
Previsão de Geração de Energia Fotovoltaica Baseada em Aprendizado de Máquina com Monitoramento de Degradação e Aprendizado Incremental

Vitor Hugo Honorato Tiago



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Vitor Hugo Honorato Tiago

**Previsão de Geração de Energia Fotovoltaica
Baseada em Aprendizado de Máquina com
Monitoramento de Degradação e Aprendizado
Incremental**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Prof. Dr. Jefferson Rodrigo de Souza
Coorientador: Prof. Dr. Gustavo Pessin

Uberlândia
2026

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

T551
2026

Tiago, Vítor Hugo Honorato, 1998-
Previsão de geração de energia fotovoltaica baseada em
aprendizado de máquina com monitoramento de degradação e
aprendizado incremental [recurso eletrônico] / Vítor Hugo
Honorato Tiago. - 2026.

Orientador: Jefferson Rodrigo de Souza.

Coorientador: Gustavo Pessin.

Dissertação (Mestrado) - Universidade Federal de Uberlândia,
Pós-graduação em Ciência da Computação.

Modo de acesso: Internet.

DOI <http://doi.org/10.14393/ufu.di.2026.677>

Inclui bibliografia.

1. Computação. I. Souza, Jefferson Rodrigo de, 1985-, (Orient.).
II. Pessin, Gustavo, 1979-, (Coorient.). III. Universidade Federal de
Uberlândia. Pós-graduação em Ciência da Computação. IV. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091

Nelson Marcos Ferreira - CRB6/3074



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Dissertação, 30/2026, PPGCO				
Data:	28 de Agosto de 2026	Hora de início:	14:00	Hora de encerramento:	15:45
Matrícula do Discente:	12412CCP031				
Nome do Discente:	Vítor Hugo Honorato Tiago				
Título do Trabalho:	Previsão de Geração de Energia Fotovoltaica Baseada em Aprendizado de Máquina com Monitoramento de Degradação e Aprendizado Incremental				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Inteligência Artificial				
Projeto de Pesquisa de vinculação:	-----				

Reuniu-se por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Marcelo Zanchetta do Nascimento - FACOM/UFU, Alan Kardek Rêgo Segundo - UFOP e Jefferson Rodrigo de Souza - FACOM/UFU, orientador(a) do(a) candidato(a).

Os examinadores participaram desde as seguintes localidades: Alan Kardek Rêgo Segundo - Ouro Preto/MG. Os outros membros da banca e o aluno(a) participaram da cidade de Uberlândia.

Iniciando os trabalhos o(a) presidente da mesa, Prof. Dr. Jefferson Rodrigo de Souza, apresentou a Comissão Examinadora e o(a) candidato(a), agradeceu a presença do público, e concedeu ao(à) Discente a palavra para a exposição do seu trabalho.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato(a):

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Jefferson Rodrigo de Souza, Professor(a) do Magistério Superior**, em 01/09/2026, às 16:33, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Marcelo Zanchetta do Nascimento, Professor(a) do Magistério Superior**, em 08/09/2026, às 09:59, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **ALAN KARDEK REGO SEGUNDO, Usuário Externo**, em 15/09/2026, às 16:32, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7612001** e o código CRC **9F5F1F6F**.

*Este trabalho é dedicado à criança
que um dia sonhou em compreender o mundo
e ao adulto que nunca deixou de perseguir esse sonho.*

“A sorte favorece a mente preparada.”
(Louis Pasteur)

“Ao vencedor, as batatas.”
(Machado de Assis)

Resumo

A crescente participação da energia fotovoltaica na matriz elétrica evidencia a necessidade de métodos de previsão capazes de fornecer estimativas precisas da geração de energia, contribuindo para o planejamento da operação do sistema elétrico. Entretanto, modelos preditivos podem sofrer degradação de desempenho ao longo do tempo em decorrência de mudanças nas características dos dados, tornando necessária sua adaptação contínua. Nesse contexto, esta dissertação propõe uma arquitetura adaptativa para previsão de geração de energia fotovoltaica baseada em aprendizado de máquina, integrando monitoramento de degradação e aprendizado incremental. Inicialmente, foram avaliadas diferentes estratégias de pré-processamento, modelos preditivos e métodos de otimização de hiperparâmetros utilizando dados históricos de uma usina fotovoltaica localizada em Minas Gerais. Foram avaliados os modelos persistência, ARIMA, Regressão Ridge, MLP e LSTM. Em seguida, um segundo conjunto de dados foi utilizado para avaliar atributos cíclicos, remoção de variáveis pouco correlacionadas e exclusão do período noturno. Por fim, foram avaliadas quatro configurações adaptativas combinando os detectores CUSUM e ADWIN às estratégias de aprendizado incremental *fine-tuning* e EWC. Os resultados demonstraram que a agregação pela média, a normalização *Min-Max* e os modelos LSTM apresentaram os melhores desempenhos preditivos, sobretudo para horizontes superiores a 15 minutos. A remoção do período noturno favoreceu o aprendizado dos modelos, enquanto a inclusão de atributos cíclicos não produziu ganhos consistentes. Entre as arquiteturas adaptativas avaliadas, a combinação ADWIN + EWC apresentou o melhor desempenho, reduzindo a frequência de atualizações do modelo e preservando sua capacidade preditiva ao longo do tempo. Além disso, o aprendizado incremental permitiu adaptar progressivamente modelos treinados com poucos dados, evidenciando seu potencial para aplicações reais.

Palavras-chave: previsão de geração fotovoltaica, aprendizado incremental, monitoramento de degradação, LSTM, energia solar, EWC, ADWIN.

Abstract

The increasing participation of photovoltaic energy in the electrical grid highlights the need for forecasting methods capable of providing accurate estimates of power generation, supporting the planning and operation of power systems. However, predictive models may experience concept drift over time due to changes in data characteristics, requiring continuous adaptation. In this context, this study proposes an adaptive architecture for photovoltaic power forecasting based on machine learning, integrating concept drift monitoring and incremental learning. Initially, different preprocessing strategies, predictive models, and hyperparameter optimization methods were evaluated using historical data from a photovoltaic power plant located in Minas Gerais, Brazil. The evaluated models included Persistence, ARIMA, Ridge Regression, MLP, and LSTM networks. Subsequently, a second dataset was used to evaluate the effects of cyclical features, the removal of weakly correlated variables, and the exclusion of nighttime data. Finally, four adaptive configurations combining the CUSUM and ADWIN concept drift detectors with the *fine-tuning* and EWC incremental learning strategies were evaluated. The results showed that mean aggregation, *Min-Max* normalization, and LSTM models achieved the best predictive performance, particularly for forecasting horizons longer than 15 minutes. Excluding nighttime data improved model learning, whereas the inclusion of cyclical features did not provide consistent performance gains. Among the evaluated adaptive architectures, the ADWIN + EWC combination achieved the best performance, reducing the frequency of model updates while preserving predictive accuracy over time. Furthermore, incremental learning enabled models trained with limited data to progressively adapt, demonstrating its potential for real-world applications.

Keywords: photovoltaic power forecasting, incremental learning, concept drift, LSTM, solar energy, EWC, ADWIN.

Lista de ilustrações

Figura 1 – Participação na geração de eletricidade por fonte no Brasil (2000–2025). Fonte: Adaptado de Our World in Data (2026).	28
Figura 2 – Esquema de uma planta solar. Fonte: Soares e Silva (2024).	36
Figura 3 – Representação da hora do dia em seno e cosseno.	42
Figura 4 – Representação simplificada de um neurônio biológico. Fonte: DeepLearning Book (2024).	47
Figura 5 – Modelo de neurônio matemático proposto por McCulloch e Pitts (1943). Fonte: Elaborado pelo autor (2026).	48
Figura 6 – Exemplo de rede neural artificial. Fonte: Tiago (2020).	49
Figura 7 – Representação de uma célula <i>Long Short-Term Memory</i> (LSTM). Fonte: Adaptada de Olah (2015).	53
Figura 8 – Método proposto para o desenvolvimento da arquitetura de predição, monitoramento de degradação e aprendizado incremental na geração de energia fotovoltaica. Fonte: Elaborado pelo autor (2026).	72
Figura 9 – Método detalhado da etapa relativa ao dataset Sol do Cerrado. Fonte: Elaborado pelo autor (2026).	73
Figura 10 – Método detalhado da etapa relativa ao conjunto de dados UNISOLAR. Fonte: Elaborado pelo autor (2026).	79
Figura 11 – Dados do dia 31-10-2023 do dataset Sol do Cerrado.	88
Figura 12 – Dados do dia 11-11-2023 do dataset Sol do Cerrado.	88
Figura 13 – Comparação das variáveis T_MOD_1 e T_MOD_2 com os limites do <i>Interquartile Range</i> (IQR).	90
Figura 14 – Matriz de correlação das variáveis do Sol do Cerrado utilizando os co- eficientes de Pearson e Spearman.	90
Figura 15 – Análise de correlação cruzada com defasagem entre as variáveis e INV_A_Measur.	91
Figura 16 – Contagem dos melhores resultados por método de escalonamento para cada modelo de previsão.	96

Figura 17 – Contagem de melhores resultados por otimizador de hiperparâmetros para cada modelo.	97
Figura 18 – Melhor configuração de cada modelo por granularidade na métrica erro absoluto médio (MAE).	98
Figura 19 – Melhor configuração de cada modelo por granularidade em MAE, raiz do erro quadrático médio (RMSE), R^2 e <i>Skill Score</i>	99
Figura 20 – Matriz de correlação de Pearson entre as <i>features</i> do dataset UNISOLAR.	101
Figura 21 – Matriz de correlação de Spearman entre as <i>features</i> do dataset UNISOLAR.	102
Figura 22 – Alarmes de degradação detectados pelo <i>Adaptive Windowing</i> (ADWIN) durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental <i>Elastic Weight Consolidation</i> (EWC).	105
Figura 23 – Alarmes de degradação detectados pelo <i>Cumulative Sum Control Chart</i> (CUSUM) durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental EWC.	106
Figura 24 – Alarmes de degradação detectados pelo ADWIN utilizando a estratégia de aprendizado incremental EWC no sítio 25.	108
Figura 25 – Contagem dos melhores escalonamentos por modelo de previsão na métrica RMSE.	124
Figura 26 – Contagem dos melhores escalonamentos por modelo de previsão na métrica R^2	125
Figura 27 – Contagem de melhores otimizadores de hiperparâmetros na métrica RMSE.	125
Figura 28 – Contagem de melhores otimizadores de hiperparâmetros na métrica R^2	126
Figura 29 – Melhor configuração de cada modelo por granularidade na métrica RMSE.	133
Figura 30 – Melhor configuração de cada modelo por granularidade na métrica R^2	134
Figura 31 – Melhor configuração de cada modelo por granularidade na métrica <i>Skill Score</i>	134
Figura 32 – Alarmes de degradação detectados pelo ADWIN durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental <i>fine-tuning</i> na variante V7.	135
Figura 33 – Alarmes de degradação detectados pelo CUSUM durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental <i>fine-tuning</i> na variante V7.	136
Figura 34 – Alarmes de degradação detectados pelo ADWIN durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental EWC na variante V5.	136

Figura 35 – Alarmes de degradação detectados pelo CUSUM durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental EWC na variante V5.	136
Figura 36 – Alarmes de degradação detectados pelo ADWIN durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental <i>fine-tuning</i> na variante V5.	137
Figura 37 – Alarmes de degradação detectados pelo CUSUM durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental <i>fine-tuning</i> na variante V5.	137

Lista de tabelas

Tabela 1	–	Resumo das configurações experimentais avaliadas.	92
Tabela 2	–	Espaço de busca dos hiperparâmetros do modelo de regressão Ridge. .	93
Tabela 3	–	Espaço de busca dos hiperparâmetros dos modelos <i>Multilayer Perceptron</i> (MLP) e LSTM.	93
Tabela 4	–	Percentual de melhores resultados por tipo de conjunto de dados considerando as 70 execuções.	94
Tabela 5	–	Número de melhores resultados por tipo de agregação obtidos por cada modelo para as métricas MAE, RMSE e R^2	94
Tabela 6	–	Número de melhores resultados por agregação obtidos para cada granularidade nas métricas MAE, RMSE e R^2	95
Tabela 7	–	Número de melhores resultados por agregação obtidos para cada técnica de escalonamento nas métricas MAE, RMSE e R^2	95
Tabela 8	–	Resumo quantitativo das etapas de limpeza e tratamento dos dados. .	100
Tabela 9	–	Variantes geradas do conjunto de dados.	103
Tabela 10	–	Espaço de busca dos hiperparâmetros do modelo LSTM.	104
Tabela 11	–	Métricas das oito variantes do modelo LSTM no conjunto de teste (kWh).104	
Tabela 12	–	Redução do MAE em kWh após cada atualização incremental acionada pelo ADWIN para as variantes V5 e V7.	106
Tabela 13	–	Desempenho final dos modelos após o monitoramento de degradação e o aprendizado incremental.	107
Tabela 14	–	Comparação entre os sítios utilizados nos experimentos.	107
Tabela 15	–	Redução do MAE após cada atualização incremental.	108
Tabela 16	–	Variantes do conjunto de dados utilizadas nos experimentos.	123
Tabela 17	–	MAE (kW) por variante e modelo.	127
Tabela 18	–	R^2 por variante e modelo.	128
Tabela 19	–	RMSE (kW) por variante e modelo.	130
Tabela 20	–	<i>Skill Score</i> por variante e modelo.	131

Lista de siglas

ADWIN *Adaptive Windowing*

AE-LSTM *Adaptive Ensemble Long Short-Term Memory*

AIC *Akaike Information Criterion*

ARIMA *Autoregressive Integrated Moving Average*

CA corrente alternada

CC corrente contínua

CJ caixas de junção

CNN redes neurais convolucionais

CUSUM *Cumulative Sum Control Chart*

DDM *Drift Detection Method*

DEN *Dynamic Expansion Network*

DNN redes neurais profundas

ECDD *Exponential Weighted Moving Average for Concept Drift Detection*

EDDM *Early Drift Detection Method*

EWC *Elastic Weight Consolidation*

EWC-DR (*EWC Done Right*)

IA Inteligência Artificial

IEA *International Energy Agency*

IQR *Interquartile Range*

KPSS Kwiatkowski–Phillips–Schmidt–Shin

LSTM *Long Short-Term Memory*

MAE erro absoluto médio

MLP *Multilayer Perceptron*

MSE erro quadrático médio

PNN *Progressive Neural Network*

POA *Plane of Array*

R² coeficiente de determinação (R^2)

ReLU *Rectified Linear Unit*

RE-SOINN *Regression Enhanced Incremental Self-Organising Neural Network*

RF *Random Forest*

RMSE raiz do erro quadrático médio

RNA redes neurais artificiais

RNN redes neurais recorrentes

SCADA sistema de supervisão e aquisição de dados

SPC controle estatístico de processo

SVM *Support Vector Machine*

UCA unidade controladora autônoma

UCR unidade controladora de rede

Sumário

1	INTRODUÇÃO	27
1.1	Objetivos	30
1.2	Hipóteses	31
1.3	Contribuições	31
1.4	Estrutura da dissertação	32
2	FUNDAMENTAÇÃO TEÓRICA	33
2.1	Geração de energia fotovoltaica	33
2.2	Pré-processamento de dados	37
2.2.1	Correlação entre <i>features</i>	37
2.2.2	Engenharia de atributos	40
2.2.3	Transformações de escala	42
2.3	Modelos clássicos de previsão	43
2.3.1	Persistência	44
2.3.2	<i>Autoregressive Integrated Moving Average</i> (ARIMA)	44
2.3.3	Regressão Ridge	45
2.4	Aprendizado de máquina	46
2.4.1	Redes neurais artificiais (RNA)	47
2.4.2	<i>Multilayer Perceptron</i> (MLP)	51
2.4.3	<i>Long Short-Term Memory</i> (LSTM)	52
2.4.4	Aprendizado incremental	55
2.5	Monitoramento de degradação	59
2.5.1	<i>Cumulative Sum Control Chart</i> (CUSUM)	59
2.5.2	<i>Adaptive Windowing</i> (ADWIN)	60
2.6	Métricas de avaliação de desempenho	61
2.7	Considerações finais	63

3	TRABALHOS RELACIONADOS	65
3.1	Previsão de energia fotovoltaica	65
3.2	Monitoramento de degradação em modelos	67
3.3	Aprendizagem incremental	68
3.4	Considerações finais	69
4	METODOLOGIA	71
4.1	Sol do Cerrado	73
4.1.1	Análise exploratória	74
4.1.2	Pré-processamento	74
4.1.3	Treinamento dos modelos de previsão	76
4.1.4	Seleção da melhor configuração de modelo	77
4.2	UNISOLAR	78
4.2.1	Pré-processamento	78
4.2.2	Treinamento dos modelos de previsão	83
4.2.3	Monitoramento de degradação e aprendizado incremental	84
4.2.4	Avaliação em um novo sítio de geração	84
4.3	Considerações finais	85
5	EXPERIMENTOS E ANÁLISE DOS RESULTADOS	87
5.1	Sol do Cerrado	87
5.1.1	Análise exploratória dos dados	87
5.1.2	Avaliação das configurações experimentais	92
5.1.3	Comparação dos modelos de previsão	97
5.2	UNISOLAR	99
5.2.1	Análise exploratória dos dados	100
5.2.2	Avaliação do modelo nas variantes do conjunto de dados	102
5.2.3	Monitoramento de degradação e aprendizado incremental	105
5.2.4	Avaliação em um novo sítio fotovoltaico	107
5.3	Considerações finais	109
6	CONCLUSÃO	111
6.1	Contribuições	113
6.2	Trabalhos futuros	114
	REFERÊNCIAS	115

APÊNDICES

121

APÊNDICE A	–	RESULTADOS COMPLETOS DOS EXPERIMENTOS COM O DATASET SOL DO CERRADO .	123
APÊNDICE B	–	RESULTADOS COMPLETOS DOS EXPERIMENTOS COM O DATASET UNISOLAR	135

Introdução

O desenvolvimento da humanidade esteve sempre atrelado ao acesso e ao uso de diferentes formas de energia. Com o crescimento populacional e o avanço das atividades humanas, a demanda por energia aumentou significativamente, abrangendo desde necessidades básicas até processos industriais complexos, com o intuito de suprir uma sociedade em constante expansão (Goldemberg; Lucon, 2007).

A demanda de energia global quase dobrou de 2000 até 2022, passando de cerca de 14,972 terawatt-hora (TWh) em 2000 para 28,510 em 2022 (Wiatros-Motyka et al., 2023). Segundo dados da *International Energy Agency* (IEA) a demanda por eletricidade crescerá 4% até 2027 e as economias emergentes deverão ser responsáveis por 85% desse crescimento (CNN Brasil, 2025).

Esse aumento na demanda por energia intensifica as preocupações em relação ao atual modelo de produção e consumo, considerado insustentável quando baseado na extração de fontes não renováveis, uma vez que as reservas são finitas e seu uso acarreta impactos ambientais significativos (International Energy Agency, 2025). Devido a um cenário de preocupação mundial com os efeitos da mudança climática, 195 países, incluindo o Brasil, assinaram o Acordo de Paris em 2015 (Iberdrola, 2024).

O Acordo de Paris inclui, entre suas principais metas, a redução das emissões de carbono, implementando um plano de transição energética (Ministério do Meio Ambiente, s.d.). A transição energética consiste na substituição de combustíveis fósseis, como o petróleo e o carvão, por fontes de energia limpas, caracterizadas por menores emissões de gases de efeito estufa. Com isso, fontes de energia de baixa emissão, como renováveis e nuclear se mostram a melhor opção para o futuro (Petrobras, 2024).

A energia renovável é obtida a partir de fontes naturais, como a luz solar, o vento, a água e a biomassa, as quais possuem capacidade de regeneração ao longo do tempo, diferentemente dos combustíveis fósseis, que são finitos e contribuem para a emissão de gases de efeito estufa (United Nations, 2026). Em 2023, a capacidade instalada de energia renovável global atingiu 3.870 gigawatts (GW), representando um aumento de 473 GW em relação ao ano anterior (IRENA, 2024). Este crescimento de aproximadamente 14%

destaca a aceleração na adoção de fontes renováveis. A IEA destaca que as fontes renováveis deverão representar 37% da geração de eletricidade mundial até 2026 (CNN Brasil, 2024).

O Brasil, por sua vez, destaca-se no cenário global, com aproximadamente 90,3% de sua capacidade instalada de geração elétrica proveniente de fontes renováveis em 2026 (ONS, 2026). Nos últimos 5 anos, o país multiplicou a geração de energia eólica e solar, que hoje supera a geração por gás natural (Petrobras, 2024).

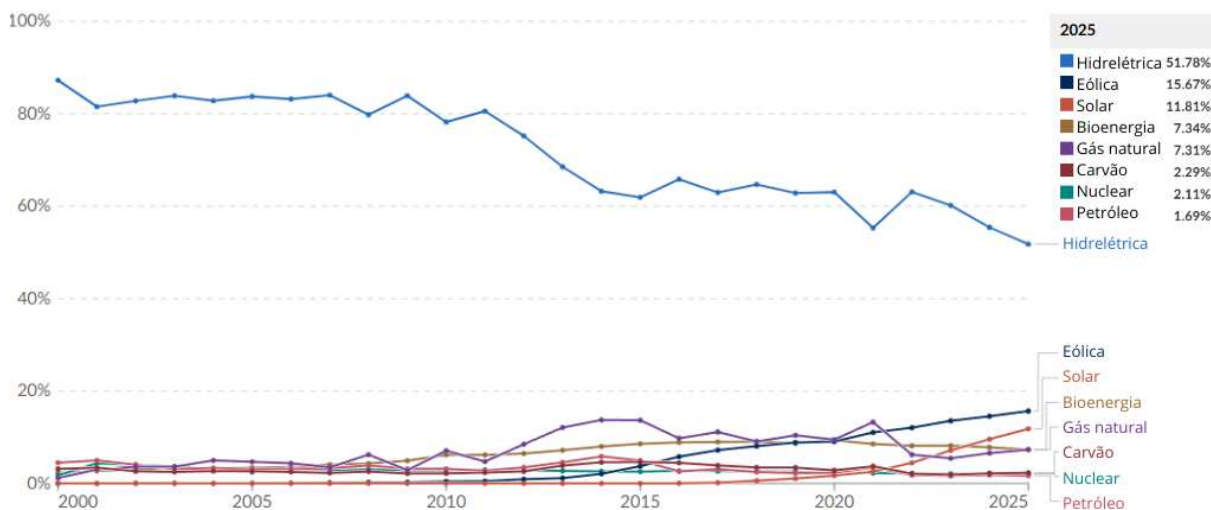


Figura 1 – Participação na geração de eletricidade por fonte no Brasil (2000–2025). Fonte: Adaptado de Our World in Data (2026).

Na Figura 1 é mostrada a participação das diferentes fontes na matriz elétrica brasileira entre 2000 e 2025. Existe uma forte predominância da energia hidrelétrica, responsável por 51,78% da geração de energia do Brasil em 2025. Entretanto, as fontes renováveis complementares tiveram um crescimento expressivo: a energia eólica chegou a 15,67%, a solar a 11,81% e a biomassa a 7,34%. Entre as fontes não renováveis, o gás natural contribuiu com 7,31%, seguido pelo carvão (2,29%), energia nuclear (2,11%) e pelo óleo combustível (1,69%). É importante ressaltar que a participação da energia solar na matriz elétrica cresceu de 1,71% em 2020 para 11,81% em 2025, evidenciando uma forte tendência de expansão dessa fonte de energia (Our World in Data, 2026).

O Brasil possui um grande potencial para a geração de energia solar, devido ao seu clima tropical com altos índices de irradiação. Nesse contexto, muitas empresas identificam diversas vantagens em adotar a energia solar. Dentre essas vantagens se destacam a maior independência energética em relação ao suprimento convencional, redução dos custos com eletricidade, acesso a incentivos fiscais concedidos pelo governo para investimentos em energia solar e a demonstração de compromisso com a sustentabilidade ambiental (Intelbras, 2022).

Uma estratégia amplamente adotada por empresas para aproveitar esse potencial é a implantação de campos geradores fotovoltaicos destinados à produção de energia para consumo próprio ou comercialização. Entretanto, a operação eficiente desses sistemas depende da capacidade de estimar a quantidade de energia que será gerada. Nesse contexto, técnicas de aprendizado de máquina se destacam por sua capacidade de modelar as relações não lineares entre variáveis meteorológicas e a geração fotovoltaica, produzindo previsões cada vez mais precisas (Tian et al., 2023; Jailani et al., 2023).

A capacidade de prever a quantidade de energia gerada por esses campos representa um ganho significativo para o planejamento e a gestão operacional da empresa. Com previsões confiáveis, torna-se possível estimar custos com aquisição de energia da rede, planejar o uso da energia produzida e antecipar situações de excedente ou déficit de geração. Dessa forma, quando a produção não é suficiente para suprir a demanda, a empresa pode prever a necessidade de aquisição de energia externa. Por outro lado, quando há excedente, a energia pode ser armazenada, redirecionada ou comercializada, sendo revertida em capital (ICLG, 2024). Em pequenas empresas, a implementação de sistemas de energia solar pode proporcionar uma redução de até 95% nas despesas com energia elétrica (Redação Homework, 2025).

Outro benefício da implantação de modelos preditivos é a capacidade de monitorar o funcionamento dos campos geradores fotovoltaicos. Esse monitoramento é realizado por meio da comparação, através de métodos estatísticos, entre os valores de energia previstos pelo modelo e os valores efetivamente medidos no sistema real. Tais métodos podem identificar quando ocorrem alterações no comportamento do sistema, como falhas em módulos, degradação de componentes, acúmulo de sujeira nas placas ou outras anomalias que comprometem a eficiência da geração de energia. Além disso, a análise desses desvios pode identificar quando o modelo preditivo já não representa o comportamento atual do sistema, evidenciando a necessidade de incorporar novos dados ao processo de aprendizagem para manter a qualidade das previsões.

O desafio de lidar com a sazonalidade na geração de energia solar, fortemente influenciada pelas condições climáticas, constitui uma das principais motivações para a adoção de técnicas de monitoramento de degradação de modelos e aprendizado incremental. Dessa forma, o acompanhamento contínuo do sistema permite identificar desvios de desempenho, apoiar decisões de manutenção preventiva, indicar a necessidade de atualização do modelo preditivo e contribuir para o aumento da confiabilidade e da eficiência energética do campo solar. Assim, torna-se importante investigar estratégias que permitam manter a qualidade das previsões mesmo diante das degradações do modelo, preservando a confiabilidade tanto da previsão quanto do monitoramento do campo fotovoltaico.

Na dissertação desenvolvida por Catarino (2024) são avaliados dados do mesmo sítio de estudo utilizado no primeiro momento do presente trabalho. O autor avalia modelos de predição de energia solar utilizando dados da irradiância solar como entrada. No trabalho

de referência foram utilizadas as técnicas de *Random Forest* (RF) e de redes neurais artificiais (RNA). Além disso, foi realizado um pré-processamento com os dados, agrupando-os em diversas frequências antes da criação dos modelos. Nos resultados de Catarino (2024), as melhores métricas de desempenho foram obtidas pela RNA com função de ativação *Rectified Linear Unit* (ReLU) e 10 neurônios, tendo coeficiente de determinação superior a 94%.

Os resultados obtidos por Catarino (2024) evidenciam o potencial do emprego de técnicas de aprendizado de máquina para a previsão da geração de energia solar. Entretanto, permanecem oportunidades para investigar arquiteturas mais adequadas à modelagem de séries temporais, avaliar o impacto da incorporação de novas variáveis meteorológicas e operacionais e desenvolver estratégias capazes de manter o desempenho do modelo ao longo do tempo diante das mudanças observadas no sistema de geração. Nesse contexto, esta pesquisa busca ampliar os avanços obtidos no trabalho de referência. Nas próximas subseções são apresentados os objetivos, as hipóteses, as contribuições e a organização desta dissertação.

1.1 Objetivos

O objetivo geral deste trabalho é avaliar um modelo de aprendizado de máquina capaz de prever a quantidade de energia a ser produzida em um campo de geração de energia fotovoltaica. Além disso, propõe-se implementar um mecanismo de monitoramento de degradação online, apto a identificar mudanças no comportamento dos dados e, se necessário, acionar uma estratégia de aprendizado incremental para atualização do modelo com novas informações.

Os objetivos específicos deste trabalho, definidos para orientar o alcance do objetivo geral, são:

- ❑ Analisar os dados de geração de energia e as variáveis meteorológicas dos sítios de estudo, investigando suas características e correlações;
- ❑ Avaliar o impacto de diferentes granularidades temporais e métodos de agrupamento dos dados no desempenho dos modelos preditivos;
- ❑ Comparar diferentes estratégias de pré-processamento dos dados, incluindo métodos de normalização, codificação cíclica das variáveis temporais e técnicas de otimização de hiperparâmetros;
- ❑ Comparar o desempenho do método de persistência e dos modelos ARIMA, regressão Ridge, MLP e LSTM para a previsão da geração de energia fotovoltaica;
- ❑ Avaliar diferentes métodos de monitoramento de degradação e estratégias de aprendizado incremental para atualização do modelo preditivo;

- ❑ Validar a capacidade de generalização da abordagem proposta por meio de sua aplicação em diferentes sítios de geração de energia fotovoltaica.

1.2 Hipóteses

Este trabalho possui as seguintes hipóteses:

- ❑ H1: É possível prever a geração de energia de um campo fotovoltaico a partir de variáveis meteorológicas e operacionais, incluindo irradiância solar, temperatura do ar, temperatura aparente, umidade relativa do ar e temperatura dos inversores, utilizando modelos de aprendizado de máquina.
- ❑ H2: A aplicação de um protocolo de monitoramento de degradação associado a estratégias de aprendizado incremental permite aprimorar o desempenho do modelo ao longo do tempo na predição da geração de energia fotovoltaica, além de possibilitar a identificação antecipada de degradações.
- ❑ H3: Um modelo preditivo treinado com uma quantidade reduzida de dados pode aprimorar progressivamente seu desempenho por meio do aprendizado incremental, tornando-se adequado para operação em campo.

As hipóteses propostas são avaliadas ao longo dessa dissertação por meio de análises estatísticas, visualizações gráficas e experimentos computacionais utilizando os dados dos campos de geração fotovoltaica de estudo de caso.

1.3 Contribuições

As principais contribuições desta dissertação concentram-se na proposição, construção e avaliação de uma arquitetura integrada de previsão, monitoramento de degradação e aprendizado incremental para campos fotovoltaicos. Dessa maneira, é realizada uma análise das características dos dados de geração de energia solar e das variáveis meteorológicas dos sítios de estudo bem como uma avaliação do impacto de diferentes granularidades temporais, métodos de agrupamento, técnicas de escalonamento de dados e estratégias de otimização de hiperparâmetros no desempenho dos modelos preditivos. Adicionalmente, é investigada a influência da representação cíclica das variáveis temporais, por meio das funções seno e cosseno, na qualidade das previsões.

Outra contribuição consiste na comparação entre diferentes abordagens para previsão da geração de energia fotovoltaica, abrangendo métodos estatísticos, modelos tradicionais de regressão e arquiteturas baseadas em aprendizado de máquina. Essa comparação permite avaliar a aplicação dessas técnicas ao problema de previsão da geração de energia

fotovoltaica, identificando as abordagens mais adequadas quanto à precisão e à capacidade de modelar séries temporais.

Por fim, a dissertação avalia métodos de monitoramento da degradação do modelo preditivo e estratégias de aprendizado incremental, permitindo que o modelo seja atualizado continuamente com novos dados sempre que forem identificadas mudanças significativas no comportamento do sistema. Além de preservar o desempenho do modelo preditivo, essa abordagem possibilita a identificação antecipada de possíveis degradações do sistema fotovoltaico, apoiando decisões de manutenção preventiva, contribuindo para a otimização da geração de energia e para a redução de perdas operacionais.

Como contribuição prática, propõe-se um modelo capaz de iniciar sua operação em campo com poucos dados de treino e aprimorar seu desempenho ao longo do tempo, validando a capacidade do modelo de se aprimorar com dados reais e, com o tempo, espera-se obter um modelo com alta precisão.

1.4 Estrutura da dissertação

Esta dissertação está organizada em seis capítulos. O Capítulo 1 apresenta a relevância do estudo voltado à geração de energia em campos fotovoltaicos juntamente com o contexto da pesquisa, sua motivação, objetivos, hipóteses e contribuições. O Capítulo 2 aborda conceitos e técnicas empregados na dissertação. O Capítulo 3 traz uma revisão da literatura com os trabalhos relacionados. O Capítulo 4 descreve a metodologia proposta para atingir o objetivo. O Capítulo 5 apresenta os experimentos realizados e a análise dos resultados obtidos. O Capítulo 6 tece as conclusões do trabalho e expõe possíveis trabalhos futuros.

Fundamentação Teórica

Neste capítulo são apresentados os conceitos teóricos que fundamentam o desenvolvimento deste trabalho, organizados em sete seções. A Seção 2.1 detalha os princípios da geração de energia fotovoltaica, descrevendo seu processo e componentes. A Seção 2.2 aborda as principais etapas de pré-processamento dos dados, incluindo análise de correlação, engenharia de atributos e transformações de escala. A Seção 2.3 introduz alguns modelos utilizados na previsão de dados. A Seção 2.4 apresenta os conceitos de aprendizado de máquina empregados nesta dissertação, contemplando redes neurais artificiais (RNA), *Multilayer Perceptron* (MLP), *Long Short-Term Memory* (LSTM), busca de hiperparâmetros e aprendizado incremental. A Seção 2.5 explora os conceitos relacionados ao monitoramento de degradação, incluindo técnicas como o CUSUM e ADWIN. A Seção 2.6 descreve as principais métricas utilizadas para avaliação do desempenho dos modelos de previsão. Por fim, a Seção 2.7 encerra o capítulo com as considerações finais.

2.1 Geração de energia fotovoltaica

Essa seção apresenta os principais equipamentos utilizados na geração de energia elétrica através da luz solar e detalha como esse processo acontece. Compreender o processo de geração de energia elétrica por meio de sistemas fotovoltaicos é fundamental para o desenvolvimento deste trabalho, uma vez que esse conhecimento permite interpretar adequadamente os dados provenientes do campo solar. Esse entendimento fornece uma base sólida para a etapa de análise e pré-processamento dos dados, permitindo compreender como cada dado registrado pelos sensores representa o comportamento do sistema. Além disso, o conhecimento sobre a origem e o significado desses dados contribui para uma aplicação mais eficaz das técnicas de aprendizado de máquina, favorecendo a construção de modelos mais consistentes.

A geração fotovoltaica ocorre em instalações conhecidas como usinas ou sítios solares, que integram diversos sistemas e componentes responsáveis pela captação da radiação solar e sua conversão em eletricidade. Trata-se de uma infraestrutura complexa, cujo fun-

cionamento eficiente depende da integração entre diferentes tecnologias (Catarino, 2024). A primeira etapa na geração de energia elétrica a partir da luz do sol consiste na captação da radiação solar por meio de painéis fotovoltaicos. Esses dispositivos são responsáveis por converter a energia luminosa em energia elétrica na forma de corrente contínua (CC), utilizando o chamado efeito fotovoltaico (BlueSol Educacional, 2016).

O efeito fotovoltaico é o fenômeno físico no qual a incidência da radiação solar sobre um material semicondutor provoca a geração de uma diferença de potencial elétrico, permitindo a conversão direta da energia luminosa em energia elétrica. Esse efeito ocorre quando os fótons provenientes da luz solar são absorvidos pelo semicondutor, fornecendo energia suficiente para excitar os elétrons e removê-los de suas posições de equilíbrio. A separação dessas cargas dá origem a um campo elétrico interno, responsável por direcionar o movimento dos elétrons. Quando um circuito externo é conectado, os elétrons fluem através dele, originando uma corrente elétrica contínua (BlueSol Educacional, 2016).

Os painéis fotovoltaicos são construídos a partir de diferentes materiais semicondutores que influenciam na durabilidade, eficiência de conversão e custo de fabricação. Os mais comuns são os de silício monocristalino, silício policristalino e os de filme fino (United States Department of Energy, 2025b). Em geral, os painéis de silício monocristalino apresentam eficiência superior, podendo converter mais de 20% da radiação solar incidente em eletricidade. Em contrapartida, os painéis de silício policristalino e os de filme fino apresentam eficiências inferiores, entretanto podem representar alternativas economicamente viáveis. Com materiais de alta durabilidade, os painéis solares mais eficientes podem permanecer em funcionamento por mais de 25 anos, além de sofrer degradação mínima ao longo do tempo (United States Department of Energy, 2025b).

Após a conversão da luz solar em energia elétrica na forma de corrente contínua (CC), ocorre a etapa de transformação dessa energia em corrente alternada (CA), que é compatível com a rede elétrica convencional. Essa conversão é realizada por equipamentos denominados inversores. Existem diferentes tipos de inversores, entre os quais se destacam os modelos *string*, centrais e híbridos, cada um com características específicas de aplicação conforme a escala e o perfil da instalação (Rampinelli; Krenzinger; Romero, 2013).

Acoplados aos inversores, os módulos de potência têm papel fundamental na conversão eficiente de energia CC — gerada pelos painéis solares — em CA compatível com a rede elétrica. Além disso, tais módulos auxiliam no controle da quantidade de energia transferida para a rede, garantindo que tensão e corrente se mantenham dentro de limites seguros e oferecendo proteção contra sobrecargas, curto-circuitos e outras condições de falha (Catarino, 2024).

Além dessas funções, os módulos de potência são projetados com foco em alta eficiência, minimizando perdas durante a conversão de energia e assegurando que a maior parte da eletricidade gerada pelos painéis seja efetivamente entregue à rede. Para manter

sua eficiência operacional, contam com sistemas de gestão térmica que dissipam o calor gerado durante o funcionamento, seja por meio de resfriamento natural ou por sistemas ativos de refrigeração (Catarino, 2024).

A energia elétrica proveniente dos inversores é, então, direcionada aos transformadores, cuja função é elevar a tensão da CA para níveis compatíveis com os requisitos de transmissão e distribuição pela rede elétrica. Essa etapa é fundamental para minimizar perdas ao longo das linhas de transmissão e garantir maior eficiência no transporte da energia. Em seguida, a energia em corrente alternada, com os níveis de tensão devidamente ajustados, é integrada ao sistema de distribuição elétrica. Nesse processo, transformadores e subestações exercem um papel essencial, assegurando que a energia gerada seja entregue de forma estável, segura e contínua aos centros de consumo (Soares, 2024).

