



MODELOS DE PREVISÃO PARA DETEÇÃO DE FALHAS EM SISTEMAS FOTOVOLTAICOS

MARIA HELENA LEAL SOARES

outubro de 2023

MODELOS DE PREVISÃO PARA DETEÇÃO DE FALHAS EM SISTEMAS FOTOVOLTAICOS

Maria Helena Leal Soares

2023

Instituto Superior de Engenharia do Porto

Departamento de Engenharia Mecânica

isen

P.PORTO

MODELOS DE PREVISÃO PARA DETEÇÃO DE FALHAS EM SISTEMAS FOTOVOLTAICOS

Maria Helena Leal Soares

Estudante n.º 1210191

Dissertação apresentada ao Instituto Superior de Engenharia do Porto para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia e Gestão Industrial, realizada sob a orientação do Doutor Telmo Manuel Sampaio Pinto Matos e coorientação da Doutora Susana Cláudia Nicola de Araújo e do Mestre Joaquim Moreira

2023

Instituto Superior de Engenharia do Porto

Departamento de Engenharia Mecânica

isen

P.PORTO

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão a todas as pessoas que contribuíram para a realização deste trabalho de pesquisa e para a conclusão desta dissertação de mestrado.

Primeiramente, gostaria de agradecer a meu orientador, Professor Doutor Telmo Matos, pela orientação, disponibilidade e apoio ao longo do processo. As suas orientações e os feedbacks construtivos foram fundamentais para o desenvolvimento desta pesquisa.

À Professora Doutora Susana Nicola, orientadora académica do Instituto Superior de Engenharia do Porto e ao Mestre Joaquim Moreira pela disponibilidade e pela orientação na realização do presente projeto.

Ao meu orientador na organização, Engenheiro Tiago Horta Nisa, pelo auxílio em qualquer questão, disponibilidade, sugestões de melhoria do projeto e pelo fornecimento de informações importantes para o desenvolvimento do projeto.

À minha família, amigos e namorado pelo apoio incondicional, encorajamento e compreensão ao longo dos anos de estudo, foram uma fonte constante de motivação.

Por fim, quero expressar minha gratidão a todas as pessoas que, de alguma forma, contribuíram para esta jornada académica. Este trabalho não teria sido possível sem o apoio e colaboração de cada um de vocês.

página propositadamente em branco

RESUMO

A crescente preocupação com a sustentabilidade energética tem conduzido a um aumento significativo na adoção de fontes de energia renovável em detrimento das fontes não renováveis, como os combustíveis fósseis. Este tópico tem ganhado cada vez mais destaque na atualidade, refletindo um movimento global em direção a práticas mais ecológicas e eficientes no uso de recursos energéticos. Os sistemas fotovoltaicos surgem como uma medida sustentável, permitindo transformar energia solar em energia elétrica. Este tipo de sistemas é bastante vantajoso, para os mais diversos setores, para além de permitir a redução de custos nas organizações, auxilia o combate nas questões climáticas.

No presente trabalho, através do conjunto de dados fornecidos contendo várias métricas, foi possível estudar uma instalação de painéis fotovoltaicos situada na zona do Porto. O principal objetivo era criar um método de diagnóstico de falhas através de métodos de previsão em linguagem *Python*. Para tal desenvolveu-se uma análise exploratória dos dados, com a finalidade de obter *insights* importantes para a seleção do modelo a desenvolver. Desta forma, de acordo com as conclusões da análise exploratória, foram selecionados dois modelos de previsão.

Inicialmente, foi implementado o modelo de previsão baseado em redes neuronais, no entanto, apenas foi aplicado para uma variável. Desta forma, após uma nova pesquisa na literatura optou-se por utilizar o modelo de previsão de séries temporais SARIMA, visto que, os dados apresentavam sazonalidade, uma característica importante para a construção deste modelo. E desta forma, foi possível também comparar os resultados dos dois modelos para uma variável em comum. O principal desafio foi a robustez do conjunto de dados disponíveis. Apesar da granularidade dos dados apresentados, apenas foram facultados dados relativos a dois anos, representando um desafio substancial, uma vez que o modelo de previsão SARIMA depende da análise do comportamento dos dados históricos para estabelecer previsões futuras confiáveis.

