Alasmari, N.; Allafi, R.; Alshahrani, H.; Elfaki, M.A.; Hamza, M.A.; Abdelmageed, A.A. **Sustainable Residential Building Energy Consumption Forecasting for Smart Cities using Optimal Weighted Voting Ensemble Learning**. Sustain. Energy Technol. Assess, 57, 103271, 2023 (<https://doi.org/10.1016/j.seta.2023.103271>)
- Awan, T. M. and Aslam, F. **Prediction of daily COVID-19 cases in European countries using Automatic ARIMA Model**. Journal of public health research, vol 9(3), 2020
- Alves, P. F, De Negri, J. A and Cavalcante, J. E. **Utilizando Aprendizado de Máquina para Estimação do Spread das Instituições Financeiras nos Empréstimos do BNDES**, 2022.
- Acosta, S. M. and Amoroso, A. L. **Aplicação da Regressão por Vetores de Relevância na Modelagem de um Processo Produtivo**. engenharia de produção: planejamento e controle da produção em foco, vol 1, 1(1), pp. 37-52, 2021.
- Agresti. Alan et al. Statistics: **The Art and Science of Learning from Data**, 5th edition, 5th edition, Pearson, 2021.

Edi, Semih *et al.* Assessing the Impact of Engineering Measures and Vegetation Restoration on Soil Erosion: A Case Study in Osmaniye, Türkiye. *Sustainability*, vol. 15(15), 12001, 2023 <https://doi.org/10.3390/su151512001>

Bendali, W.; Saber, I.; Bourachdi, B.; Boussetta, M.; Mourad, Y. **Deep Learning Using Genetic Algorithm Optimization for Short Term Solar Irradiance Forecasting**. In Proceedings of the Fourth International Conference on Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 21–23 October 2020. <https://doi.org/10.1109/ICDS50568.2020.9268682>.

Brownlee, Jason. **Deep Learning For Times Series Forecasting: Predict the Future with MLPs, CNNs, and LSTM in Python**. Machine Learning Mastery, 2020

Bradshaw, A. L. **Electricity Market Reforms and Renewable Energy: The Case of Wind and Solar in Brazil**. PhD Thesis, Columbia University, USA, 2018.

Borsato, R. e Corso, L.L. **Aplicação de Inteligência Artificial e ARIMA na Previsão de Demanda no Setor Metal Mecânico**. *Scientia cum Industria*, vol 7(2), pp. 165-176. 2019.

Belentani, Y. **Previsão de Preços de Energia no Mercado de Curto Prazo: uma análise combinada de séries temporais e redes neurais artificiais**. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, Brasil, 2023.

CNN Brasil, 2023. **Estudo quantifica efeitos das mudanças climáticas na sobrevivência de ursos polares**. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/>. Acesso em: 16/04/2024

CCEE. **Câmara de Comercialização de Energia Elétrica**, 2023. Análises do Comportamento dos Preços, Metodologias de Cálculo. Disponível em: <https://www.ccee.org.br/>. Acesso em: 11/04/2023

Cao, Luga *et al.* **Forecasting of Water Consumption by Integrating Spatial and Temporal Characteristics of Short-Term Water use in Cities**. *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 130, p. 103390, 2023 (<https://doi.org/10.1016/j.pce.2023.103390>)

Carneiro, Tatiane *et al.* **Ridge Regression Ensemble of Machine Learning Models Applied to Solar and Wind Forecasting in Brazil and Spain**. *Appl. Energy*, vol. 314, p. 118936. 2022. (<https://doi.org/10.1016/j.apenergy.2022.118936>).

Carvalho, André *et al.* **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. Rio de Janeiro: LTC, 2011. 394 p.

Chen, Wei *et al.* **Combining Evolutionary Algorithms and Machine Learning Models in Landslide Susceptibility Assessments**, vol. 12(23), 2020. (<https://doi.org/10.3390/rs12233854>).