A geração de energia solar fotovoltaica é altamente influenciada pelas condições ambientais do local de instalação. Fatores como poeira, vento, umidade, temperatura e irradiância solar exercem impacto direto sobre o desempenho dos módulos e a eficiência do sistema. A presença de poeira sobre a superfície dos painéis pode reduzir significativamente a incidência direta da radiação solar, ocasionando perdas de eficiência na conversão energética (Araujo et al., 2019). Da mesma forma, a umidade e a precipitação podem afetar a durabilidade dos componentes, favorecendo processos de corrosão e falhas nos contatos elétricos. Assim, a adoção de materiais resistentes à corrosão e de sistemas automatizados de limpeza pode ser necessária em regiões com altos índices de umidade e chuva (Catarino, 2024).

A temperatura ambiente também desempenha papel relevante, uma vez que a eficiência das células fotovoltaicas tende a diminuir com o aumento da temperatura. Por isso, é importante considerar a temperatura média do local e, quando necessário, empregar soluções que favoreçam o resfriamento natural ou forçado dos painéis. O vento, por sua vez, apresenta um efeito duplo: contribui para o resfriamento dos módulos, melhorando o desempenho térmico, mas, se intenso, pode causar danos mecânicos às estruturas de suporte e aos próprios painéis. Assim, a utilização de estruturas robustas e adequadamente projetadas é indispensável para garantir a estabilidade do sistema (United States Department of Energy, 2025a).

Além disso, a irradiância solar é um dos fatores mais determinantes para o dimensionamento e o desempenho da planta fotovoltaica. Regiões com altos índices de irradiância são preferenciais, pois permitem maior aproveitamento da energia solar disponível por unidade de área. Diante dessa variabilidade climática, o monitoramento contínuo das condições ambientais é essencial para otimizar a operação e a manutenção da planta.

Nesse contexto, as estações solarimétricas desempenham papel fundamental, pois realizam a medição de parâmetros como irradiância solar, temperatura, velocidade e direção do vento, fornecendo dados críticos para o sistema de supervisão e aquisição de dados (SCADA). Os sistemas de geração de energia utilizados nesse trabalho contam

com monitoramentos e controle SCADA que desempenham um papel essencial na operação e no monitoramento das plantas solares fotovoltaicas. Esse sistema integra os diversos componentes da planta, permitindo a coleta, o processamento e a visualização de dados em tempo real (Catarino, 2024).

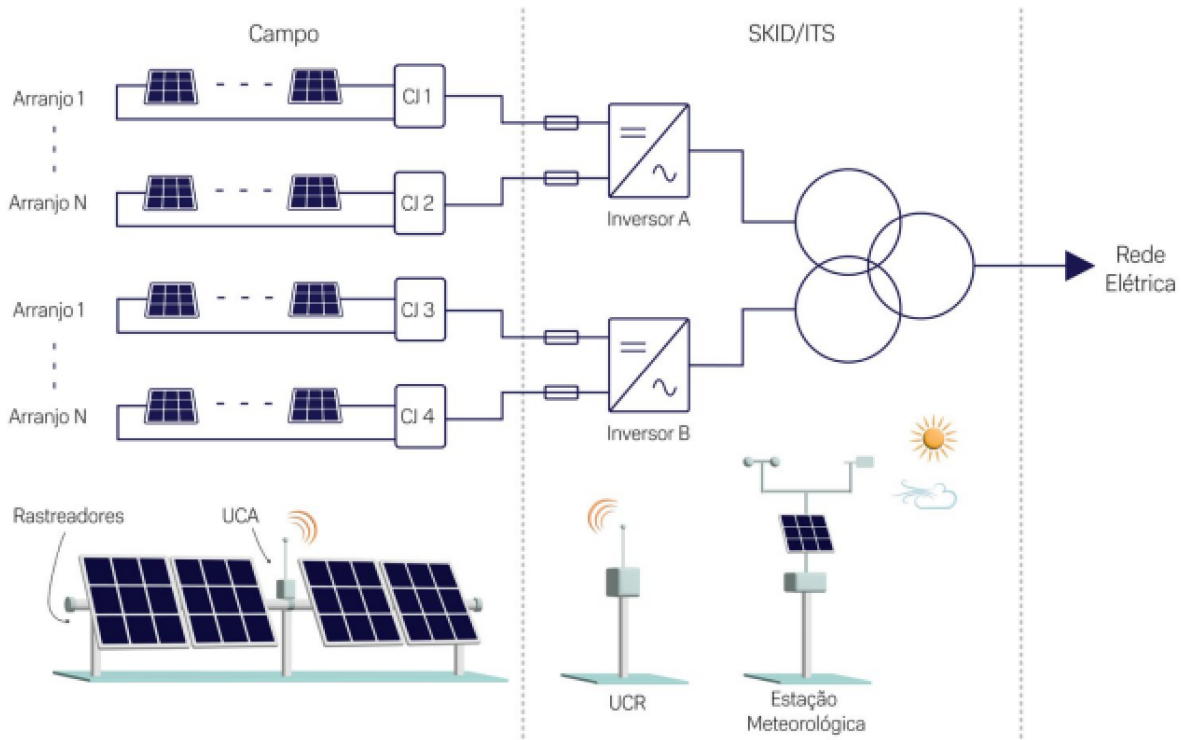


Figura 2 – Esquema de uma planta solar. Fonte: Soares e Silva (2024).

A Figura 2 apresenta o esquema simplificado de uma planta solar fotovoltaica, ilustrando o fluxo de energia desde os painéis fotovoltaicos até a rede elétrica. O campo e os arranjos representam os conjuntos de painéis responsáveis pela captação da radiação solar e conversão em energia elétrica de corrente contínua. Essa energia é conduzida até as caixas de junção (CJ), que têm a função de reunir as correntes provenientes dos diferentes módulos *string*. As CJ estão conectadas aos inversores, responsáveis pela conversão da energia em corrente alternada. Em seguida, os transformadores elevam a tensão para os níveis necessários à injeção na rede. Além disso, observa-se na figura a presença da estação meteorológica, sendo também o ponto de integração do sistema SCADA. Na Figura 2, a unidade controladora autônoma (UCA) é instalada nos suportes dos módulos, enquanto a unidade controladora de rede (UCR) é responsável pelo controle e configuração do rastreamento de cada conjunto de equipamentos.

A compreensão do processo de geração de energia solar fotovoltaica e de seus principais componentes fornece a base necessária para a etapa seguinte deste trabalho, voltada ao desenvolvimento de modelos de previsão e análise de desempenho. O entendimento das

variáveis envolvidas na conversão da energia solar em elétrica permite relacionar o comportamento físico do sistema às técnicas computacionais empregadas para sua modelagem. Dessa forma, a próxima seção apresenta os fundamentos teóricos do pré processamento dos dados, etapa essencial para o desenvolvimento de modelos preditivos mais precisos e robustos.

2.2 Pré-processamento de dados

O pré-processamento de dados constitui uma das etapas mais importantes no desenvolvimento de modelos preditivos. Seu objetivo é analisar e preparar os dados antes de sua utilização pelos algoritmos de aprendizado, de modo a melhorar sua qualidade, reduzir inconsistências e tornar as informações mais adequadas para a etapa de modelagem. Além de contribuir para a construção de modelos mais precisos e robustos, o pré-processamento também facilita a interpretação das características dos dados e a identificação de padrões relevantes (Carvalho; Menezes; Bonidia, 2024).

Essa seção compreende alguns métodos utilizados durante o pré-processamento dos dados, organizando-os em três etapas principais. Inicialmente, são apresentados os conceitos relacionados a análise de correlação entre *features*. Em seguida, são introduzidas técnicas de engenharia de atributos, utilizadas para representar de forma mais adequada informações relevantes para os modelos de previsão. Por fim, são discutidas transformações de escala dos dados, empregadas para adequar a escala das variáveis de entrada ao processo de treinamento dos modelos de aprendizado de máquina.

2.2.1 Correlação entre *features*

A análise de correlação entre *features* consiste em uma técnica estatística utilizada para quantificar o grau de associação entre duas variáveis. De acordo com Barbetta, Reis e Bornia (2019), um exemplo clássico é a relação entre altura e peso de indivíduos, na qual pessoas mais altas tendem, em média, a apresentar maior peso corporal. Nesse caso, a análise de correlação permite medir a intensidade da associação entre essas variáveis, sem necessariamente estabelecer uma relação de causa e efeito.

A intensidade da correlação pode ser classificada como forte, moderada, fraca ou inexistente. Correlações fortes indicam uma elevada associação entre as variáveis, enquanto correlações fracas ou inexistentes sugerem que a variação observada em uma delas apresenta pouca ou nenhuma associação com a outra (Montgomery; Runger, 2021). Além da intensidade, a correlação também pode ser caracterizada quanto à sua direção. Uma correlação positiva indica que as variáveis tendem a variar no mesmo sentido, enquanto uma correlação negativa indica que elas tendem a variar em sentidos opostos (Barbetta; Reis; Bornia, 2019). Diversos coeficientes e técnicas podem ser empregados para quantificar a correlação entre variáveis. Entre os métodos mais utilizados destacam-se os coeficientes

de Pearson e de Spearman, além de abordagens específicas para dados circulares e análises com defasagem (*lag*), adequadas a diferentes características dos dados.

Correlação de Pearson

O coeficiente de correlação de Pearson é uma medida estatística amplamente utilizada para quantificar o grau de associação linear entre duas variáveis quantitativas de uma amostra (Montgomery; Runger, 2021). De acordo com Montgomery e Runger (2021), o coeficiente de correlação de Pearson é calculado conforme descrito na Equação 1.

$$r_{xy} = \frac{\sum_{i=1}^n y_i(x_i - \bar{x})}{\sqrt{\left[\sum_{i=1}^n (y_i - \bar{y})^2 \right] \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]}} \quad (1)$$

Na Equação 1, x_i e y_i representam as observações das variáveis X e Y , respectivamente, $\bar{x}$ e $\bar{y}$ correspondem às médias amostrais dessas variáveis e n representa o número de observações. O coeficiente de correlação de Pearson assume valores no intervalo de -1 a 1 . Valores próximos de 1 indicam uma forte correlação linear positiva, enquanto valores próximos de -1 indicam uma forte correlação linear negativa. Por sua vez, valores próximos de zero sugerem ausência de associação linear entre as variáveis (Montgomery; Runger, 2021).

Correlação de Spearman

A correlação de Spearman é uma medida estatística utilizada para avaliar a intensidade e a direção da associação monotônica entre duas variáveis. Por essa razão, é amplamente empregada em situações nas quais os dados não atendem às premissas de normalidade ou apresentam distribuições assimétricas e valores discrepantes (Siegel; Jr., 2006).

O cálculo do coeficiente de Spearman é realizado a partir dos postos (*ranks*) das observações. Inicialmente, os valores de cada variável são ordenados, atribuindo-se um posto a cada observação. Em seguida, compara-se a posição ocupada por cada elemento nas duas ordenações. Como exemplo didático, no livro Siegel e Jr. (2006) os autores apresentam o caso de estudantes classificados segundo suas colocações em uma avaliação em sala de aula e em um exame vestibular. Dessa maneira, cada aluno possui um posto em cada classificação, e o coeficiente de Spearman verifica se aqueles que obtiveram melhor desempenho na avaliação em sala tendem a ocupar também as primeiras posições no vestibular. Quanto mais semelhantes forem essas ordenações, maior será a associação entre as variáveis. O coeficiente de correlação de Spearman é descrito na Equação 2.

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2)$$

Na Equação 2, o r_s indica o coeficiente de correlação de Spearman, n representa a quantidade de observações e d_i a distância entre os postos ocupados por uma mesma observação nas duas ordenações. No exemplo dado, d_i corresponde à diferença entre as posições do aluno i na avaliação em sala e no vestibular. Quanto menores forem essas diferenças, mais próximo de 1 será o coeficiente de Spearman, indicando forte associação positiva. Por outro lado, diferenças elevadas reduzem o valor de r_s indicando uma associação mais fraca (Siegel; Jr., 2006).

Assim como o coeficiente de Pearson, o coeficiente de Spearman assume valores no intervalo de -1 a 1 . Valores próximos de 1 indicam forte associação positiva, valores próximos de -1 representam forte associação negativa e valores próximos de zero sugerem ausência de associação. Embora ambos os coeficientes variem no mesmo intervalo e possuam interpretações semelhantes quanto à intensidade e à direção da associação, diferem na natureza da relação avaliada.

Em uma relação linear, a variação entre as variáveis pode ser representada aproximadamente por uma reta. Já uma relação monotônica caracteriza-se apenas pela manutenção da tendência de crescimento ou de decrescimento entre as variáveis, independentemente de a relação seguir uma forma linear. Assim, enquanto o coeficiente de Pearson mede o grau de associação linear, o coeficiente de Spearman mede o grau de associação monotônica por meio da ordenação das observações.

Correlação circular de Mardia

Algumas variáveis apresentam natureza circular, isto é, seus valores representam posições em um ciclo e não em uma escala linear. Nesses casos, valores numericamente distantes podem representar direções ou instantes muito próximos, como ocorre com horas do dia (23 h e 0 h) ou dias do ano (31 de dezembro e 1° de janeiro). Nessas situações, coeficientes de correlação tradicionais, como os de Pearson e Spearman, podem produzir resultados inadequados por desconsiderarem a periodicidade dos dados. Para contornar essa limitação, Mardia propôs um coeficiente de correlação circular-linear capaz de quantificar a associação entre uma variável linear e outra circular preservando sua natureza periódica (Mardia, 1976).

Considerando x uma variável linear e θ uma variável circular expressa em radianos. O coeficiente de correlação circular-linear de Mardia é definido pela Equação 3.

$$\rho_{cl} = \sqrt{\frac{r_{xc}^2 + r_{xs}^2 - 2r_{xc}r_{xs}r_{cs}}{1 - r_{cs}^2}} \quad (3)$$

Na Equação 3, r_{xc} representa o coeficiente de correlação de Pearson entre x e $\cos(\theta)$, r_{xs} corresponde ao coeficiente de correlação entre x e $\sin(\theta)$ e r_{cs} representa a correlação entre $\cos(\theta)$ e $\sin(\theta)$. O coeficiente ρ_{cl} assume valores entre 0 e 1, onde próximo de zero

indica ausência de associação entre as variáveis, enquanto valores próximos de um indicam forte associação entre a variável linear e a variável circular (Mardia, 1976).

Correlação com defasagem (*lag*)

Em séries temporais, nem sempre a influência de uma variável sobre outra ocorre de forma instantânea. Em muitos fenômenos, uma alteração em determinada variável pode produzir efeitos somente após um intervalo de tempo, fazendo com que a maior associação entre as séries não ocorra entre observações registradas no mesmo instante. Dessa forma, a análise da correlação apenas entre valores simultâneos pode não representar adequadamente a dependência existente entre as variáveis, tornando necessária a investigação de possíveis defasagens temporais em sua relação (Morettin; Toloi, 2018).

A correlação com defasagem consiste em calcular uma medida de correlação entre duas séries temporais considerando diferentes deslocamentos temporais (*lags*). Diferentemente da correlação convencional, em que cada observação é comparada apenas com a observação correspondente no mesmo instante, nessa abordagem uma das séries é sucessivamente deslocada para frente ou para trás no tempo, permitindo avaliar se existe uma associação mais forte quando uma variável antecede ou sucede a outra (Morettin; Toloi, 2018).

Essa estratégia pode ser aplicada utilizando diferentes coeficientes de correlação, como Pearson, Spearman ou Kendall (Morettin; Toloi, 2018). Para cada valor de *lag*, calcula-se o coeficiente de correlação entre as séries deslocadas, sendo possível identificar a defasagem temporal na qual a associação entre as variáveis é máxima. O cálculo da correlação com defasagem é definido na Equação 4.

$$r(\tau) = \text{Corr}(X_t, Y_{t+\tau}) \quad (4)$$

Na Equação 4, $r(\tau)$ representa o coeficiente de correlação para uma determinada defasagem τ , correspondente ao *lag* analisado, podendo assumir valores positivos, negativos ou nulos. Corr representa a medida de correlação empregada, como os coeficientes de Pearson, Spearman ou Kendall. Quando $\tau = 0$, não há deslocamento entre as séries temporais, de modo que a correlação é calculada entre observações do mesmo instante.

A análise de correlação constitui uma etapa importante do pré-processamento, pois permite compreender as relações existentes entre as variáveis. Nesse contexto, a próxima etapa consiste na engenharia de atributos, responsável por transformar e criar atributos capazes de representar de forma mais eficiente as características do problema.

2.2.2 Engenharia de atributos

A qualidade das variáveis utilizadas como entrada exerce influência direta no desempenho de modelos de aprendizado de máquina. Dessa forma, além da seleção das variáveis mais relevantes, é comum realizar transformações que permitam representar de maneira

mais adequada as características presentes nos dados. Esse processo é conhecido como engenharia de atributos e consiste na criação, modificação ou representação de váraiveis com o objetivo de facilitar a identificação de padrões pelos algoritmos de aprendizado, tornando as informações mais representativas para a tarefa de modelagem (Bansal; Balaji; Lalani, 2025).

Representação de variáveis cíclicas

Em problemas de aprendizado de máquina, diferentes tipos de variáveis podem ser utilizados como entrada para o treinamento dos modelos. Algumas variáveis possuem comportamento cíclico, de modo que sua representação numérica convencional não preserva adequadamente a relação existente entre seus valores. Exemplos desse tipo de variável incluem a hora do dia, o dia do ano, o mês e a direção do vento.

Quando essas variáveis são representadas apenas por valores inteiros, o modelo tende a interpretá-las como pertencentes a uma escala linear. Por exemplo, o modelo tende a considerar que a distância entre as horas 23:00 e 00:00 é igual a 23 unidades, quando, na realidade, a hora 00:00 é consecutiva a hora 23:00. De maneira análoga, na direção do vento, os ângulos de 359° e 0° representam direções praticamente idênticas, embora seus valores numéricos sejam bastante distintos. Essa representação pode dificultar o aprendizado de padrões associados à natureza periódica dessas variáveis (Bansal; Balaji; Lalani, 2025).

Para contornar esse problema, é comum empregar transformações capazes de preservar sua periodicidade antes do treinamento dos modelos. Uma das abordagens mais utilizadas consiste em representar a variável por meio das funções seno e cosseno. Como essas funções descrevem uma trajetória circular ao longo de um período completo, elas mantêm a continuidade entre o final e o início do ciclo, preservando a proximidade entre valores consecutivos. Considerando uma variável cíclica normalizada em um período P , sua representação pode ser obtida por meio das Equações 5 e 6.

$$x_{\text{sen}} = \sin\left(\frac{2\pi x}{P}\right) \quad (5)$$

$$x_{\text{cos}} = \cos\left(\frac{2\pi x}{P}\right) \quad (6)$$

Nas Equações 5 e 6, x representa o valor da variável cíclica e P corresponde ao período completo do ciclo, como 24 para horas do dia, 12 para meses do ano ou 360 para direções em graus (Bansal; Balaji; Lalani, 2025). A utilização simultânea das funções seno e cosseno permite que cada valor da variável seja representado por um ponto em uma circunferência unitária, preservando corretamente as distâncias angulares entre as observações. Dessa forma, instantes como 23:00 e 0:00 passam a possuir representações próximas entre si refletindo adequadamente a continuidade existente nessas variáveis cíclicas.

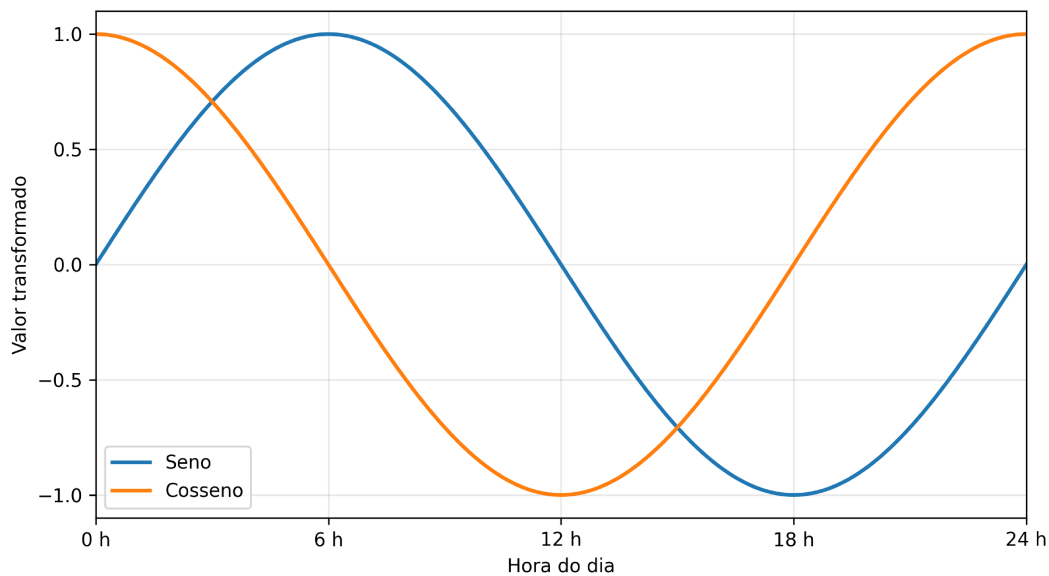


Figura 3 – Representação da hora do dia em seno e cosseno.

A Figura 3 apresenta o comportamento das funções seno e cosseno ao longo de um período completo, evidenciando como ambas descrevem continuamente um ciclo da hora em um dia e retornam ao ponto inicial sem descontinuidades. Além da engenharia de atributos, outra etapa importante do pré-processamento consiste na adequação da escala das variáveis. A subseção seguinte apresenta as técnicas de transformação de escala utilizadas neste trabalho.

2.2.3 Transformações de escala

Na construção de um modelo de aprendizado de máquina é comum a utilização de vários dados de entrada e, frequentemente, tais dados podem apresentar escalas distintas. Enquanto uma variável pode assumir valores entre 0 e 1, outra pode variar entre centenas ou milhares de unidades. Essa diferença de magnitude pode dificultar o treinamento de diversos algoritmos, uma vez que atributos com valores numéricos elevados tendem a exercer maior influência durante o processo de treinamento do modelo (Carvalho; Menezes; Bonidia, 2024).

Com o objetivo de tornar as variáveis comparáveis e facilitar o processo de aprendizagem, é comum aplicar técnicas de transformação de escala aos dados de entrada. Essas técnicas modificam a representação numérica das variáveis sem alterar sua distribuição relativa ou as relações existentes entre as observações. Entre os métodos mais utilizados destacam-se a normalização *Min-Max* e a padronização *Z-score*.

Normalização *Min-Max*

O método de normalização, também conhecido como *Min-Max*, consiste em transformar os valores de uma variável para um intervalo predefinido entre 0 e 1. Essa transformação é realizada utilizando os valores mínimo e máximo observados na própria variável, conforme apresentado na Equação 7.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

Na Equação 7, x representa o valor original da observação; $x_{\min}$ e $x_{\max}$ correspondem, respectivamente, aos menores e maiores valores observados para a variável; e x' representa o valor normalizado. Como resultado, o menor valor da variável passa a ser representado por 0, o maior por 1, e os demais valores são distribuídos proporcionalmente dentro desse intervalo (Carvalho; Menezes; Bonidia, 2024).

Padronização *Z-score*

A padronização, também conhecida como *Z-score*, transforma os dados considerando a média e o desvio padrão da variável. Diferentemente da normalização, esse método não restringe os valores a um intervalo fixo, mas produz uma nova escala em que a média é igual a 0 e o desvio padrão é igual a 1. A transformação é dada pela Equação 8.

$$z = \frac{x - \mu}{\sigma} \quad (8)$$

Na Equação 8, x representa o valor original da observação, μ corresponde à média da variável, σ representa o desvio padrão e z é o valor padronizado. Após essa transformação, valores positivos indicam observações acima da média, enquanto valores negativos representam observações abaixo da média (Carvalho; Menezes; Bonidia, 2024).

2.3 Modelos clássicos de previsão

Modelos de previsão são utilizados para estimar valores futuros de uma variável com base em observações históricas. Para isso, esses modelos buscam identificar e representar os padrões e as relações existentes nos dados históricos, permitindo inferir o comportamento futuro da variável de interesse. Dependendo da abordagem empregada, esses padrões podem ser modelados por meio de hipóteses estatísticas, técnicas de aprendizado de máquina ou estratégias simplificadas de referência (Morettin; Toloi, 2018).

Nesta seção são apresentados três modelos de previsão empregados neste trabalho. O modelo de persistência foi adotado como referência por representar uma estratégia simples e amplamente utilizada na previsão de geração fotovoltaica. O modelo ARIMA é um representante dos métodos estatísticos para séries temporais, enquanto a regressão

Ridge constitui uma abordagem de aprendizado de máquina baseada em regressão linear regularizada. Vale ressaltar que além desses modelos, a Seção 2.4 apresenta outros dois modelos baseados em redes neurais artificiais: *Multilayer Perceptron* (MLP) e *Long Short-Term Memory* (LSTM).

2.3.1 Persistência

O modelo de persistência é amplamente utilizado como referência na avaliação de modelos de previsão. Esse modelo assume que o valor previsto para o próximo instante da série temporal é igual ao último valor observado, não realizando qualquer modelagem dos padrões presentes na série. Em outras palavras, considera-se que não haverá variação entre o instante atual e o seguinte (Yang, 2019).

Devido à sua simplicidade, o modelo de persistência não requer etapa de treinamento, ajuste de parâmetros ou processamento computacional significativo, apresentando custo computacional praticamente nulo. Por esse motivo, é frequentemente empregado como base de comparação para modelos mais sofisticados, permitindo avaliar se o aumento da complexidade computacional e do esforço de modelagem resulta, de fato, em ganhos relevantes de desempenho preditivo. Dessa forma, a persistência estabelece um limite mínimo de desempenho esperado, servindo como um importante indicador da relação entre custo computacional e qualidade das previsões obtidas (Yang, 2019).

$$\hat{y}_{t+1} = y_t \quad (9)$$

A Equação 9 define o modelo de persistência, em que $\hat{y}_{t+1}$ representa a geração prevista para o próximo instante e y_t corresponde ao valor observado no instante atual.

2.3.2 Autoregressive Integrated Moving Average (ARIMA)

O modelo *Autoregressive Integrated Moving Average* (ARIMA) é um dos métodos estatísticos mais tradicionais para previsão de séries temporais. Proposto por Box e Jenkins (1976), esse modelo busca representar a estrutura temporal de uma série por meio da combinação de componentes autorregressivos, de diferenciação e de médias móveis, sendo amplamente empregado em problemas de previsão devido à sua capacidade de modelar dependências temporais presentes nos dados (Morettin; Toloi, 2018).

A notação $ARIMA(p, d, q)$ indica a configuração do modelo, em que p representa a ordem da componente autorregressiva (AR), d corresponde ao número de diferenciações (I) aplicadas à série para torná-la estacionária e q representa a ordem da componente de médias móveis (MA). A etapa de diferenciação tem como objetivo estabilizar a média da série temporal, permitindo que as componentes AR e MA sejam adequadamente ajustadas (Morettin; Toloi, 2018).

A componente autorregressiva modela a relação entre o valor atual da série e seus valores passados, enquanto a componente de médias móveis avalia a influência dos erros de previsão observados em instantes anteriores. A combinação dessas três componentes permite representar diferentes padrões temporais, tornando o modelo adequado para uma ampla variedade de aplicações em previsão. Um modelo $ARIMA(p, d, q)$ pode ser representado pela Equação 10.

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t, \quad (10)$$

Na Equação 10, y_t representa o valor da série temporal no instante t , B é o operador de defasagem, $\phi(B)$ e $\theta(B)$ correspondem, respectivamente, aos polinômios autorregressivo e de médias móveis, d é a ordem de diferenciação, c é uma constante e ε_t representa um ruído branco com média zero e variância constante.

Embora o ARIMA apresente bom desempenho para séries temporais aproximadamente lineares e estacionárias, sua capacidade de representar relações altamente não lineares é limitada. Ainda assim, o modelo permanece válido para comparação com abordagens baseadas em aprendizado de máquina.

2.3.3 Regressão Ridge

Os modelos de regressão são métodos estatísticos utilizados para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes, permitindo estimar o comportamento da variável de interesse a partir das informações fornecidas pelas variáveis preditoras (Runkler, 2012).

Na regressão linear, a estimativa dos coeficientes é obtida pela minimização da soma dos quadrados dos resíduos, buscando atingir a função objetivo apresentada na Equação 11. Essa abordagem pretende encontrar os coeficientes que minimizam a diferença entre os valores observados e os valores estimados pelo modelo.

$$\min_{\beta} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \quad (11)$$

Na Equação 11, y_i representa o valor observado da variável dependente para a i -ésima observação, x_{ij} corresponde ao valor da j -ésima variável preditora da i -ésima observação, β_0 é o intercepto do modelo e β_j representa o coeficiente associado à j -ésima variável independente (Hastie; Tibshirani; Friedman, 2009).

Entretanto, quando existe multicolinearidade entre as variáveis preditoras ou há risco de *overfitting*, os coeficientes estimados pela regressão linear podem apresentar elevada variância, tornando o modelo sensível a pequenas variações nos dados. Para contornar esse problema, a regressão Ridge modifica a função objetivo da regressão linear incluindo um termo de regularização.

O termo de regularização é somado à função de custo, e é responsável por penalizar coeficientes de grande magnitude. Essa penalização reduz a variância do modelo, tornando as previsões mais estáveis e melhorando sua capacidade de generalização para novos dados. O termo de penalização é mostrado na Equação 12 enquanto a função objetivo final na regressão Ridge é descrita pela Equação 13 (Hastie; Tibshirani; Friedman, 2009).

$$\lambda \sum_{j=1}^p \beta_j^2 \quad (12)$$

$$\min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (13)$$

Nas Equações 12 e 13, λ é o parâmetro de regularização responsável por controlar a intensidade da penalização aplicada aos coeficientes do modelo. Quanto maior o valor de λ , maior será a penalização sobre coeficientes de grande magnitude, reduzindo a variância do modelo e aumentando sua capacidade de generalização.

A regressão Ridge apresenta baixo custo computacional e boa capacidade de generalização, o que a torna um bom modelo de referência em problemas de previsão. Dessa forma, seu desempenho pode ser comparado ao de abordagens mais sofisticadas, como os modelos baseados em aprendizado de máquina. Nesse contexto, a próxima seção apresenta os principais métodos de aprendizado de máquina, destacando sua capacidade de construir modelos mais inteligentes e adaptativos.

2.4 Aprendizado de máquina

Aprendizado de máquina é uma área de estudo da Inteligência Artificial (IA) dedicada ao desenvolvimento de algoritmos e técnicas que conferem a modelos computacionais a capacidade de aprender a partir de dados, sem que sejam explicitamente programados para cada tarefa. Segundo Mitchell (1997):

Um programa de computador é dito aprender com a experiência E, com relação a alguma classe de tarefas T e medida de desempenho P, se seu desempenho em tarefas de T, medido por P, melhora com a experiência E.

A aprendizagem de modelos pode ser descrita por 4 etapas principais: definição de qual componente do modelo deve melhorar, ou seja, aprender, representação do conhecimento prévio que o agente já possui, representação dos novos dados disponíveis para a aprendizagem e disponibilização de um *feedback* para o agente (Russell; Norvig, 2013).

Existem três principais tipos de aprendizado baseado no *feedback* fornecido ao agente, são eles: aprendizagem não supervisionada, aprendizagem por reforço e aprendizagem supervisionada. Na aprendizagem não supervisionada o agente aprende por padrões na

entrada sem ter um *feedback* explícito, tendo como exemplo mais comum tarefas de agrupamento. Na aprendizagem por reforço o agente recebe recompensas ou punições como *feedback* dependendo de seu desempenho, como exemplos existem jogos competitivos onde o agente ganha ou perde ao final. Entretanto, cabe a ele decidir quais ações levaram ao ganho ou perda da partida. Por fim, existe a aprendizagem supervisionada, onde o agente recebe tanto um conjunto de entrada quanto um conjunto de saída e assim consegue construir uma relação entre eles (Russell; Norvig, 2013).

As técnicas de aprendizado de máquina são amplamente empregadas em tarefas como classificação, regressão, agrupamento, entre outras. No contexto deste trabalho, a previsão da geração de energia solar enquadra-se como um problema de regressão, cujo objetivo é estimar valores contínuos de potência gerada a partir dos dados de entrada em uma série temporal. Dentre as técnicas de aprendizado de máquina, destacam-se as redes neurais artificiais (RNA), que serão utilizadas neste trabalho sob a abordagem de aprendizado supervisionado. As RNA demonstram bom desempenho em problemas de séries temporais, sendo capazes de capturar dependências complexas e padrões não lineares nos dados. A seguir, apresenta-se o referencial teórico sobre redes neurais artificiais, contemplando sua estrutura, funcionamento e principais características.

2.4.1 Redes neurais artificiais (RNA)

As redes neurais artificiais (RNA) são modelos computacionais do campo de aprendizado de máquina, desenvolvidos com o objetivo de simular, de forma simplificada, o funcionamento do cérebro humano e sua capacidade de aprender. O cérebro realiza o processamento de informações por meio de uma complexa sequência de interações eletroquímicas entre neurônios. De maneira análoga, uma rede neural artificial é composta por unidades denominadas neurônios artificiais, interconectadas entre si, que processam e transmitem informações por meio de conexões ponderadas, simulando o comportamento das redes neurais biológicas (Russell; Norvig, 2013).

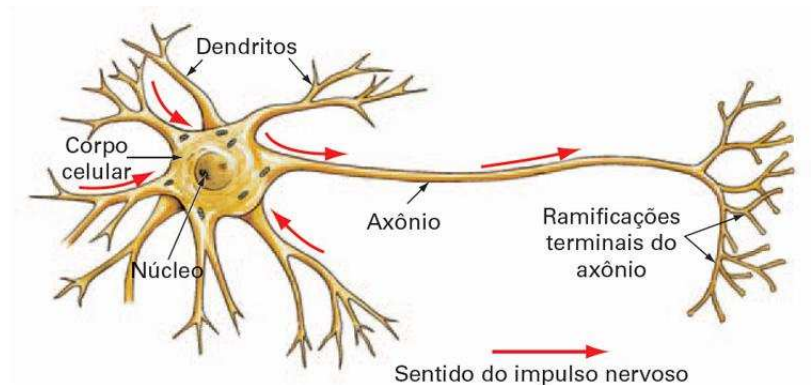


Figura 4 – Representação simplificada de um neurônio biológico. Fonte: Deeplearning Book (2024).

A Figura 4 representa um modelo de neurônio humano. O neurônio é a unidade básica do funcionamento do sistema nervoso humano, responsável por receber, processar e propagar impulsos elétricos. Um neurônio é composto por três partes principais: dendritos, corpo celular e axônio. Os dendritos são ramificações finas responsáveis por captar impulsos elétricos provenientes do ambiente ou de outras células e conduzi-los até o corpo celular. O corpo celular é a região onde se localiza o núcleo da célula nervosa, sendo responsável, principalmente, por processar os impulsos recebidos e determinar quais estímulos devem ser transmitidos adiante. O axônio, por sua vez, recebe os impulsos do corpo celular e os conduz, distribuindo-os para o ambiente ou para outras células (KhanAcademy, 2020).

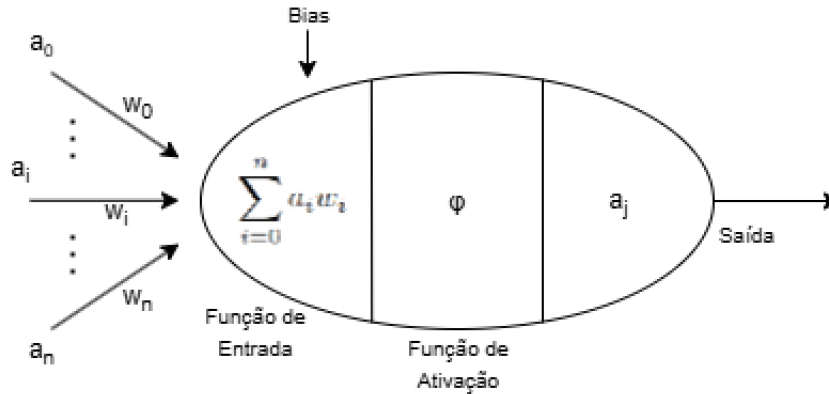


Figura 5 – Modelo de neurônio matemático proposto por McCulloch e Pitts (1943). Fonte: Elaborado pelo autor (2026).

A Figura 5 apresenta o primeiro modelo matemático, desenvolvido em 1943 por McCulloch e Pitts (1943), para representação e simulação artificial de um neurônio. O neurônio artificial é a unidade básica de uma RNA e é composto por três elementos principais: a entrada dos dados, a função de ativação e a saída dos dados. É possível estabelecer um paralelo entre o neurônio biológico e o artificial, dessa maneira, os dendritos correspondem à entrada de dados, o corpo celular à função de ativação e o axônio à saída (Russell; Norvig, 2013).

O funcionamento de um neurônio artificial tem início na entrada de dados. Nessa etapa, o neurônio recebe múltiplos valores de entrada, cada um associado a um peso específico, e realiza uma combinação linear por meio de uma função somatório. Além das entradas ponderadas, é adicionado um termo chamado *bias*, que desloca a função de ativação no espaço de saída. Esse termo aumenta a flexibilidade do modelo, permitindo que o neurônio gere uma saída não nula mesmo quando todas as entradas possuem valor zero, além de facilitar o ajuste do limiar de ativação durante o treinamento. O somatório da equação de entrada é especificado na Equação 14.

$$z = \sum_{i=1}^n a_i w_i + bias \quad (14)$$

Após a combinação linear feita na entrada dos dados, o neurônio possui um único valor a ser processado. Esse valor é então submetido a função de ativação. A função de ativação tem como principal objetivo introduzir não linearidade na transformação realizada pelo neurônio, ampliando sua capacidade de aprender padrões complexos e de executar tarefas que exigem modelagem não linear. Existem diversos tipos de funções de ativação, cada uma mais adequada a contextos e problemas específicos, como, por exemplo, as funções linear, sigmoide, ReLU, entre outras. A função de ativação é usualmente representada pelo símbolo φ , conforme expresso na Equação 15.

$$a_j = \varphi(z) \quad (15)$$

Concluído o processamento pela função de ativação, a saída do neurônio é responsável por encaminhar o resultado produzido às demais partes da rede. Uma rede neural artificial é composta por vários neurônios interconectados por ligações direcionadas em camadas que permitem o processamento progressivo das informações.