Os dois modelos permitiram desenvolver valores de previsões futuros sobre diversas variáveis relevantes para a identificação de potenciais falhas no sistema fotovoltaico. Os resultados das previsões indicam que, para o próximo ano, os valores das variáveis devem manter-se semelhantes aos dados históricos e, portanto, não são esperadas falhas. Estes resultados eram esperados, uma vez que os fornecedores garantem um alto desempenho no início da vida útil do sistema.

Os sistemas fotovoltaicos representam uma solução definitiva e sustentável e de grande relevância na atualidade. Quando unimos esse tópico à previsão, ele torna-se ainda mais fascinante para explorar. Além disso, a integração de questões relacionadas às energias renováveis com aspetos de engenharia, como a programação de métodos de previsão, ofereceu uma abordagem muito interessante de desenvolver.

PALAVRAS-CHAVE

Painéis Fotovoltaicos, análise de dados, séries temporais, SARIMA, Redes Neuronais

página propositadamente em branco

ABSTRACT

The growing concern about energy sustainability has led to a significant increase in the adoption of renewable energy sources to the detriment of non-renewable sources, such as fossil fuels. This topic has gained increasing prominence today, reflecting a global movement towards more ecological and efficient practices in the use of energy resources. Photovoltaic systems appear as a sustainable measure, allowing solar energy to be transformed into electrical energy. This type of systems is quite advantageous for the most diverse sectors, in addition to allowing cost reduction in organizations, it helps combat climate issues.

In the present work, through the set of data provided containing several metrics, it was possible to study a photovoltaic panel installation located in the Porto area. The main objective was to create a fault diagnosis method using prediction methods in Python language. To this end, an exploratory analysis of the data was developed, with the purpose of obtaining important insights for the selection of the model to be developed. Therefore, according to the conclusions of the exploratory analysis, two forecast models were selected.

Initially, the prediction model based on neural networks was implemented, however, it was only applied to one variable. Therefore, after a new search in the literature, it was decided to use the SARIMA time series forecasting model, since the data presented seasonality, an important characteristic for the construction of this model. This way, it was also possible to compare the results of the two models for a common variable. The main challenge was the robustness of the available data set. Despite the granularity of the data presented, data was only provided for two years, representing a

substantial challenge, since the SARIMA forecast model relies on analyzing the behavior of historical data to establish reliable future forecasts.

The two models made it possible to develop future prediction values on several variables relevant to identifying potential failures in the photovoltaic system. The forecast results indicate that, for the next year, the values of the variables should remain similar to historical data and, therefore, failures are not expected. These results were expected, as suppliers guarantee high performance at the beginning of the system's useful life.

Photovoltaic systems represent a definitive and sustainable solution and are of great relevance today. When we combine this topic with prediction, it becomes even more fascinating to explore. Furthermore, the integration of issues related to renewable energy with engineering aspects, such as programming forecasting methods, offered a very interesting approach to develop.