Conte, Thiago. **Modelagem Neural da Resistência Elétrica dos Fornos de Redução do Alumínio**. 2015. 54 f. Dissertação (Mestrado em Computação Aplicada) - Programa em Pós Graduação em Engenharia Elétrica, Universidade Federal do Pará, Belém, 2015.

Chen, C., Liu, Hui. **Medium-Term Wind Power Forecasting based on Multi-Resolution Multi-Learner Ensemble and Adaptive Model Selection.** Energy Conversion and Management, vol. 206, 112492, 2020 (<https://doi.org/10.1016/j.enconman.2020.112492>)

Christoph, E.; Raviv, E.; Roetzer, G. **Forecast Combinations in R using the Forecast Comb Package.** RJ., 10, 262, 2019 (<https://doi.org/10.32614/RJ-2018-052>)

Camelo, H.; Lúcio, P.; Júnior, J.; Carvalho, P. **Métodos de Previsão de Séries Temporais e Modelagem Híbrida Aplicados à Velocidade Média Mensal do Vento para as Regiões Nordeste do Brasil.** Braz. Meteorol. Mag., 32, 565–574, 2017. (<https://doi.org/10.1590/0102-7786324005>)

Cao, Y.; Liu, G.; Luo, D.; Bavirisetti, D.P.; Xiao, G. **Multi-Timescale Photovoltaic Power Forecasting using an Improved Stacking Ensemble Algorithm based LSTM-Informer Model.** Energy, 283, 128669, 2023 (<https://doi.org/10.1016/j.energy.2023.128669>)

Cervantes, jair *et al.* **A comprehensive survey on Support Vector Machine Classification: Applications, challenges and Trends.** Neurocomputing, vol. 408, p. 189-215, setembro 2020. (<https://doi.org/10.1016/j.neucom.2019.10.118>).

Cao, Jian; LI, Zhi; LI, Jian. **Financial Time Series Forecasting Model based on CEEMDAN and LSTM.** Physica: Statistical Mechanics and its Applications, v. 519, p. 127-139, abril 2019.

Cemiloglu, Ahmed et al. **Support Vector Machine (SVM) Application for Uniaxial Compression Strength (UCS) Prediction: A Case Study for Maragheh Limestone.** Appl. Sci., 13(4), 2217, 2023 (<https://doi.org/10.3390/app13042217>)

Dasgupta, Aniruddha. **Ação pelo clima: 4 grandes questões para acompanhar em 2024,** 2024. Disponível em: <https://www.wribrasil.org.br/>. Acesso em: 11/04/2023.

DATA SCIENCE ACADEMY. **Deep Learning Book Brasil:** Tecnologias mais revolucionárias do nosso tempo. 2018. Disponível em: <https://www.datascienceacademy.com.br>. Acesso em: 15 agosto 2019.

Dumitru, C. and Gligor, A. **Daily Average Wind Energy Forecasting Using Artificial Neural Networks.** Procedia Engineering, vol. 181, pp. 829-836, 2017. (ISSN 1877-7058).

Dominguez *et al.* **Estimation of Linear Regression with the Dimensional Analysis Method.** Mathematics, 10(10), 1645, 2022 (<https://doi.org/10.3390/math10101645>)

Elsken, Thomaz, Metzen, Jan and Hutter, Frank. **Neural Architecture Search: A Survey.** Faouzi, johann. Time Series Classification: A review of Algorithms and Implementations. Machine Learning (Emerging Trends and Applications), 2022 (ISBN 978–1–8381524–1–3)

Forootan, Mohammad et al. **Machine Learning and Deep Learning in Energy Systems: A Review.** Annual Report 2022 released, vol 14 (8). 2022 (<https://doi.org/10.3390/su14084832>)

Fangyu, Z., Shen, L., Jie, Z., Zhang, W. and Liu, W. **A Sufficient Condition for Convergences of Adam and RMSProp.** IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June, Long Beach, CA, USA, 2019. (DOI: 10.1109/CVPR.2019.01).