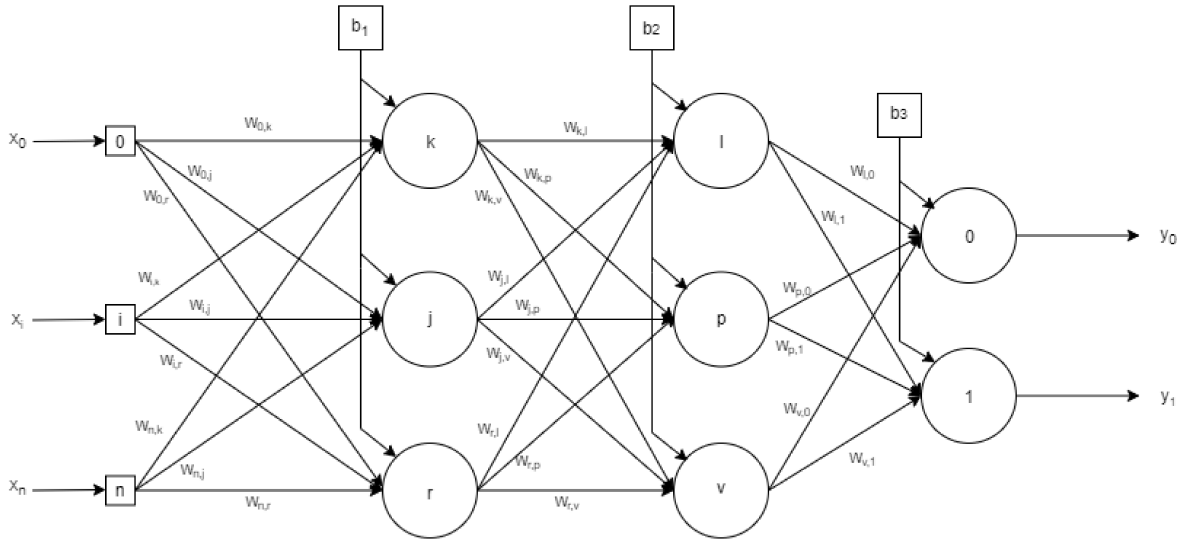


Figura 6 – Exemplo de rede neural artificial. Fonte: Tiago (2020).

Na Figura 6 é representado o modelo genérico de uma RNA. A primeira camada é denominada camada de entrada, responsável por receber os dados brutos e repassá-los à rede para processamento. Em seguida, encontram-se as camadas ocultas, que realizam as transformações sobre os dados por meio de combinações ponderadas e funções de ativação, extraindo padrões e relações complexas. Por fim, a camada de saída fornece o resultado final do processamento, que pode corresponder, por exemplo, a uma classificação ou previsão numérica.

As redes neurais podem ser classificadas de acordo com sua arquitetura e fluxo de informação. As redes *feedforward* são as mais comuns, nas quais os dados fluem apenas em uma direção. Já as redes neurais recorrentes (RNN) possuem conexões que permitem o retorno de informações, tornando-as adequadas para o processamento de dados sequenciais e temporais. Existem ainda redes especializadas, como as redes neurais convolucionais (CNN), amplamente utilizadas em visão computacional, e as redes neurais profundas (DNN), que empregam múltiplas camadas ocultas para capturar padrões complexos e abstratos (Goodfellow; Bengio; Courville, 2016). A arquitetura da rede e o fluxo de informações influenciam tanto na eficácia da rede quanto na escolha da técnica de aprendizado incremental adotada neste trabalho.

As RNA possuem dois conjuntos de elementos essenciais ao seu funcionamento: os parâmetros e os hiperparâmetros. Ambos exercem influência direta sobre a capacidade de aprendizado e o desempenho do modelo, embora desempenhem papéis distintos. Os parâmetros correspondem aos valores ajustados automaticamente durante o processo de treinamento, sendo os principais os pesos e os *bias*. Os pesos determinam a influência de cada entrada sobre o neurônio, enquanto o *bias* permite ajustar o limiar de ativação. Já os hiperparâmetros são definidos previamente para o modelo e controlam aspectos estruturais e de aprendizado, como a taxa de aprendizado, o número de camadas ocultas, a quantidade de neurônios por camada, a função de ativação e o número de épocas de treinamento (Goodfellow; Bengio; Courville, 2016).

Como os hiperparâmetros não são aprendidos automaticamente durante o treinamento, sua definição requer um processo de escolha. Tradicionalmente, essa tarefa é realizada por meio de um *grid search*, que avalia exaustivamente todas as combinações possíveis de um conjunto previamente definido de valores. Embora simples e amplamente utilizada, essa abordagem apresenta elevado custo computacional à medida que aumenta o número de hiperparâmetros e de valores avaliados. Em função dessa limitação, *frameworks* mais recentes, como o *Optuna*¹ e o *Ray Tune*², empregam estratégias de busca e otimização mais eficientes, capazes de explorar o espaço de hiperparâmetros de forma adaptativa e reduzir o custo computacional necessário para encontrar configurações de alto desempenho.

Para que uma rede neural artificial seja capaz de aprender uma tarefa e identificar padrões nos dados, é necessário que ela passe por um processo de treinamento. Antes dessa etapa, são definidos os hiperparâmetros que controlam o processo de aprendizado e permanecem fixos durante todo o treinamento. Em seguida, os pesos e *bias* de cada neurônio são inicializados. Em geral, os pesos recebem valores aleatórios, enquanto os *bias* costumam ser definidos com valores próximos de zero. Uma boa inicialização acelera o processo de aprendizado e favorece a obtenção de melhores resultados.

O treinamento de uma rede neural, baseado no aprendizado supervisionado, consiste no

¹ <<https://optuna.readthedocs.io/en/stable/>>

² <<https://docs.ray.io/en/latest/tune/index.html>>

ajuste iterativo dos parâmetros a partir de um conjunto de dados rotulado. O método mais utilizado para treinamento é o *backpropagation*. O processo inicia-se com a propagação direta, em que as entradas percorrem as camadas da rede até a obtenção de uma saída estimada. Em seguida, calcula-se o erro, definido como a diferença entre a saída obtida pela rede e o valor esperado nos dados reais. Esse erro é então propagado de volta pela rede, camada por camada, ajustando os pesos de forma proporcional ao gradiente do erro em relação a cada parâmetro. O ajuste é realizado por meio de um algoritmo de otimização, geralmente o gradiente descendente, e controlado pela taxa de aprendizado. O processo se repete por várias épocas até que o erro seja minimizado e a rede alcance um desempenho satisfatório.

Após o treinamento, a rede neural é submetida a uma fase de teste, na qual seu desempenho é avaliado sobre um conjunto de dados que não foi utilizado durante o treinamento. Essa etapa é essencial para verificar a capacidade do modelo de produzir respostas corretas para dados não vistos. O desempenho obtido na fase de testes permite identificar se a rede aprendeu padrões reais dos dados ou apenas memorizou o conjunto de treinamento, fenômeno indesejado conhecido como *overfitting*.

A avaliação do desempenho de uma rede neural é realizada por meio de métricas quantitativas que variam conforme o tipo de tarefa. Em problemas de classificação, são comumente utilizadas métricas como acurácia, precisão e a *F1 score*, que expressam a proporção de acertos. Já em problemas de regressão, como o do presente trabalho, são utilizadas medidas como a raiz do erro quadrático médio (RMSE), o erro absoluto médio (MAE) e o coeficiente de determinação (R^2) abordados na Seção 2.6.

Dentre as diferentes arquiteturas de redes neurais artificiais, destacam-se para o presente trabalho a *Multilayer Perceptron* (MLP) e as redes neurais recorrentes (RNN). A MLP é uma arquitetura do tipo *feedforward*, amplamente empregada em problemas de classificação e regressão, sendo apresentada na subseção seguinte. Já as RNN são capazes de processar informações sequenciais por meio da utilização de estados internos que armazenam informações de instantes anteriores, tornando-os adequados para aplicações envolvendo séries temporais. Entre as arquiteturas de RNN, a *Long Short-Term Memory* (LSTM) será utilizada nessa dissertação e abordada posteriormente.

2.4.2 *Multilayer Perceptron* (MLP)

A *Multilayer Perceptron* (MLP) é uma arquitetura de rede neural artificial do tipo *feedforward*, caracterizada pela organização dos neurônios em uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Nessa arquitetura, os neurônios de cada camada são conectados a todos os neurônios da camada seguinte, formando uma estrutura totalmente conectada. As informações percorrem a rede apenas no sentido da entrada para a saída, não existindo conexões recorrentes ou mecanismos de memória (Russell; Norvig, 2013).

Durante o processamento, cada camada recebe as ativações produzidas pela camada anterior, realiza uma combinação linear das entradas ponderadas pelos respectivos pesos, adiciona o termo de *bias* e aplica uma função de ativação não linear. Esse processo é repetido sucessivamente até a camada de saída, responsável por produzir a predição final. Os parâmetros da rede são ajustados durante o treinamento por meio do algoritmo de *backpropagation*, conforme descrito na Seção 2.4.1.

As MLP são amplamente empregadas em problemas de classificação e regressão, apresentando boa capacidade para modelar relações não lineares entre as variáveis de entrada e saída (Russell; Norvig, 2013). Entretanto, por processarem cada amostra de forma independente, essas redes não são capazes de representar explicitamente dependências temporais presentes em dados sequenciais. Em aplicações envolvendo séries temporais, como a previsão de geração fotovoltaica abordada neste trabalho, essa limitação pode comprometer a modelagem de relações de longo prazo, motivando o emprego de arquiteturas recorrentes, em especial a *Long Short-Term Memory* (LSTM), apresentada a seguir.

2.4.3 *Long Short-Term Memory* (LSTM)

Os seres humanos possuem a capacidade de armazenar e recuperar informações relevantes mesmo após longos períodos de tempo. Por exemplo, durante a leitura de um texto, uma informação apresentada nos primeiros parágrafos pode ser utilizada para compreender ou responder a uma pergunta ao final da leitura, ainda que essa informação não tenha sido explicitamente utilizada ao longo do restante do texto. Em contraste, redes neurais recorrentes (RNN) convencionais apresentam dificuldades para preservar informações por longos intervalos temporais devido a um problema conhecido como desaparecimento do gradiente (Bengio; Simard; Frasconi, 1994).

Com o objetivo de superar essa limitação, foi proposta por Hochreiter e Schmidhuber (1997) a arquitetura *Long Short-Term Memory*. A LSTM consiste em um tipo especializado de rede neural recorrente capaz de aprender dependências de curto e longo prazo em dados sequenciais. Sua principal característica é a utilização de mecanismos denominados *gates*, responsáveis por controlar o fluxo de informações ao longo da rede, determinando quais informações devem ser armazenadas, atualizadas ou descartadas. Esse mecanismo seletivo permite que a LSTM preserve informações relevantes por períodos mais longos. A unidade básica dessa arquitetura é denominada célula e é ilustrada na Figura 7.

A Figura 7 apresenta a estrutura interna de uma célula LSTM. A seta horizontal superior representa o estado da célula, denominada C_t , responsável por armazenar o contexto acumulado ao longo da sequência temporal. Como as células LSTM são interconectadas ao longo do tempo, esse estado atua como uma memória capaz de transportar informações relevantes entre diferentes instantes da série temporal. A entrada da célula é composta pelo vetor de entrada atual (x_t) e pelo estado oculto da célula anterior (h_{t-1}), representados pela seta de entrada na parte inferior da figura.

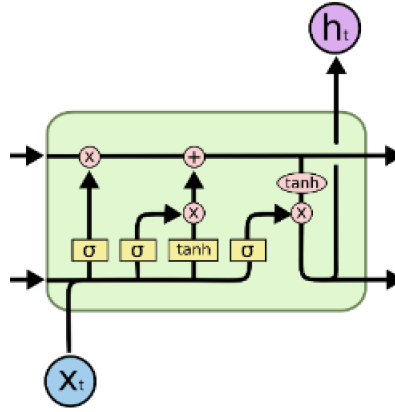


Figura 7 – Representação de uma célula LSTM. Fonte: Adaptada de Olah (2015).

As caixas amarelas representam camadas neurais responsáveis pelo processamento das informações, enquanto os círculos representam operações matemáticas, como multiplicação elemento a elemento, soma e funções de ativação. Os blocos identificados por σ correspondem aos *gates*, cujas saídas assumem valores entre 0 e 1. Dessa forma, cada *gate* atua como um filtro, determinando a proporção da informação que será propagada pela rede (Olah, 2015). A seguir, é apresentado o funcionamento interno de uma célula LSTM, detalhando cada um de seus componentes e as equações responsáveis pelo controle do fluxo de informações ao longo da rede.

O primeiro *gate*, da esquerda para a direita, é denominado *forget gate* e é responsável por controlar quais informações armazenadas no estado da célula anterior (C_{t-1}) devem ser preservadas ou descartadas. Seu funcionamento é descrito pela Equação 16. Nessa operação, uma função sigmoide é aplicada à combinação linear entre o estado oculto da célula anterior (h_{t-1}) e a entrada atual (x_t), ponderados por pesos e somados a um *bias*.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (16)$$

Em seguida, o *input gate* define quanto das novas informações extraídas da entrada atual será incorporado ao estado da célula. Assim como o *forget gate*, esse mecanismo utiliza uma função sigmoide para produzir um vetor de ativação que controla a inclusão de novas informações na memória da rede, conforme apresentado na Equação 17.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (17)$$

Da mesma maneira, uma função *tanh* é utilizada para gerar o estado candidato da célula ($\hat{C}_t$), representado pela Equação 18. Esse vetor contém as novas informações extraídas da entrada atual e que poderão ser incorporadas à memória da rede.

$$\hat{C}_t = \tanh(W_C[h * t - 1, x_t] + b_C) \quad (18)$$

A atualização do estado da célula ocorre pela combinação entre a memória preservada pelo *forget gate* e as novas informações selecionadas pelo *input gate*, resultando no novo estado da célula (C_t), conforme apresentado na Equação 19.

$$C_t = f_t \odot C_{t-1} + i_t \odot \hat{C}_t \quad (19)$$

Após a atualização da memória, o *output gate* determina quais informações do estado da célula serão disponibilizadas na saída da rede. Seu cálculo é realizado por meio de uma função sigmoide aplicada às entradas da célula, conforme descrito na Equação 20.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (20)$$

Por fim, o estado oculto da célula (h_t) é obtido pela combinação entre a saída do *output gate* e o estado atualizado da célula, após a aplicação da função *tanh*, conforme apresentado na Equação 21. A utilização da função *tanh* contribui para manter os valores da rede em uma faixa limitada, permitindo uma representação mais controlada das informações armazenadas e propagadas pela célula.

$$h_t = o_t \odot \tanh(C_t) \quad (21)$$

A capacidade da LSTM de armazenar e selecionar informações relevantes ao longo do tempo torna essa arquitetura particularmente adequada para problemas envolvendo séries temporais. Diferentemente de modelos tradicionais que analisam observações de forma isolada, a LSTM é capaz de explorar dependências temporais presentes nos dados, permitindo que eventos passados influenciem as previsões futuras. Essa característica é especialmente importante em aplicações de previsão de geração fotovoltaica, nas quais fatores como condições meteorológicas, sazonalidade e padrões operacionais exercem influência sobre o comportamento da série. Dessa forma, a utilização de redes LSTM possibilita a construção de modelos mais capazes de representar a dinâmica temporal do sistema e, consequentemente, produzir previsões mais precisas (Kandananond; Chotewiwat; Wong-sathan, 2021).

Apesar dessas vantagens, após o processo de treinamento a LSTM mantém seus parâmetros fixos, tornando-se um modelo estático. Embora essa característica seja adequada em diversos cenários, ela pode representar uma limitação em aplicações cujos dados apresentam mudanças ao longo do tempo. No contexto da geração de energia solar, alterações climáticas, sazonais, operacionais e ambientais podem modificar gradualmente o comportamento da série, reduzindo a capacidade de generalização de modelos treinados apenas com dados históricos. Nesse cenário, torna-se necessário empregar estratégias que permitam a adaptação contínua do modelo às novas condições observadas. Entre essas estratégias destaca-se o aprendizado incremental, apresentado na próxima subseção, que possibilita a atualização do conhecimento da rede à medida que novos dados se tornam disponíveis.

2.4.4 Aprendizado incremental

Uma das principais características da aprendizagem humana é a capacidade de adaptar-se continuamente a novas informações e mudanças no ambiente ao longo do tempo. O cérebro desenvolveu a habilidade de adquirir e armazenar conhecimento de forma incremental, adaptando-se continuamente a novas tarefas e experiências. A teoria dos sistemas de aprendizado complementares (McClelland; McNaughton; O'Reilly, 1995) postula que o hipocampo é responsável por um aprendizado rápido e adaptativo de curto prazo, enquanto o córtex cerebral atua de forma mais lenta, porém com maior capacidade de retenção e generalização, formando a base do aprendizado de longo prazo.

Dessa maneira, os princípios biológicos do cérebro inspiram o desenvolvimento de abordagens computacionais capazes de reproduzir essa forma de aprendizado contínuo e adaptativo. Os modelos de aprendizado de máquina tendem a ser treinados, de forma convencional, utilizando conjuntos de dados estáticos, bem rotulados e identicamente distribuídos, o que limita sua capacidade de adaptação a novos contextos. Assim, o aprendizado incremental surge como uma abordagem voltada a permitir que um modelo continue aprendendo de forma contínua a partir de novos dados que são gerados. A importância do aprendizado incremental surge na representação de fenômenos do mundo real, especialmente em séries temporais, nas quais os dados são constantemente atualizados (Pua et al., 2021).

Dessa forma, o modelo é capaz de assimilar novos padrões sem esquecer o conhecimento previamente adquirido. Nos últimos anos, essa abordagem tem ganhado destaque em diversas áreas, como robótica inteligente, condução autônoma e veículos aéreos não tripulados, consolidando-se como um dos grandes avanços em sistemas de IA contínua (Luo et al., 2020).

De acordo com a taxonomia proposta por Ven, Tuytelaars e Tolias (2022) o aprendizado incremental pode ser classificado em três cenários: tarefa, domínio e classe. No cenário por tarefa, o modelo aprende um conjunto de tarefas distintas e mantém uma separação clara entre elas. Já no cenário por classe, o modelo deve aprender novas classes durante o processo de aprendizagem incremental, exigindo que reconheça todas as classes aprendidas de maneira unificada. Por fim, no cenário por domínio, o modelo enfrenta uma única tarefa, mas os dados são provenientes de domínios que mudam ao longo do tempo, como diferentes condições ambientais ou variações sazonais.

O presente trabalho enquadra-se em um aprendizado incremental por domínio pois o objetivo é desenvolver um modelo capaz de se adaptar continuamente às mudanças nas condições de geração de energia solar, sem a necessidade de recomeçar o treinamento a cada nova atualização de dados. Nesse contexto, a tarefa de previsão permanece a mesma, mas o domínio dos dados varia em função de fatores como a sazonalidade, a irradiância solar, a temperatura e as condições atmosféricas. Assim, o modelo precisa ajustar-se dinamicamente a esses novos padrões, preservando o conhecimento adquirido em períodos

anteriores e garantindo previsões consistentes ao longo do tempo.

Dois tipos de redes neurais artificiais serão utilizadas para previsão da geração de energia solar nesse trabalho. Sendo assim, é necessário considerar as dificuldades de adaptação que esses modelos apresentam quando inseridos em um processo de aprendizado incremental. Durante o treinamento, uma rede neural ajusta iterativamente os seus parâmetros por meio do algoritmo *backpropagation*, conforme detalhado na Seção 2.4.1. No entanto, no contexto do aprendizado incremental, emerge um desafio particular conhecido como esquecimento catastrófico. Esse fenômeno ocorre quando o reajuste dos pesos da rede, necessário para incorporar novos conhecimentos, acaba por modificar representações previamente consolidadas, resultando na perda de informações anteriormente aprendidas (Luo et al., 2020).

A mitigação do esquecimento catastrófico exige que o modelo mantenha um equilíbrio entre duas propriedades: plasticidade e estabilidade. A plasticidade confere ao modelo a capacidade de adaptar-se a novos dados e incorporar informações recentes, enquanto a estabilidade garante a retenção do conhecimento previamente aprendido. No entanto, conciliar essas duas características simultaneamente é um desafio, conhecido como dilema estabilidade-plasticidade (Abraham; Robins, 2005). Se os pesos da rede permanecerem excessivamente estáveis, o modelo tende a não aprender novos padrões, por outro lado, uma grande plasticidade pode causar a rápida degradação do conhecimento previamente adquirido.

O dilema estabilidade-plasticidade também é observado nos sistemas biológicos de aprendizado, nos quais o cérebro busca equilibrar a incorporação de novas informações e a preservação do conhecimento consolidado. Estudos em neurociência indicam que existem mecanismos responsáveis por regular esse equilíbrio em diferentes áreas cerebrais, e que os sistemas cognitivos se desenvolvem de acordo com os estímulos externos recebidos (Zenke; Gerstner; Ganguli, 2017).

Com base nas observações em sistemas biológicos, diversas abordagens foram propostas para implementar o aprendizado incremental em redes neurais artificiais, cada uma buscando equilibrar a estabilidade e a plasticidade de maneiras distintas. Entre elas, destacam-se as estratégias baseadas em ajuste contínuo dos parâmetros do modelo (*fine-tuning*), as abordagens arquiteturais, que expandem dinamicamente a estrutura da rede para acomodar novos conhecimentos, as técnicas baseadas em regularização, que restringem alterações em parâmetros considerados importantes para tarefas anteriores, e os métodos de relembramento (*rehearsal*), que preservam experiências passadas por meio do armazenamento ou da reconstrução de amostras previamente observadas (Luo et al., 2020).

As abordagens arquiteturais incluem métodos como as *Progressive Neural Network* (PNN) (Rusu et al., 2022) e as *Dynamic Expansion Network* (DEN) (Yoon et al., 2018). Entre as técnicas baseadas em regularização, destaca-se o EWC (Kirkpatrick et al., 2017),

enquanto os métodos de *rehearsal* compreendem estratégias de *replay* e *pseudo-replay* (Park et al., 2025). Neste trabalho, serão adotadas as abordagens de *fine-tuning* e *Elastic Weight Consolidation* (EWC).

Fine-tuning

O *fine-tuning* é uma das abordagens mais simples e amplamente utilizadas em aprendizado incremental. Nessa estratégia, um modelo previamente treinado continua seu processo de aprendizado à medida que novos dados se tornam disponíveis, utilizando os parâmetros já aprendidos como ponto de partida para novas atualizações. Em vez de reinicializar os pesos da rede e realizar um novo treinamento completo, o algoritmo de *backpropagation* é executado novamente utilizando apenas os novos dados.

Nesse processo, entretanto, os parâmetros do modelo não são inicializados aleatoriamente, mas sim com os valores aprendidos durante o treinamento anterior. A partir desses parâmetros, a rede realiza novas atualizações iterativas por meio do novo treinamento, ajustando seus pesos de acordo com os novos dados disponíveis. Dessa forma, preserva-se parte do conhecimento previamente adquirido e permite que o modelo se adapte gradualmente às mudanças presentes nos dados (Parisi et al., 2019).

Apesar de sua simplicidade e eficiência computacional, o *fine-tuning* é particularmente suscetível ao fenômeno do esquecimento catastrófico, uma vez que a atualização dos parâmetros com base em novos dados pode comprometer conhecimentos adquiridos em etapas anteriores. Embora a utilização de taxas de aprendizado reduzidas torne as atualizações mais conservadoras e atenuie parcialmente esse problema, ela não impede que informações relevantes sejam sobrescritas ao longo do treinamento. Por esse motivo, o *fine-tuning* é frequentemente utilizado como abordagem de referência para comparação com métodos mais sofisticados de aprendizado incremental, como aqueles baseados em regularização.

***Elastic Weight Consolidation* (EWC)**

O *Elastic Weight Consolidation* (EWC) foi proposto por Kirkpatrick et al. (2017) como uma estratégia para mitigar o problema do esquecimento catastrófico em redes neurais. Diferentemente do *fine-tuning*, no qual todos os parâmetros do modelo são atualizados livremente durante o novo treinamento, o EWC busca preservar o conhecimento previamente adquirido restringindo alterações em parâmetros considerados importantes para as tarefas anteriormente aprendidas.

Para isso, o método assume que nem todos os pesos da rede contribuem igualmente para o desempenho do modelo. Durante o treinamento inicial, alguns parâmetros tornam-se mais relevantes para a representação do conhecimento aprendido, enquanto outros exercem menor influência. Assim, ao iniciar o treinamento com novos dados, o EWC aplica uma penalização às alterações realizadas nos parâmetros mais importantes, permitindo

maior flexibilidade apenas na atualização daqueles considerados menos relevantes. Dessa forma, o modelo é capaz de incorporar novas informações enquanto reduz a degradação do conhecimento previamente adquirido.

Essa restrição é incorporada à função de perda por meio de um termo de regularização que penaliza desvios em relação aos valores previamente aprendidos, sendo a intensidade dessa penalização proporcional à importância estimada de cada parâmetro. A Equação 22 expõe a função de perda total somada ao termo de regularização do EWC.

$$L(\theta) = L_{new}(\theta) + \frac{\lambda}{2} \sum_i F_i (\theta_i - \theta_i^*)^2 \quad (22)$$

Na Equação 22, $L(\theta)$ representa a função de perda total do modelo em função dos seus parâmetros; $L_{new}(\theta)$ representa a função de perda associada aos novos dados utilizados no treinamento incremental; θ corresponde ao conjunto de parâmetros do modelo durante a etapa de atualização; θ_i^* representa o valor do parâmetro i obtido após o treinamento anterior; F_i corresponde ao elemento diagonal da matriz de Informação de Fisher, utilizada para estimar a importância de cada parâmetro para o conhecimento previamente aprendido; e λ é um hiperparâmetro que controla a intensidade da regularização, estabelecendo o equilíbrio entre a preservação do conhecimento anterior e a incorporação de novas informações (Kirkpatrick et al., 2017).

Observa-se que o primeiro termo da Equação 22 corresponde à função de perda calculada sobre os novos dados, enquanto o segundo termo atua como um termo de regularização que penaliza alterações nos parâmetros previamente aprendidos. Quanto maior o valor de F_i , maior é a importância atribuída ao parâmetro i e, consequentemente, maior será a penalização caso seu valor seja modificado durante o treinamento incremental. Por outro lado, parâmetros com baixos valores de F_i podem ser ajustados com maior liberdade, permitindo que o modelo incorpore novos conhecimentos sem comprometer significativamente o desempenho em tarefas aprendidas anteriormente (Kirkpatrick et al., 2017).

Embora o aprendizado incremental permita que modelos se adaptem continuamente a novos dados, a realização de atualizações frequentes nem sempre é desejável. Dessa forma, torna-se importante estabelecer um protocolo capaz de identificar os momentos em que a atualização do modelo é realmente necessária. Nesse sentido, um protocolo de monitoramento de degradação pode detectar reduções significativas na acurácia das previsões, indicando que o modelo se tornou obsoleto diante de mudanças nas condições operacionais ou ambientais. A partir dessa detecção, o processo de aprendizado incremental pode ser acionado para atualizar a rede com os novos dados, restaurando seu desempenho. Essa abordagem integrada permite que o sistema mantenha previsões consistentes e estáveis ao longo do tempo. Na Seção 2.5 é apresentado o referencial teórico sobre monitoramento de degradação.

2.5 Monitoramento de degradação

O monitoramento de degradação de modelos de aprendizado de máquina consiste em um processo contínuo de avaliação do desempenho preditivo de um modelo ao longo do tempo, com o objetivo de identificar perdas de acurácia ou desvios em relação ao comportamento esperado. Essa degradação pode ocorrer devido à alteração gradual das condições que regem o fenômeno modelado, denominada mudança de conceito (*concept drift*) (Agrahari; Singh, 2022).

Em cenários reais, como na previsão de geração de energia solar, fatores externos como variações sazonais, envelhecimento dos equipamentos, acúmulo de sujeira nos módulos fotovoltaicos ou mudanças nas condições atmosféricas podem modificar a relação entre as variáveis de entrada e saída, tornando o modelo previamente treinado menos eficiente. Assim, o monitoramento de degradação atua como uma etapa essencial para garantir a confiabilidade e a adaptabilidade dos modelos, permitindo a detecção precoce de obsolescência e o acionamento de estratégias de atualização, como o aprendizado incremental.

Um dos métodos mais tradicionais para o monitoramento da estabilidade é o controle estatístico de processo (SPC), introduzido por Shewhart (1931). O SPC baseia-se na utilização de gráficos de controle para distinguir variações naturais do processo de alterações que indiquem perda de estabilidade ou degradação de desempenho. O princípio fundamental consiste em comparar observações correntes com limites estatísticos obtidos a partir de um período de operação considerado normal. No gráfico de controle de Shewhart, esses limites são normalmente estabelecidos em três desvios-padrão acima e abaixo da média do processo (Shewhart, 1931).

Apesar de sua ampla utilização, os gráficos de Shewhart apresentam sensibilidade limitada à detecção de pequenas mudanças persistentes no processo. Em aplicações de previsão de geração fotovoltaica, a degradação do desempenho dos modelos frequentemente ocorre de forma gradual, em decorrência de mudanças sazonais, alterações nas condições meteorológicas ou deriva dos dados de entrada. Nesses cenários, pequenas elevações sucessivas das métricas de erro podem demorar a ser detectadas pelos limites de controle tradicionais.

Devido às limitações dos gráficos de Shewhart na detecção de mudanças graduais, diferentes métodos de monitoramento foram propostos com o objetivo de identificar alterações no comportamento dos dados de forma mais sensível. Entre eles, destacam-se o *Cumulative Sum Control Chart* (CUSUM) e o *Adaptive Windowing* (ADWIN), adotados nesta dissertação para o monitoramento da degradação dos modelos.

2.5.1 *Cumulative Sum Control Chart* (CUSUM)

O método CUSUM, proposto por Page (1954) foi desenvolvido especificamente para identificar mudanças pequenas e persistentes na média de um processo. Diferentemente

dos gráficos de Shewhart, que analisam cada observação individualmente, o CUSUM acumula os desvios observados ao longo do tempo, tornando-se mais sensível à degradação contínua do desempenho.

$$S_t = \max(0, S_{t-1} + (e_t - \mu_0 - k)) \quad (23)$$

$$S_t \geq h \quad (24)$$

A estatística acumulada do CUSUM é definida pela Equação 23, em que S_t representa a soma acumulada no instante t , e_t corresponde à métrica de erro monitorada, μ_0 é o valor de referência obtido durante o período de operação normal e k é um parâmetro de sensibilidade responsável por definir a magnitude mínima de mudança desejada para detecção (Page, 1954).

Desse modo, um alarme de degradação é emitido quando a estatística acumulada ultrapassa um limiar previamente definido, especificado pela Equação 24, onde h representa o limite de decisão do método. Quando a condição é satisfeita, considera-se que ocorreu uma mudança estatisticamente significativa no comportamento do modelo e um alerta do monitoramento de degradação é emitido.

A principal vantagem do CUSUM é que pequenos aumentos sucessivos nas métricas de erro, que poderiam permanecer por longos períodos sem ultrapassar os limites de controle definidos pela regra de Shewhart, são acumulados pela Equação 23, permitindo a detecção antecipada de fenômenos de deriva e obsolescência do modelo (K.; Kumar, 2020).

2.5.2 *Adaptive Windowing* (ADWIN)

O ADWIN, proposto por Bifet e Gavalda (2007), é um algoritmo de detecção de mudanças desenvolvido para aplicações em fluxos de dados. Diferentemente do CUSUM, que depende da definição prévia de parâmetros como o valor de referência e os limiares de decisão, o ADWIN adapta automaticamente sua janela de observação em função das características dos dados monitorados, dispensando o conhecimento prévio da distribuição do processo.

O funcionamento do ADWIN consiste em manter uma janela deslizante de tamanho variável contendo as observações mais recentes. A cada nova amostra recebida, a janela é dividida em duas subjanelas, correspondentes aos conjuntos de dados mais antigos e mais recentes. Em seguida, as médias dessas subjanelas são comparadas por meio de um teste estatístico baseado no limite de Hoeffding. Caso a diferença entre as médias seja estatisticamente significativa, conclui-se que ocorreu uma mudança na distribuição dos dados, sendo a porção mais antiga da janela descartada e um evento de detecção de mudança sinalizado (Gama, 2010).

Supondo $W = W_0 \cup W_1$ uma janela contendo n observações, em que W_0 representa a porção mais antiga da janela e W_1 a mais recente. O ADWIN sinaliza uma mudança quando a condição da Equação 25 é satisfeita.

$$|\mu_{W_0} - \mu_{W_1}| > \epsilon, \quad (25)$$

Na Equação 25, μ_{W_0} e μ_{W_1} representam as médias das duas subjanelas e ϵ corresponde ao limite estatístico calculado a partir da desigualdade de Hoeffding, considerando o tamanho da janela e o nível de confiança adotado.

A principal vantagem do ADWIN é sua capacidade de adaptar automaticamente o tamanho da janela de monitoramento às mudanças observadas no fluxo de dados sem a necessidade de definir limiares fixos. Essa característica torna o algoritmo particularmente adequado para cenários em que o comportamento dos dados pode variar ao longo do tempo, como no aprendizado incremental e no monitoramento de degradação de modelos de aprendizado de máquina (Gama, 2010).

A implementação de técnicas de monitoramento contínuo, como o CUSUM e o ADWIN constitui um elemento fundamental para garantir a confiabilidade de modelos preditivos em ambientes dinâmicos. Na Seção 2.6, são apresentadas as métricas utilizadas nesse trabalho para avaliação de desempenho de modelos.

2.6 Métricas de avaliação de desempenho

Nesta seção são apresentadas quatro métricas utilizadas neste trabalho para avaliar o desempenho dos modelos de previsão: erro absoluto médio (MAE), raiz do erro quadrático médio (RMSE), coeficiente de determinação (R^2) e *Skill Score*. Tais métricas permitem avaliar o desempenho dos modelos sob diferentes perspectivas, incluindo a magnitude dos erros, a capacidade explicativa e o desempenho em relação a um modelo de referência. Cada métrica é detalhada a seguir de acordo com as definições de James et al. (2013) e Hastie, Tibshirani e Friedman (2009).

Erro absoluto médio (MAE)

O Erro absoluto médio (MAE) corresponde à média dos erros absolutos entre os valores observados e preditos pelo modelo. Por considerar o valor absoluto dos erros, todos contribuem de forma proporcional para o cálculo da métrica, tornando-a menos sensível à influência de erros de grande magnitude. Dessa forma, o MAE fornece uma medida intuitiva da grandeza média dos erros de previsão. O MAE é definido pela Equação 26.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (26)$$

Na Equação 26, y_i representa o valor observado, $\hat{y}_i$ representa o valor predito pelo modelo e n corresponde ao número total de observações avaliadas.

Raiz do erro quadrático médio (RMSE)

A Raiz do erro quadrático médio (RMSE) é uma das métricas mais utilizadas para avaliar o desempenho de modelos de regressão. Ela é obtida pela raiz quadrada da média dos quadrados das diferenças entre os valores observados e os preditos pelo modelo. Como os erros são elevados ao quadrado antes do cálculo da média, a métrica atribui maior penalização a erros de maior magnitude, tornando-se particularmente sensível a previsões com grandes desvios. Além disso, a aplicação da raiz quadrada faz com que o resultado seja expresso na mesma unidade da variável de interesse, o que facilita a interpretação da magnitude média dos erros de previsão. O RMSE é definido pela Equação 27.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (27)$$

Na Equação 27, y_i representa o valor observado, $\hat{y}_i$ representa o valor predito pelo modelo e n corresponde ao número total de observações avaliadas.

Coefficiente de determinação (R^2)

O Coeficiente de determinação (R^2) é uma métrica que avalia a capacidade de modelos de regressão em explicar a variabilidade dos dados observados. Essa métrica quantifica a proporção da variância da variável dependente que é explicada pelas previsões do modelo em relação à variância total dos dados. O valor de R^2 varia, em geral, entre 0 e 1, sendo que valores próximos de 1 indicam maior capacidade explicativa do modelo, enquanto valores próximos de 0 indicam baixo poder de explicação. O R^2 é definido pela Equação 28.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (28)$$

Na Equação 28, y_i representa o valor observado, $\hat{y}_i$ representa o valor predito pelo modelo, $\bar{y}$ corresponde à média dos valores observados e n representa o número total de observações avaliadas.

Skill Score

O *Skill Score* é uma métrica utilizada para quantificar o ganho de desempenho de um modelo de previsão em relação a um modelo de referência. Diferentemente de métricas como MAE e RMSE, que avaliam diretamente a magnitude dos erros, o *Skill Score* expressa a melhoria relativa obtida pelo modelo em comparação com outra abordagem. Valores positivos indicam que o modelo apresentou desempenho superior a referência,

enquanto valores negativos indicam desempenho inferior. Um valor igual a zero indica que ambos os modelos obtiveram o mesmo desempenho. Nesta dissertação, o modelo de persistência foi adotado como referência e a métrica de erro utilizada para comparação foi o RMSE. Dessa forma, o *Skill Score* é definido pela Equação 29.

$$SS = 1 - \frac{RMSE_{\text{modelo}}}{RMSE_{\text{persistência}}} \quad (29)$$

2.7 Considerações finais

Este capítulo teve como objetivo apresentar os principais conceitos teóricos que fundamentam o desenvolvimento deste trabalho. A Seção 2.1 descreveu o processo de geração de energia solar fotovoltaica, abordando os componentes e as etapas envolvidas na conversão da radiação solar em energia elétrica.

A Seção 2.2 abordou as etapas do pré-processamento dos dados, incluindo a análise de correlação de *features*, engenharia de atributos e transformações de escala. Em seguida, a Seção 2.3 introduziu os modelos clássicos de previsão utilizados neste trabalho, detalhando os modelos de persistência, ARIMA e regressão ridge.

Po sua vez, a Seção 2.4 apresentou os fundamentos do aprendizado de máquina, com destaque para as redes neurais artificiais e seu funcionamento. Além disso, foram discutidos duas arquiteturas de RNA empregadas nesta pesquisa: MLP e LSTM. Também foram introduzidos os conceitos relacionados ao aprendizado incremental, detalhando as técnicas de *fine-tuning* e EWC.

A Seção 2.5 abordou os conceitos relacionados à degradação de modelos de aprendizado de máquina, descrevendo técnicas de monitoramento estatístico capazes de identificar perdas de desempenho ao longo do tempo. Em particular, foi discutido os métodos CUSUM e ADWIN, adotados neste trabalho. Por fim, a Seção 2.6 encerrou o capítulo apresentando às principais métricas utilizadas nesta dissertação para a avaliação do desempenho dos modelos preditivos.