KEYWORDS

Photovoltaic Panels, data analysis, time series, SARIMA, Neural Networks

página propositadamente em branco

ÍNDICE

ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABELAS	XV
LISTAS DE SIGLAS E SÍMBOLOS.....	XIX
1. INTRODUÇÃO	1
1.1. Enquadramento e pertinência	1
1.2. Questão e objetivos de investigação.....	2
1.3. Opções metodológicas	2
1.4. Apresentação da empresa.....	3
1.5. Estrutura do trabalho	3
2. REVISÃO BIBLIOGRÁFICA.....	4
2.1. <i>Big Data</i> e análise de dados na gestão industrial.....	4
2.1.1. Características de Big data	5
2.1.2. Vantagens de Big data.....	5
2.1.3. O impacto da análise de <i>Big Data</i> nas empresas.....	5
2.1.4. Limitações de Big data	6
2.2. Panorama nacional da energia e sustentabilidade energética	7
2.2.1. Sustentabilidade energética	7
2.2.2. Panorama energético em Portugal	8
2.3. Sistemas fotovoltaicos.....	10
2.3.1. Tipos de sistemas fotovoltaicos	10
2.3.2. Constituição de um sistema fotovoltaico	13
2.3.3. Funcionamento dos sistemas fotovoltaicos.....	13
2.3.4. Origem e classificação das falhas nos sistemas fotovoltaicos	14
2.3.5. Exemplos de falhas.....	15
2.3.6. Modelos de deteção e diagnóstico de falhas nos sistemas fotovoltaicos	17
2.3.7. Manutenção e monitorização de PF	18
2.4. Métodos de previsão.....	20
2.4.1. Análise de séries temporais	21
2.4.2. Redes neuronais.....	21
2.4.3. SARIMA.....	23
2.5. Métricas de avaliação dos modelos de previsão.....	28
3. METODOLOGIA E TÉCNICAS APLICADAS.....	31
3.1. Descrição da unidade de produção de autoconsumo.....	31
3.2. Software utilizado	32
3.3. Descrição e tratamento dos dados.....	33
3.3.1. Recolha e descrição dos dados	33
3.3.2. Tratamento dos dados	33

3.4. Análise exploratória.....	39
3.5. Pré-processamento	41
3.6. Redes Neurais.....	41
3.7. SARIMA.....	47
4. RESULTADOS E ANÁLISES	62
4.1. Resultados da análise nulos	62
4.2. Resultados da análise exploratória dos dados	63
4.3. Resultados Redes neuronais	71
4.4. Resultados SARIMA	80
4.4.1. Variável 1: Temperatura média ($T_MEDIA [^{\circ}C]$)	81
4.4.2. Variável 2: Irradiância média ($IRR_MEDIA [Wh/m^2]$).....	87
4.4.3. Variável 3: Potencia do lado DC inversor A ($P_DC_INV_A [W]$).....	93
4.4.4. Variável 4: Potencia do lado AC inversor A ($P_AC_INV_A [W]$)	99
4.4.5. Variável 5: Temperatura do inversor A ($T_INV_A [^{\circ}C]$).....	105
4.4.6. Variável 6: Potencia do lado DC do inversor B ($P_DC_INV_B [W]$)	110
4.4.7. Variável 7: Potencia do lado AC do inversor B ($P_AC_INV_B [W]$)	115
4.4.8. Variável 8: Temperatura do inversor B ($T_INV_B [^{\circ}C]$)	120
4.5. Discussão de resultados	126
5. CONCLUSÃO	129
5.1. Principais conclusões.....	129
5.2. Desafios encontrados.....	130
5.3. Trabalhos futuros	131
REFERÊNCIAS BIBLIOGRÁFICAS	133
ANEXO A – Funções análise nulos	139
ANEXO B – Funções análise exploratória	143
ANEXO C – Funções pré-processamento	151
ANEXO D – Funções redes neuronais.....	153
ANEXO E – Funções sarima	155
ANEXO F – Resultados análise exploratória	164