Filho, J. C. Affonso and Oliveira, R. **Energy price prediction multi-step ahead using Hybrid Model in the Brazilian Market**. Electric Power Systems Research. vol. 117, pp. 115-122, 2014

G1, 2024. Meio Ambiente. Disponível em: <https://g1.globo.com/meio-ambiente>. Acesso em: 16/04/2024

GONTIJO, Marília. **A Produção Científica sobre Inteligência Artificial e seus Impactos: Análise de indicadores bibliométricos e altimétricos**. 2020. Dissertação (Mestrado em Ciência da Informação) - Programa em Pós Graduação em Gestão e Organização do Conhecimento, Universidade Federal de Minas, Belo Horizonte, 2020.

Goyal, Anil, Kundu, Amit. **ARIMA and Indian Stock Market Forecasting**. Architecture & Technology XII, p. 60-70, March 2020. (DOI:10.37896/JXAT12(s).03/807).

Gordim, João et al. **Support Vector Machines in Smile Detection: A comparison of auto-tuning standard processes in Gaussian kernel**. SBCOPENLIB. Anais do Workshop de Visão Computacional (WVC), 2021

Géron, A. **Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**, 2nd ed.; O'Reilly Media: Sebastopol, CA, USA, 2020; pp. 62–70.

Géron, A. **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow**. Alta Books, 2019.

Hassanien, Aboul *et al.* **Deep Learning: Research and Applications**. De Gruyter, 2020. 161 p.

Hernandez, Alejandro, Nieuwenhuys, Ineke and Rojas, Sebastian. **A Survey on Multi-Objective Hyperparameter Optimization Algorithms for Machine Learning**. Springer, December 2022.

Heydari, A., Nezhad, M., Pirshayan, E., Garcia, D., Keynia, F. and De Santoli, L. **Short-term Electricity Price and Load Forecasting in Isolated Power Grids based on Composite Neural Network and Gravitational Search Optimization Algorithm**, Applied Energy, vol. 277, pp. 115-503, 2020. (<https://doi.org/10.1016/j.apenergy.2020.115503>).

Huang, C.; Karimi, H.R.; Mei, P.; Yang, D.; Shi, Q. **Evolving Long Short-Term Memory Neural Network for Wind Speed Forecasting**. Inf. Sci. 2023, 632, 390–410. (<https://doi.org/10.1016/j.ins.2023.03.031>)

Hahn, Yannik et al. **Time Series Dataset Survey for Forecasting with Deep Learning**. MDPI, vol 5(1), March 2023 (<https://doi.org/10.3390/forecast5010017>)

HAYKIN, Simon. **Neural Networks and Learning Machines**. 3 ed. Canada: Prentice Hall, 2008. 906 p.

Horak, Jakub, Vrbka, Jaromir and Suler, Petr. **Support Vector Machine Methods and Artificial Neural Networks Used for the Development of Bankruptcy Prediction Models and their Comparison**. vol. 13(3), p. 60, march 2020. (<https://doi.org/10.3390/jrfm13030060>)

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D. and Zareipour, H. **Energy Forecasting: A Review and Outlook**, IEEE Open Access Journal of Power and Energy, Early Access Article, 2020. (DOI:10.1109/OAJPE.2020.3029979).

Iliadis, L., Papaleonidas, A., Angelov, P., Jayne, C. **Artificial Neural Networks and Machine Learning**. ICANN, v. 4, September 26-29, 2023.

Izidio, D.M.F.; de Mattos Neto, P.S.G.; Barbosa, L.; de Oliveira, J.F.L.; Marinho, M.H.d.N.; Rissi, G.F. **Evolutionary Hybrid System for Energy Consumption Forecasting for Smart Meters**. Energies, 14, 1794, 2021 (<https://doi.org/10.3390/en14071794>)

Junior, K. **Genetic Algorithms and Deep Learning Based on Long Short-Term Memory (LSTM) Recurrent Neural Networks for Medical Diagnosis Assistance**. Doctoral Thesis, University of São Paulo, Ribeirão Preto, SP, Brazil, 2023. (<https://doi.org/10.11606/D.59.2023.tde-20062023-152537>)

Kramar, Vadim; Alchakov, Vasily. **Time-Series Forecasting of Seasonal Data Using Machine Learning Methods**, Algorithms, vol. 16(5), p. 248, 2023 (<https://doi.org/10.3390/a16050248>)

KERAS DOCUMENTATION. **Keras: The Python Deep Learning Library**. 2023. Disponível em: <https://keras.io>. Acesso em: 18 abril 2023.