Os conceitos apresentados neste capítulo fornecem a fundamentação teórica necessária para a compreensão da metodologia proposta nesta pesquisa. O Capítulo 3 apresenta os trabalhos relacionados encontrados na literatura, destacando as principais abordagens empregadas na previsão de geração fotovoltaica, no monitoramento de degradação de modelos e nas técnicas de aprendizado incremental.

Trabalhos Relacionados

Este capítulo apresenta os principais trabalhos relacionados a esta dissertação encontrados na literatura. Esses trabalhos estão organizados em três seções temáticas. A Seção 3.1 aborda estudos voltados à previsão de geração de energia elétrica fotovoltaica. A Seção 3.2 discorre sobre trabalhos relacionados à degradação de modelos. A Seção 3.3 apresenta estudos que empregam técnicas de aprendizagem incremental com o objetivo de promover a atualização contínua e o aprimoramento de modelos. Por fim, a Seção 3.4 traz as considerações finais do capítulo.

A busca por trabalhos relacionados foi feita por palavras-chave em português e inglês nos sites Google Acadêmico¹, ACM *Digital Library*² e *Research Gate*³. Algumas das palavras-chave utilizadas foram 'previsão de geração fotovoltaica', 'aprendizado incremental', 'monitoramento de degradação', *concept drift*, *solar power forecasting*, *incremental learning*, entre outras.

3.1 Previsão de energia fotovoltaica

O trabalho de Catarino (2024) propõe a aplicação de técnicas de aprendizado de máquina para a previsão da geração de energia elétrica no mesmo complexo solar utilizado como estudo de caso na primeira etapa dessa dissertação – Sol do Cerrado. O autor avaliou diferentes arquiteturas de RNA e modelos baseados em RF, utilizando como variável de entrada exclusivamente a irradiância solar. Além disso, investigou-se a influência da frequência temporal dos dados (15, 30 e 60 minutos) sobre a acurácia das previsões, empregando as métricas MAE, erro quadrático médio (MSE) e R^2 . Os resultados indicaram elevado desempenho dos modelos avaliados, com destaque para uma arquitetura de RNA utilizando a função de ativação *ReLU*, além de evidenciar maior variabilidade das previsões durante períodos de elevada geração solar.

¹ <https://scholar.google.com.br/>

² <https://dl.acm.org/>

³ <https://www.researchgate.net/>

O estudo também mostrou que não houve diferença estatisticamente significativa entre as granularidades temporais analisadas. Como trabalhos futuros, os autores sugerem a incorporação de novos dados operacionais, a avaliação de outras resoluções temporais e o desenvolvimento de soluções capazes de operar continuamente com dados da planta. Essa dissertação dá continuidade a esse estudo ao incorporar variáveis meteorológicas adicionais, investigar novos intervalos de granularidade, diferentes técnicas de pré-processamento e integrar métodos de aprendizado incremental para adaptação contínua dos modelos às mudanças nos dados.

Em Kandananond, Chotewiwat e Wongsathan (2021), os autores propõem um modelo baseado em redes neurais recorrentes (RNN) do tipo *Long Short-Term Memory* (LSTM) para previsão de potência de sistemas fotovoltaicos. O objetivo principal é realizar previsões de curto prazo, 1,5 h à frente, utilizando exclusivamente dados históricos de potência, sem fontes meteorológicas externas. Foram utilizados dados reais de uma planta fotovoltaica localizada em Nicósia, Chipre, abrangendo o período de setembro de 2016 a janeiro de 2019, com granularidade de 15 minutos. O pré-processamento envolveu remoção de valores noturnos, interpolação de lacunas e normalização dos dados. Para a previsão, foram utilizadas janelas de entrada com 192 instantes temporais (3,2 dias de dados) para prever os 6 instantes subsequentes, correspondentes a 1,5 h de geração futura. Os resultados demonstraram que o modelo apresentou elevada precisão para o horizonte de previsão considerado, além de boa capacidade de generalização. Os autores concluem que redes LSTM são capazes de modelar adequadamente as relações temporais presentes na geração fotovoltaica utilizando apenas dados históricos de potência, evidenciando o potencial dessa abordagem para aplicações de previsão de curto prazo.

A pesquisa desenvolvida nesse trabalho incorpora algumas das direções apontadas por Kandananond, Chotewiwat e Wongsathan (2021). Em particular, foram investigadas diferentes configurações de hiperparâmetros por meio de técnicas de otimização, empregadas arquiteturas baseadas em LSTM para previsão de geração fotovoltaica e analisados horizontes temporais de curto prazo. Além disso, algumas estratégias de pré-processamento da dissertação foram inspiradas no trabalho, como remoção dos dados noturnos, interpolação e normalização.

Em Tian et al. (2023), Jailani et al. (2023), os autores apresentam revisões da literatura sobre a aplicação de técnicas de aprendizado de máquina e aprendizado profundo na previsão de irradiância solar e geração fotovoltaica. Os estudos destacam o uso de modelos como RNA, CNN, LSTM, *Support Vector Machine* (SVM), RF e arquiteturas híbridas, evidenciando sua eficácia na modelagem de padrões complexos em dados meteorológicos. Além disso, é ressaltado que modelos híbridos, como CNN-LSTM, tendem a apresentar desempenho superior às arquiteturas LSTM isoladas, em virtude de sua capacidade de capturar simultaneamente características espaciais e temporais, embora com maior custo computacional. Esses trabalhos fundamentam a escolha dos modelos investigados nesta

dissertação, em especial a utilização de arquiteturas baseadas em LSTM para previsão da geração fotovoltaica, bem como a avaliação de diferentes estratégias de pré-processamento e configuração dos modelos.

3.2 Monitoramento de degradação em modelos

O estudo de Agrahari e Singh (2022) apresenta uma revisão da literatura sobre detecção de *concept drift* em fluxos contínuos de dados, destacando que alterações na distribuição estatística ao longo do tempo podem degradar significativamente o desempenho de modelos de aprendizado de máquina. O método de detecção é organizado em diferentes categorias, incluindo abordagens baseadas em análise estatística, janelas deslizantes, distribuição dos dados, limites de decisão e algoritmos dependentes do modelo, discutindo suas vantagens, limitações e cenários de aplicação. Além disso, são revisados métodos amplamente utilizados, como *Cumulative Sum Control Chart* (CUSUM), *Drift Detection Method* (DDM), *Early Drift Detection Method* (EDDM), *Adaptive Windowing* (ADWIN) e *Exponential Weighted Moving Average for Concept Drift Detection* (ECDD), bem como estratégias de adaptação baseadas em mecanismos de esquecimento e janelas deslizantes. Como principal contribuição, os autores sintetizam o estado da arte da área, identificam desafios relacionados à escalabilidade, ao tratamento de dados ruidosos e desbalanceados e apontam direções para o desenvolvimento de métodos mais robustos de detecção de desvio de conceito.

Em Azeem et al. (2025) os autores investigam o impacto do *concept drift* na previsão de carga elétrica de curto prazo em redes inteligentes, propondo o *framework* adaptativo *Adaptive Ensemble Long Short-Term Memory* (AE-LSTM), baseado em aprendizado em conjunto e redes LSTM. O modelo incorpora mecanismos de atenção dupla e seleção dinâmica de características, permitindo atualizar continuamente as variáveis mais relevantes e adaptar-se às mudanças na distribuição dos dados sem a necessidade de re-treinamento completo. Avaliado em bases de dados reais de geração de energia provenientes de diferentes fontes, como carvão, gás natural, hidrelétrica e solar, o AE-LSTM apresentou desempenho superior aos modelos RF, SVM e LSTM convencional, alcançando melhorias de até 55,61% no coeficiente de determinação (R^2).

Os trabalhos de Agrahari e Singh (2022), Azeem et al. (2025) são relevante para esta dissertação ao fornecer uma visão abrangente das técnicas de detecção de *concept drift*, fundamentando a escolha de métodos de monitoramento de degradação empregados neste estudo.

3.3 Aprendizagem incremental

Em relação ao aprendizado incremental, o trabalho de Puah et al. (2021) propõe uma abordagem não supervisionada para previsão de irradiância solar baseada no algoritmo denominado *Regression Enhanced Incremental Self-Organising Neural Network* (RE-SOINN). Os autores consideram que a irradiância apresenta componentes de baixa e alta frequência, correspondentes, respectivamente, à tendência diária da curva solar e às rápidas variações provocadas por condições atmosféricas. Para explorar essas características, a série temporal é decomposta em dois sinais, processados separadamente por modelos incrementais capazes de se adaptar continuamente à chegada de novos dados, sem necessidade de re-treinamento completo.

Os resultados demonstram que o RE-SOINN apresentou desempenho superior aos modelos de persistência, *Exponential Smoothing* e RNA convencionais. Além disso, a decomposição da série em componentes de baixa e alta frequência contribuiu para reduzir a influência do ruído e melhorar a representação tanto da tendência diária quanto das variações rápidas da irradiância. O trabalho de Puah et al. (2021) é relevante para esta dissertação por demonstrar a viabilidade do aprendizado incremental em problemas de previsão de irradiância solar, evidenciando a capacidade de adaptação contínua do modelo frente às mudanças observadas nos dados.

O trabalho de Liu e Chang (2026) revisita o método EWC e demonstra limitações na forma tradicional de estimar a importância dos parâmetros, propondo uma variante denominada (*EWC Done Right*) (EWC-DR). Os resultados indicam que a nova abordagem reduz o esquecimento catastrófico de maneira mais eficaz que o EWC original e outras técnicas de regularização, evidenciando que ainda existem oportunidades para aprimorar mecanismos de preservação de conhecimento em aprendizado contínuo. A presente dissertação emprega o EWC como estratégia de aprendizado incremental para reduzir o esquecimento catastrófico durante a atualização contínua do modelo. Nesse contexto, o trabalho de Liu e Chang (2026) é relevante por demonstrar que ainda existem oportunidades de aprimoramento desse método, reforçando a importância do EWC como uma abordagem consolidada e evidenciando possíveis direções para trabalhos futuros.

Por fim, os autores de Al-Dahidi et al. (2024) apresentam uma revisão da literatura e propõem um *framework* orientado por dados para previsão da produção de energia solar fotovoltaica. A abordagem é organizada em etapas de preparação dos dados, desenvolvimento e avaliação dos modelos e melhorias avançadas, incluindo detecção de mudanças de contexto, aprendizado incremental e quantificação de incertezas. Como principal contribuição, o trabalho integra esses componentes em uma única estrutura, buscando aumentar a precisão, a robustez e a confiabilidade das previsões em aplicações reais. Nesta dissertação são integrados etapas de Al-Dahidi et al. (2024), incluindo pré-processamento, desenvolvimento e avaliação de modelos, monitoramento de degradação e aprendizado incremental, adaptando-os ao contexto da previsão de geração de energia fotovoltaica.

3.4 Considerações finais

Este capítulo teve como objetivo apresentar os principais trabalhos relacionados à previsão da geração de energia fotovoltaica por meio de técnicas de aprendizado de máquina, bem como estudos voltados ao monitoramento de degradação e ao aprendizado incremental em modelos preditivos.

A partir da análise dos trabalhos relacionados, identificam-se lacunas na integração entre previsão de geração fotovoltaica, monitoramento da degradação dos modelos e aprendizado incremental em um mesmo fluxo de operação. Embora estudos individuais abordem técnicas de previsão, detecção de *concept drift* ou aprendizado contínuo, observa-se a necessidade de uma abordagem que combine essas etapas e as avalie. Além disso, permanece relevante investigar de forma conjunta a influência de diferentes granularidades temporais, técnicas de agregação, métodos de escalonamento e configurações de modelos sobre o desempenho preditivo na geração de energia fotovoltaica.

Nesse contexto, a pesquisa proposta diferencia-se dos trabalhos relacionados por: i) avaliar diferentes modelos de previsão, incluindo persistência, ARIMA, regressão Ridge, MLP e LSTM; ii) investigar o impacto de diferentes granularidades temporais, métodos de agregação e técnicas de escalonamento de dados no desempenho dos modelos; iii) integrar técnicas de monitoramento de degradação e aprendizado incremental em um mesmo fluxo de atualização do modelo, permitindo a adaptação contínua às mudanças nos dados; e iv) validar a abordagem em dados reais de diferentes sítios fotovoltaicos.

Metodologia Proposta para Predição e Monitoramento da Geração de Energia Fotovoltaica

Esta dissertação propõe uma arquitetura composta de etapas de predição, monitoramento de degradação e aprendizado incremental, aplicada ao problema de previsão da geração de energia fotovoltaica. A metodologia apresentada neste capítulo descreve os procedimentos adotados para o desenvolvimento da pesquisa e o alcance do objetivo. A Figura 8 apresenta uma visão geral da metodologia proposta, destacando o fluxo de execução da arquitetura e a relação entre as diferentes etapas do trabalho. Cada um dos blocos representados na Figura 8 é detalhado nas seções seguintes.

Este capítulo está organizado em três seções, definidas de acordo com os conjuntos de dados empregados em cada etapa da pesquisa. A Seção 4.1 apresenta o dataset Sol do Cerrado, proveniente de um sítio de geração de energia fotovoltaica localizado no norte de Minas Gerais e pertencente à mineradora VALE, cujos dados foram cedidos para a realização deste estudo. Esse conjunto de dados é composto por 19 dias de medições com granularidade de um segundo.

O dataset Sol do Cerrado é utilizado para a análise de diferentes estratégias de pré-processamento, incluindo a avaliação de distintas granularidades temporais, métodos de agregação e técnicas de escalonamento dos dados. Além disso, ele é empregado na construção, configuração e comparação dos modelos de previsão, sendo responsável pela seleção da abordagem com melhor desempenho preditivo. Entretanto, devido ao curto período de dias disponível no dataset Sol do Cerrado, foi necessária a utilização de um segundo conjunto de dados, contendo um período significativamente maior de observações, viabilizando a implementação e a avaliação do monitoramento de degradação e do aprendizado incremental.

Dessa forma, a Seção 4.2 apresenta o dataset UNISOLAR, disponível publicamente¹,

¹ <<https://www.kaggle.com/datasets/cdaclab/unisolar>>

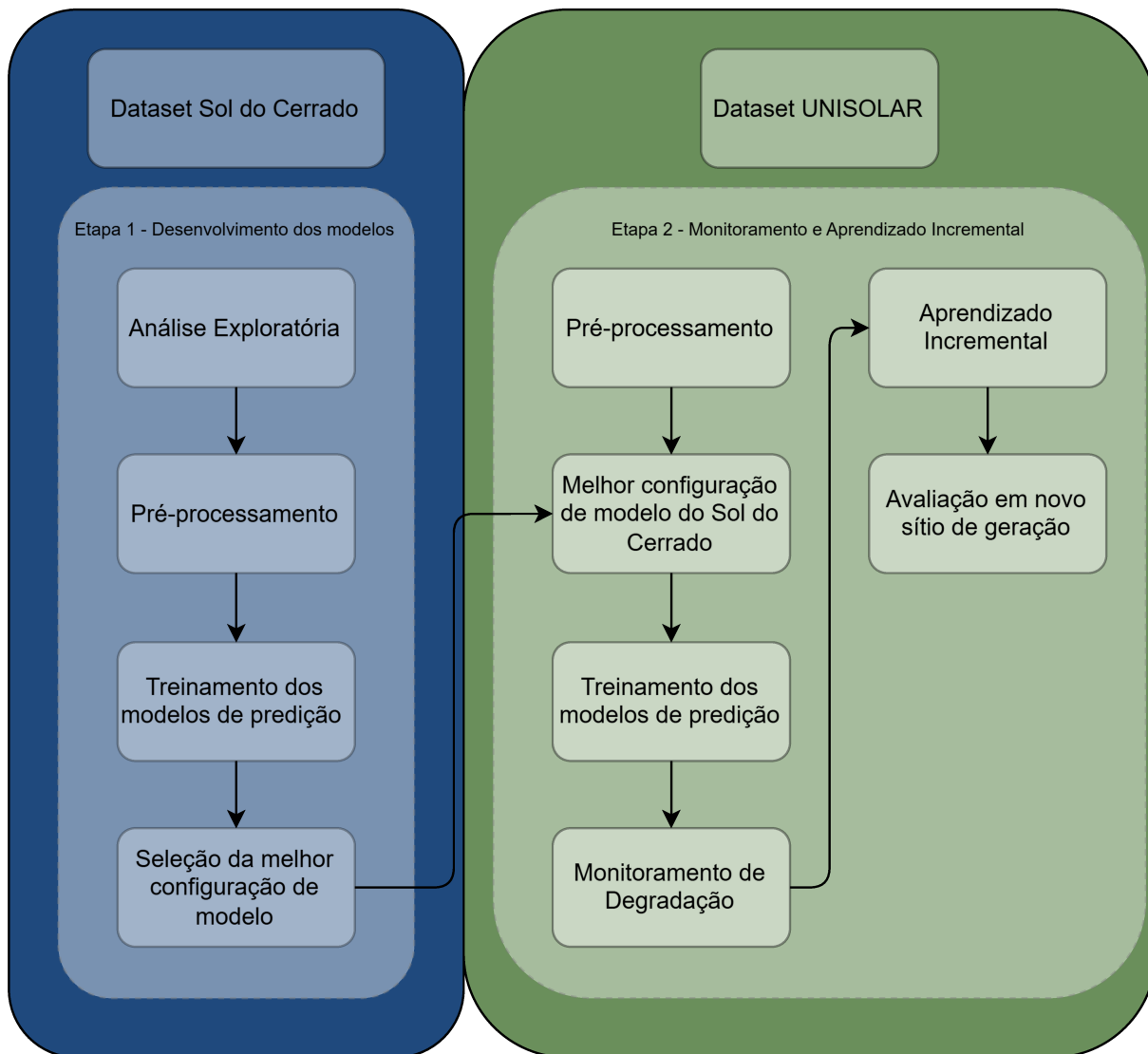


Figura 8 – Método proposto para o desenvolvimento da arquitetura de predição, monitoramento de degradação e aprendizado incremental na geração de energia fotovoltaica. Fonte: Elaborado pelo autor (2026).

utilizado na segunda fase deste trabalho. Nessa etapa, são adotadas as configurações de modelo obtidas a partir dos experimentos realizados com o dataset Sol do Cerrado. A partir delas, são avaliadas diferentes estratégias de engenharia de atributos, incluindo a comparação entre variantes com e sem registros do período noturno, bem como diferentes estratégias de monitoramento de degradação e técnicas de aprendizado incremental.

Por fim, a Seção 4.3 apresenta as considerações finais do capítulo, sintetizando os principais procedimentos metodológicos descritos.

4.1 Sol do Cerrado

Esta seção descreve a primeira etapa da metodologia proposta, relativa ao dataset Sol do Cerrado. O conjunto de dados foi obtido por meio de medições realizadas em um sítio de geração de energia solar, pertencente a mineradora VALE, localizado no norte do estado de Minas Gerais. As medições foram conduzidas ao longo de um período de 19 dias — 27/10/2023 a 14/11/2023 —, com frequência de amostragem de um dado por segundo, resultando em um total de 1.641.600 registros.

Cada registro do conjunto de dados contém as seguintes informações: o **timestamp** correspondente ao momento da coleta; a irradiância solar no plano dos módulos, denominada *Plane of Array* (POA), que representa a quantidade de energia solar incidente por unidade de área; as temperaturas registradas pelos sensores instalados nos inversores, **T_MOD_1** e **T_MOD_2**; a precipitação acumulada, identificada como **PluAcum**; e, por fim, a quantidade de energia efetivamente gerada pelos painéis solares **INV_A_Measur**.

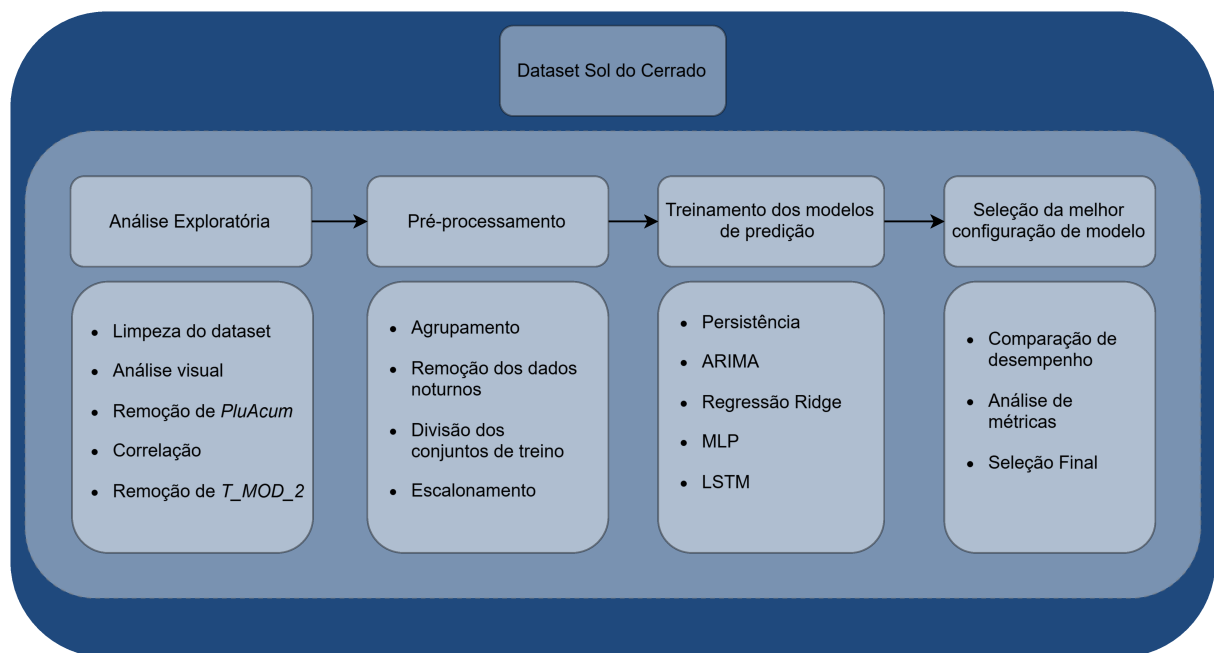


Figura 9 – Método detalhado da etapa relativa ao dataset Sol do Cerrado. Fonte: Elaborado pelo autor (2026).

A Figura 9 representa a metodologia detalhada relativa ao dataset Sol do Cerrado. Esta etapa foi dividida em quatro fases: análise exploratória, pré-processamento dos dados, treinamento dos modelos de previsão e seleção da melhor configuração de modelo. O objetivo desta etapa é investigar diferentes configurações para a construção de modelos de previsão da geração de energia fotovoltaica, avaliando o impacto das estratégias de pré-processamento, das arquiteturas e de seus respectivos hiperparâmetros sobre o desempenho preditivo. Ao final dessa etapa, a configuração que apresentar os melhores

resultados será utilizada no treinamento do modelo empregado nas etapas de monitoramento de degradação e aprendizado incremental.

4.1.1 Análise exploratória

Inicialmente, foi realizada uma etapa de limpeza do conjunto de dados, mantendo apenas as colunas: `timestamp`, `POA`, `T_MOD_1`, `T_MOD_2`, `PluAcum` e `INV_A_Measur`. Esse procedimento teve como finalidade eliminar redundâncias e facilitar a leitura e manipulação das informações, uma vez que o arquivo original apresentava o `timestamp` repetido para cada sensor, o que resultava em uma quantidade considerável de duplicidade. Durante o processo de limpeza dos dados foram removidas todas as linhas com valores ausentes.

Após o processo de limpeza, foram gerados gráficos diários de todas as variáveis registradas pelos sensores, com o propósito de realizar uma análise visual preliminar e compreender o comportamento dos dados. Tal análise permitiu identificar que a variável `PluAcum` permaneceu praticamente constante durante todo o período analisado, isso porque durante o período de criação do dataset não houve chuvas na região do sítio de geração de energia solar. Consequentemente, por não apresentar variabilidade nem contribuir com informações relevantes para os modelos preditivos, o atributo `PluAcum` foi removido.

Também foi identificado que a variável `T_MOD_2` apresentava comportamentos incompatíveis com o esperado para a temperatura dos módulos fotovoltaicos, indicando uma possível falha de aquisição pelo sensor. Essa observação motivou uma investigação mais detalhada, cujos resultados são apresentados no Capítulo 5. Com base nessa análise, o atributo `T_MOD_2` também foi removido do conjunto de dados.

Paralelamente, foi desenvolvido um estudo de correlação entre as *features* do dataset. Dessa maneira, foram calculados os coeficientes de correlação de Pearson e Spearman, bem como a correlação cruzada com defasagem de 0 a 30 amostras com o objetivo de identificar relações lineares, monotônicas e temporais entre as variáveis. Com base nos coeficientes, foram construídas matrizes de correlação para ambos os métodos, possibilitando a inspeção visual da intensidade e da direção das correlações entre os atributos. Os resultados dessa análise embasaram a seleção das variáveis utilizadas no desenvolvimento dos modelos de previsão.

4.1.2 Pré-processamento

A etapa seguinte consistiu no pré-processamento dos dados, com o intuito de prepará-los para o treinamento dos modelos de previsão. Nessa etapa, foram geradas diferentes variantes do dataset, permitindo a comparação entre estratégias de escalonamento e agregação temporal. O pré-processamento foi estruturado em quatro etapas: (i) agregação temporal; (ii) filtragem do período noturno; (iii) criação dos conjuntos de treinamento,

validação e teste; e (iv) escalonamento dos dados. Cada uma dessas etapas é detalhada a seguir.

Agregação temporal

Os dados originais do sítio de geração de energia fotovoltaica apresentam frequência de amostragem de um segundo, o que resulta em elevada resolução temporal e alta variabilidade nas medições. Embora essa frequência permita capturar variações instantâneas do sistema, ela também torna o conjunto de dados mais suscetível a ruídos e flutuações locais, que podem afetar o desempenho dos modelos de previsão. Dessa forma, com o objetivo de investigar a influência da resolução na qualidade das previsões, foram construídas diferentes variantes do conjunto de dados por meio de agregação temporal.

Os registros do dataset foram agregados em intervalos de 1, 10, 15, 30 e 60 minutos, empregando três métodos distintos de agregação: média, mediana e primeiro valor da janela. Dessa forma, para cada granularidade temporal foram geradas três versões do conjunto de dados, totalizando quinze variantes. Essa estratégia permitiu avaliar o impacto tanto da granularidade temporal quanto do método de agregação sobre o desempenho dos modelos de previsão.

Filtragem do período noturno

Após a agregação temporal, todas as variantes do conjunto de dados passaram por uma etapa de filtragem do período noturno. Para isso, foram removidos os registros em que a irradiância POA era inferior a 5 W/m^2 . Esse limiar foi adotado por representar condições nas quais não há incidência significativa de radiação solar e, conseqüentemente, a geração de energia é nula ou desprezível.

A remoção desses registros tem como objetivo concentrar a análise no desempenho dos modelos durante os períodos de operação do sistema fotovoltaico. A inclusão de observações referentes ao período noturno poderia superestimar o desempenho dos modelos, uma vez que a previsão de valores próximos de zero durante a noite constitui uma tarefa trivial e pouco representativa da capacidade preditiva do modelo em condições reais de operação.

Criação dos conjuntos de treinamento, validação e teste

Nessa etapa cada uma das variantes do conjunto de dados, criadas anteriormente, foi dividida cronologicamente em três subconjuntos destinados ao treinamento, à validação e ao teste dos modelos. A divisão respeitou a ordem temporal das observações, sem qualquer tipo de embaralhamento, de forma a preservar a dependência temporal e evitar o vazamento de informações futuras para o treinamento. Para cada variante, 70% das

amostras foram destinadas ao conjunto de treinamento, 15% ao conjunto de validação e os 15% restantes ao conjunto de teste.

Escalonamento dos dados

Com o objetivo de avaliar a influência da escala dos dados sobre o desempenho dos modelos de previsão, foram consideradas duas estratégias de escalonamento dos atributos: *Min-Max* e *Z-score*. Cada uma das quinze variantes do conjunto de dados geradas na etapa de agregação foi processada utilizando ambos os métodos, resultando em um total de trinta conjuntos de dados para a realização dos experimentos.

Os parâmetros de escalonamento foram estimados exclusivamente a partir do conjunto de treinamento. Em seguida, esses mesmos parâmetros foram aplicados aos conjuntos de validação e teste, prevenindo o vazamento de dados. Os parâmetros de escalonamento obtidos pelos métodos *Min-Max* e *Z-score* foram armazenados para cada variante, possibilitando a posterior reversão do escalonamento das previsões para a escala original.

4.1.3 Treinamento dos modelos de previsão

Após a conclusão das etapas de pré-processamento, cada uma das trinta variantes do conjunto de dados foi utilizada no treinamento e avaliação dos modelos de previsão. Foram utilizados cinco modelos distintos: persistência, *Autoregressive Integrated Moving Average* (ARIMA), regressão Ridge, *Multilayer Perceptron* (MLP) e *Long Short-Term Memory* (LSTM). O modelo de persistência foi utilizado como método de referência, o ARIMA foi empregado como método estatístico, enquanto a Regressão Ridge representa uma abordagem linear regularizada. Por fim, as arquiteturas MLP e LSTM constituem os modelos baseados em aprendizado de máquina avaliados neste trabalho.

Embora todos os modelos tenham sido treinados e avaliados sobre as mesmas variantes do conjunto de dados, existem diferenças na forma como cada abordagem utiliza as informações temporais. Os modelos Ridge, MLP e LSTM utilizam janelas temporais de entrada, permitindo que a previsão seja realizada a partir de uma sequência de observações anteriores, denominada, janelas de *lookback*. Em contrapartida, o modelo de persistência utiliza apenas o valor observado no instante imediatamente anterior como previsão para o próximo instante, enquanto o modelo ARIMA realiza previsões utilizando sua própria estrutura autorregressiva e de médias móveis, não empregando um tamanho de janela fixo. As previsões tanto do modelo de persistência quanto do ARIMA foram realizadas considerando apenas observações pertencentes ao mesmo dia, evitando que valores do final de um dia fossem utilizados para prever o início do dia seguinte.

Nos modelos Ridge, MLP e LSTM, onde há entrada de sequências temporais, cada conjunto de entrada foi inicialmente segmentado por dia, garantindo que as sequências fossem construídas apenas entre observações pertencentes ao mesmo dia de geração. Em

seguida, foi empregada uma janela deslizante de tamanho lb , responsável por gerar pares de entrada e saída para o treinamento dos modelos. Cada entrada é composta pelas lb observações anteriores de todas as variáveis utilizadas, enquanto a saída corresponde ao valor da potência gerada, `INV_A_Measur`, no instante imediatamente seguinte. O tamanho da janela temporal foi tratado como um hiperparâmetro do processo de otimização, sendo avaliados os valores de 10, 30 e 60 observações.

Em relação à otimização dos parâmetros, o modelo ARIMA teve seus três parâmetros estimados por meio da *Akaike Information Criterion* (AIC), através da função `auto_arima` da biblioteca `pmdarima`². Para a regressão Ridge, cujo único hiperparâmetro corresponde ao coeficiente de regularização α , foi empregada uma *grid search* sobre um conjunto predefinido de valores, selecionando a configuração com menor erro no conjunto de validação.

A otimização dos hiperparâmetros das arquiteturas MLP e LSTM foi realizada utilizando duas abordagens distintas: *Optuna*³ e *Ray Tune*⁴. Ambas exploraram o mesmo espaço de busca, permitindo comparar diretamente a eficiência das estratégias de otimização empregadas. O espaço de busca contemplou o número de neurônios ocultos, o número de camadas na LSTM ou a taxa de *dropout* na MLP, a taxa de aprendizado, o tamanho do lote e o tamanho do *lookback*.

Para cada variante do conjunto de dados, foram executados cinquenta *trials*, utilizando como critério de seleção a configuração que apresentou o menor erro no conjunto de validação. Devido à inicialização aleatória dos pesos das arquiteturas MLP e LSTM, a configuração de hiperparâmetros vencedora foi treinada três vezes utilizando diferentes inicializações. Ao final, foi selecionado e armazenado o modelo que apresentou o menor erro no conjunto de validação, reduzindo a influência da aleatoriedade inerente ao processo de treinamento.

4.1.4 Seleção da melhor configuração de modelo

Após o treinamento de todas as variantes, foi realizada uma análise comparativa com o objetivo de identificar a combinação de estratégias que apresentou o melhor desempenho preditivo. Foram avaliadas as diferentes variantes do conjunto de dados, considerando os métodos de agregação temporal, as granularidades, as estratégias de escalonamento, os modelos de previsão e os métodos de otimização de hiperparâmetros empregados.

O desempenho dos modelos foi avaliado exclusivamente sobre o conjunto de teste por meio das métricas MAE, RMSE, coeficiente de determinação (R^2) e *Skill Score* (relativo ao método de persistência). Os valores de MAE, RMSE foram transformados de volta para a escala original da variável alvo (kW), utilizando os parâmetros de esca-

² <https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html>

³ <<https://optuna.readthedocs.io/en/stable/>>

⁴ <<https://docs.ray.io/en/latest/tune/index.html>>

namento estimados na etapa de pré-processamento a partir do conjunto de treinamento. Vale ressaltar que as métricas R^2 e *Skill Score* são independentes da escala dos dados. Esse procedimento permitiu comparar diretamente os resultados obtidos por variantes construídas com diferentes métodos de escalonamento.

Ao final, foi selecionada a configuração experimental que apresentou o melhor desempenho global. A configuração selecionada foi então empregada na segunda etapa deste trabalho, utilizando o conjunto de dados UNISOLAR. Nessa etapa, seu desempenho foi avaliado em um cenário distinto daquele utilizado durante o processo de seleção, servindo como base para a implementação e avaliação das estratégias de monitoramento de degradação e aprendizado incremental propostas nesta dissertação.

4.2 UNISOLAR

A segunda etapa da metodologia proposta neste trabalho utiliza o conjunto de dados UNISOLAR para o desenvolvimento e avaliação de uma estratégia de monitoramento de degradação e aprendizado incremental, empregando a configuração experimental selecionada na etapa anterior, realizada com o conjunto de dados Sol do Cerrado.

UNISOLAR é uma base de dados pública⁵ que reúne medições de geração de energia fotovoltaica provenientes de 42 sítios distribuídos em cinco campi da La Trobe University, localizada na Austrália. Essa base foi selecionada para a segunda etapa deste trabalho por abranger aproximadamente dois anos de registros, compreendendo o período de 08/01/2020 a 23/04/2022, com resolução temporal de 15 minutos. A Figura 10 ilustra a metodologia aplicada ao conjunto de dados UNISOLAR, composta pelas etapas de análise dos dados e pré-processamento, treinamento do modelo de previsão utilizando a configuração selecionada na etapa do conjunto de dados Sol do Cerrado, monitoramento de degradação, aprendizado incremental e teste em um novo sítio de produção de energia fotovoltaica. Cada uma dessas etapas é descrita nas seções seguintes.

4.2.1 Pré-processamento

O pré-processamento do conjunto de dados da UNISOLAR foi estruturado em seis etapas: (i) análise exploratória dos dados; (ii) limpeza do conjunto de dados; (iii) engenharia de atributos; (iv) análise de correlação entre as *features*; (v) divisão nos conjuntos de treinamento, validação, teste e simulação; e (vi) normalização. Cada etapa do pré-processamento é detalhada a seguir.

⁵ <<https://www.kaggle.com/datasets/cdaclab/unisolar>>

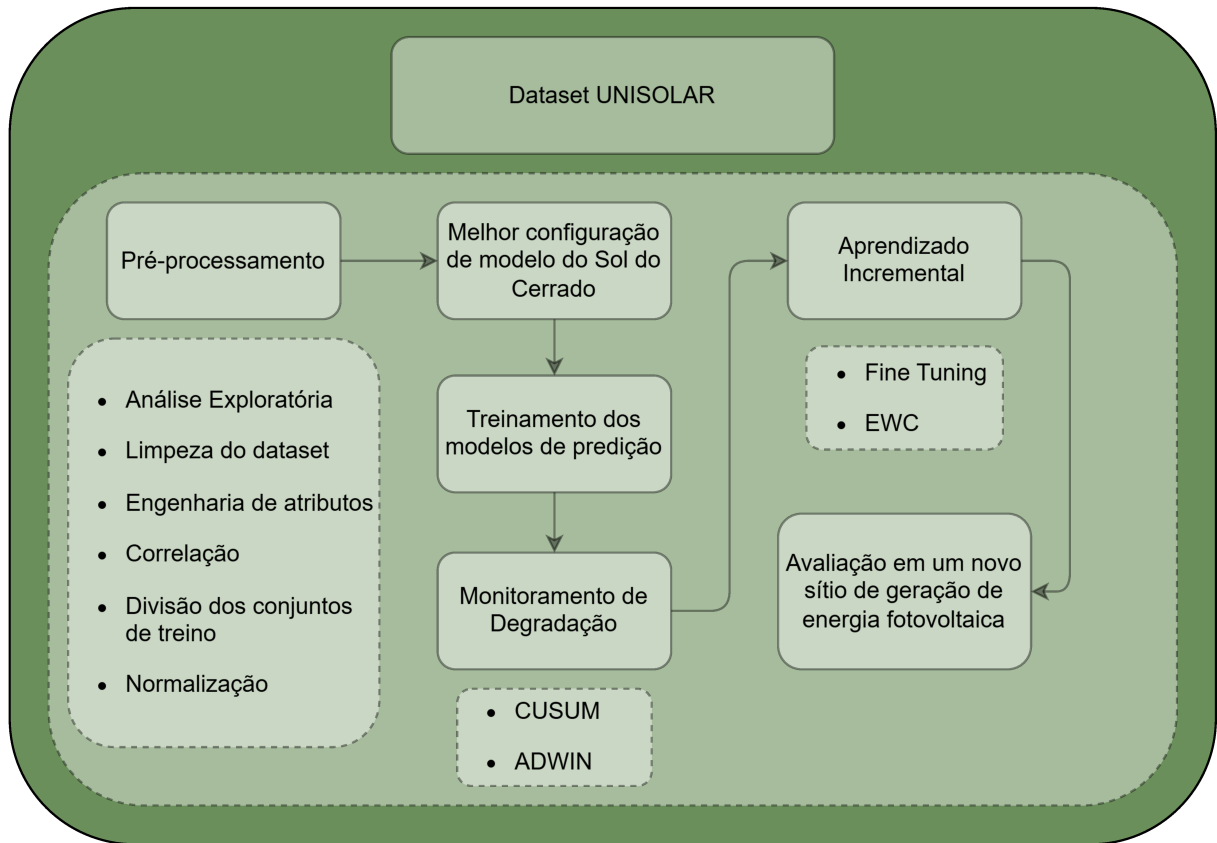


Figura 10 – Método detalhado da etapa relativa ao conjunto de dados UNISOLAR. Fonte: Elaborado pelo autor (2026).