página propositadamente em branco

ÍNDICE DE FIGURAS

Figura 1 : Comparação de consumos energéticos em Portugal entre 1999 e 2019 (da Energia, 2021)	8
Figura 2: Consumo de energia final por setor de atividade em 1999 e 2019 (da Energia, 2021)	8
Figura 3: Consumo de energia final por habitante na união europeia em 2019 (da Energia, 2021)	9
Figura 4: Emissões de GEE por habitante em 2019 na união europeia (da Energia, 2021)	10
Figura 5 : Funcionamento de uma UPAC (Moreira, 2016)	12
Figura 6: Modo de funcionamento de uma UPP (Moreira, 2016)	12
Figura 7 : Modo de funcionamento de um sistema isolado (Francisco et al., n.d.)	13
Figura 8 : Configuração de um sistema fotovoltaico (Miguel et al., 2023)	14
Figura 9 : Exemplo de uma falha entre linhas num sistema Fotovoltaico (Miguel et al., 2023)	16
Figura 10 : Diferentes tipos de falha (Madeti & Singh, 2017)	17
Figura 11: Periodicidade de manutenção dos elementos de um sistema fotovoltaico (Conrado, 2021)	20
Figura 12 : Estrutura geral de uma rede neuronal (Runge & Zmeureanu, 2019)	22
Figura 13 : Principais etapas de desenvolvimento do método de Box & Jenkins (Kirchner, 2006)	25
Figura 14 : Ambiente de programação python no IDE pycharm	32
Figura 15 : Função calcular_eficiencia_inversor em Python e aplicação	36
Figura 16 : Do lado esquerdo vemos um exemplo de alerta de identificação de falha de comunicação, do lado direito um exemplo de alerta da não existência de falhas	38
Figura 17 : Síntese dos Ficheiros criados para desenvolver a análise exploratória dos dados	39
Figura 18 : Divisão do conjunto de dados em treino e teste da variável X e Y	44
Figura 19 : Divisão do conjunto de dados em treino e teste da variável X e Y com comando shuffle	44
Figura 20 : Parâmetros de análise da estrutura da rede neuronal	45
Figura 21 : Fases de compilar o modelo e treinar da rede	46
Figura 22 : Parâmetros selecionados através de testes	47
Figura 23 : Metodologia aplicada	48
Figura 24: Estrutura de implementação do método	49
Figura 25: Linhas de código para calcular parâmetro “d”	52
Figura 26 : Exemplo de aplicação das diferenciações e teste ADF pela 2ª vez	53
Figura 27 : Fluxograma do teste de estacionariedade da série	54
Figura 28 : Método ADF	55
Figura 29: Parâmetros a analisar para ACF e PACF	56
Figura 30 : Funções para o desenvolvimento das previsões	57
Figura 31: Utilização da biblioteca <i>auto_arima</i> para seleção do modelo	58
Figura 32: Estrutura dos gráficos de diagnostico para admitir características de ruído branco	60
Figura 33 : Análise dos valores médios mensais da temperatura dos módulos a Este para os dois anos	63
Figura 34 : Análise dos valores médios mensais da temperatura dos módulos a Oeste para os dois anos	64
Figura 35 : Análise dos valores médios mensais da radiação a Este para os dois anos	65
Figura 36 : Análise dos valores médios mensais da radiação a Oeste para os dois anos	65

Figura 37 : Análise dos valores médios mensais da energia acumulada convertida pelo inversor A em 2021 (lado esquerdo) e 2022 (lado direito)	66
Figura 38 : Valores médios mensais da potência do lado DC em 2021 (canto superior esquerdo), em 2022 (canto superior direito) , Valores médios mensais da potência do lado AC em 2021 (canto inferior esquerdo), em 2022 (canto inferior direito) para o inversor A.....	67
Figura 39 : Análise dos valores de temperatura médios mensais do Inversor A em 2021 (lado esquerdo) em 2022 (lado direito)	67
Figura 40 : Valores médios mensais da potência do lado DC em 2021 (canto superior esquerdo), em 2022 (canto superior direito), Valores médios mensais da potência do lado AC em 2021 (canto inferior esquerdo), em 2022 (canto inferior direito) do Inversor B.....	68
Figura 41 : Análise dos valores de temperatura médios mensais do Inversor B em 2021 (lado esquerdo) em 2022 (lado direito)	69
Figura 42 : Análise dos valores médios mensais de irradiância média em 2021 (lado esquerdo) e 2022 (lado direito).....	69
Figura 43 : Análise dos valores médios mensais de temperatura média em 2021 (lado esquerdo) e 2022 (lado direito).....	70
Figura 44 : Influência da temperatura na potência do lado AC ao longo de 2021 (em cima lado esquerdo) e 2022 (em cima lado direito) e influência da irradiância na potência do lado AC ao longo de 2021 (em cima baixo esquerdo) e 2022 (em baixo lado direito)	71
Figura 45 : Etapas de construção de uma rede neuronal multicamadas (MLP)	72
Figura 46 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do primeiro teste	73
Figura 47 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do segundo teste.....	74
Figura 48 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do terceiro teste.....	76
Figura 49 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do quarto teste.....	77
Figura 50 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do quinto teste.....	78
Figura 51 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável T_MEDIA [°C].....	81
Figura 52 : Resultados teste ADF para a variável T_MEDIA [°C]	82
Figura 53 : Resultados teste ADF para a variável T_MEDIA [°C] após diferenciação sazonal da série	82
Figura 54: Gráficos ACF e PACF para a variável T_MEDIA [°C]	83
Figura 55 :Diagnóstico do modelo	84
Figura 56: Comparação dados reais e previstos para a variável T_MEDIA [°C]	85
Figura 57: Previsões futuras da temperatura média para a variável T_MEDIA [°C]	86
Figura 58 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo para a variável IRR_MEDIA [Wh/m ²]	87
Figura 59 : Resultados teste ADF para a variável IRR_MEDIA [Wh/m ²].....	87
Figura 60 : Resultados teste ADF para a variável IRR_MEDIA [Wh/m ²] após diferenciação sazonal da série.....	88