Kirchoff, D. F. **Avaliação de Técnicas de Aprendizado de Máquina para Previsão de Cargas de Trabalho Aplicadas para Otimizar o Provisionamento de Recursos em Nuvens Computacionais**. Dissertação de Mestrado - Programa de Pós-Graduação em Ciência da Computação, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2019.

Kochenderfer, Mykel; Wheeler Tim. **Algorithms for Optimization**. MIT Press, 2019.

Ludermir, Teresa. **Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências**. Scielo - Estudos Avançados, vol. 35 (101), 2021. (DOI: <https://doi.org/10.1590/s0103-4014.2021.35101.007>)

Lim, Bryan and Zohren, Stefan. **Time-series Forecasting with Deep Learning: a survey**. Royal Society, vol. 379, 2021 (ISSN 1471-2962).

Luo, X., Oyedele, L., Ajayi, A., Akinade, O., Delgado, J., Owolabi, H. and Ahmed, A. **Genetic Algorithm-Determined Deep Feedforward Neural Network Architecture for Predicting Electricity Consumption in Real Buildings**. Energy and AI, vol. 2, 2020. (<https://doi.org/10.1016/j.egyai.2020.100015>).

Li, W.; Zang, C.; Liu, D.; Zeng, P. **Short-term Load Forecasting of Long-short Term Memory Neural Network Based on Genetic Algorithm**. In Proceedings of the 4th IEEE Conference on Energy Internet and Energy System Integration (EI2), Wuhan, China, 30 October–1 November 2020. (<https://doi.org/10.1109/EI250167.2020.9346907>)

Liu, Peiqi. **Time Series Forecasting Based on ARIMA and LSTM**. Advances in Economics, Business and Management Research, vol 656, 2022.

- Lv, S.; Wang, L.; Wang, S. **A Hybrid Neural Network Model for Short-Term Wind Speed Forecasting**, *Energies*, 16, 1841, 2023 (<https://doi.org/10.3390/en16041841>)
- Lee, S.; Kim, J.; Kang, H.; Kang, D. **Genetic Algorithm Based Deep Learning Neural Network Structure and Hyperparameter Optimization**. *Appl. Sci.*, 11, 774, 2021(<https://doi.org/10.3390/app11020744>)
- Lim, B. and Zohren, S. **Time-series forecasting with deep learning: a survey**. *Philosophical Transactions of the Royal Society A*, vol. 379(2194), 20200209, 2021.
- MCTI. **Estimativas Anuais de Emissões de Gases de Efeito Estufa no Brasil**, 6º ed., 2022.
- MME. **Ministério de Minas e Energia**, 2023. Preço de Liquidação de Diferenças (PLD). Disponível em : <https://www.gov.br/mme/>. Acesso em: 11/04/2023
- MATPLOTLIB DOCUMENTATION. **Matplotlib: users guide**. 2023. Disponível em: <https://keras.io>. Acesso em: 18 abril 2023.
- Kotyrba, Martin *et al.* **The Influence of Genetic Algorithms on Learning Possibilities of Artificial Neural Networks**. *Computers*, vol. 11(5), 70, 2022 (<https://doi.org/10.3390/computers11050070>)
- Nielsen, Ailenn. **Análise Prática de Séries Temporais: Predição com Estatística e Aprendizado de Máquina**. Alta Books, 1 ed, 2021
- NEVES, J. M. M. **Otimização de Hiperparâmetros em Machine Learning utilizando uma Surrogate e Algoritmos Evolutivos** (Bachelor's thesis, Universidade Tecnológica Federal do Paraná), 2020.
- Otosaka, I. *et al.* **Mass balance of the Greenland and Antarctic ice sheets from 1992 to 2020**. *Earth System Science Data*, 15(4), 1597-1616, 2023 (<https://doi.org/10.5194/essd-15-1597-2023>).
- ONU Programa para o meio ambiente. **O oceano está mais quente do que nunca; veja por quê**, 2023. Disponível: <https://www.unep.org/pt-br/>. Acesso em: 16/04/2024
- Ozcanli, A.; Yaprakdal, F.; Baysal, M. **Deep learning Methods and Applications for Electrical Power Systems: A comprehensive review**. *Energy Res.*, 44, 7136–7157, 2020 (<https://doi.org/10.1002/er.5331>).
- OLIVEIRA, Mônica de. **Reconhecimento DE padrões em Biomecânica de Corrida usando Aprendizado de Máquina**. 2018. 61 f. Trabalho de Conclusão de Curso (Engenharia de Computação) - Universidade Federal de Ouro Preto, João Monlevade, 2018.
- Ospina, Raydonal *et al.* **An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil**. *Mathematics*, 11(14), 3069; 2023 (<https://doi.org/10.3390/math11143069>)
- Priscilla, Basseto. **Análise Comparativa de Modelos Lineares para Previsão do Preço do Açúcar**. Dissertação (Mestrado em Engenharia da Produção) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2022.