Análise exploratória dos dados

O conjunto de dados UNISOLAR disponibiliza medições de geração de energia fotovoltaica, irradiância solar e variáveis meteorológicas provenientes de 42 sistemas fotovoltaicos distribuídos em cinco campi da La Trobe University. Os dados são disponibilizados em duas planilhas distintas: uma contendo os registros de geração de energia fotovoltaica e outra contendo as variáveis meteorológicas, ambas indexadas por **timestamp**. Dessa forma, foi realizada a integração dessas planilhas por meio da correspondência entre seus **timestamps**, resultando em um único conjunto de dados contendo todas as variáveis desejadas.

Como a base original é composta por diferentes combinações de campi e sítios, com distintos níveis de qualidade e completude dos dados, foi realizada uma análise preliminar para identificar o sítio mais adequado ao desenvolvimento deste trabalho. O sítio selecionado corresponde ao **CampusKey 1** e **SiteKey 27**, localizado em Wagga Wagga, New South Wales, por apresentar a maior quantidade de registros disponíveis e o menor número de lacunas temporais, proporcionando uma base de dados mais consistente.

Para o sítio selecionado, a variável de geração de energia fotovoltaica, denominada

SolarGeneration, foi definida como variável alvo, enquanto as variáveis meteorológicas **ApparentTemperature**, **AirTemperature**, **DewPointTemperature** e **RelativeHumidity** foram utilizadas como atributos de entrada dos modelos. O conjunto de dados também disponibiliza as variáveis **WindSpeed** e **WindDirection**; entretanto, ambas foram descartadas devido à elevada proporção de valores ausentes.

Limpeza dos dados

Após a análise exploratória apresentada anteriormente, foi realizada a etapa de limpeza do conjunto de dados, com o objetivo de aumentar sua consistência e adequá-lo às etapas subsequentes de processamento. O dataset representa a ausência de geração de energia por meio de valores NaN, em vez de utilizar o valor zero. Dessa forma, foi necessário aplicar um conjunto de regras de limpeza para tratar esses registros, preenchendo-os ou removendo-os conforme o contexto de cada ocorrência. Ao todo, foram aplicadas seis regras de limpeza, descritas a seguir.

A primeira regra tem como objetivo preencher valores ausentes isolados nas variáveis meteorológicas **ApparentTemperature**, **AirTemperature**, **DewPointTemperature** e **RelativeHumidity**. Para isso, cada valor ausente é substituído pela média aritmética entre as observações imediatamente anterior e posterior, procedimento conhecido como interpolação. Essa regra é aplicada apenas quando ambas as observações vizinhas estão separadas por exatamente 15 minutos, caracterizando uma lacuna pontual e não uma sequência de valores ausentes. Essa estratégia não é aplicada à variável de geração de energia, pois diferentemente das variáveis meteorológicas a geração de energia pode sofrer variações abruptas em decorrência de fatores como a passagem de nuvens. Nessas situações, a média entre observações vizinhas pode não representar o valor real.

A segunda regra trata os valores ausentes da variável **SolarGeneration** durante o período noturno, quando não ocorre geração de energia fotovoltaica. Para isso, todo valor NaN registrado antes das 06:00 horas ou a partir das 20:00 horas é preenchido com zero. A terceira regra propõe identificar e remover períodos de falha prolongada do sensor. Para isso, um dia é considerado suspeito quando a variável **SolarGeneration** permanece sem registro às 10:00, horário em que já é esperada geração de energia. Em seguida, os dias suspeitos são agrupados e considerados falhas de sensor quando possuem duração igual ou superior a dois dias consecutivos. Os períodos identificados como falha são removidos integralmente do conjunto de dados até à primeira leitura válida no dia em que o sensor volta a funcionar.

A regra quatro preenche qualquer valor de **SolarGeneration** ainda ausente. Como a segunda regra tratou dados noturnos e a terceira falhas no sensor, a interpretação adotada é que as lacunas remanescentes são de falta de geração de energia por nebulosidade ou tempestade, e o valor é preenchido com zero. Na regra cinco qualquer linha que ainda contenha um valor ausente em qualquer uma das cinco colunas é removida por completo.

Por fim, a sexta e última regra foi criada para gerar uma variante do dataset sem o período noturno. Para isso a identificação do período diurno é realizada individualmente para cada dia, utilizando como referência o primeiro e o último instante em que a variável **SolarGeneration** apresenta valores superiores a 1 kWh. O início do período diurno é definido por essa primeira ocorrência, desde que ela aconteça até às 09:00; caso contrário, adota-se o horário fixo de 06:00. De forma análoga, o término do período diurno é determinado pela última ocorrência superior a 1 kWh, desde que ela ocorra a partir das 17:00; caso contrário, utiliza-se o horário fixo de 21:00. Em dias em que não há nenhuma leitura superior a 1 kWh, considera-se diretamente o intervalo fixo entre 06:00 e 21:00.

A seguir, apresenta-se um resumo das seis regras de limpeza aplicadas ao conjunto de dados.

1. Preenchimento de lacunas nas variáveis meteorológicas por interpolação;
2. Preenchimento de valores ausentes da geração de energia no período noturno com zero;
3. Remoção de períodos de falha prolongada do sensor de geração de energia;
4. Preenchimento de valores ausentes da geração de energia no período diurno com zero;
5. Remoção de registros contendo valores ausentes remanescentes;
6. Geração de uma variante do conjunto de dados sem registros do período noturno.

Ao final da etapa de limpeza, são obtidas duas variantes do conjunto de dados. A primeira mantém os registros do período noturno, enquanto a segunda contém apenas os registros do período diurno. Ambas resultam da aplicação das cinco primeiras regras de limpeza, diferindo apenas pela aplicação da sexta regra, responsável pela remoção dos registros noturnos.

Engenharia de atributos

Na etapa de engenharia de atributos foi analisada a possibilidade de incorporar ao conjunto de dados informações temporais que auxiliem os modelos na identificação de padrões sazonais da geração de energia fotovoltaica. Para isso, foi utilizada a variável do **timestamp** sendo decomposta em hora do dia e dia do ano, que representam, respectivamente, os ciclos diários e anuais da incidência de radiação solar. Como essas variáveis possuem natureza cíclica ambas as variáveis foram transformadas em duas componentes, seno e cosseno, conforme as Equações 30 e 31.

$$x_{\text{sen}} = \sin\left(2\pi \frac{x}{P}\right) \quad (30)$$

$$x_{\cos} = \cos\left(2\pi \frac{x}{P}\right) \quad (31)$$

Nas Equações 30 e 31, x representa o valor da variável temporal e P corresponde ao seu período. Para a variável hora do dia, utiliza-se $P = 24$, enquanto para o dia do ano utiliza-se $P = 365$.

Ao final dessa etapa, cada uma das duas variantes obtidas na etapa de limpeza origina duas novas versões: uma mantendo apenas as variáveis temporais originais e outra contendo suas representações cíclicas. Dessa forma, passam a estar disponíveis quatro variantes do conjunto de dados para as etapas subsequentes.

Correlação de *features*

Após a etapa de engenharia de atributos, foi realizada uma análise de correlação entre as *features*, com o objetivo de identificar o grau de associação entre elas e avaliar sua relevância para a previsão da geração de energia fotovoltaica. Para essa análise, foram empregados o coeficiente de correlação de Pearson, o coeficiente de correlação de Spearman e a correlação circular proposta por Mardia (1976) para as variáveis cíclicas geradas na etapa anterior.

Os resultados mostraram que a variável `DewPointTemperature` apresentou, de forma consistente, os menores níveis de associação com a variável `SolarGeneration`. Desse modo, com o objetivo de avaliar o impacto dessa variável no desempenho dos modelos, foi gerada uma nova variante do conjunto de dados sem sua utilização.

Ao final dessa etapa, cada uma das quatro variantes obtidas anteriormente passa a possuir uma versão adicional sem a variável `DewPointTemperature`, totalizando oito variantes do conjunto de dados que serão utilizadas nos experimentos descritos no Capítulo 5.

Divisão dos conjuntos de dados

Após a geração das oito variantes do conjunto de dados, cada uma delas foi dividida em quatro subconjuntos: treinamento, validação, teste e simulação. A divisão foi realizada utilizando datas fixas, de forma a preservar a ordem cronológica das observações e evitar o uso de informações futuras durante o treinamento dos modelos.

O conjunto de treinamento compreende os primeiros 31 dias de medições, correspondendo a aproximadamente um mês de dados. Em seguida, foram reservados 15 dias para validação e outros 15 dias para teste. Todo o período restante, correspondente a aproximadamente dois anos e 46 dias de registros, foi destinado ao conjunto de simulação.

Normalização

A etapa final do pré-processamento consiste no escalonamento dos dados. Nesta parte da pesquisa, foi adotada a normalização *Min-Max* por ter apresentado o melhor desempenho nos experimentos realizados com o conjunto de dados Sol do Cerrado. A comparação entre os diferentes métodos de escalonamento, que levou a essa decisão, é apresentada no Capítulo 5.

Para evitar vazamento de informações entre os subconjuntos de dados, os parâmetros da normalização são calculados exclusivamente a partir do conjunto de treinamento. Em seguida, esses mesmos parâmetros são utilizados para transformar os conjuntos de validação, teste e simulação. As variáveis cíclicas representadas pelas componentes seno e cosseno não são submetidas ao processo de normalização, pois seus valores já se encontram limitados entre -1 e 1.

Ao final dessa etapa, cada uma das oito variantes do conjunto de dados possui seus respectivos subconjuntos de treinamento, validação, teste e simulação devidamente normalizados. Além disso, os parâmetros utilizados na normalização são armazenados para possibilitar que as métricas de desempenho dos modelos sejam analisadas na escala original da variável de geração de energia.

4.2.2 Treinamento dos modelos de previsão

Antes do treinamento dos modelos, cada uma das oito variantes do conjunto de dados foi convertida em sequências temporais por meio de janelas deslizantes. O tamanho da janela de entrada (*look-back*) foi tratado como um hiperparâmetro, enquanto o horizonte de previsão foi fixado em um único passo à frente. Para preservar a continuidade temporal das sequências, foram descartadas todas as janelas que continham intervalos superiores a 15 minutos entre registros consecutivos.

Como modelo de referência, foi utilizada a persistência. Esse modelo foi avaliado apenas em sequências com intervalo de 15 minutos entre observações consecutivas, utilizando os conjuntos de validação e teste. Considerando os resultados obtidos no dataset Sol do Cerrado, apresentados no Capítulo 5, a LSTM apresentou o melhor desempenho entre os modelos avaliados e, por esse motivo, foi selecionada para esta etapa da pesquisa. A avaliação dos modelos foi realizada por meio das métricas MAE_{diurno} , $RMSE_{diurno}$ e R^2_{diurno} , calculadas sobre os conjuntos de validação e teste. Além das métricas individuais, o desempenho da LSTM foi comparado ao modelo de persistência utilizando o *Skill Score*, calculado a partir do $RMSE_{diurno}$. Os dois melhores modelos foram selecionados para serem utilizados nas etapas a seguir de monitoramento de degradação e aprendizado incremental.

4.2.3 Monitoramento de degradação e aprendizado incremental

Após a seleção dos melhores modelos de previsão, foi realizada uma etapa de simulação para o monitoramento de degradação integrado ao aprendizado incremental. O objetivo dessa etapa é identificar perdas de desempenho ao longo da simulação e, quando necessário, atualizar o modelo para adaptá-lo às mudanças observadas nos dados.

O monitoramento foi realizado diariamente utilizando o MAE_{diurno} . Ao final de cada dia de simulação, o valor obtido é utilizado para atualizar o método de monitoramento. Quando um processo de degradação é detectado, é acionado um procedimento de aprendizado incremental, responsável por atualizar o modelo utilizando os dados mais recentes disponíveis. Foram avaliadas duas estratégias de monitoramento de degradação, *Cumulative Sum Control Chart* (CUSUM) e *Adaptive Windowing* (ADWIN), combinadas com duas estratégias de aprendizado incremental: *Fine-Tuning* e *Elastic Weight Consolidation* (EWC). Dessa forma, para cada um dos dois modelos selecionados na etapa anterior, foram consideradas quatro configurações experimentais.

Em ambos os métodos de aprendizado incremental, o retreinamento é realizado utilizando uma janela móvel contendo os 30 dias anteriores ao dia em que o monitoramento identifica a degradação. O processo de atualização emprega a função de perda MSE e, após cada retreinamento, é registrada a variação do MAE obtida, permitindo avaliar o impacto de cada atualização no desempenho do modelo ao longo da simulação.

4.2.4 Avaliação em um novo sítio de geração

A melhor configuração de monitoramento de degradação e aprendizado incremental obtida na etapa anterior foi avaliada em um segundo sítio de geração pertencente ao mesmo campus do conjunto de dados UNISOLAR. O objetivo dessa etapa é verificar a capacidade de generalização da estratégia proposta quando aplicada a uma usina com características distintas daquela utilizada durante o treinamento, mas que compartilha os mesmos dados de clima por ser no mesmo campus. Como os dois sítios possuem capacidades instaladas diferentes, foi inicialmente calculado um fator de escala para tornar as grandezas comparáveis. Esse fator foi definido pela razão entre a potência instalada do sítio de destino e a potência instalada do sítio de origem, conforme a Equação 32.

$$kwp_{ratio} = \frac{kwp_{destino}}{kwp_{origem}} \quad (32)$$

Em seguida, o conjunto de dados do novo sítio foi submetido às mesmas etapas de pré-processamento empregadas no sítio de origem, incluindo limpeza dos dados e engenharia de atributos. Posteriormente, foi aplicado o processo de normalização utilizando os mesmos parâmetros obtidos durante o treinamento no sítio de origem, ajustados pelo fator de escala calculado anteriormente. Antes do início da simulação, foi adotado um período de *warm-up* de sete dias. Após esse período, todo o restante do conjunto de

dados foi utilizado na simulação, durante a qual foram executados o monitoramento de degradação e o aprendizado incremental.

Com a conclusão das etapas de pré-processamento, treinamento dos modelos, monitoramento de degradação, aprendizado incremental e avaliação em um novo sítio de geração, encerra-se a metodologia adotada para os experimentos realizados com o conjunto de dados UNISOLAR. Na seção seguinte, são apresentadas as considerações finais deste capítulo, sintetizando a metodologia proposta para a dissertação.

4.3 Considerações finais

Este capítulo apresentou a metodologia proposta nesta dissertação de mestrado. Inicialmente, foi descrita a metodologia desenvolvida a partir do conjunto de dados Sol do Cerrado, contemplando as etapas de análise exploratória, pré-processamento, divisão dos dados, treinamento dos modelos de previsão e seleção da configuração experimental de melhor desempenho. Nessa fase, foram comparados diferentes modelos de previsão, estratégias de agregação temporal, métodos de escalonamento e técnicas de otimização de hiperparâmetros, permitindo identificar a configuração mais adequada para o problema estudado.

Na sequência, foi apresentada a metodologia aplicada ao conjunto de dados UNISOLAR, na qual a configuração experimental selecionada anteriormente foi utilizada como base para o desenvolvimento da arquitetura proposta neste trabalho. Essa etapa compreende o monitoramento de degradação, a atualização incremental do modelo e a avaliação de sua capacidade de generalização em um novo sítio de geração. O próximo capítulo apresenta os experimentos realizados e discute os resultados obtidos em cada uma das etapas da metodologia proposta.

Experimentos e Análise dos Resultados

Este capítulo apresenta os experimentos realizados e os resultados obtidos nesta dissertação. Sua organização segue a mesma estrutura apresentada no Capítulo 4 para a metodologia. A Seção 5.1 apresenta os resultados obtidos com o conjunto de dados Sol do Cerrado, incluindo a análise exploratória, a comparação entre os modelos de previsão, as estratégias de agregação temporal, os métodos de escalonamento e as técnicas de otimização de hiperparâmetros.

Em seguida, a Seção 5.2 apresenta os experimentos realizados com o conjunto de dados UNISOLAR, contemplando a avaliação das variantes do conjunto de dados, das estratégias de monitoramento de degradação, dos métodos de aprendizado incremental e da capacidade de generalização para um novo sítio de geração. Por fim, a Seção 5.3 apresenta as considerações finais do capítulo, sintetizando os principais resultados obtidos.

5.1 Sol do Cerrado

Esta seção apresenta os resultados obtidos com o conjunto de dados Sol do Cerrado. Inicialmente, são apresentados os resultados da análise exploratória dos dados, incluindo a análise de correlação entre as variáveis. Em seguida, são discutidos os experimentos realizados para avaliar diferentes configurações dos modelos de previsão, contemplando distintas granularidades temporais, estratégias de agregação, métodos de escalonamento dos dados e técnicas de otimização de hiperparâmetros. Por fim, é realizada uma comparação entre os modelos avaliados, destacando os hiperparâmetros selecionados e a configuração que apresentou o melhor desempenho.

5.1.1 Análise exploratória dos dados

A princípio, foi realizada uma verificação da qualidade do conjunto de dados, com o objetivo de identificar a existência de valores ausentes. Como nenhum registro apresentou dados faltantes, o conjunto de dados foi mantido integralmente, totalizando 1.641.600

registros. Em seguida, foram gerados gráficos diários para todas as variáveis monitoradas pelos sensores, com o objetivo de realizar uma análise exploratória visual e compreender o comportamento temporal das medições.

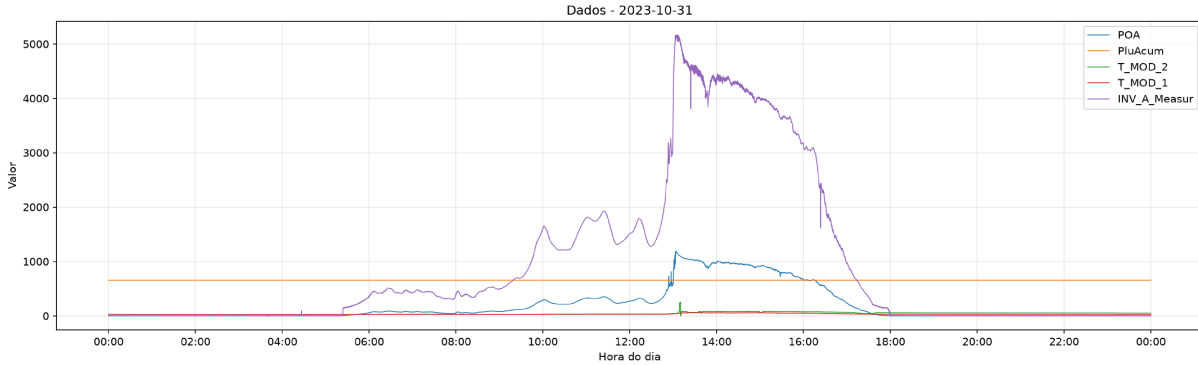


Figura 11 – Dados do dia 31-10-2023 do dataset Sol do Cerrado.

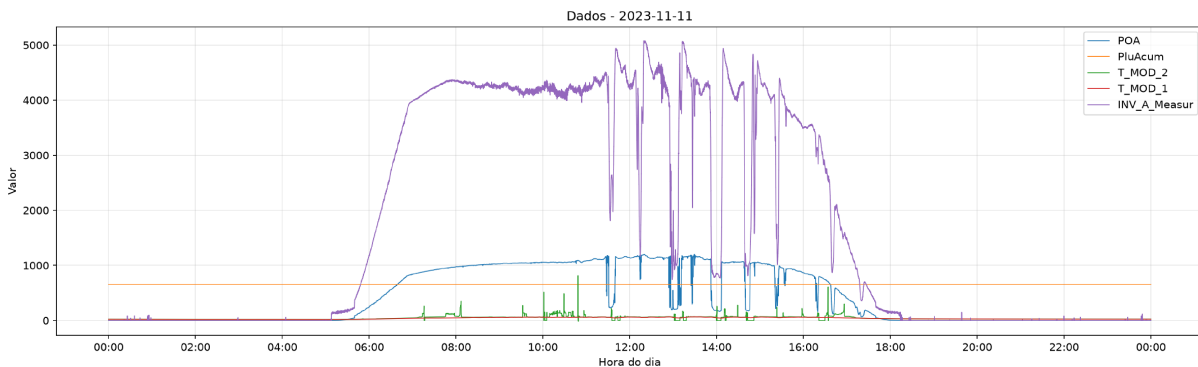


Figura 12 – Dados do dia 11-11-2023 do dataset Sol do Cerrado.

As Figuras 11 e 12 mostram o comportamento das variáveis monitoradas em dois dias distintos do conjunto de dados Sol do Cerrado. Essa análise visual permitiu identificar que a variável **PluAcum** permanecia constante ao longo dos dias analisados. Posteriormente, verificou-se que esse comportamento se repetia em todo o conjunto de dados, motivando sua remoção antes da construção dos modelos.

Outro aspecto observado foi a forte relação entre a variável **POA**, que indica a irradiância solar, e **INV_A_Measur**, que representa a geração de energia fotovoltaica. Na Figura 11, observa-se um dia com baixa irradiância durante o período da manhã, resultando no atraso do início da geração de energia. Em contraste, a Figura 12 apresenta um dia com geração iniciada nas primeiras horas da manhã e comportamento relativamente estável até o meio do dia, quando passam a ocorrer oscilações mais acentuadas, possivelmente associadas à presença de nuvens transitórias.

Também foi possível observar diferenças entre os sensores de temperatura dos módulos fotovoltaicos. A variável T_MOD_1 apresentou comportamento mais estável ao longo do tempo, enquanto T_MOD_2 exibiu maior oscilação. Essa observação motivou uma investigação mais detalhada sobre o comportamento de T_MOD_2, apresentada a seguir.

Análise do comportamento de T_MOD_2

Diante do comportamento da variável T_MOD_2 observado nas Figuras 11 e 12, foi realizada uma investigação mais detalhada sobre esse sensor. Para isso, empregou-se o método IQR, amplamente utilizado para identificação de *outliers*. Inicialmente, foram calculados o primeiro quartil (Q_1), correspondente ao 25º percentil, o terceiro quartil (Q_3), correspondente ao 75º percentil, e o intervalo interquartil, definido pela Equação 33.

$$IQR = Q_3 - Q_1 \quad (33)$$

$$L_{inf} = Q_1 - 1,5 \times IQR \quad (34)$$

$$L_{sup} = Q_3 + 1,5 \times IQR \quad (35)$$

Os limites definidos pelas Equações 34 e 35 foram utilizados para identificar valores discrepantes da variável T_MOD_2. A Figura 13 apresenta uma comparação entre os sensores T_MOD_1 e T_MOD_2. Observa-se que a variável T_MOD_2 apresenta variabilidade significativamente superior à observada em T_MOD_1, além da ocorrência frequente de valores extremos incompatíveis com o comportamento esperado para a temperatura dos módulos fotovoltaicos.

A análise estatística confirma esse comportamento. O conjunto de dados possui 1.641.600 observações, média de 47,73 °C e desvio-padrão de 32,72 °C. Entretanto, foram registrados valores entre -130,57 °C e 813,90 °C, evidenciando medições fisicamente inconsistentes. Considerando os limites calculados pelo método IQR, foram identificados 86.515 valores discrepantes, correspondendo a aproximadamente 5,27% das observações, sendo praticamente todos acima do limite superior (86.501 registros), enquanto apenas 14 observações ficaram abaixo do limite inferior. Esses resultados indicam que o sensor associado à variável T_MOD_2 não apresenta confiabilidade suficiente para ser utilizada nos experimentos, motivando sua exclusão do conjunto de dados antes da etapa de treinamento dos modelos.

Análise de correlação

Paralelamente à análise da variável T_MOD_2, foi realizada uma análise de correlação entre as variáveis do conjunto de dados, com o objetivo de identificar o grau de associação

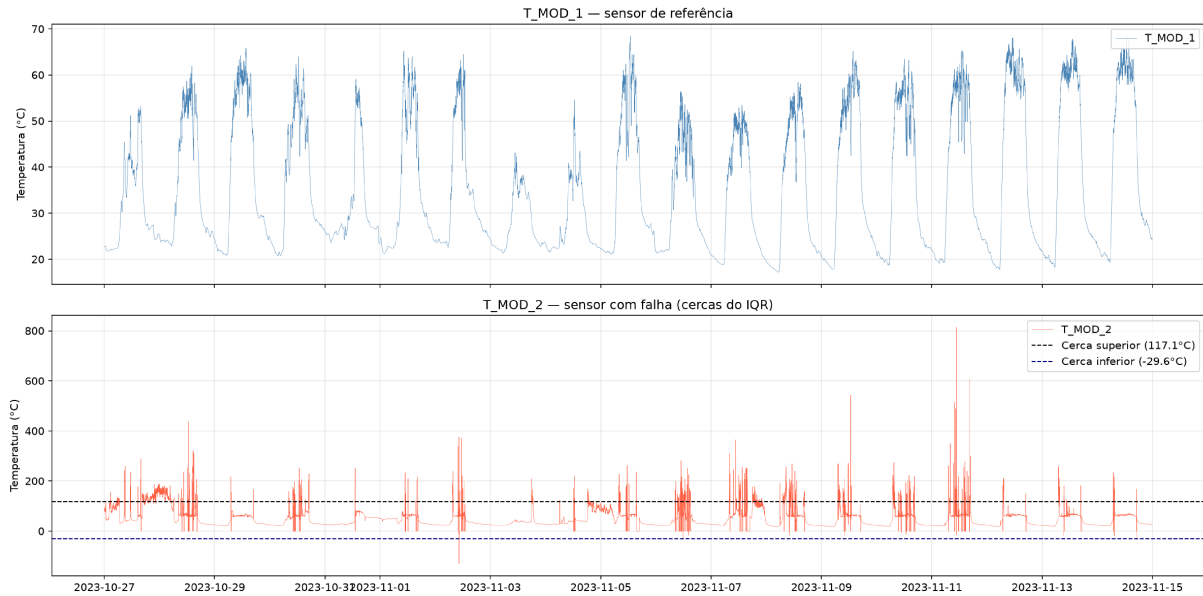


Figura 13 – Comparação das variáveis T_MOD_1 e T_MOD_2 com os limites do IQR.

entre elas. Para isso, foram calculados os coeficientes de correlação de Pearson e de Spearman, além da correlação cruzada com defasagem. Os coeficientes de Pearson e de Spearman são apresentados na matriz de correlação da Figura 14.

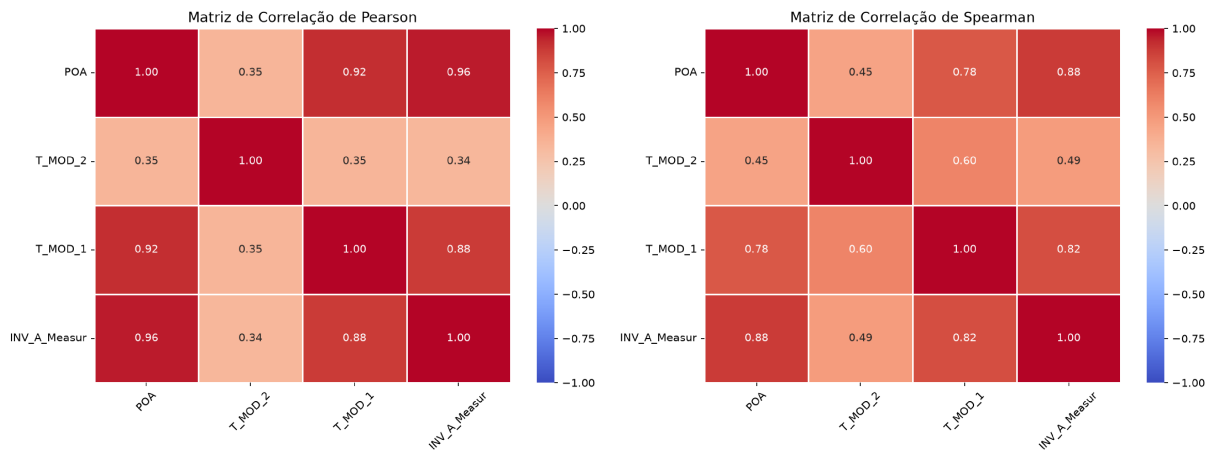


Figura 14 – Matriz de correlação das variáveis do Sol do Cerrado utilizando os coeficientes de Pearson e Spearman.

Conforme observado na Figura 14, a variável POA apresenta a maior correlação com a variável INV_A_Measur, seguido por T_MOD_1. Enquanto isso, T_MOD_2 apresenta a menor correlação com INV_A_Measur. Esperava-se que ambas — T_MOD_1 e T_MOD_2 — apresentassem coeficientes de correlação de magnitudes semelhantes em relação à variável INV_A_Measur. Entretanto, esse comportamento não foi observado para T_MOD_2, reforçando as evidências obtidas na análise estatística apresentada anteriormente e corro-

borando para a decisão de remover T_MOD_2.

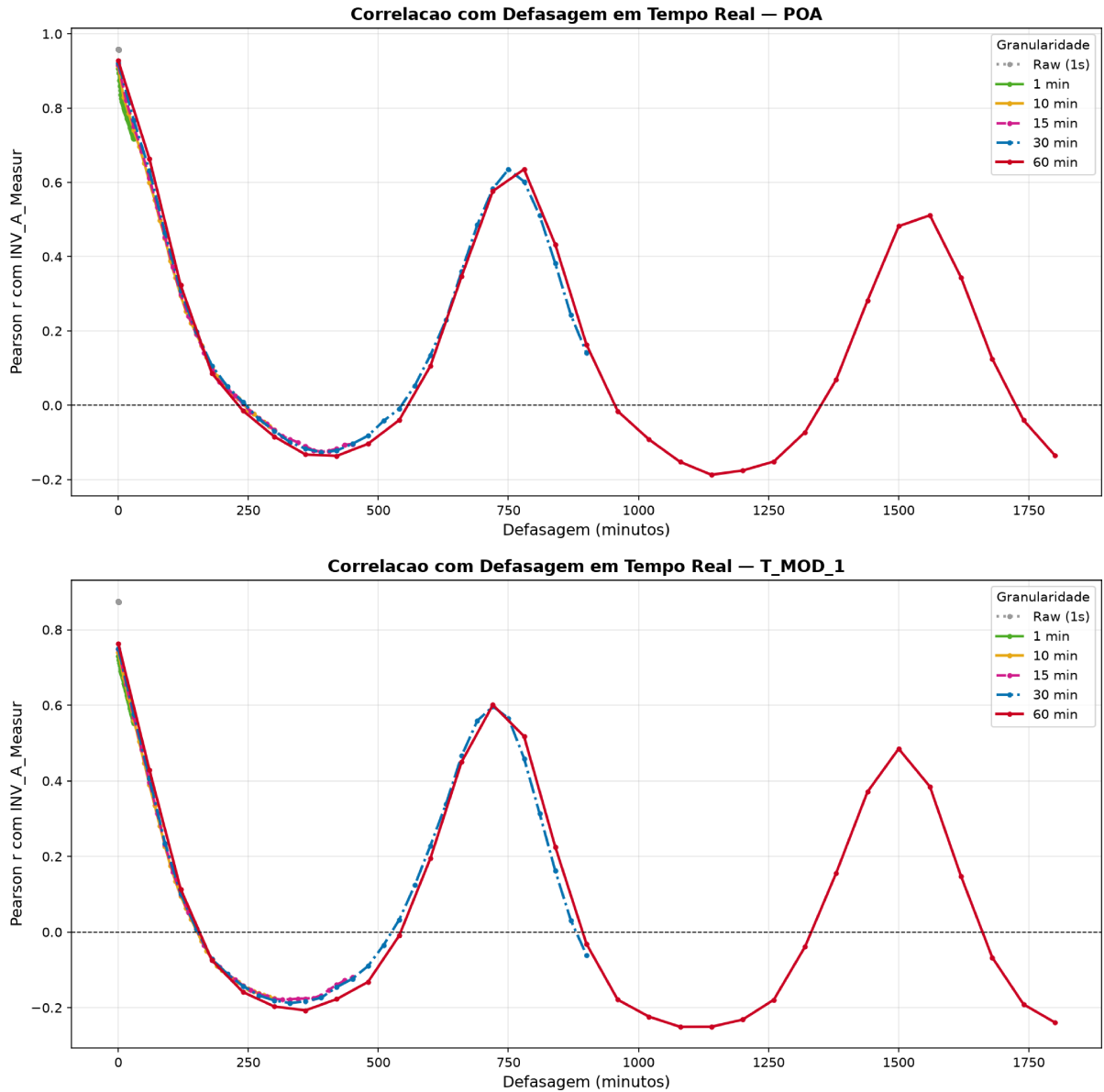


Figura 15 – Análise de correlação cruzada com defasagem entre as variáveis e INV_A_Measur.

A Figura 15 apresenta a análise de correlação por defasagem utilizando o coeficiente de correlação de Pearson, considerando até 30 pontos de defasagem. Observa-se que o maior coeficiente de correlação ocorre em *lag* igual a zero, indicando a maior associação entre observações simultâneas. Além disso, à medida que a defasagem aumenta, nota-se um comportamento aproximadamente senoidal da correlação, evidenciando a sazonalidade presente na série temporal, especialmente associada ao ciclo diário de geração de energia fotovoltaica. Esse comportamento foi observado em todas as granularidades temporais avaliadas, indicando que a estrutura temporal dos dados é preservada independentemente do intervalo de agregação considerado.

Após a realização da análise exploratória dos dados, foi conduzido o pré-processamento do conjunto de dados e o treinamento dos modelos utilizando as diferentes variantes geradas. A próxima seção descreve o processo de treinamento dos modelos e apresenta uma análise comparativa dos resultados obtidos para cada modelo e para cada variante considerada.

5.1.2 Avaliação das configurações experimentais

Foram avaliadas 30 variantes do conjunto de dados original, obtidas pela combinação de diferentes granularidades temporais, métodos de agregação e técnicas de escalonamento. Além disso, foram considerados cinco modelos de previsão: persistência, ARIMA, regressão Ridge, MLP e LSTM. Para os modelos MLP e LSTM, foram utilizadas duas estratégias de otimização de hiperparâmetros, *Optuna* e *Ray Tune*, resultando em sete configurações de modelos. As configurações experimentais são resumidas na Tabela 1. A relação completa das 30 variantes do conjunto de dados, bem como suas respectivas denominações, é apresentada na Tabela 16 do Apêndice A.

Tabela 1 – Resumo das configurações experimentais avaliadas.

Fator experimental	Configurações avaliadas
Granularidade temporal	1, 10, 15, 30 e 60 minutos
Método de agregação	Dado bruto, Média e Mediana
Escalonamento dos dados	<i>Min-Max</i> e <i>Z-score</i>
Modelos de previsão	Persistência, ARIMA, Ridge, MLP e LSTM
Otimização de hiperparâmetros	<i>Optuna</i> e <i>Ray Tune</i>

Na sequência, são apresentados os detalhes do processo de treinamento dos modelos e realizada uma série de análises comparativas das diferentes configurações experimentais. São investigados os efeitos do método de agregação, da técnica de escalonamento dos dados e da estratégia de otimização de hiperparâmetros sobre o desempenho dos modelos de previsão.

Processo de treinamento

O modelo de persistência não requer etapa de treinamento, uma vez que a previsão para o instante seguinte é obtida diretamente a partir da última observação disponível. As previsões foram realizadas considerando apenas observações pertencentes ao mesmo dia, evitando que valores do final de um dia fossem utilizados para prever o início do dia seguinte.

Para o modelo ARIMA, a seleção automática da estrutura (p, d, q) foi realizada utilizando a função *auto_arima* da biblioteca *pmdarima*¹. A busca pelos hiperparâmetros em-

¹ <https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html>

pregou o algoritmo de Hyndman-Khandakar, ativado por meio da opção *stepwise=True*. Esse algoritmo realiza uma busca gulosa baseada no *Akaike Information Criterion* (AIC). Foram delimitados valores máximos de $p = q = 12$. A ordem de diferenciação d foi determinada automaticamente pela biblioteca por meio do teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS), utilizando sua configuração padrão.

Em relação ao modelo de regressão Ridge foi realizada uma *grid search* sobre os hiperparâmetros *look-back* e coeficiente de regularização α . O espaço de busca é apresentado na Tabela 2. O erro quadrático médio (MSE) do conjunto de validação foi adotado como critério para a seleção do melhor valor de α .

Tabela 2 – Espaço de busca dos hiperparâmetros do modelo de regressão Ridge.

Hiperparâmetro	Valores avaliados
<i>look-back</i>	10, 30, 60
α	10^{-3} , 10^{-2} , 10^{-1} , 10^0 , 10^1 , 10^2 , 10^3

Por fim, os modelos MLP e LSTM tiveram seus hiperparâmetros ajustados utilizando duas estratégias distintas de otimização: *Optuna*² e *Ray Tune*³. Para cada variante do conjunto de dados foram executadas 50 *trials*, considerando diferentes combinações de hiperparâmetros. O espaço de busca dos hiperparâmetros é apresentado na Tabela 3.

Tabela 3 – Espaço de busca dos hiperparâmetros dos modelos MLP e LSTM.

Hiperparâmetro	LSTM	MLP
Tamanho da camada oculta	32, 64, 128, 256	64, 128, 256, 512
Número de camadas / <i>dropout</i>	1, 2	0, 0.5
Taxa de aprendizado	Log-uniforme [10^{-4} , 10^{-1}]	Log-uniforme [10^{-4} , 10^{-1}]
<i>Batch size</i>	16, 32, 64, 128	16, 32, 64, 128
<i>Look-back</i>	10, 30, 60	10, 30, 60

Todos os modelos foram treinados utilizando o otimizador Adam e a função de perda MSE, empregando *gradient clipping* com norma máxima igual a 1,0 para aumentar a estabilidade do treinamento. O treinamento foi limitado a 100 épocas, com *early stopping* após dez épocas consecutivas sem melhora. Após a etapa de otimização, a melhor configuração de hiperparâmetros encontrada foi retreinada três vezes com diferentes inicializações aleatórias dos pesos. Entre essas execuções, foi selecionado o modelo que apresentou a menor perda no conjunto de validação.

² <<https://optuna.readthedocs.io/en/stable/>>

³ <<https://docs.ray.io/en/latest/tune/index.html>>

Análise de agregação

Após o treinamento dos modelos, o conjunto de teste foi utilizado para calcular as métricas de desempenho e realizar as comparações entre as diferentes configurações avaliadas. Inicialmente, foi analisada a influência da estratégia de agregação dos dados sobre o desempenho dos modelos. Nesse contexto, a Tabela 4 apresenta a frequência com que cada estratégia de agregação — dado bruto, média e mediana — obteve o melhor desempenho em cada métrica.

Tabela 4 – Percentual de melhores resultados por tipo de conjunto de dados considerando as 70 execuções.