Figura 61 : Resultados teste ADF para a variável IRR_MEDIA [Wh/m ²] após diferenciação sazonal e não sazonal da série.....	88
Figura 62 : Gráficos ACF e PACF para a variável IRR_MEDIA [Wh/m ²]	89
Figura 63 : Diagnóstico do modelo.....	90
Figura 64 : Comparação dados reais e previstos para a variável IRR_MEDIA [Wh/m ²].....	91
Figura 65 : Previsões futuras da temperatura média para a variável IRR_MEDIA [Wh/m ²].....	92
Figura 66 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável P_DC_INV_A [W].....	93
Figura 67 : Resultados teste ADF para a variável P_DC_INV_A [W].....	93
Figura 68 : Resultados teste ADF para a variável P_DC_INV_A [W] após diferenciação sazonal	94
Figura 69 : Resultados teste ADF para a variável P_DC_INV_A [W] após diferenciação sazonal e não sazonal.....	94
Figura 70 : Gráficos ACF e PACF para a variável P_DC_INV_A [W]	95
Figura 71 : Diagnóstico do modelo.....	96
Figura 72 : Comparação dados reais e previstos para a variável P_DC_INV_A [W].....	97
Figura 73 : Previsões futuras da temperatura média para a variável P_DC_INV_A [W].....	98
Figura 74 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável P_AC_INV_A [W]	99
Figura 75 : Resultados teste ADF para a variável P_AC_INV_A [W].....	99
Figura 76 : Resultados teste ADF para a variável P_AC_INV_A [W] após diferenciação sazonal ..	100
Figura 77 : Resultados teste ADF para a variável P_AC_INV_A [W] após diferenciação sazonal e não sazonal.....	100
Figura 78 : Gráficos ACF e PACF para a variável P_AC_INV_A [W]	101
Figura 79 : Diagnóstico do modelo.....	102
Figura 80 : Comparação dados reais e previstos para a variável P_AC_INV_A [W].....	103
Figura 81 : Previsões futuras da temperatura média para a variável P_AC_INV_A [W].....	104
Figura 82 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável T_INV_A [°C].....	105
Figura 83 : Resultados teste ADF para a variável T_INV_A [°C]	105
Figura 84 : Resultados teste ADF para a variável T_INV_A [°C] após diferenciação sazonal da série	106
Figura 85 : Gráficos ACF e PACF para a variável T_INV_A [°C]	106
Figura 86 : Diagnóstico do modelo.....	107
Figura 87 : Comparação dados reais e previstos para a variável T_INV_A [°C]	108
Figura 88 : Previsões futuras da temperatura média para a variável T_MEDIA [°C]	110
Figura 89 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo para a variável P_DC_INV_B [W]	111
Figura 90 : Resultados teste ADF para a variável P_DC_INV_B [W].....	111
Figura 91 : Gráficos ACF e PACF para a variável P_DC_INV_B [W]	112
Figura 92 : Diagnóstico do modelo.....	113
Figura 93 : Comparação dados reais e previstos para a variável P_DC_INV_B [W].....	114
Figura 94 : Previsões futuras da temperatura média para a variável P_DC_INV_B [W].....	115
Figura 95 : Gráfico da média móvel e desvio padrão móvel da série ao longo do tempo da variável P_AC_INV_B [W]	116
Figura 96 : Resultados teste ADF para a variável P_AC_INV_B [W].....	116