Papastefanopoulos, V., Linardatos, P., Kotsiantis, S. **COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population**, Applied Sciences, vol. 10(11), 3880, 2020 (<https://doi.org/10.3390/app10113880>)

PUHEIM, M. and MADARÁSZ, L. **Normalization of inputs and outputs of neural network based robotic arm controller in role of inverse kinematic model**. IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMi), January, Herl'any, Slovakia, 2014. (DOI: 10.1109/SAMI.2014.6822439)

PANDAS DOCUMENTATION. **Panda**: users guide 2023. Disponível em: <https://keras.io>. Acesso em: 18 abril 2023.

PMDARIMA DOCUMENTATION. **Pmdarima**: users guide 2023. Disponível em: <https://pypi.org/project/pmdarima/>. Acesso em: 18 abril 2023.

Ribeiro, M.; da Silva, G.; Canton, C.; Fraccanabbi, N.; Mariani, C.; Coelho, S. **Electricity Energy Price Forecasting based on Hybrid multi-stage Heterogeneous Ensemble: Brazilian commercial and residential cases**. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020. (<https://doi.org/10.1109/IJCNN48605.2020.9206691>)

SEEG. **Análise das emissões brasileiras de Gases de Efeito Estufa e suas aplicações para as metas climáticas do Brasil (1970-2020)**, outubro 2021

Sagi, O.; Rokach, L. **Ensemble Learning: A survey**. WIREs Data Mining Knowl Discov. vol. 8, p. e1249, 2018. (<https://doi.org/10.1002/widm.1249>).

Shahid, F.; Zameer, A.; Muneeb, M. **A Novel Genetic LSTM Model for Wind Power Forecast**. Energy 2021, 223, 120069. (<https://doi.org/10.1016/j.energy.2021.120069>).

Shi, J.; Teh, J. **Load Forecasting for Regional Integrated Energy System based on Complementary Ensemble Empirical Mode de Composition and Multi-Model Fusion**. Appl. Energy, 353, 122146, 2024 (<https://doi.org/10.1016/j.apenergy.2023.122146>)

SIVANANDAM, S. N.; DEEPA, S.N. **Introduction to Genetic Algorithms**. 1 ed., Springer-Verlag Berlin, p. 442, 2008

SCIKIT-LEARN DOCUMENTATION. **Scikit-Learn**: Machine Learning in Python. 2023. Disponível em: <https://keras.io>. Acesso em: 18 abril 2023.

Sales, J., Silva, J., Dias, V., Firmino, P. **Utilização do Método de Box-Jenkins para Modelagem e Previsão da Produção Brasileira de Autoveículos**. Latin American Journal of Business Management, 2018.