Agregação	MAE	RMSE	R^2
Dado bruto	11,4%	12,9%	14,3%
Média	55,7%	61,4%	61,4%
Mediana	32,9%	25,7%	24,3%

Pode-se observar na Tabela 4 que a agregação por média apresentou a maior frequência de melhores resultados nas três métricas avaliadas, indicando um desempenho superior em relação ao dado bruto e à mediana. A fim de aprofundar essa análise, a Tabela 5 apresenta a distribuição dessas vitórias considerando cada modelo de previsão. Os valores apresentados correspondem ao número absoluto de vitórias, totalizando 10 para cada modelo, resultado da combinação entre as cinco granularidades e as duas técnicas de normalização.

Tabela 5 – Número de melhores resultados por tipo de agregação obtidos por cada modelo para as métricas MAE, RMSE e R^2 .

Modelo	MAE			RMSE			R^2		
	DB	M	Md	DB	M	Md	DB	M	Md
Persistência	0	4	6	0	8	2	0	8	2
ARIMA	0	7	3	1	7	2	1	8	1
Ridge	0	9	1	1	7	2	2	6	2
MLP + <i>Optuna</i>	2	7	1	2	7	1	2	7	1
MLP + <i>Ray Tune</i>	4	4	2	4	4	2	4	4	2
LSTM + <i>Optuna</i>	1	5	4	1	6	3	1	6	3
LSTM + <i>Ray Tune</i>	1	3	6	0	4	6	0	4	6

A Tabela 5 evidencia o predomínio da agregação por média nos melhores resultados na maior parte das configurações analisadas. As exceções são observadas no modelo LSTM otimizado com *Ray Tune*, em que a agregação por mediana apresentou melhor desempenho em todas as métricas, e no modelo de Persistência, no qual a mediana foi superior apenas para a métrica MAE. Além disso, observa-se um empate entre o dado

bruto e a média no modelo MLP otimizado com *Ray Tune*, independentemente da métrica considerada. Na sequência, a Tabela 6 apresenta a frequência de melhores resultados para cada granularidade temporal. Nesse caso, os valores de cada linha por métrica totalizam 14, correspondentes às combinações entre os sete modelos avaliados e as duas técnicas de normalização.

Tabela 6 – Número de melhores resultados por agregação obtidos para cada granularidade nas métricas MAE, RMSE e R^2 .

Granularidade	MAE			RMSE			R^2		
	DB	M	Md	DB	M	Md	DB	M	Md
1 min	4	9	1	4	10	0	4	10	0
10 min	2	10	2	2	10	2	3	9	2
15 min	0	7	7	0	9	5	0	9	5
30 min	0	6	8	0	8	6	0	10	4
60 min	2	7	5	3	6	5	3	5	6

A Tabela 6 reforça o predomínio da agregação por média nos melhores resultados entre as diferentes granularidades temporais. As únicas exceções são observadas na granularidade de 30 minutos, considerando a métrica MAE, e na granularidade de 60 minutos, considerando a métrica R^2 , nas quais a agregação por mediana apresentou a maior frequência de melhores resultados. Por sua vez, a Tabela 7 apresenta a distribuição dos melhores resultados de agregação em função da técnica de escalonamento empregada, evidenciando que a agregação por média predominou em ambos escalonamentos e em todas as métricas avaliadas.

Tabela 7 – Número de melhores resultados por agregação obtidos para cada técnica de escalonamento nas métricas MAE, RMSE e R^2 .

Escalonamento	MAE			RMSE			R^2		
	DB	M	Md	DB	M	Md	DB	M	Md
<i>Z-score</i>	4	20	11	5	21	9	4	22	9
<i>Min-Max</i>	4	19	12	4	22	9	6	21	8

Por fim, a análise da estratégia de agregação mostrou que a utilização da média proporcionou, de forma consistente, a maior frequência de melhores resultados entre os modelos avaliados. Em seguida, foi analisada a influência dos diferentes escalonamentos de dados no desempenho dos modelos.

Análise de escalonamento dos dados

A próxima etapa consistiu na análise do impacto da técnica de escalonamento dos dados sobre o desempenho dos modelos, considerando todas as variantes avaliadas. A

Figura 16 apresenta, para a métrica MAE, a frequência com que cada técnica de normalização obteve o melhor desempenho em cada modelo. Optou-se por apresentar apenas essa métrica, uma vez que as análises realizadas com RMSE e R^2 conduziram às mesmas conclusões quanto à técnica de escalonamento predominante, diferindo apenas na quantidade de vitórias obtidas. Os resultados completos para as métricas RMSE e R^2 encontram-se na Figura 25 e 26, respectivamente, no Apêndice A.

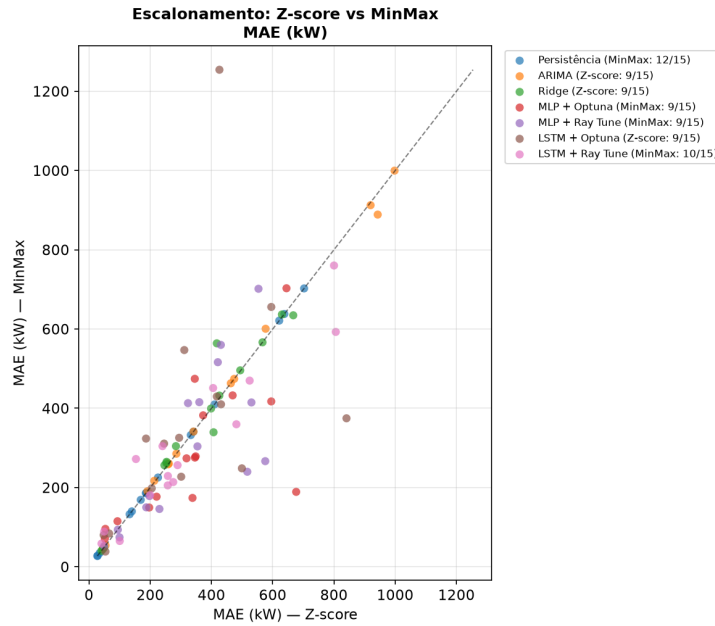


Figura 16 – Contagem dos melhores resultados por método de escalonamento para cada modelo de previsão.

A Figura 16 mostra que o *Min-Max* apresentou melhor desempenho nos modelos de persistência, MLP otimizado com *Optuna*, MLP otimizado com *Ray Tune* e LSTM otimizado com *Ray Tune*. Por outro lado, a padronização *Z-score* obteve melhores resultados nos modelos ARIMA, Ridge e LSTM otimizado com *Optuna*. Considerando apenas os modelos de aprendizado de máquina, o *Min-Max* foi superior em três dos quatro modelos avaliados, indicando uma tendência de melhor adaptação dessa técnica às arquiteturas baseadas em redes neurais utilizadas neste trabalho. Além disso, observa-se que a maioria dos pontos encontra-se próxima da linha de identidade, indicando que, embora uma das técnicas apresente desempenho superior em determinadas configurações, as diferenças de MAE entre as duas estratégias de escalonamento são, em geral, pequenas.

Análise do otimizador de hiperparâmetros

Posteriormente, foi analisada a influência do método de otimização de hiperparâmetros no desempenho dos modelos LSTM e MLP. A Figura 17 apresenta a comparação entre os algoritmos *Ray Tune* e *Optuna* considerando a métrica MAE. Para o modelo

LSTM, observa-se um equilíbrio entre os dois métodos de otimização, sem predominância significativa de um deles. Já para o modelo MLP, o *Ray Tune* apresentou melhor desempenho por 1 modelo, obtendo 16 melhores das 30 configurações avaliadas. Os resultados obtidos para as métricas RMSE e R^2 confirmam a análise da Figura 17 e são apresentados na Figura 27 e 28, respectivamente, no Apêndice A.

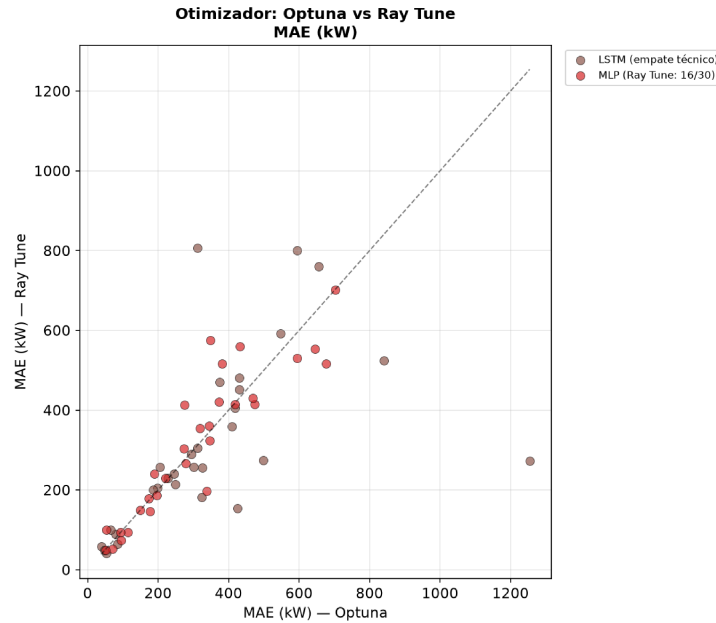


Figura 17 – Contagem de melhores resultados por otimizador de hiperparâmetros para cada modelo.

5.1.3 Comparação dos modelos de previsão

Por fim, foi realizada uma comparação dos modelos de previsão em todas as variantes montadas. Com essa comparação é possível analisar qual modelo foi melhor para cada granularidade dos dados. A Figura 18 apresenta o desempenho da melhor configuração obtida para cada modelo em cada granularidade temporal, considerando a métrica MAE. Os resultados completos de todas as configurações avaliadas, considerando todas as métricas computadas, são apresentados nas Tabelas 17, 18, 19 e 20, localizadas no Apêndice A.

A persistência apresentou o menor MAE nas granularidades de 1 e 10 minutos. Entretanto, nas granularidades de 15, 30 e 60 minutos, o modelo LSTM com *Ray Tune* obteve os menores valores de MAE. Observa-se ainda que, à medida que a granularidade aumenta, o desempenho dos modelos persistência, ARIMA e Ridge se deteriora em relação aos modelos MLP e LSTM. Os *heatmaps* das métricas RMSE, R^2 e *Skill Score* são apresentados nas Figuras 29, 30 e 31, respectivamente, no Apêndice A.

A Figura 19 apresenta a melhor configuração de cada modelo para cada granularidade, considerando conjuntamente as quatro métricas de avaliação: MAE, RMSE, R^2 e *Skill*

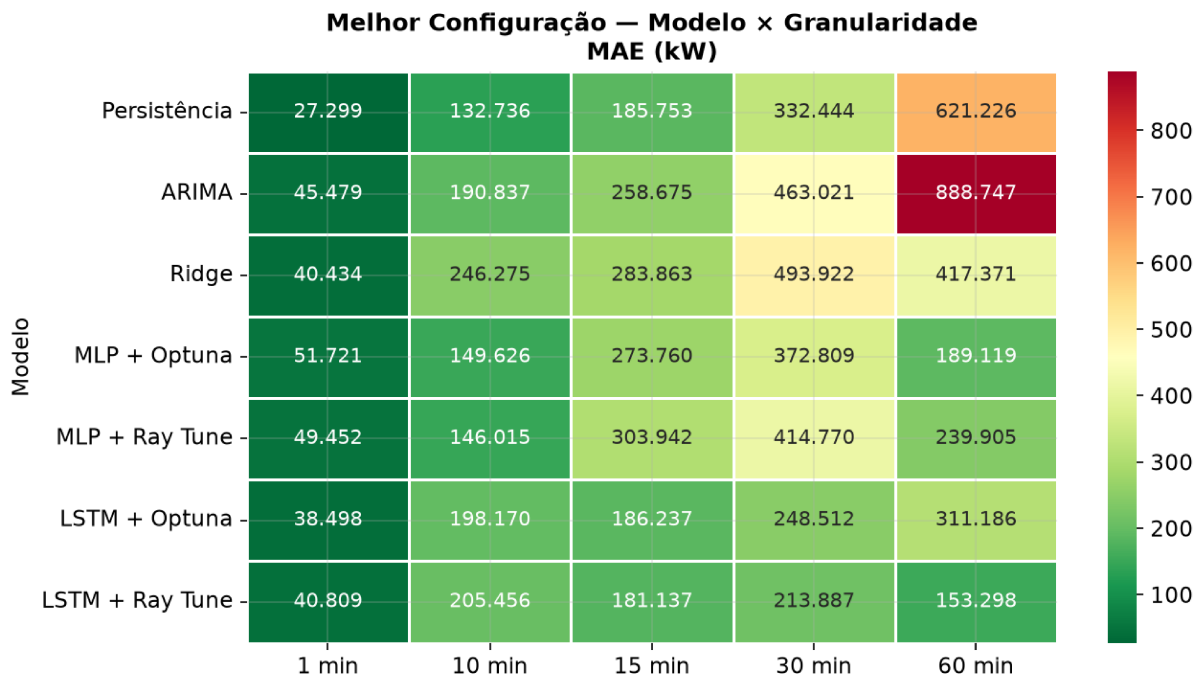


Figura 18 – Melhor configuração de cada modelo por granularidade na métrica MAE.

Score. Os modelos apresentam desempenho superior nas menores granularidades (1, 10 e 15 minutos), refletido pelos menores valores de MAE e RMSE e pelos maiores valores de R^2 . À medida que a granularidade aumenta para 30 e 60 minutos, verifica-se uma degradação do desempenho para a maioria dos modelos.

Entre os modelos avaliados, a persistência apresentou os menores erros nas granularidades de 1 e 10 minutos, demonstrando elevada competitividade em previsões de curto prazo. Entretanto, seu desempenho deteriora-se significativamente nas granularidades mais elevadas, especialmente em 60 minutos, onde o MAE ultrapassa 620 kW. Já os modelos baseados em aprendizado de máquina mostraram maior capacidade de generalização nas maiores granularidades. Destacam-se o LSTM otimizado com *Ray Tune* e o MLP otimizado com *Optuna*, que obtiveram os menores valores de MAE e RMSE em 60 minutos, além dos maiores valores de *Skill Score*, indicando desempenho superior ao modelo de persistência.

A métrica *Skill Score* reforça esse comportamento, pois os modelos ARIMA e Ridge apresentam valores negativos em diversas configurações, indicando desempenho inferior a persistência. Em contrapartida, os modelos MLP e LSTM obtêm *Skill Scores* positivos na maior parte das granularidades.

Com isso, encerra-se a análise do conjunto de dados Sol do Cerrado. Os resultados obtidos indicam que a agregação por média apresentou desempenho superior, o escalonamento *Min-Max* foi o mais adequado para os modelos de aprendizado de máquina e que o otimizador *Ray Tune* apresentou uma vantagem discreta em relação ao *Optuna*. Entre os

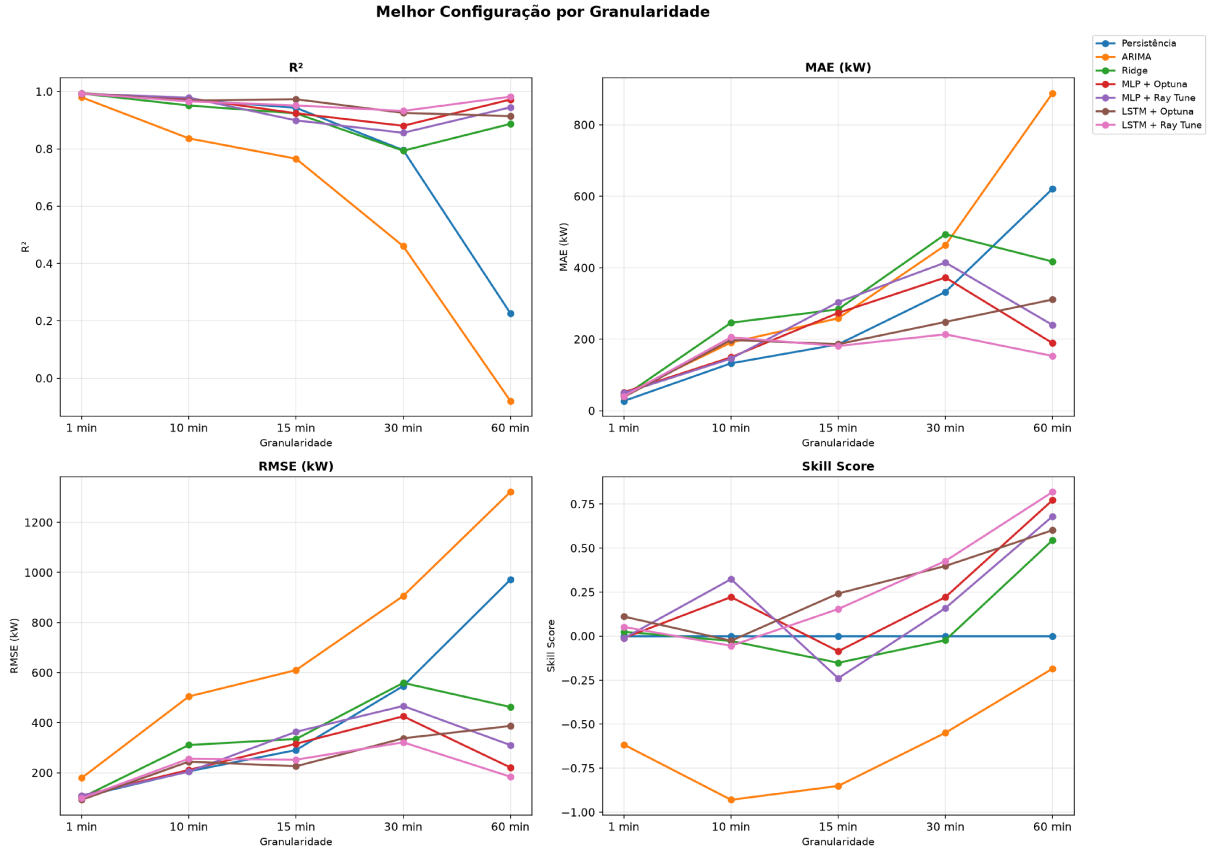


Figura 19 – Melhor configuração de cada modelo por granularidade em MAE, RMSE, R^2 e *Skill Score*.

modelos avaliados, o LSTM destacou-se por apresentar o melhor desempenho na maioria das granularidades analisadas. Dessa forma, essas configurações foram adotadas na etapa seguinte da pesquisa, referente ao conjunto de dados UNISOLAR, cujos experimentos e resultados são apresentados na sequência.

5.2 UNISOLAR

A próxima etapa dos experimentos foi conduzida utilizando o conjunto de dados UNISOLAR. Os experimentos e as análises foram organizados em quatro etapas. A primeira consiste na análise exploratória do conjunto de dados, abrangendo as etapas de limpeza dos dados e análise de correlação entre as *features*. Em seguida, é realizado o treinamento dos modelos utilizando as diferentes variantes do conjunto de dados. Na terceira etapa, são conduzidos os experimentos de monitoramento de degradação e aprendizado incremental. Por fim, o melhor modelo é avaliado em um sítio distinto pertencente ao mesmo campus, com o objetivo de verificar a capacidade de generalização da abordagem proposta.

5.2.1 Análise exploratória dos dados

Em um primeiro momento, foi realizada a junção entre os dados de geração de energia solar e as variáveis meteorológicas correspondentes. A partir do conjunto de dados resultante, o **CampusKey** igual a 1 e o **SiteKey** igual a 27 foram selecionados para a realização dos experimentos por apresentarem a maior quantidade de registros disponíveis. Ao todo, o conjunto de dados selecionado é composto por 79.798 registros com granularidade temporal de 15 minutos e sua capacidade de energia instalada é 539,80 kWp.

Limpeza dos dados

Entre os registros do **SiteKey** 27 e **CampusKey** 1, 29.293 não continham valores para as variáveis **WindSpeed** e **WindDirection**, chegando a 36,71% dos dados. Devido a essa grande quantidade de valores ausentes, essas duas variáveis foram removidas do conjunto de dados. Dessa forma, cada registro passou a ser composto pelas variáveis **ApparentTemperature**, **AirTemperature**, **DewPointTemperature**, **RelativeHumidity** e **SolarGeneration**. Após isso, foram aplicadas as regras de limpeza ao conjunto de dados. A Tabela 8 apresenta um resumo das etapas de tratamento realizadas e seus impactos nos registros.

Tabela 8 – Resumo quantitativo das etapas de limpeza e tratamento dos dados.

Etapas	Resultado
Conjunto de dados original	79 798 registros
Preenchimento de variáveis meteorológicas (Regra 1)	14 520 registros preenchidos
Preenchimento de geração noturna com zero (Regra 2)	32 238 registros preenchidos
Remoção de períodos com falha de sensor (Regra 3)	8 128 registros removidos
Preenchimento de geração durante o dia (Regra 4)	8 043 registros preenchidos
Remoção de registros com valores faltantes (Regra 5)	78 registros removidos
Conjunto de dados limpo	71 592 registros
Remoção dos registros noturnos (Regra 6)	35 806 registros removidos
Conjunto de dados sem período noturno	35 786 registros

Ao término dessa etapa, foram obtidas duas versões do conjunto de dados: uma contendo o período noturno, composta por 71.592 registros, e outra sem o período noturno, totalizando 35.786 registros. Em seguida, foi realizada a etapa de engenharia de atributos, resultando em duas variantes para cada conjunto de dados: uma com a inclusão das variáveis cíclicas correspondentes à hora do dia e ao dia do ano, e outra mantendo apenas as variáveis originais. Esse processo produziu quatro variantes do conjunto de dados.

Análise de correlação de *features*

Como próximo passo, foi realizada uma análise de correlação entre as *features*. Essa análise foi conduzida no conjunto de dados sem o período noturno, uma vez que a inclusão desses registros poderia introduzir uma correlação artificial decorrente do ciclo diário nas variáveis temporais e a geração nula de energia durante a noite. Dessa forma, a análise de correlação representa de maneira mais fiel as relações existentes durante o período efetivo de geração de energia.

A Figura 20 apresenta a correlação de Pearson, enquanto a Figura 21 expõe a correlação de Spearman.

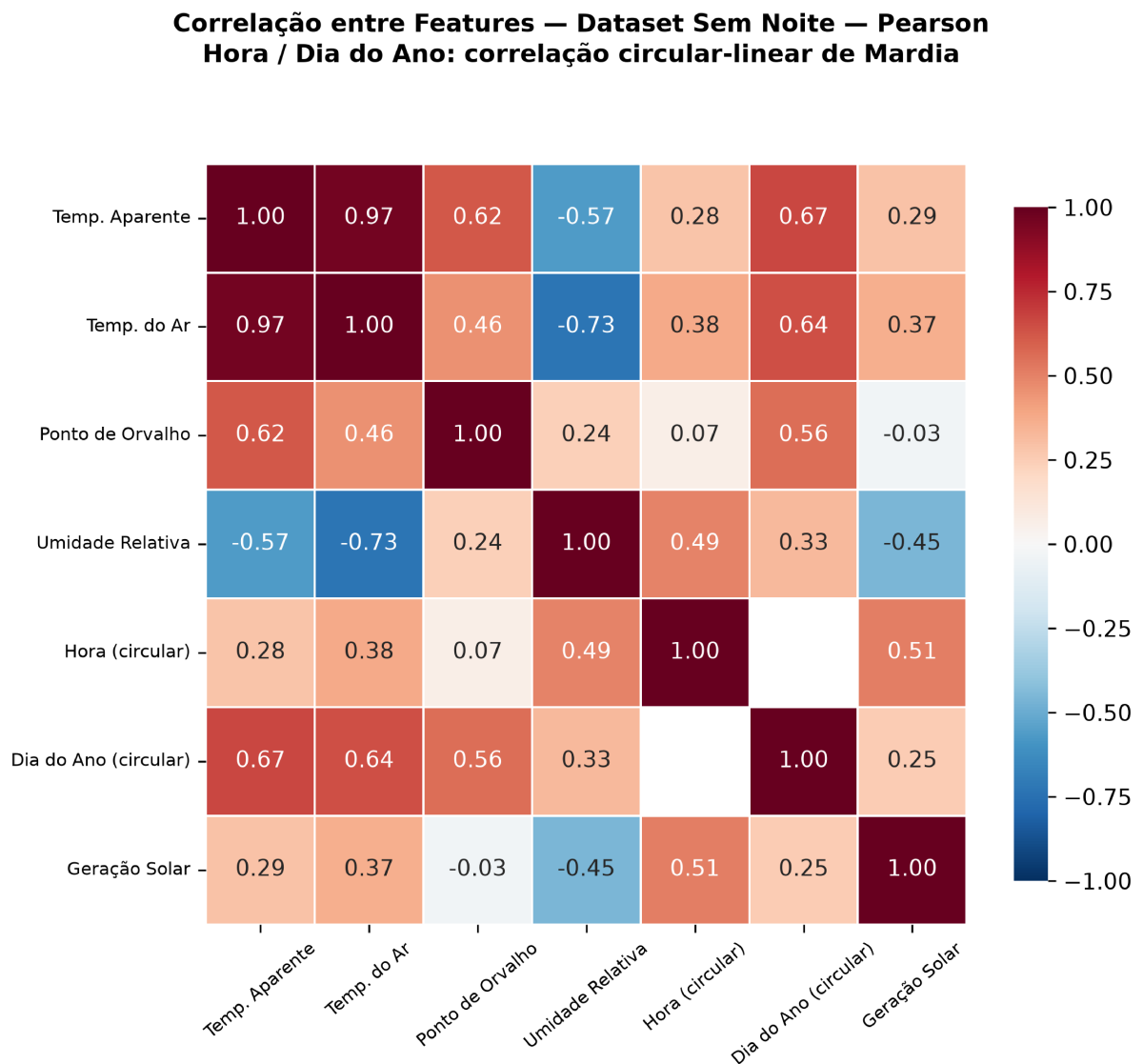


Figura 20 – Matriz de correlação de Pearson entre as *features* do dataset UNISOLAR.

A variável `DewPointTemperature` apresentou a menor correlação com a geração de energia solar entre as variáveis analisadas, tanto na correlação de Pearson quanto na

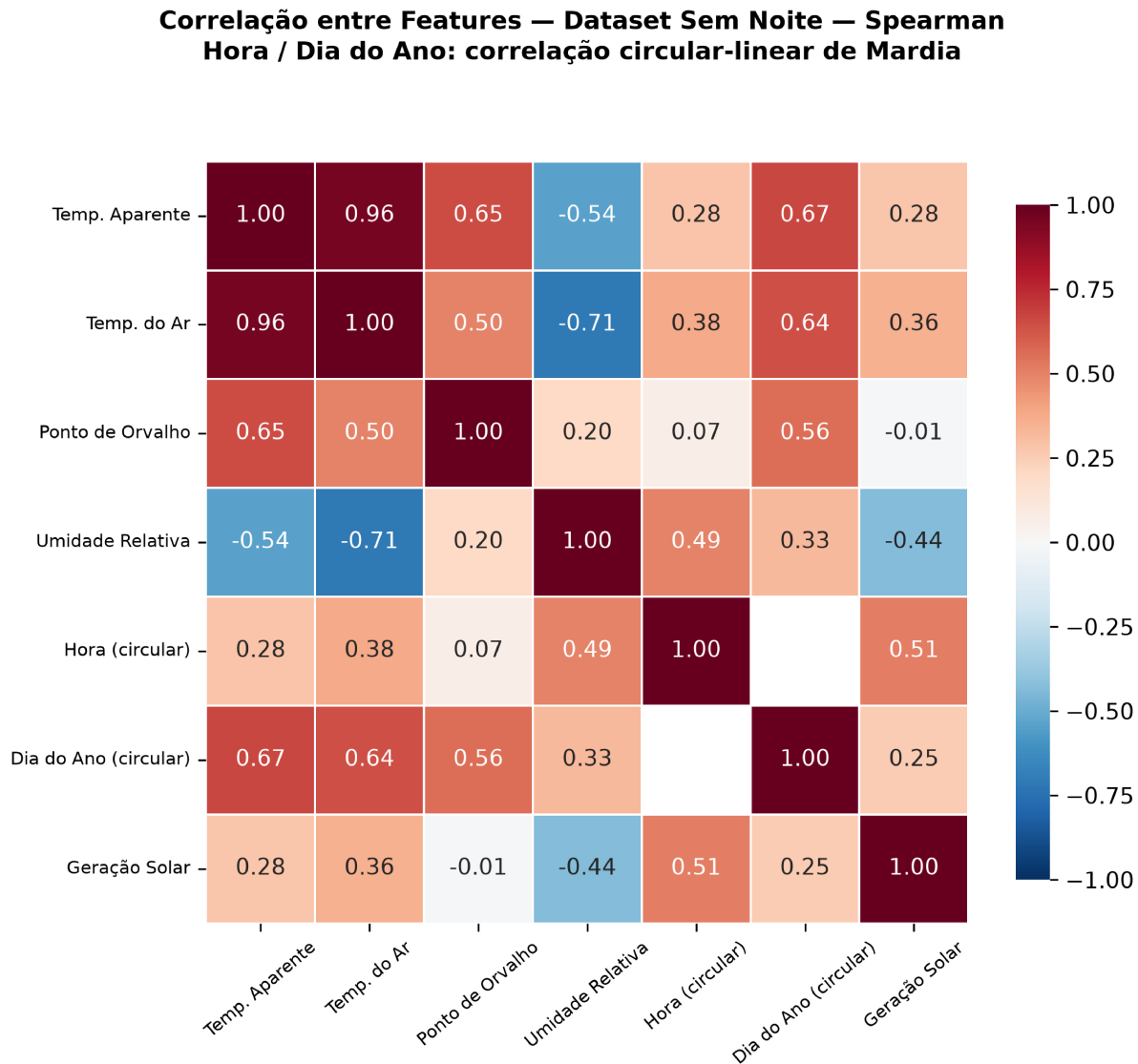


Figura 21 – Matriz de correlação de Spearman entre as *features* do dataset UNISOLAR.

de Spearman. Por esse motivo, foi criada uma variante do conjunto de dados sem essa variável, com o objetivo de avaliar seu impacto no desempenho dos modelos de previsão. Dessa forma, cada uma das quatro variantes do conjunto de dados foi replicada em uma versão com e outra sem a variável `DewPointTemperature`, resultando em um total de oito variantes utilizadas nos experimentos. As variantes são especificadas na Tabela 9.

5.2.2 Avaliação do modelo nas variantes do conjunto de dados

Após a geração das oito variantes, cada uma delas foi dividida cronologicamente em quatro subconjuntos: treinamento (31 dias), validação (15 dias), teste (15 dias) e simulação, composta pelos registros remanescentes. Em seguida, foi aplicada a normalização *Min-Max*, selecionada com base nos resultados obtidos no conjunto de dados Sol do Cer-

Tabela 9 – Variantes geradas do conjunto de dados.

Variante	Variáveis cíclicas	DewPointTemperature	Período noturno
V1	Sim	Sim	Incluído
V2	Sim	Não	Incluído
V3	Não	Sim	Incluído
V4	Não	Não	Incluído
V5	Sim	Sim	Removido
V6	Sim	Não	Removido
V7	Não	Sim	Removido
V8	Não	Não	Removido

rado, onde apresentou o melhor desempenho para os modelos baseados em aprendizado de máquina.

Os experimentos foram conduzidos utilizando o modelo LSTM, por ter apresentado o melhor desempenho geral no dataset Sol do Cerrado. Dessa forma, a primeira análise concentrou-se no impacto das diferentes variantes do conjunto de dados sobre o desempenho do modelo de previsão. Como referência, foi utilizado o modelo de persistência e para sua avaliação, foram considerados apenas registros separados por 15 minutos, e métricas calculadas exclusivamente sobre registros diurnos.

O treinamento do modelo LSTM foi realizado utilizando janelas deslizantes sobre a série temporal, nas quais o tamanho da sequência foi definido pelo hiperparâmetro *look-back*. Durante a construção das sequências, janelas contendo descontinuidades temporais foram descartadas, garantindo que cada amostra representasse intervalos consecutivos de 15 minutos.

A otimização dos hiperparâmetros foi conduzida com o *Ray Tune*, em razão de seu desempenho ligeiramente superior ao *Optuna* no dataset Sol do Cerrado. Para cada uma das oito variantes foram realizadas três execuções independentes de otimização, cada uma composta por 50 *trials*, totalizando 150 configurações avaliadas por variante. A melhor configuração foi selecionada com base no MAE_{diurno} obtido no conjunto de validação.

Em cada avaliação foi utilizado o otimizador Adam, a função de perda MSE, parada antecipada com paciência de 10 épocas e um limite máximo de 100 épocas de treinamento. Ao término da otimização, a melhor configuração encontrada foi retreinada e empregada na avaliação final sobre os conjuntos de validação e teste. A Tabela 10 mostra o espaço de busca dos hiperparâmetros.

A Tabela 11 apresenta o desempenho do modelo LSTM para todas as variantes, considerando as métricas MAE_{diurno} , $RMSE_{\text{diurno}}$, R^2_{diurno} e *Skill Score*. De forma geral, observa-se que as variantes sem registros noturnos obtiveram melhor desempenho, evidenciado pela redução dos valores de MAE e RMSE e pelo aumento do R^2 . Esse comportamento pode ser explicado pelo fato de que o período noturno é composto predominantemente por dados de baixa atividade solar.

Tabela 10 – Espaço de busca dos hiperparâmetros do modelo LSTM.

Hiperparâmetro	Espaço de busca
<i>Look-back</i>	10, 30, 60
Tamanho da camada oculta	32, 64, 128
Número de camadas	1, 2
<i>Dropout</i>	0.1, 0.2, 0.3
<i>Batch size</i>	32, 64
Taxa de aprendizado	Log-uniforme $[10^{-4}, 10^{-2}]$

Tabela 11 – Métricas das oito variantes do modelo LSTM no conjunto de teste (kWh).

V	Variante	MAE_{diurno}	$RMSE_{\text{diurno}}$	R^2_{diurno}	<i>Skill Score</i>
Com período noturno					
V1	Cíclicas	9,70	16,44	0,652	0,092
V2	Cíclicas sem DewPoint	9,63	16,57	0,646	0,085
V3	Sem cíclicas	9,99	16,90	0,632	0,067
V4	Sem cíclicas e sem DewPoint	9,89	16,89	0,633	0,067
Sem período noturno (Regra 6)					
V5	Cíclicas	8,09	14,05	0,711	0,231
V6	Cíclicas sem DewPoint	8,76	14,31	0,701	0,218
V7	Sem cíclicas	8,00	14,15	0,707	0,226
V8	Sem cíclicas e sem DewPoint	8,14	14,07	0,710	0,231

temente por valores nulos de geração fotovoltaica, contribuindo com pouca informação para o aprendizado da dinâmica de geração durante o dia.

Além disso, as variantes que mantiveram a variável **DewPointTemperature** apresentaram desempenho superior às variantes em que essa variável foi removida, indicando que, mesmo apresentando baixa correlação, sua presença contribui para o processo de aprendizagem quando combinada com as demais variáveis de entrada. Por outro lado, a inclusão das variáveis temporais cíclicas não resultou em ganhos consistentes de desempenho. Por fim, todas as variantes do modelo LSTM apresentaram *Skill Score* positivo, indicando desempenho superior ao modelo de persistência utilizado como referência.

As duas variantes que apresentaram o melhor desempenho geral foram selecionadas para a continuidade dos experimentos — V5 e V7. Destaca-se, entretanto, que a comparação entre V5 e V8 não evidencia diferença estatisticamente significativa nas métricas R^2 e *Skill Score*, considerando a incerteza associada aos resultados. Ambas as variantes selecionadas utilizam exclusivamente os registros do período diurno e mantêm a variável **DewPointTemperature**, diferenciando-se pela utilização de atributos cíclicos. Os modelos treinados com essas variantes foram, então, utilizados na etapa seguinte da pesquisa, dedicada ao monitoramento da degradação e à avaliação de estratégias de aprendizado incremental.

5.2.3 Monitoramento de degradação e aprendizado incremental

Após a seleção dos dois modelos de previsão com melhor desempenho (V5 e V7), foram avaliadas quatro configurações adaptativas obtidas pela combinação dos dois detectores de degradação (CUSUM e ADWIN) com as duas estratégias de aprendizado incremental (*Fine-tuning* e EWC). Cada configuração foi simulada durante 685 dias, permitindo que o módulo de monitoramento identificasse degradações no desempenho do modelo e acionasse atualizações incrementais sempre que necessário.

As Figuras 22 e 23 apresentam o comportamento do monitoramento obtido pela variante V7 pelos detectores ADWIN e CUSUM, respectivamente. Em ambos os casos foi empregada a estratégia de aprendizado incremental EWC, por apresentar uma média maior de redução no erro após a atualização incremental, conforme mostrado na Tabela 12. Os resultados correspondentes às demais configurações avaliadas, bem como a variante V5, apresentam comportamento semelhante aos discutidos nessa seção. Por esse motivo são apresentados nas Figuras 32, 33, 34, 35, 36, 37 do Apêndice B.

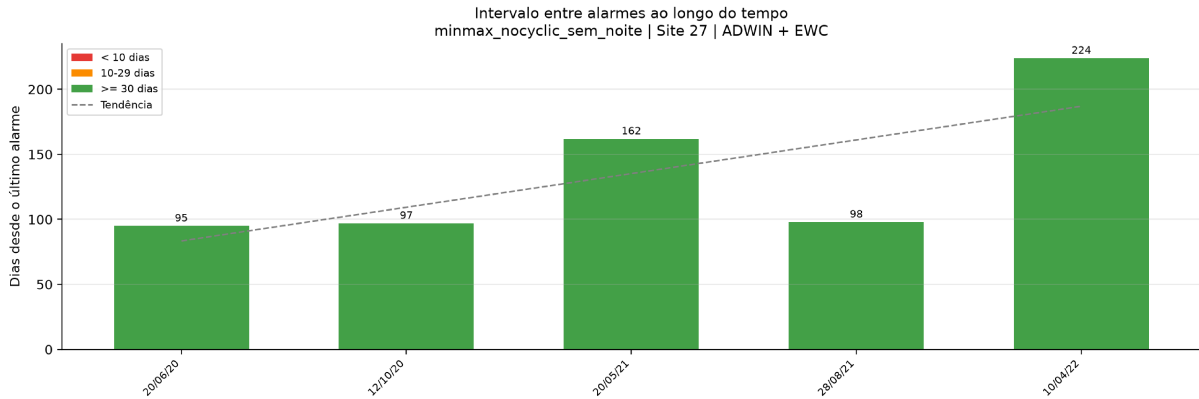


Figura 22 – Alarmes de degradação detectados pelo ADWIN durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental EWC.