Figura 97 : Gráficos ACF e PACF para a variável P_AC_INV_B [W].....	117
Figura 98 : Diagnóstico do modelo.....	118
Figura 99 : Comparação dados reais e previstos para a variável P_AC_INV_B [W].....	119
Figura 100 : Previsões futuras da temperatura média para a variável P_AC_INV_B [W].....	120
Figura 101 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável T_INV_B [°C].....	121
Figura 102 : Resultados teste ADF para a variável T_INV_B [°C]	121
Figura 103 : Resultados teste ADF para a variável T_INV_B [°C] após diferenciação sazonal da série	122
Figura 104 : Gráficos ACF e PACF para a variável T_INV_B [°C]	122
Figura 105 : Diagnóstico do modelo.....	123
Figura 106 : Comparação dados reais e previstos para a variável T_INV_B [°C]	124
Figura 107 : Previsões futuras da temperatura média para a variável T_INV_B [°C]	125
Figura 108 : Função contar_nulos em <i>Python</i>	139
Figura 109 : Continuação da função contar_nulos em <i>Python</i>	140
Figura 110 : Função get_daytime em <i>Python</i>	141
Figura 111 : Função substituir_nulos em <i>Python</i>	142
Figura 112 : Função criar_grafico_linhas em <i>Python</i>	143
Figura 113 : Continuação da função criar_grafico_linhas em <i>Python</i>	144
Figura 114 : Função criar_matriz_correlacao em <i>Python</i>	145
Figura 115 : Função criar_grafico_barras em <i>Python</i>	146
Figura 116 : Continuação da função criar_grafico_barras em <i>Python</i>	147
Figura 117 : Função criar_grafico_barras_eficiencia_word em <i>Python</i>	149
Figura 118 : Função media_mensal em <i>Python</i>	150
Figura 119 : Linhas de código do Ficheiro pre-processamento	151
Figura 120 : Continuação das linhas de código do Ficheiro pre-processamento	152
Figura 121 : Linhas de código de implementação da rede neuronal MLP para a variável de temperatura do inversor A em graus celsius	153
Figura 122 : Continuação da implementação da rede neuronal MLP para a variável de temperatura do inversor A em graus celsius.....	154
Figura 123: Função pre_processamento_SARIMA em <i>Python</i>	155
Figura 124: Função pre_processamento_SARIMA em <i>Python</i>	156
Figura 125: Função teste_dickey_fuller em <i>Python</i>	157
Figura 126: Função plot_acf_pacf_to_word em <i>Python</i>	158
Figura 127: Função Previsao_Sarima_treino_teste em <i>Python</i>	160
Figura 128: Continuação da função Previsao_Sarima_treino_teste em <i>Python</i>	161
Figura 129: Função fazer_previsoes_futuras em <i>Python</i>	163
Figura 130 : Análise da temperatura do módulo a Este em 2021 (lado esquerdo) e 2022 (lado direito)	164
Figura 131: Análise da temperatura do módulo a Oeste em 2021 (lado esquerdo) e 2022 (lado direito).....	164
Figura 132: Análise dos valores médios mensais da radiação a Este em 2021 (lado esquerdo) e 2022 (lado direito)	165
Figura 133: Análise dos valores médios mensais da radiação a Oeste em 2021 (lado esquerdo) e 2022 (lado direito).....	165