Silva, J., Monteiro, Ana. **Uso de Aprendizado de Máquina para Predizer o Comportamento do Doador de Sangue: Uma Revisão da Literatura**. Anais do XV encontro unificado de computação do Piauí, 2022 (<https://doi.org/10.5753/enucompi.2022.222533>)

Santos, C.; Silva, L.; Castro, R.; Marques, R. **Application of Machine Learning to Project the Hourly Price for Settlement of Differences to Support Electricity Trading Strategies**. Braz. Energy Mag., 28, 243–279, 2022 (<https://doi.org/10.47168/rbe.v28i1.671>)

Solano, E.S.; Affonso, C.M. **Solar Irradiation Forecasting Using Ensemble Voting Based on Machine Learning Algorithms.** Sustainability, 15, 7943, 2023. (<https://doi.org/10.3390/su15107943>)

SONDA Homepage, <http://sonda.ccst.inpe.br/basedados/>, last accessed 2020/05/06.

Ul Islam, B.; Baharudin, Z.; Nallagownden, P.; Raza, M.Q. **A Hybrid Neuro-Genetic Approach for STLF: A Comparative Analysis of Model Parameter Variations.** In Proceedings of the IEEE 8th International Power Engineering and Optimization Conference (PEOCO2014), Langkawi, Malaysia, 24–25 March 2014. (<https://doi.org/10.1109/PEOCO.2014.6814485>)

Veja. Após ano mais quente, janeiro de 2024 também bate recorde de calor, 2024. Disponível: <https://veja.abril.com.br/>. Acesso em: 16/04/2024

Valencia-Arias, Alejandro *et al.* **Machine Learning and Blockchain: A Bibliometric Study on Security and Privacy.** Information, 15(1), 65, 2024 (<https://doi.org/10.3390/info15010065>)

Vicari, R. **Influências das Tecnologias da Inteligência Artificial no Ensino.** Scielo-Estudos avançados, vol. 35, n.101, 2021 (<https://doi.org/10.1590/s0103-4014.2021.35101.006>)

Villanueva, W. **Comitê de Máquinas na Previsão de Séries Temporais.** Dissertação de Mestrado, Universidade Estadual de Campinas, Campinas, SP, Brasil, 2006.

Venkatachalam, K. et al. **DWFH: An improved data-driven deep weather forecasting hybrid model using Transductive Long Short Term Memory (T-LSTM),** Expert Systems with Applications, vol. 213, 119270, 2023 (DOI: <https://doi.org/10.1016/j.eswa.2022.119270>)

Wang, S.; Zhang, N.; Wu, L.; Wang, Y. **Wind Speed Forecasting based on the Hybrid Ensemble Empirical Mode Decomposition and GA-BP Neural Network Method.** Renew. Energy, 94, 629e636, 2016 (<https://doi.org/10.1016/j.renene.2016.03.103>)

Wu, Z.; Wang, B. **An Ensemble Neural Network Based on Variational Mode Decomposition and an Improved Sparrow Search Algorithm for Wind and Solar Power Forecasting.** IEEE Access, 9, 166709, 2021 (<https://doi.org/10.1109/ACCESS.2021.3136387>)

Wang, L. **Advanced Multivariate Time Series Forecasting Models.** J. Math. Stat, 14, 253–260., 2018 (<https://doi.org/10.3844/jmssp.2018.253.260>)

YANG, X. **Nature-Inspired Optimization Algorithms**, 2nd Edition (Paperback), Academic Press, 2021.