A Figura 22 mostra que o ADWIN gerou cinco alarmes de degradação ao longo dos 685 dias de simulação. Além disso, o intervalo entre alarmes consecutivos aumentou gradualmente ao longo do tempo, a tendência de crescimento é representada pela linha tracejada na Figura 22. Isso indica que o modelo de previsão passou a necessitar de um número cada vez menor de atualizações à medida que o aprendizado incremental o adaptava às condições de operação. Esse comportamento sugere que a abordagem proposta converge para um modelo de previsão mais estável.

A Figura 23 apresenta os resultados obtidos com o detector CUSUM. Durante o mesmo período de simulação foram realizadas 25 atualizações incrementais, número cinco vezes superior ao observado com o ADWIN. Além disso, diversos alarmes ocorreram em intervalos curtos de tempo, indicando que o CUSUM foi significativamente mais sensível às variações no erro de previsão.

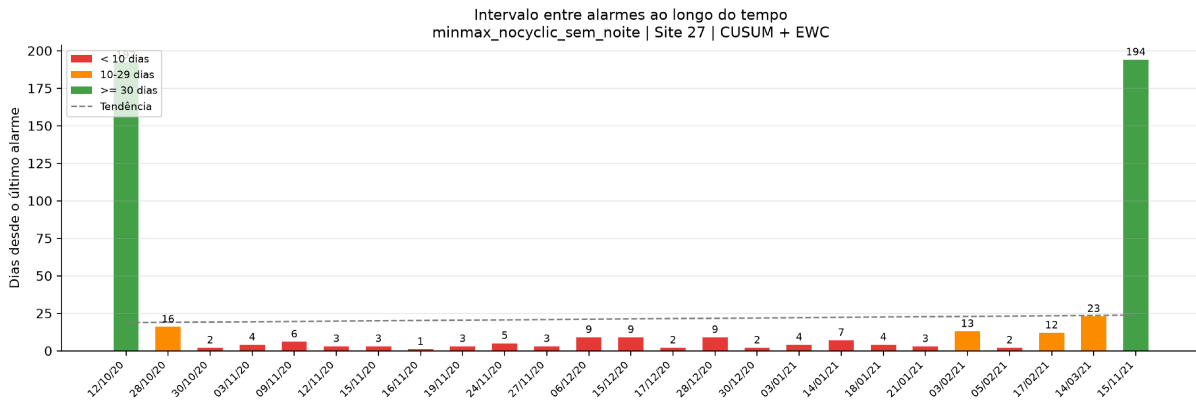


Figura 23 – Alarmes de degradação detectados pelo CUSUM durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental EWC.

A Tabela 12 apresenta a redução do MAE obtida após cada atualização incremental acionada pelo ADWIN para a variante V5 e V7. Embora ambas as estratégias de aprendizado incremental tenham reduzido o erro de previsão na maior parte das atualizações, o EWC apresentou reduções superiores em quatro das cinco atualizações realizadas em V7, além de apresentar uma maior redução na média geral. Em relação a variante V5, o EWC apresentou reduções superiores somente em duas das cinco atualizações realizadas, mas ainda assim se prova melhor na média geral de redução do erro. Desse modo, os resultados demonstram que o EWC constitui um mecanismo de adaptação relativamente mais eficaz do que o *fine-tuning* convencional.

Tabela 12 – Redução do MAE em kWh após cada atualização incremental acionada pelo ADWIN para as variantes V5 e V7.

Atualização	V5		V7	
	Fine-tuning	EWC	Fine-tuning	EWC
1	12.03	11.97	1.96	2.01
2	7.31	7.65	0.01	-0.07
3	2.05	1.73	0.37	0.47
4	3.06	2.79	0.33	0.14
5	6.54	7.56	0.38	0.58
Média	6.20	6.34	0.61	0.63

Por fim, a eficácia de cada configuração foi avaliada por meio do MAE médio calculado nos últimos dez dias da simulação. Os resultados, apresentados na Tabela 13, estão organizados da melhor para a pior configuração. Observa-se um padrão consistente: as configurações baseadas no detector ADWIN apresentaram desempenho superior às baseadas no CUSUM; entre os modelos de previsão avaliados, a variante V7 obteve menores erros de previsão do que a V5; e, de maneira geral, a estratégia EWC superou o *fine-tuning*.

Tabela 13 – Desempenho final dos modelos após o monitoramento de degradação e o aprendizado incremental.

Variante	Configuração	MAE médio (últimos 10 dias)
V7	ADWIN + EWC	2.01 kWh
V7	ADWIN + fine-tuning	2.07 kWh
V5	ADWIN + EWC	2.99 kWh
V5	ADWIN + fine-tuning	3.02 kWh
V7	CUSUM + EWC	3.10 kWh
V7	CUSUM + fine-tuning	3.45 kWh
V5	CUSUM + EWC	3.81 kWh
V5	CUSUM + fine-tuning	6.97 kWh

Em síntese, a configuração de melhor desempenho utilizou a variante sem atributos cíclicos e sem registros do período noturno, manteve a variável `DewPointTemperature` e combinou o detector ADWIN com a estratégia EWC para o aprendizado incremental. Após isso, essa melhor configuração foi testada em um novo sítio de geração de energia fotovoltaica do mesmo campus, utilizando o mesmo período de tempo, os resultados são detalhados a seguir.

5.2.4 Avaliação em um novo sítio fotovoltaico

Para avaliar a capacidade de generalização do modelo proposto e analisar seu comportamento durante um período mais longo de monitoramento de degradação, o modelo foi aplicado a um segundo sítio fotovoltaico. O sítio selecionado foi o `siteKey` 25, por apresentar características semelhantes às do `siteKey` 27, utilizado no desenvolvimento do modelo. A Tabela 14 apresenta uma comparação entre os dois sítios.

Tabela 14 – Comparação entre os sítios utilizados nos experimentos.

Característica	Site 27	Site 25
Potência instalada (kWp)	539,80	384,12
Registros brutos	79.798	62.901
Após limpeza (Regras 1–5)	71.592	61.352
Após remoção do período noturno	35.786	30.571

Os experimentos foram realizados utilizando o modelo da variante V7, após a etapa de monitoramento de degradação e aprendizado incremental empregando o ADWIN e o EWC. Neste novo sítio, foi avaliada apenas essa configuração, por ter apresentado o melhor desempenho nos experimentos anteriores e com o objetivo de verificar se a melhoria obtida também se mantém em um ambiente distinto ao longo do tempo. A Figura 24 apresenta os alarmes gerados durante o monitoramento dos dados do `siteKey` 25.

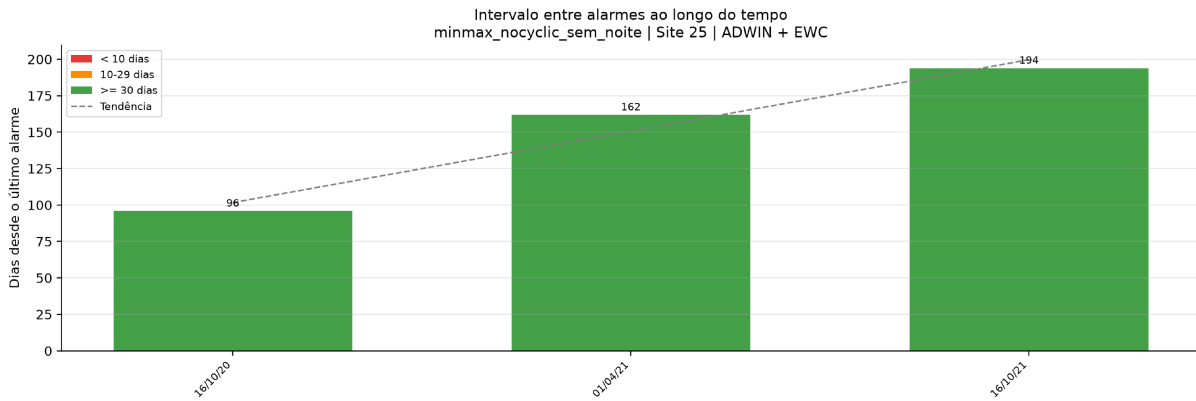


Figura 24 – Alarmes de degradação detectados pelo ADWIN utilizando a estratégia de aprendizado incremental EWC no sítio 25.

Na Figura 24, observa-se que o monitoramento de degradação gerou apenas três alarmes ao longo de 636 dias de simulação, número inferior ao observado no `siteKey` 27. Além disso, verifica-se um aumento progressivo no intervalo entre os alarmes, indicando que as atualizações incrementais tendem a preservar o desempenho do modelo por períodos cada vez maiores. Esse comportamento é evidenciado pela linha de tendência apresentada na Figura 24, que representa o aumento do espaçamento entre as atualizações ao longo do tempo. Desse modo, os resultados indicam que o monitoramento de degradação aliado ao aprendizado incremental foi efetivo em melhorar o desempenho do modelo.

Por fim, a Tabela 15 apresenta a redução do MAE obtida após cada atualização incremental. Observa-se que as duas primeiras atualizações reduziram o erro de previsão em 1,3 kWh e 0,3 kWh, respectivamente. Na terceira atualização, houve um aumento de 0,2 kWh no MAE em relação a segunda atualização, indicando que, nesse momento, a atualização incremental não proporcionou melhoria no desempenho. Ainda assim, considerando as três atualizações realizadas, foi obtida uma redução média de 0,47 kWh, reforçando que a combinação entre o monitoramento de degradação e o aprendizado incremental foi capaz de preservar o desempenho do modelo quando aplicada a um sítio fotovoltaico diferente daquele utilizado em seu desenvolvimento.

Tabela 15 – Redução do MAE após cada atualização incremental.

Atualização	EWC
1	1.3
2	0.3
3	-0.2
Média	0.47

5.3 Considerações finais

Este capítulo apresentou os experimentos e as análises dos resultados do método proposto neste trabalho. A Seção 5.1 avaliou os resultados do dataset Sol do Cerrado, comparando diferentes modelos de previsão, configurações de agregação temporal, técnicas de escalonamento dos dados e métodos de otimização de hiperparâmetros, permitindo identificar a configuração mais adequada para o problema estudado.

Por sua vez, a Seção 5.2 expôs os experimentos realizados com o conjunto de dados UNISOLAR. Nessa etapa, foram avaliadas diferentes variantes do conjunto de dados e, posteriormente, analisadas as estratégias de monitoramento de degradação e aprendizado incremental. Por fim, a configuração selecionada foi aplicada a um segundo sítio fotovoltaico, permitindo avaliar sua capacidade de generalização.

O próximo capítulo apresenta as conclusões deste trabalho, destacando as principais contribuições, limitações e perspectivas para pesquisas futuras.

Conclusão

O objetivo desta dissertação foi desenvolver um modelo de aprendizado de máquina capaz de prever a geração de energia em um campo fotovoltaico, propondo uma arquitetura composta por um monitoramento de degradação em tempo real e uma estratégia de aprendizado incremental, de forma a preservar o desempenho preditivo do modelo ao longo do tempo.

Para atingir esse objetivo, foram conduzidos experimentos utilizando dois datasets distintos. No primeiro dataset, foram investigados os efeitos de diferentes estratégias de agregação, granularidades temporais, métodos de escalonamento dos dados, modelos de previsão e técnicas de otimização de hiperparâmetros. No segundo dataset, foram avaliados o impacto da utilização de atributos cíclicos, da remoção da variável de menor correlação com o problema, bem como a combinação de dois métodos de monitoramento de degradação — *Cumulative Sum Control Chart* (CUSUM) e *Adaptive Windowing* (ADWIN) — com duas estratégias de aprendizado incremental — *fine-tuning* e *Elastic Weight Consolidation* (EWC).

Durante os experimentos, observou-se que as decisões de pré-processamento exerceram influência significativa sobre o desempenho dos modelos. De modo geral, a agregação temporal por média apresentou os melhores resultados, a normalização *Min-Max* foi responsável pelos melhores modelos de aprendizado de máquina e a estratégia de otimização de hiperparâmetros *Ray Tune* apresentou um ganho de desempenho em relação ao *Optuna*. Na comparação entre os modelos de previsão, os resultados mostraram que, para granularidades temporais de 1 e 10 minutos, o modelo de persistência apresentou os menores valores de MAE. Entretanto, a partir da granularidade de 15 minutos, o modelo LSTM passou a apresentar os melhores resultados, evidenciando sua capacidade de aprender padrões temporais mais complexos. A análise do *Skill Score* mostrou que quando considerado o RMSE, o modelo LSTM só perde para a persistência na granularidade de 10 minutos. Dessa forma, conclui-se que, para previsões de curtíssimo prazo, a elevada complexidade dos modelos de aprendizado de máquina pode não ser justificada, enquanto, para horizontes de previsão mais longos, esses modelos tornam-se significativamente mais

eficientes por conseguirem aprender padrões presentes na série histórica.

No segundo conjunto de dados, foram avaliadas diferentes estratégias de engenharia de atributos. Os resultados mostraram que a inclusão de atributos cíclicos, representando o horário do dia e o dia do ano, não esteve presente na configuração de melhor desempenho. Também foi avaliado o impacto da remoção da variável com menor coeficiente de correlação com a variável alvo. Embora essa variável apresentasse baixa correlação individual, sua remoção resultou em pior desempenho do modelo, indicando que redes neurais artificiais são capazes de explorar relações não lineares entre as variáveis que não são capturadas pelas análises de correlação. Além disso, os modelos treinados utilizando apenas amostras do período diurno apresentaram desempenho superior aos modelos treinados com o conjunto de dados completo. Esse comportamento pode ser explicado pelo fato de que, durante o período noturno, a geração fotovoltaica permanece praticamente constante e igual a zero, não contribuindo para o aprendizado do comportamento da geração durante o período em que efetivamente ocorre produção de energia.

Em relação ao modelo de degradação e a aprendizagem incremental, a combinação entre ADWIN e EWC apresentou o melhor desempenho. O ADWIN demonstrou maior estabilidade, acionando o módulo de aprendizado incremental aproximadamente cinco vezes menos que o CUSUM. Entre as estratégias de aprendizado incremental, o EWC apresentou desempenho ligeiramente superior ao *fine-tuning*. A arquitetura incremental contribuiu para preservar o desempenho do modelo ao longo do tempo, evidenciado pelo aumento gradual do intervalo entre ativações consecutivas do monitoramento de degradação, indicando que o aprendizado incremental foi capaz de incorporar novos padrões aos modelos, retardando sua obsolescência.

Por fim, a arquitetura proposta foi avaliada em um segundo sítio fotovoltaico pertencente ao mesmo campus do conjunto de dados utilizado no desenvolvimento do modelo. Os resultados obtidos mostraram comportamento semelhante ao observado nos experimentos anteriores, com o detector ADWIN acionando o aprendizado incremental em menor frequência e mantendo a tendência de aumento do intervalo entre ativações. Além disso, na maior parte das atualizações realizadas, o aprendizado incremental promoveu redução do MAE do modelo, demonstrando que a arquitetura proposta é capaz de adaptar-se às mudanças presentes em novos dados e preservar o desempenho preditivo ao longo do tempo.

Os resultados obtidos ao longo desta dissertação permitem afirmar que as hipóteses propostas pelo trabalho foram corroboradas. Os modelos de aprendizado de máquina avaliados, em especial a LSTM, demonstraram ser capazes de prever a geração de energia fotovoltaica a partir de variáveis meteorológicas e operacionais. Os experimentos evidenciaram que esses modelos conseguem aprender padrões presentes na série temporal, superando os modelos tradicionais.

Além disso, a arquitetura proposta, composta por um monitoramento de degradação

associado a estratégias de aprendizado incremental, foi capaz de identificar degradações no desempenho do modelo e promover sua adaptação ao longo do tempo. Os resultados mostraram que ambas as estratégias de aprendizado incremental contribuíram para preservar o desempenho preditivo, enquanto o detector ADWIN apresentou maior estabilidade ao reduzir significativamente a frequência de atualizações quando comparado ao CUSUM.

Por fim, os experimentos demonstraram que um modelo treinado inicialmente com uma quantidade reduzida de dados pode aprimorar progressivamente seu desempenho por meio do aprendizado incremental. O aumento gradual do intervalo entre ativações consecutivas do módulo de monitoramento indica que o modelo incorporou novos conhecimentos ao longo da operação, tornando-se progressivamente mais adaptado às características do ambiente e evidenciando o potencial da arquitetura proposta para aplicações em campo.

6.1 Contribuições

As principais contribuições deste trabalho são:

- Realização de uma investigação abrangente sobre estratégias de construção de modelos de previsão de geração de energia fotovoltaica, envolvendo a avaliação de diferentes técnicas de pré-processamento, engenharia de atributos, agregação e granularidade temporal, métodos de escalonamento, modelos preditivos e estratégias de otimização de hiperparâmetros, identificando as configurações mais adequadas para o problema.
- Proposição e avaliação de uma arquitetura adaptativa para previsão de geração de energia fotovoltaica, integrando modelos de aprendizado de máquina, monitoramento de degradação e aprendizado incremental. A arquitetura foi concebida para permitir sua implantação com uma quantidade reduzida de dados e sua adaptação contínua ao longo da operação. Além disso, foi realizada uma avaliação comparativa entre os monitoramentos de degradação CUSUM e ADWIN e as estratégias de aprendizado incremental *fine-tuning* e EWC, identificando a combinação ADWIN + EWC como a configuração de melhor desempenho e demonstrando que o aprendizado incremental preserva e, em diversos momentos, aprimora o desempenho do modelo ao longo do tempo.
- Validação da arquitetura proposta em um segundo sítio fotovoltaico, demonstrando sua capacidade de adaptação e seu potencial de aplicação em diferentes sistemas de geração de energia fotovoltaica.

6.2 Trabalhos futuros

Os seguintes trabalhos futuros são propostos com o objetivo de explorar novas questões e ampliar as investigações realizadas nesta dissertação:

- ❑ Investigar novas estratégias de engenharia de atributos capazes de representar informações temporais de forma mais eficiente, explorando alternativas aos atributos cíclicos utilizados neste trabalho.
- ❑ Avaliar arquiteturas híbridas de aprendizado profundo, como CNN-LSTM e modelos baseados em *Transformers*, analisando o compromisso entre custo computacional e melhora no desempenho preditivo.
- ❑ Avaliar diferentes configurações do monitoramento de degradação CUSUM, investigando o impacto de seus parâmetros na sensibilidade à detecção de degradação e na frequência de atualizações do modelo.
- ❑ Investigar a utilização de outros monitoramentos de degradação, como DDM, EDDM e ECDD, comparando sua eficiência com os métodos CUSUM e ADWIN.
- ❑ Investigar outras estratégias de aprendizado incremental, comparando seu desempenho com o *fine-tuning* e o EWC.
- ❑ Desenvolver mecanismos de transferência de conhecimento entre modelos treinados em diferentes sítios fotovoltaicos, permitindo o compartilhamento de informações e reduzindo a necessidade de re-treinamento completo.
- ❑ Validar a arquitetura proposta em usinas fotovoltaicas localizadas em diferentes regiões geográficas e sob diferentes condições climáticas, avaliando sua capacidade de generalização.
- ❑ Estender a arquitetura para múltiplos horizontes de previsão, permitindo estimar a geração de energia em instantes mais distantes no futuro.
- ❑ Implementar e avaliar a arquitetura em um ambiente de operação em tempo real, considerando restrições computacionais, latência e disponibilidade dos dados.

Referências

- ABRAHAM, W. C.; ROBINS, A. Memory retention–the synaptic stability versus plasticity dilemma. **Trends in Neurosciences**, v. 28, p. 73–78, 2005. ISSN 0166-2236. Disponível em: <<https://doi.org/10.1016/j.tins.2004.12.003>>.
- AGRAHARI, S.; SINGH, A. K. Concept drift detection in data stream mining : A literature review. **Journal of King Saud University - Computer and Information Sciences**, v. 34, n. 10, Part B, p. 9523–9540, 2022. ISSN 1319-1578. Disponível em: <<https://doi.org/10.1016/j.jksuci.2021.11.006>>.
- AL-DAHIDI, S. et al. Forecasting solar photovoltaic power production: A comprehensive review and innovative data-driven modeling framework. **Energies**, v. 17, n. 16, p. 4145, 2024. Disponível em: <<https://doi.org/10.3390/en17164145>>.
- ARAUJO, D. N.; SILVA, F. M.; SOUZA, R. A.; OLIVEIRA, L. P. Efeitos da acumulação de sujeira sobre o desempenho de módulos fotovoltaicos: Experiências brasileiras e internacionais. **Revista Tecnologia (Fortaleza)**, v. 30, n. 2, p. 01–10, 2019. Acesso em: 28 out. 2025. Disponível em: <<https://doi.org/10.5020/23180730.2019.9414>>.
- AZEEM, A.; ISMAIL, I.; MOHANI, S. S.; DANYARO, K. U.; HUSSAIN, U.; SHABBIR, S.; JUSOH, R. Z. B. Mitigating concept drift challenges in evolving smart grids: An adaptive ensemble lstm for enhanced load forecasting. **Energy Reports**, Elsevier, v. 13, p. 1369–1383, 2025. Disponível em: <<https://doi.org/10.1016/j.egyr.2024.12.078>>.
- BANSAL, A.; BALAJI, K.; LALANI, Z. **Temporal Encoding Strategies for Energy Time Series Prediction**. 2025. Disponível em: <<https://arxiv.org/abs/2503.15456>>.
- BARBETTA, P. A.; REIS, M. M.; BORNIA, A. C. **Estatística para Cursos de Engenharia, Computação e Ciência de Dados**. 4. ed. [S.l.]: LTC, 2019.
- BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. **IEEE Transactions on Neural Networks**, v. 5, n. 2, p. 157–166, 1994. Disponível em: <<https://doi.org/10.1109/72.279181>>.
- BIFET, A.; GAVALDA, R. Learning from time-changing data with adaptive windowing. In: **Proceedings of the 2007 SIAM International Conference on Data Mining (SDM)**. Society for Industrial and Applied Mathematics (SIAM), 2007. p. 443–448. Disponível em: <<https://doi.org/10.1137/1.9781611972771.42>>.

BlueSol Educacional. **Os sistemas de energia solar fotovoltaica**. Florianópolis, SC: [s.n.], 2016. Acesso em: 28 out. 2025. Disponível em: <<https://docente.ifsc.edu.br/klunger.beck/energia/extra/livros/Livro-Digital-de-Introdu%C3%A7%C3%A3o-aos-Sistemas-Solares-novo.pdf>>.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. Revised. San Francisco: Holden-Day, 1976.

CARVALHO, A. C. P. L. F. d.; MENEZES, A. G.; BONIDIA, R. P. **Ciência de Dados: Fundamentos e Aplicações**. Rio de Janeiro: LTC, 2024. ISBN 9788521638759.

CATARINO, V. H. O. **Predição da Geração de Energia Elétrica Baseado na Irradiância Solar**. Dissertação (Dissertação de Mestrado) — Universidade Federal de Ouro Preto, Ouro Preto, Minas Gerais, Brasil, 2024. Orientador: Luciano Perdigão Cota.

CNN Brasil. **Fontes renováveis serão responsáveis por 37% da geração de eletricidade no mundo até 2026**. 2024. Disponível em: <<https://www.cnnbrasil.com.br/economia/macroeconomia/fontes-renovaveis-serao-responsaveis-por-37-da-geracao-de-eletricidade-no-mundo-ate-2026-diz-aie/>>.

_____. **Demanda global por eletricidade crescerá 4% até 2027, diz IEA**. 2025. Atualizado em 14/02/25 às 13:53. Disponível em: <<https://www.cnnbrasil.com.br/economia/macroeconomia/demanda-global-por-eletricidade-crescera-4-ate-2027-diz-iea/>>.

Deeplearning Book. **O neurônio biológico e matemático**. 2024. Disponível em: <<https://www.deeplearningbook.com.br/o-neuronio-biologico-e-matematico/>>.

GAMA, J. **Knowledge Discovery from Data Streams**. Boca Raton, FL: Chapman Hall/CRC, 2010. (Data Mining and Knowledge Discovery Series). ISBN 9781439826119. Disponível em: <<https://doi.org/10.1201/EBK1439826119>>.

GOLDEMBERG, J.; LUCON, O. Energias renováveis: um futuro sustentável. **Revista USP**, n. 72, p. 6–15, fev. 2007. Disponível em: <<https://doi.org/10.11606/issn.2316-9036.v0i72p6-15>>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York: Springer, 2009. ISBN 9780387848570. Disponível em: <<https://doi.org/10.1007/978-0-387-84858-7>>.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 11 1997. ISSN 0899-7667. Disponível em: <<https://doi.org/10.1162/neco.1997.9.8.1735>>.

Iberdrola. **Paris Agreement: what it is and why it is important in the fight against climate change**. 2024. Acesso em: 3 set. 2025. Disponível em: <<https://www.iberdrola.com/sustainability/energy-transition/paris-agreement>>.

- ICLG. **Renewable Energy Laws and Regulations Brazil 2024**. 2024. Acesso em: 28 out. 2025. Disponível em: <<https://www.lexology.com/library/detail.aspx?g=06cc458f-1a52-4595-bb53-31a2f9059149>>.
- Intelbras. **Por que investir em energia solar**. 2022. Blog Intelbras. Publicado em 19 set. 2022; acesso em 3 set. 2025. Disponível em: <<https://blog.intelbras.com.br/por-que-investir-em-energia-solar/>>.
- International Energy Agency. **Global Energy Review 2025**. Paris, 2025. Lead authors: Alex Martinos and Thomas Spencer. Disponível em: <<https://www.iea.org>>.
- IRENA. **Record Growth in Renewables, but Progress Needs to be Equitable**. 2024. Disponível em: <<https://www.irena.org/News/pressreleases/2024/Mar/Record-Growth-in-Renewables-but-Progress-Needs-to-be-Equitable-PT>>.
- JAILANI, N. L. M.; DHANASEGARAN, J. K.; ALKAWSI, G.; ALKAHTANI, A. A.; PHING, C. C.; CAPRETZ, L. F.; AL-SHETWI, A. Q.; TIONG, S. K. Investigating the power of lstm-based models in solar energy forecasting. **Processes**, v. 11, n. 5, p. 1382, 2023. Disponível em: <<https://doi.org/10.3390/pr11051382>>.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning: with Applications in R**. New York: Springer, 2013. (Springer Texts in Statistics). ISBN 9781461471370. Disponível em: <<https://doi.org/10.1007/978-1-4614-7138-7>>.
- K., N.; KUMAR, G. S. Cusum based concept drift detector for data stream clustering. In: **Proceedings of the 2020 4th International Conference on Big Data and Internet of Things**. New York, NY, USA: Association for Computing Machinery, 2020. (BDIOT '20), p. 90–95. ISBN 9781450375504. Disponível em: <<https://doi.org/10.1145/3421537.3421548>>.
- KANDANANOND, K.; CHOTEWIWAT, K.; WONGSATHAN, R. Solar photovoltaic forecasting of power output using lstm networks. **Atmosphere**, MDPI, v. 12, n. 2, p. 124, 2021. Accessed: 29 Oct 2025. Disponível em: <<https://doi.org/10.3390/atmos12010124>>.
- KhanAcademy. **O neurônio, estrutura e função**. 2020. Disponível em: <<https://pt.khanacademy.org/science/6-ano/vida-e-evolucao-os-sistemas-do-corpo-humano/os-neuronios/a/o-neuronio-estrutura-e-funcao>>.
- KIRKPATRICK, J.; PASCANU, R.; RABINOWITZ, N.; VENESS, J.; DESJARDINS, G.; RUSU, A. A.; MILAN, K.; QUAN, J.; RAMALHO, T.; GRABSKA-BARWINSKA, A.; HASSABIS, D.; CLOPATH, C.; KUMARAN, D.; HADSELL, R. Overcoming catastrophic forgetting in neural networks. **Proceedings of the National Academy of Sciences**, v. 114, n. 13, p. 3521–3526, 2017. Disponível em: <<https://doi.org/10.1073/pnas.1611835114>>.
- LIU, X.; CHANG, X. Elastic weight consolidation done right for continual learning. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2026. p. 3930–3940.
- LUO, Y.; YIN, L.; BAI, W.; MAO, K. An appraisal of incremental learning methods. **Entropy**, v. 22, n. 11, 2020. ISSN 1099-4300. Disponível em: <<https://www.mdpi.com/1099-4300/22/11/1190>>.

MARDIA, K. V. Linear-circular correlation coefficients and rhythmometry. **Biometrika**, v. 63, n. 2, p. 403–405, 1976. Disponível em: <<https://doi.org/10.2307/2335637>>.

MCCLELLAND, J.; MCNAUGHTON, B. L.; O'REILLY, R. C. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. **Psychological Review**, v. 102, p. 419–457, 1995. Disponível em: <<https://doi.org/10.1037/0033-295X.102.3.419>>.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, 1943. Disponível em: <<https://doi.org/10.1007/BF02478259>>.

Ministério do Meio Ambiente. **Acordo de Paris**. s.d. Acesso em: 28 set. 2025. Disponível em: <<https://antigo.mma.gov.br/clima/convencao-das-nacoes-unidas/acordo-de-paris.html>>.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**. 7. ed. [S.l.]: LTC, 2021.

MORETTIN, P. A.; TOLOI, C. M. de C. **Análise de Séries Temporais**. 3. ed. São Paulo: Blucher, 2018. ISBN 9788521214472.

OLAH, C. **Understanding LSTM Networks**. 2015. <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em: 21 jun. 2026.

ONS. **O Sistema em Números**. 2026. Acesso em: 11 jul. 2026. Disponível em: <<https://www.ons.org.br/paginas/sobre-o-sin/o-sistema-em-numeros>>.

Our World in Data. **Electricity Mix**. 2026. Acesso em: 11 jul. 2026. Disponível em: <<https://ourworldindata.org/electricity-mix>>.

PAGE, E. S. Continuous inspection schemes. **Biometrika**, v. 41, n. 1/2, p. 100–115, 1954. Disponível em: <<https://doi.org/10.1093/biomet/41.1-2.100>>.

PARISI, G. I.; KEMKER, R.; PART, J. L.; KANAN, C.; WERMTER, S. Continual lifelong learning with neural networks: A review. **Neural Networks**, v. 113, p. 54–71, 2019. Disponível em: <<https://doi.org/10.1016/j.neunet.2019.01.012>>.

PARK, H.-S.; CHU, H.-C.; SUNG, M.-K.; KIM, C.; LEE, J.; KIM, D.-W.; LEE, J. Survey on replay-based continual learning and empirical validation on feasibility in diverse edge devices using a representative method. **Mathematics**, v. 13, n. 14, 2025. ISSN 2227-7390. Disponível em: <<https://doi.org/10.3390/math13142257>>.

Petrobras. **Guia completo sobre Transição Energética**. 2024. Nossa Energia – site institucional da Petrobras. Atualizado em 25/08/2025; postado em 14/06/2024. Disponível em: <<https://nossaenergia.petrobras.com.br/w/transicao-energetica/tudo-sobre-transicao-energetica-o-que-e-qual-a-importancia-principais-beneficios-e-mais>>.

PUAH, B. K.; CHONG, L. W.; WONG, Y. W.; BEGAM, K.; KHAN, N.; JUMAN, M. A.; RAJKUMAR, R. K. A regression unsupervised incremental learning algorithm for solar irradiance prediction. **Renewable Energy**, v. 164, p. 908–925, 2021. ISSN 0960-1481. Disponível em: <<https://doi.org/10.1016/j.renene.2020.09.080>>.

RAMPINELLI, G. A.; KRENZINGER, A.; ROMERO, F. C. Descrição e análise de inversores utilizados em sistemas fotovoltaicos. **Revista Ciências Exatas e Naturais**, v. 15, n. 1, p. 31–42, 2013.

Redação Homework. **Energia solar reduz custos e aumenta lucro de pequenas empresas**. 2025. Disponível em: <<https://www.terra.com.br/economia/meu-negocio/energia-solar-reduz-custos-e-aumenta-lucro-de-pequenas-empresas,c4dd0540ae28b50dd8910b5690fcac44cd089t67.html>>.

RUNKLER, T. A. **Data Analytics: Models and Algorithms for Intelligent Data Analysis**. 1. ed. Wiesbaden: Vieweg+Teubner Verlag, 2012. ISBN 978-3-8348-2588-9. Disponível em: <<https://doi.org/10.1007/978-3-8348-2589-6>>.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial: Uma abordagem moderna**. 3. ed. Rio de Janeiro: Elsevier: Campus, 2013. Tradução da 3ª edição norte-americana. ISBN 978-85-352-6976-6.

RUSU, A. A.; RABINOWITZ, N. C.; DESJARDINS, G.; SOYER, H.; KIRKPATRICK, J.; KAVUKCUOGLU, K.; PASCANU, R.; HADSELL, R. **Progressive Neural Networks**. 2022. Disponível em: <<https://arxiv.org/abs/1606.04671>>.

SHEWHART, W. A. **Economic Control of Quality of Manufactured Product**. New York: D. Van Nostrand Company, 1931.

SIEGEL, S.; JR., N. J. C. **Estatística Não Paramétrica para Ciências do Comportamento**. 2. ed. [S.l.]: Artmed, 2006.

SOARES, P. H. V. Introdução à digitalização de subestações e energias renováveis. **Revista O Setor Elétrico**, v. 1, n. 200, p. 22–25, 2024. (a).

SOARES, P. H. V.; SILVA, C. R. Comissionamento de centrais geradoras fotovoltaicas. **Revista O Setor Elétrico**, n. 203, p. 24–30, 2024. (b).

TIAGO, V. H. H. **Predição de Proteínas Exportadas por Meio de Redes Neurais Artificiais**. Uberlândia, MG: [s.n.], 2020. 46 f. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Universidade Federal de Uberlândia, 2021.

TIAN, J. et al. Multi-scale solar radiation and photovoltaic power forecasting with machine learning algorithms in urban environment: A state-of-the-art review. **Journal of Cleaner Production**, v. 426, p. 139040, 2023. Disponível em: <<https://doi.org/10.1016/j.jclepro.2023.139040>>.

United Nations. **What is renewable energy?** 2026. Web page. Acessado em 20 de maio de 2026. Disponível em: <<https://www.un.org/en/climatechange/what-is-renewable-energy>>.

United States Department of Energy. **Solar Performance and Efficiency**. 2025. Acesso em: 28 out. 2025. Disponível em: <<https://www.energy.gov/eere/solar/solar-performance-and-efficiency>>.

_____. **Solar Photovoltaic Cell Basics**. 2025. Acesso em: 28 out. 2025. Disponível em: <<https://www.energy.gov/cmei/systems/solar-photovoltaic-cell-basics>>.

- VEN, G. M. van de; TUYTELAARS, T.; TOLIAS, A. S. Three types of incremental learning. **Nature Machine Intelligence**, v. 4, n. 12, p. 1185–1197, 2022. ISSN 2522-5839. Disponível em: <<https://doi.org/10.1038/s42256-022-00568-3>>.
- WIATROS-MOTYKA, M.; JONES, D.; BROADBENT, H.; FULGHUM, N.; BRUCE-LOCKHART, C.; DIZON, R.; MACDONALD, P.; MOORE, C.; CANDLIN, A.; LEE, U.; COPSEY, L.; HAWKINS, S.; EWEN, M.; WORTHINGTON, B.; BENHAM, H.; TRUEMAN, M.; YANG, M.; LOLLA, A.; EDIANTO, A. S.; CZYŻAK, P.; BROWN, S.; ROSSLOWE, C.; BLACK, R. **Global Electricity Review 2023**. [S.l.], 2023. Disponível em: <<https://ember-climate.org>>.
- YANG, D. Making reference solar forecasts with climatology, persistence, and their optimal convex combination. **Solar Energy**, v. 193, p. 981–985, 2019. Disponível em: <<https://doi.org/10.1016/j.solener.2019.10.006>>.
- YOON, J.; YANG, E.; LEE, J.; HWANG, S. J. Lifelong learning with dynamically expandable networks. In: **International Conference on Learning Representations**. [s.n.], 2018. Disponível em: <<https://openreview.net/forum?id=Sk7KsfW0->>.
- ZENKE, F.; GERSTNER, W.; GANGULI, S. The temporal paradox of hebbian learning and homeostatic plasticity. **Current Opinion in Neurobiology**, v. 43, p. 166–176, 2017. ISSN 0959-4388. Neurobiology of Learning and Plasticity. Disponível em: <<https://doi.org/10.1016/j.conb.2017.03.015>>.

Apêndices

Resultados completos dos experimentos com o dataset Sol do Cerrado

Este apêndice reúne os resultados completos dos experimentos realizados com o dataset Sol do Cerrado. Ao longo do Capítulo 5, em especial na Seção 5.1, foram apresentados apenas os resultados necessários para fundamentar as análises e discussões desenvolvidas. As tabelas reunidas neste apêndice complementam essas análises, disponibilizando os resultados completos de todas as configurações avaliadas. Vale ressaltar que todos os resultados apresentados nesse apêndice corroboram com as análises feitas no texto principal da dissertação.

Tabela 16 – Variantes do conjunto de dados utilizadas nos experimentos.