Figura 134: Análise dos valores médios mensais da eficiência do inversor A em 2021 (lado esquerdo) e 2022 (lado direito).....	166
Figura 135: Análise dos valores médios mensais da eficiência do inversor B em 2021 (lado esquerdo) e 2022 (lado direito).....	166

página propositadamente em branco

ÍNDICE DE TABELAS

Tabela 1: Variáveis utilizadas, sua definição e abreviatura	34
Tabela 2 : Identificação da variável em estudo, tipo de falha e abreviatura da variável.....	50
Tabela 3 : Desenvolvimento de funções método SARIMA no Ficheiro SARIIMA_FUNCOES	51
Tabela 4 : Ficheiros Python para cada variável para aplicação do método SARIMA	51
Tabela 5 : Análise da quantidade de valores nulos do conjunto original de dados por variável.....	62
Tabela 6 : Parâmetros da rede do primeiro teste	73
Tabela 7 : Valores de MAE e MSE para o primeiro teste	73
Tabela 8 : Parâmetros da rede do segundo teste	74
Tabela 9 : Valores de MAE e MSE para o segundo teste.....	74
Tabela 10 : Parâmetros da rede do terceiro teste	75
Tabela 11 : Valores de MAE e MSE para o terceiro teste.....	75
Tabela 12 : Parâmetros da rede do quarto teste	76
Tabela 13 : Valores de MAE e MSE para o quarto teste	77
Tabela 14 : Parâmetros da rede do quinto teste	78
Tabela 15 : Valores de MAE e MSE para o quinto teste.....	78
Tabela 16 : Alteração do parâmetro batch_size para 64	79
Tabela 17 : Alteração do parâmetro batch_size para 32	79
Tabela 18 : Alteração do parâmetro epochs para 50.....	80
Tabela 19 : Parâmetros modelo SARIMA e valor de AIC.....	83
Tabela 20 : Valores das previsões para o conjunto de teste para a variável T_MEDIA [°C]	84
Tabela 21: Valores de MAE e RMSE para modelo SARIMA para a variável T_MEDIA [°C].....	85
Tabela 22 : Valores previsões futuras com intervalos de confiança para a variável T_MEDIA [°C].	86
Tabela 23 : Parâmetros modelo SARIMA e valor de AIC.....	89
Tabela 24 : Valores das previsões para o conjunto de teste para a variável IRR_MEDIA [Wh/m ²].	90
Tabela 25 :Valores de MAE e RMSE para modelo SARIMA para a variável IRR_MEDIA [Wh/m ²]...	91
Tabela 26 : Valores previsões futuras com intervalos de confiança para a variável IRR_MEDIA [Wh/m ²]	92
Tabela 27 : Parâmetros modelo SARIMA e valor de AIC.....	95
Tabela 28 : Valores das previsões para o conjunto de teste para a variável P_DC_INV_A [W].....	96
Tabela 29 : Valores de MAE e RMSE para modelo SARIMA para a variável P_DC_INV_A [W].....	97
Tabela 30 : Valores previsões futuras com intervalos de confiança para a variável P_DC_INV_A [W]	98
Tabela 31 : Parâmetros modelo SARIMA e valor de AIC.....	101
Tabela 32 : Valores das previsões para o conjunto de teste para a variável P_AC_INV_A [W].....	102
Tabela 33 : Valores de MAE e RMSE para modelo SARIMA para a variável P_AC_INV_A [W]	103
Tabela 34 : Valores previsões futuras com intervalos de confiança para a variável P_AC_INV_A [W]	104
Tabela 35 : Parâmetros modelo SARIMA e valor de AIC.....	107
Tabela 36 : Valores das previsões para o conjunto de teste para a variável T_INV_A [°C]	108
Tabela 37 : Valores de MAE e RMSE para modelo SARIMA para a T_INV_A [°C]	109
Tabela 38 : Valores previsões futuras com intervalos de confiança para a variável T_INV_A [°C]	109
Tabela 39 : Parâmetros modelo SARIMA e valor de AIC.....	112