Zhao, Miao; Ye, Ning. **High-Dimensional Ensemble Learning Classification: An Ensemble Learning Classification Algorithm Based on High-Dimensional Feature Space Reconstruction.** Appl. Sci., 14(5), 1956, 2024 (<https://doi.org/10.3390/app14051956>)

Zulfiqar, M.; Rasheed, M.B. **Short-Term Load Forecasting using Long Short Term Memory Optimized by Genetic Algorithm.** In Proceedings of the IEEE Sustainable Power and Energy Conference (iSPEC), December, Perth, Australia, 4–7 December 2022. (<https://doi.org/10.1109/iSPEC54162.2022.10033074>)

Zhou, Q.; Wang, C.; Zhang, G. **A Combined Forecasting System based on Modified Multi-Objective Optimization and Sub-Model Selection Strategy for Short-Term Wind Speed.** Appl. Soft Comput, 94, 106463, 2020 (<https://doi.org/10.1016/j.asoc.2020.106463>)

Zolpakar, Nor *et al.* **A review: use of evolutionary algorithm for optimization of machining parameters.** The International Journal of Advanced Manufacturing Technology, 115, p. 31-47, 2021. <https://doi.org/10.1007/s00170-021-07155-7>

---

### Resultados das Métricas de Avaliação para 1 Passo à Frente dos Modelos

Neste apêndice, fornecemos os resultados das métricas de avaliação - Erro Médio Absoluto (MAE), Erro Quadrático Médio (MSE) e Raiz do Erro Quadrático Médio (RMSE) - referentes à previsão de 1 passo à frente dos modelos analisados. Essa análise foi conduzida utilizando as bases de dados do Preço de Liquidação das Diferenças (PLD) da região norte e da Velocidade do Vento em Macau. Vale ressaltar que, no contexto do PLD, 1 passo representa a média semanal, enquanto para a velocidade do vento, 1 passo corresponde a um período de um dia.

Os valores das métricas são apresentados para cada modelo testado, fornecendo uma visão abrangente do desempenho de cada abordagem em relação à previsão de curto prazo nessas duas diferentes bases de dados.

A seguir, os resultados das métricas para cada modelo:

#### Modelo 1: MLP

<b>Base de Dados</b>	<b>MAE</b>	<b>MSE</b>	<b>RMSE</b>
PLD	0,0695	0,00242	0,0492
Velocidade do Vento	0,1719	0,01476	0,1215

#### Modelo 2: LSTM

<b>Base de Dados</b>	<b>MAE</b>	<b>MSE</b>	<b>RMSE</b>
PLD	0,0731	0,00268	0,0518

Velocidade do Vento	0,1920	0,01844	0,1358
---------------------	--------	---------	--------

## Modelo 3: ARIMA

<b>Base de Dados</b>	<b>MAE</b>	<b>MSE</b>	<b>RMSE</b>
PLD	12,687	80,398	8,969
Velocidade do Vento	0,5940	0,177	0,4204

## Modelo 4: SVM

<b>Base de Dados</b>	<b>MAE</b>	<b>MSE</b>	<b>RMSE</b>
PLD	32,981	543,180	23,309
Velocidade do Vento	0,6547	0,214	0,4629

## Modelo 5: AG + MLP

<b>Base de Dados</b>	<b>MAE</b>	<b>MSE</b>	<b>RMSE</b>
PLD	0,0605	0,00183	0,0428
Velocidade do Vento	0,169	0,01429	0,1195

## Modelo 6: AG + LSTM

<b>Base de Dados</b>	<b>MAE</b>	<b>MSE</b>	<b>RMSE</b>
PLD	0,04493	0,001012	0,03181
Velocidade do Vento	0,1598	0,01306	0,113

## Modelo 7: Ensemble 01 (MLP, Árvore de Decisão, Regressão Linear, SVM)

Base de Dados	MAE		MSE		RMSE	
	VOA	VOWA	VOA	VOWA	VOA	VOWA
PLD	0,6149	0,5498	0,18823	0,15082	0,4342	0,3888
Velocidade do Vento	0,7751	0,6156	0,30056	0,18988	0,5482	0,4356

Modelo 8: Ensemble 02 (MLP, LSTM, SVM, ARIMA)

Base de Dados	MAE		MSE		RMSE	
	VOA	VOWA	VOA	VOWA	VOA	VOWA
PLD	0,3125	0,5800	0,22115	0,16823	0,4703	0,4101
Velocidade do Vento	0,459	0,3785	0,10561	0,07191	0,325	0,268