#	Granularidade	Agregação	Escalonamento	Nome da variante
1	1 min	Média	Z-Score	1min_mean_zscore
2	1 min	Média	Min-Max	1min_mean_minmax
3	1 min	Mediana	Z-Score	1min_median_zscore
4	1 min	Mediana	Min-Max	1min_median_minmax
5	1 min	Dado bruto	Z-Score	1min_raw_zscore
6	1 min	Dado bruto	Min-Max	1min_raw_minmax
7	10 min	Média	Z-Score	10min_mean_zscore
8	10 min	Média	Min-Max	10min_mean_minmax
9	10 min	Mediana	Z-Score	10min_median_zscore
10	10 min	Mediana	Min-Max	10min_median_minmax
11	10 min	Dado bruto	Z-Score	10min_raw_zscore
12	10 min	Dado bruto	Min-Max	10min_raw_minmax
13	15 min	Média	Z-Score	15min_mean_zscore
14	15 min	Média	Min-Max	15min_mean_minmax

Continuação da Tabela 16

#	Granularidade	Agregação	Escalonamento	Nome da variante
15	15 min	Mediana	Z-Score	15min_median_zscore
16	15 min	Mediana	Min-Max	15min_median_minmax
17	15 min	Dado bruto	Z-Score	15min_raw_zscore
18	15 min	Dado bruto	Min-Max	15min_raw_minmax
19	30 min	Média	Z-Score	30min_mean_zscore
20	30 min	Média	Min-Max	30min_mean_minmax
21	30 min	Mediana	Z-Score	30min_median_zscore
22	30 min	Mediana	Min-Max	30min_median_minmax
23	30 min	Dado bruto	Z-Score	30min_raw_zscore
24	30 min	Dado bruto	Min-Max	30min_raw_minmax
25	60 min	Média	Z-Score	60min_mean_zscore
26	60 min	Média	Min-Max	60min_mean_minmax
27	60 min	Mediana	Z-Score	60min_median_zscore
28	60 min	Mediana	Min-Max	60min_median_minmax
29	60 min	Dado bruto	Z-Score	60min_raw_zscore
30	60 min	Dado bruto	Min-Max	60min_raw_minmax

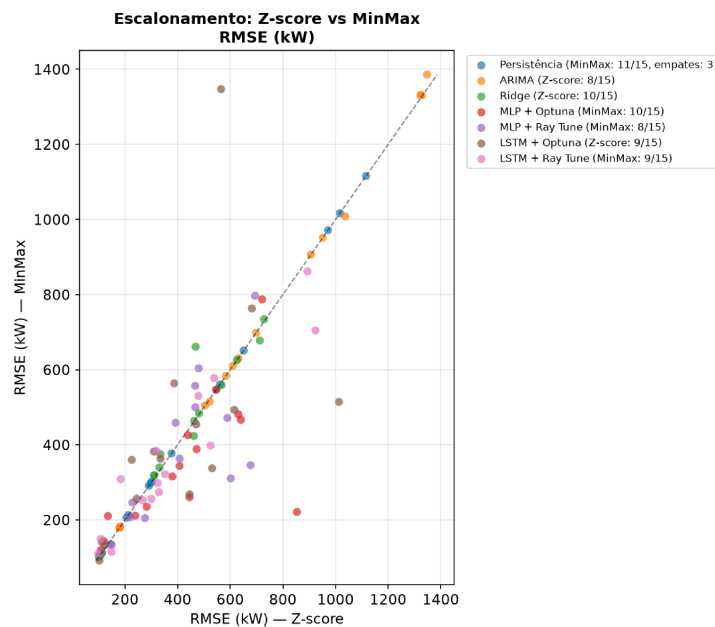


Figura 25 – Contagem dos melhores escalonamentos por modelo de previsão na métrica RMSE.

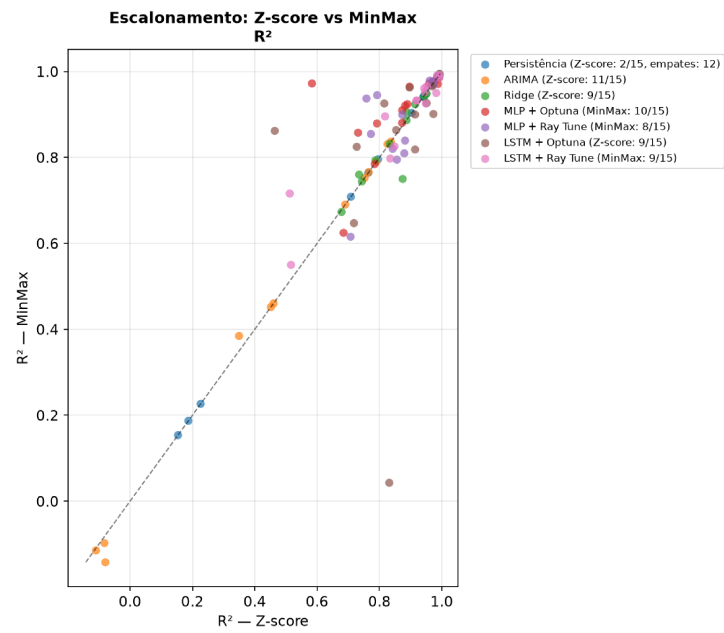


Figura 26 – Contagem dos melhores escalonamentos por modelo de previsão na métrica R^2 .

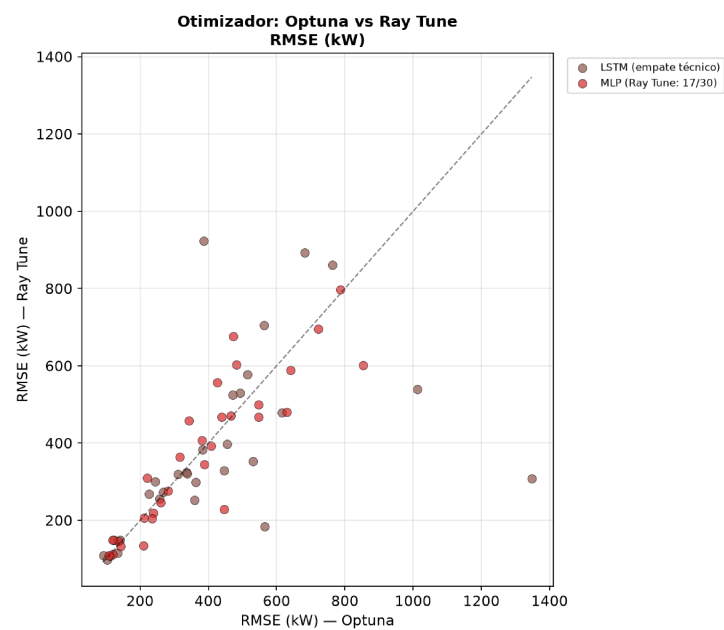


Figura 27 – Contagem de melhores otimizadores de hiperparâmetros na métrica RMSE.

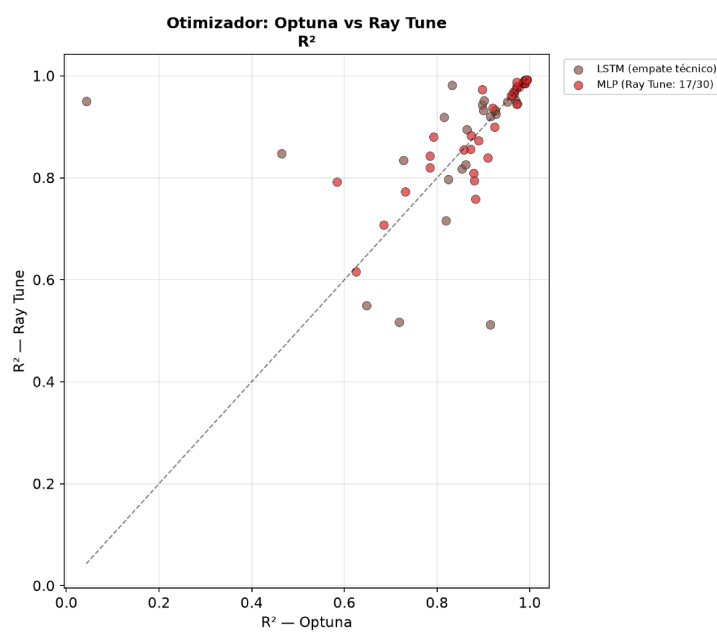


Figura 28 – Contagem de melhores otimizadores de hiperparâmetros na métrica R^2 .

Tabela 17 – MAE (kW) por variante e modelo.

Variante	Persistência	ARIMA	Ridge	MLP + <i>Optuna</i>	MLP + <i>Ray Tune</i>	LSTM + <i>Optuna</i>	LSTM + <i>Ray Tune</i>
1min_mean_minmax	27,30	45,54	40,43	115,04	93,69	38,50	58,55
1min_mean_zscore	27,30	45,48	40,59	93,05	94,16	53,68	40,81
1min_median_minmax	28,07	47,27	44,22	95,74	74,43	80,12	88,83
1min_median_zscore	28,07	47,13	44,51	53,43	99,73	47,95	49,62
1min_raw_minmax	34,88	54,14	49,13	71,09	52,61	84,12	65,24
1min_raw_zscore	34,88	53,98	49,37	51,72	49,45	65,89	100,40
10min_mean_minmax	139,95	190,84	256,36	149,63	150,02	198,17	205,46
10min_mean_zscore	139,95	190,84	246,28	196,58	187,02	204,95	257,14
10min_median_minmax	132,74	217,04	261,81	173,86	178,61	227,35	229,14
10min_median_zscore	132,74	213,07	252,65	338,00	197,45	301,08	257,61
10min_raw_minmax	168,92	285,37	264,99	176,96	146,02	310,91	304,51
10min_raw_zscore	168,92	285,32	253,31	220,76	230,02	245,28	240,00
15min_mean_minmax	186,97	258,67	304,54	273,76	303,94	325,58	256,53
15min_mean_zscore	186,97	258,67	283,86	318,53	354,18	294,39	289,30
15min_median_minmax	185,75	260,66	339,51	275,56	413,01	323,56	181,14
15min_median_zscore	185,75	260,66	406,72	345,90	323,32	186,24	199,99
15min_raw_minmax	225,35	341,24	399,46	474,42	415,41	429,76	451,16
15min_raw_zscore	225,35	341,29	398,64	345,47	360,18	417,46	405,29
30min_mean_minmax	340,95	474,24	495,71	382,23	516,30	410,07	359,70
30min_mean_zscore	340,95	474,09	493,92	372,81	420,69	431,15	481,27
30min_median_minmax	332,44	463,02	566,50	417,36	414,77	248,51	213,89
30min_median_zscore	332,44	463,02	566,64	595,16	530,46	499,23	274,98

Tabela 17 – MAE (kW) por variante e modelo (continuação).

Variante	Persistência	ARIMA	Ridge	MLP + <i>Optuna</i>	MLP + <i>Ray Tune</i>	LSTM + <i>Optuna</i>	LSTM + <i>Ray Tune</i>
30min_raw_minmax	409,67	600,67	636,74	702,87	701,54	655,66	760,27
30min_raw_zscore	409,67	576,81	629,93	644,91	553,49	595,26	800,10
60min_mean_minmax	621,23	912,45	563,93	189,12	239,91	546,94	592,74
60min_mean_zscore	621,23	919,60	417,37	676,58	516,89	311,19	805,86
60min_median_minmax	637,99	888,75	432,13	278,74	266,62	1254,17	272,04
60min_median_zscore	637,99	942,84	425,80	348,39	575,55	425,94	153,30
60min_raw_minmax	702,48	999,60	634,62	432,60	559,98	374,76	469,84
60min_raw_zscore	702,48	997,85	667,06	469,09	430,60	840,84	524,60

Tabela 18 – R^2 por variante e modelo.

Variante	Persistência	ARIMA	Ridge	MLP + <i>Optuna</i>	MLP + <i>Ray Tune</i>	LSTM + <i>Optuna</i>	LSTM + <i>Ray Tune</i>
1min_mean_minmax	0,993	0,980	0,993	0,971	0,988	0,995	0,992
1min_mean_zscore	0,993	0,980	0,993	0,988	0,986	0,992	0,993
1min_median_minmax	0,992	0,979	0,992	0,987	0,988	0,987	0,985
1min_median_zscore	0,992	0,979	0,992	0,991	0,986	0,992	0,992
1min_raw_minmax	0,993	0,980	0,992	0,991	0,992	0,988	0,991
1min_raw_zscore	0,993	0,980	0,992	0,992	0,992	0,989	0,986
10min_mean_minmax	0,970	0,837	0,947	0,977	0,978	0,966	0,966
10min_mean_zscore	0,970	0,837	0,949	0,970	0,975	0,969	0,953
10min_median_minmax	0,972	0,831	0,940	0,965	0,968	0,963	0,961
10min_median_zscore	0,972	0,826	0,943	0,897	0,973	0,897	0,944

Tabela 18 – R^2 por variante e modelo (continuação).

Variante	Persistência	ARIMA	Ridge	MLP + <i>Optuna</i>	MLP + <i>Ray Tune</i>	LSTM + <i>Optuna</i>	LSTM + <i>Ray Tune</i>
10min_raw_minmax	0,941	0,790	0,949	0,972	0,979	0,926	0,926
10min_raw_zscore	0,941	0,790	0,951	0,960	0,962	0,951	0,949
15min_mean_minmax	0,943	0,766	0,924	0,924	0,899	0,900	0,932
15min_mean_zscore	0,943	0,766	0,915	0,890	0,873	0,914	0,920
15min_median_minmax	0,941	0,752	0,905	0,910	0,839	0,901	0,951
15min_median_zscore	0,941	0,752	0,887	0,873	0,882	0,973	0,945
15min_raw_minmax	0,903	0,690	0,831	0,784	0,820	0,825	0,797
15min_raw_zscore	0,903	0,690	0,833	0,785	0,843	0,727	0,835
30min_mean_minmax	0,796	0,461	0,793	0,880	0,795	0,864	0,895
30min_mean_zscore	0,796	0,461	0,787	0,873	0,856	0,854	0,818
30min_median_minmax	0,764	0,452	0,744	0,858	0,854	0,925	0,932
30min_median_zscore	0,764	0,452	0,744	0,732	0,773	0,815	0,919
30min_raw_minmax	0,709	0,385	0,673	0,624	0,615	0,647	0,550
30min_raw_zscore	0,709	0,350	0,679	0,685	0,708	0,718	0,516
60min_mean_minmax	0,226	-0,115	0,750	0,972	0,945	0,818	0,716
60min_mean_zscore	0,226	-0,109	0,875	0,584	0,793	0,914	0,512
60min_median_minmax	0,187	-0,142	0,887	0,920	0,937	0,043	0,950
60min_median_zscore	0,187	-0,079	0,887	0,882	0,758	0,832	0,982
60min_raw_minmax	0,154	-0,098	0,760	0,879	0,809	0,862	0,826
60min_raw_zscore	0,154	-0,082	0,735	0,792	0,880	0,464	0,848

Tabela 19 – RMSE (kW) por variante e modelo.

Variante	Persistência	ARIMA	Ridge	MLP + <i>Optuna</i>	MLP + <i>Ray Tune</i>	LSTM + <i>Optuna</i>	LSTM + <i>Ray Tune</i>
1min_mean_minmax	103,37	179,23	100,84	210,13	134,86	91,89	109,45
1min_mean_zscore	103,37	179,21	100,83	135,30	146,55	102,63	97,97
1min_median_minmax	112,68	182,17	112,43	143,44	133,43	141,38	149,22
1min_median_zscore	112,68	182,11	112,42	119,48	148,67	112,41	108,07
1min_raw_minmax	107,36	179,52	107,76	120,27	112,85	133,16	115,11
1min_raw_zscore	107,36	179,43	107,81	108,50	108,70	123,95	148,50
10min_mean_minmax	212,91	504,76	317,83	211,15	207,14	256,44	256,05
10min_mean_zscore	212,91	504,76	311,09	239,00	219,53	244,39	300,42
10min_median_minmax	205,64	514,77	339,62	260,35	246,32	267,97	273,72
10min_median_zscore	205,64	522,50	330,98	445,35	228,43	445,68	329,08
10min_raw_minmax	302,36	583,51	319,18	235,26	204,53	382,04	383,93
10min_raw_zscore	302,36	583,41	310,70	282,57	276,03	310,33	318,91
15min_mean_minmax	290,69	610,06	375,14	315,70	363,53	362,73	298,11
15min_mean_zscore	290,69	610,06	334,69	380,22	407,72	335,12	323,78
15min_median_minmax	298,02	630,70	423,38	343,68	458,69	359,93	252,17
15min_median_zscore	298,02	630,70	461,72	407,44	392,66	225,90	267,69
15min_raw_minmax	377,21	698,12	483,64	546,86	499,75	493,21	530,15
15min_raw_zscore	377,21	698,17	481,47	546,60	467,50	615,10	478,95
30min_mean_minmax	546,42	951,69	559,07	425,46	557,23	454,59	398,00
30min_mean_zscore	546,42	951,52	567,31	439,26	466,84	470,62	525,14
30min_median_minmax	560,78	906,33	625,21	466,58	471,55	337,45	321,51
30min_median_zscore	560,78	906,33	625,54	640,40	589,39	531,01	352,37

Tabela 19 – RMSE (kW) por variante e modelo (continuação).

Variante	Persistência	ARIMA	Ridge	MLP + <i>Optuna</i>	MLP + <i>Ray Tune</i>	LSTM + <i>Optuna</i>	LSTM + <i>Ray Tune</i>
30min_raw_minmax	651,06	1008,42	734,30	787,17	796,59	763,17	861,75
30min_raw_zscore	651,06	1036,54	728,23	721,05	694,70	681,84	893,23
60min_mean_minmax	971,42	1330,80	661,27	221,00	310,49	563,45	704,76
60min_mean_zscore	971,42	1327,62	468,46	853,28	602,11	387,02	923,39
60min_median_minmax	1016,80	1386,12	463,51	388,46	345,58	1347,07	308,48
60min_median_zscore	1016,80	1347,48	462,54	472,12	677,06	565,18	184,11
60min_raw_minmax	1115,74	1331,30	677,73	481,12	603,79	514,25	577,75
60min_raw_zscore	1115,74	1321,83	712,28	630,60	479,78	1012,60	538,85

Tabela 20 – *Skill Score* por variante e modelo.

Variante	Persistência	ARIMA	Ridge	MLP + <i>Optuna</i>	MLP + <i>Ray Tune</i>	LSTM + <i>Optuna</i>	LSTM + <i>Ray Tune</i>
1min_mean_minmax	0,000	-0,734	0,025	-1,033	-0,305	0,111	-0,059
1min_mean_zscore	0,000	-0,734	0,025	-0,309	-0,418	0,007	0,052
1min_median_minmax	0,000	-0,617	0,002	-0,273	-0,184	-0,255	-0,324
1min_median_zscore	0,000	-0,616	0,002	-0,060	-0,319	0,002	0,041
1min_raw_minmax	0,000	-0,672	-0,004	-0,120	-0,051	-0,240	-0,072
1min_raw_zscore	0,000	-0,671	-0,004	-0,011	-0,012	-0,154	-0,383
10min_mean_minmax	0,000	-1,371	-0,493	0,008	0,027	-0,204	-0,203
10min_mean_zscore	0,000	-1,371	-0,461	-0,122	-0,031	-0,148	-0,411
10min_median_minmax	0,000	-1,503	-0,652	-0,266	-0,198	-0,303	-0,331

Tabela 20 – *Skill Score* por variante e modelo (continuação).

Variante	Persistência	ARIMA	Ridge	MLP + <i>Optuna</i>	MLP + <i>Ray Tune</i>	LSTM + <i>Optuna</i>	LSTM + <i>Ray Tune</i>
10min_median_zscore	0,000	-1,541	-0,610	-1,166	-0,111	-1,167	-0,600
10min_raw_minmax	0,000	-0,930	-0,056	0,222	0,324	-0,264	-0,270
10min_raw_zscore	0,000	-0,930	-0,028	0,065	0,087	-0,026	-0,055
15min_mean_minmax	0,000	-1,099	-0,290	-0,086	-0,251	-0,248	-0,025
15min_mean_zscore	0,000	-1,099	-0,151	-0,308	-0,403	-0,153	-0,114
15min_median_minmax	0,000	-1,116	-0,421	-0,153	-0,539	-0,208	0,154
15min_median_zscore	0,000	-1,116	-0,549	-0,367	-0,318	0,242	0,102
15min_raw_minmax	0,000	-0,851	-0,282	-0,450	-0,325	-0,307	-0,405
15min_raw_zscore	0,000	-0,851	-0,276	-0,449	-0,239	-0,631	-0,270
30min_mean_minmax	0,000	-0,742	-0,023	0,221	-0,020	0,168	0,272
30min_mean_zscore	0,000	-0,741	-0,038	0,196	0,146	0,139	0,039
30min_median_minmax	0,000	-0,616	-0,115	0,168	0,159	0,398	0,427
30min_median_zscore	0,000	-0,616	-0,116	-0,142	-0,051	0,053	0,372
30min_raw_minmax	0,000	-0,549	-0,128	-0,209	-0,224	-0,172	-0,324
30min_raw_zscore	0,000	-0,592	-0,118	-0,107	-0,067	-0,047	-0,372
60min_mean_minmax	0,000	-0,370	0,319	0,772	0,680	0,420	0,275
60min_mean_zscore	0,000	-0,367	0,518	0,122	0,380	0,602	0,049
60min_median_minmax	0,000	-0,363	0,544	0,618	0,660	-0,325	0,697
60min_median_zscore	0,000	-0,325	0,545	0,536	0,334	0,444	0,819
60min_raw_minmax	0,000	-0,193	0,393	0,569	0,459	0,539	0,482
60min_raw_zscore	0,000	-0,185	0,362	0,435	0,570	0,092	0,517

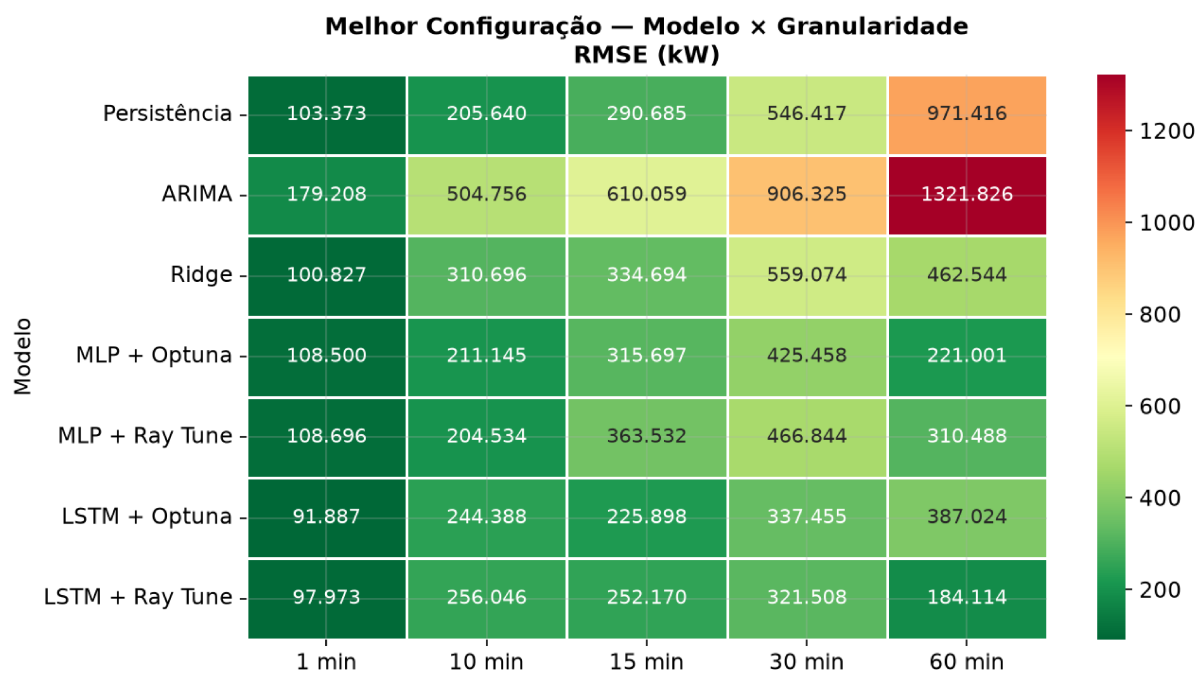


Figura 29 – Melhor configuração de cada modelo por granularidade na métrica RMSE.

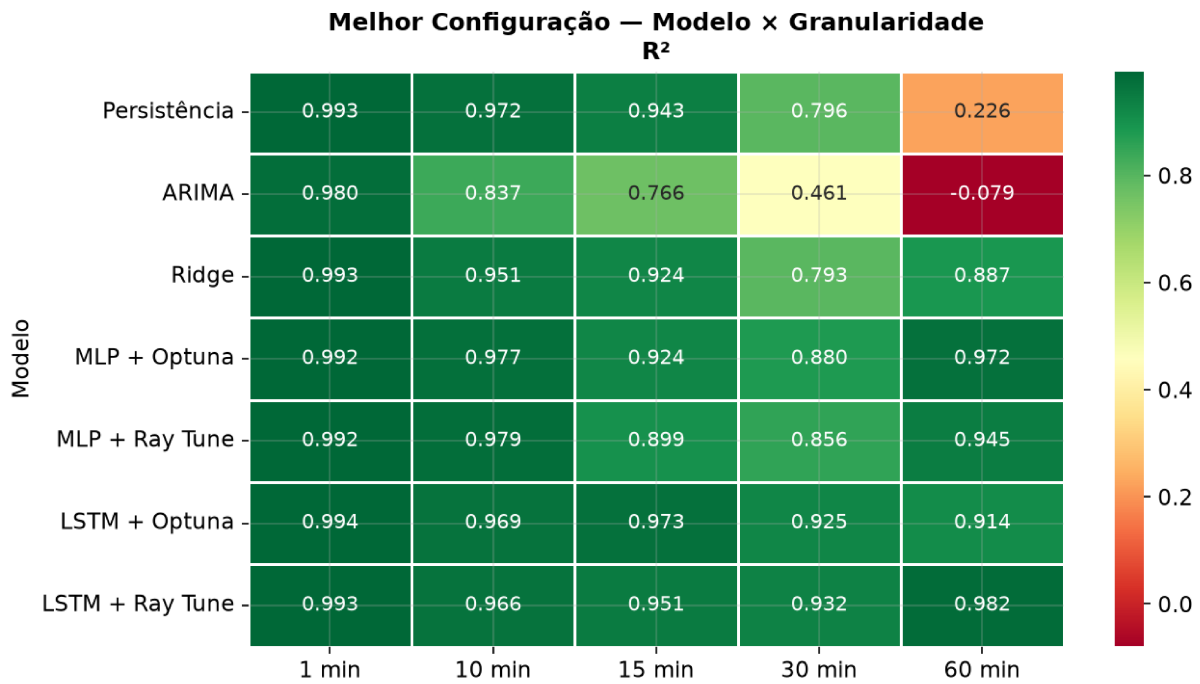


Figura 30 – Melhor configuração de cada modelo por granularidade na métrica R^2 .

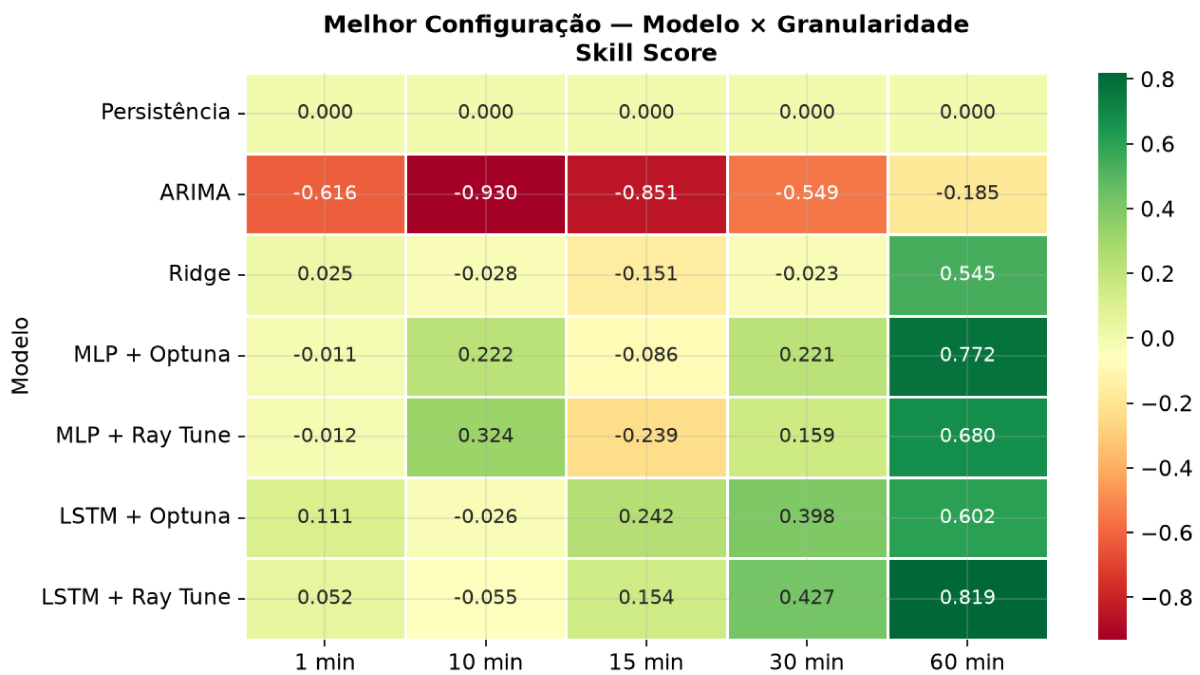


Figura 31 – Melhor configuração de cada modelo por granularidade na métrica *Skill Score*.

Resultados completos dos experimentos com o dataset UNISOLAR

Este apêndice reúne os resultados completos dos experimentos realizados com o dataset UNISOLAR. Ao longo do Capítulo 5, em especial na Seção 5.2, foram apresentados apenas os resultados necessários para fundamentar as análises e discussões desenvolvidas. As tabelas reunidas neste apêndice complementam essas análises, disponibilizando os resultados completos de todas as configurações avaliadas. Vale ressaltar que todos os resultados apresentados nesse apêndice corroboram com as análises feitas no texto principal da dissertação.

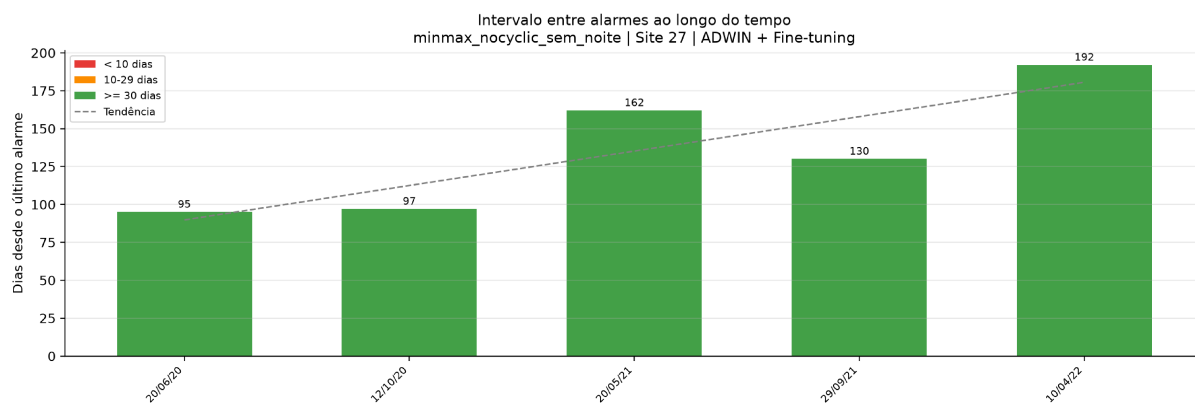


Figura 32 – Alarmes de degradação detectados pelo ADWIN durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental *fine-tuning* na variante V7.

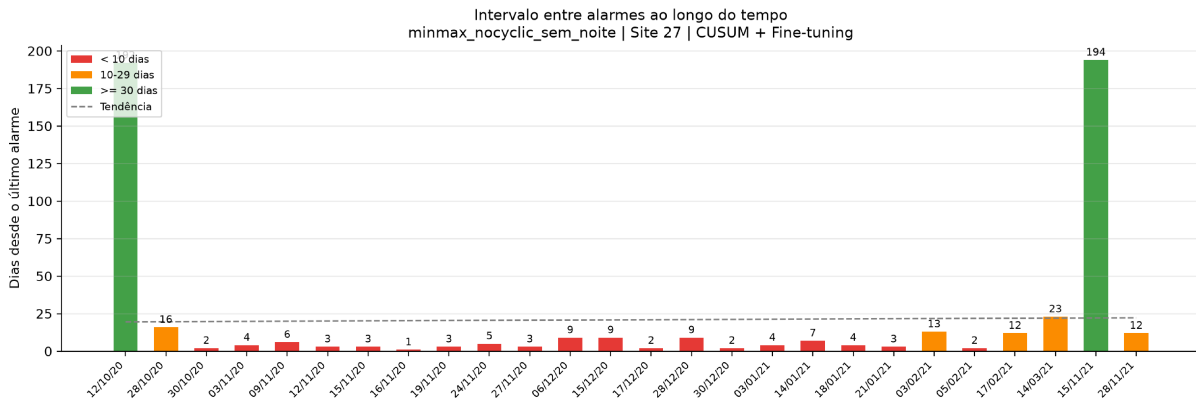


Figura 33 – Alarmes de degradação detectados pelo CUSUM durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental *fine-tuning* na variante V7.

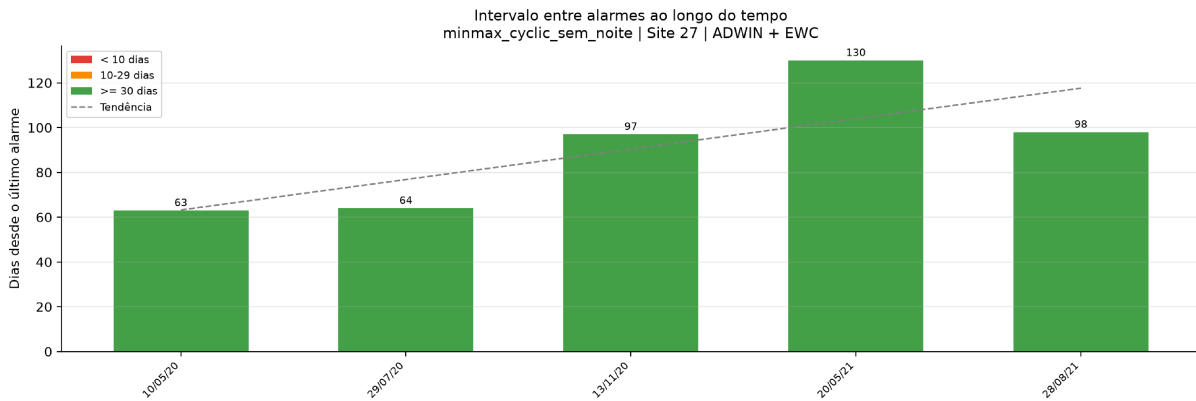


Figura 34 – Alarmes de degradação detectados pelo ADWIN durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental EWC na variante V5.

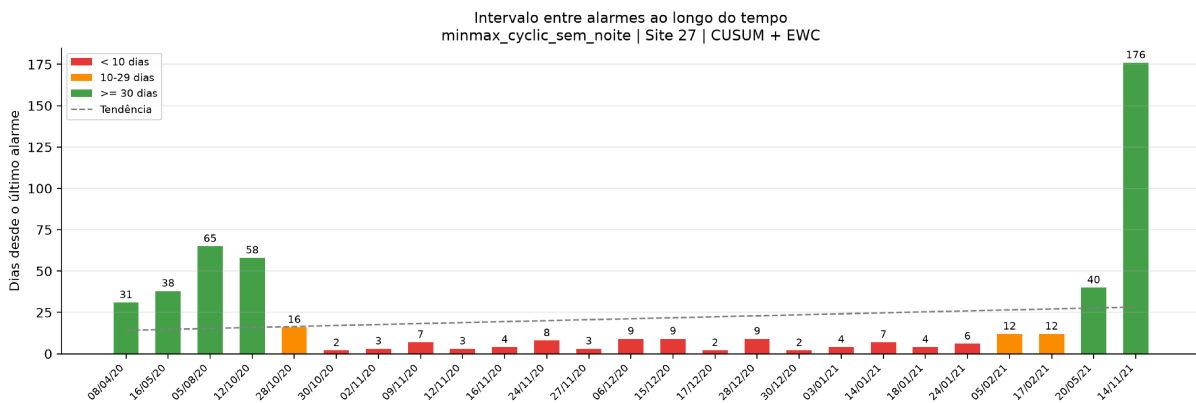


Figura 35 – Alarmes de degradação detectados pelo CUSUM durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental EWC na variante V5.

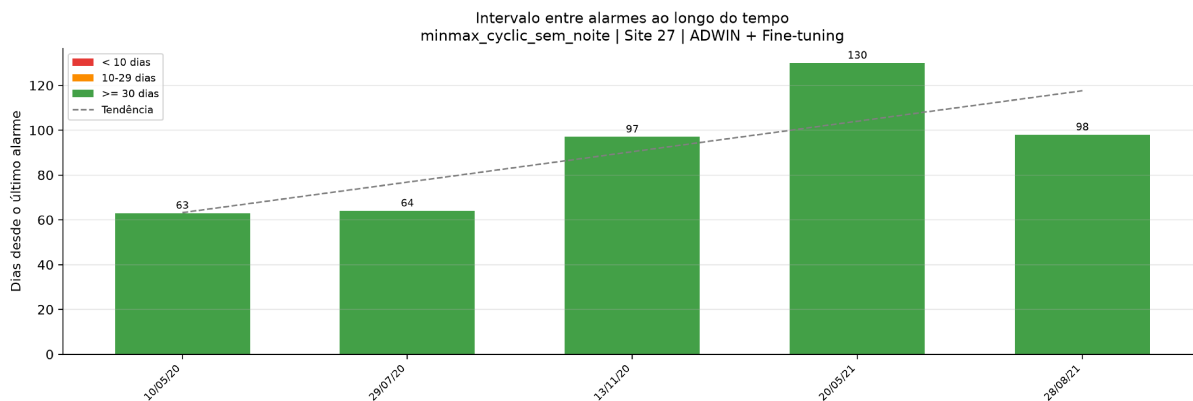


Figura 36 – Alarmes de degradação detectados pelo ADWIN durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental *fine-tuning* na variante V5.

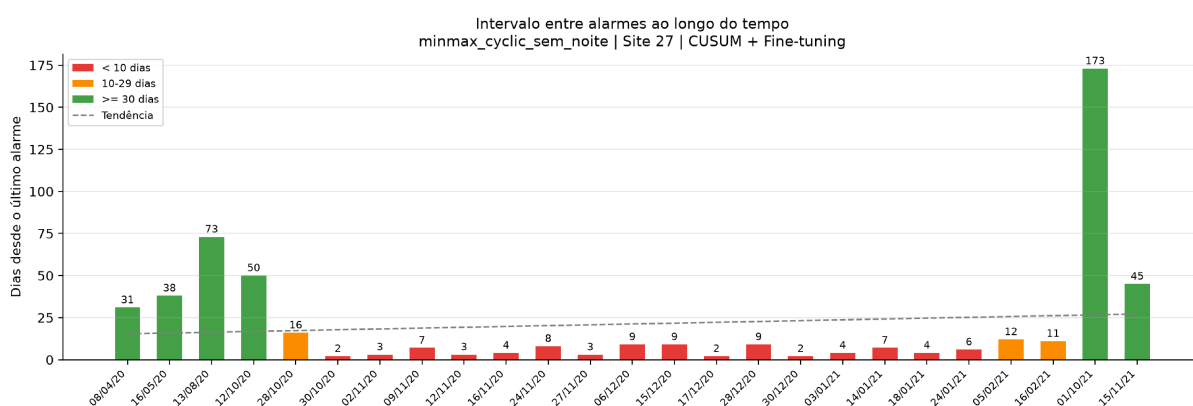


Figura 37 – Alarmes de degradação detectados pelo CUSUM durante os 685 dias de simulação utilizando a estratégia de aprendizado incremental *fine-tuning* na variante V5.