Tabela 40 : Valores das previsões para o conjunto de teste para a variável P_DC_INV_B [W].....	114
Tabela 41 : Valores de MAE e RMSE para modelo SARIMA para a variável P_DC_INV_B [W]	114
Tabela 42 : Valores previsões futuras com intervalos de confiança para a variável P_DC_INV_B [W]	115
Tabela 43 : Parâmetros modelo SARIMA e valor de AIC.....	117
Tabela 44 : Valores das previsões para o conjunto de teste para a variável P_AC_INV_B [W].....	119
Tabela 45 : Valores de MAE e RMSE para modelo SARIMA para a variável P_AC_INV_B [W]	119
Tabela 46 : Valores previsões futuras com intervalos de confiança para a variável P_AC_INV_B [W]	120
Tabela 47 : Parâmetros modelo SARIMA e valor de AIC.....	123
Tabela 48 : Valores das previsões para o conjunto de teste para a variável T_INV_B [°C]	124
Tabela 49 : Valores de MAE e RMSE para modelo SARIMA para a variável T_INV_B [°C].....	124
Tabela 50 : Valores previsões futuras com intervalos de confiança para a variável T_INV_B [°C]	125
Tabela 51: Comparação do valor de MAE entre os dois algoritmos para a variável relativa à temperatura do inversor A.....	126

página propositadamente em branco

LISTAS DE SIGLAS E SÍMBOLOS

Lista de Siglas

ISEP	Instituto Superior de Engenharia do Porto
P.Porto	Instituto Politécnico do Porto
UPAC	Unidade de produção de autoconsumo
ADF	Augmented Dickey-Fuller
AIC	Akaike information criterion
BIC	Bayesian information criterion
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
DC	Corrente Contínua
AC	Corrente Alternada
IDE	Integrated Development Environment
RN	Rede Neuronal
MLP	Multi-Layer Perceptron
ARIMA	AutoRegressive Integrated Moving Average
PACF	Partial Auto-Correlation Function
ACF	Auto-Correlation Function
HIC	Hannan-Quinn Information Criterion
NMBE	Normalized Mean Bias Error
MBE	Mean Bias Error

página propositadamente em branco

1. INTRODUÇÃO

No presente capítulo é descrito de uma forma geral o contexto de desenvolvimento do projeto. De seguida, é abordado o enquadramento do tema bem como o problema de investigação, são também enunciados os objetivos que se pretendem alcançar no presente projeto e por fim é abordado a metodologia utilizada.

1.1. Enquadramento e pertinência

A Presente dissertação foi desenvolvida em regime empresarial, em particular, na SENS – Iqony Solar Energy Solutions. Durante este projeto, foi concebida uma ferramenta abrangente de análise de dados e implementação de dois modelos de previsão. Esta ferramenta representa uma vantagem significativa para a empresa, uma vez que, até ao momento, não havia sido realizado um estudo dessa natureza. Para além disso, possibilita a obtenção de insights a partir da análise geral dos dados, facilita a deteção de falhas de comunicação por meio de alertas, desencadeando ações de manutenção preventiva. Além disso, esta abordagem ajuda a evitar despesas substanciais em equipamentos, bem como permite um monitoramento regular do desempenho desses equipamentos.

O principal objetivo de qualquer organização é otimizar a produção, garantindo maior quantidade, qualidade consistente e custos mínimos. No entanto, as falhas nos processos produtivos podem representar uma ameaça significativa a esses objetivos, muitas vezes resultando em despesas não planeadas. A gestão proativa dessas falhas tornou-se uma prioridade crescente, uma vez que permite a prevenção de gastos substanciais, tanto em termos de reparos de equipamentos quanto de interrupções na produção devido a falhas.

Atualmente, o desenvolvimento de previsões é amplamente utilizado em diversas áreas, desde a engenharia, à gestão. No caso da indústria, esta abordagem permite a análise de tendências, podemos obter informações preliminares sobre diversos processos, observações, reações e operações, e ainda, auxilia na tomada de decisões (Buchatskaya et al., 2015).

A indústria de produção de energia elétrica a partir de sistemas fotovoltaicos, é um mercado altamente competitivo, onde apenas pequenas perdas de eficiência traduzem-se em valores avultados nas margens de lucro (Gradwohl et al., 2021). Desta forma, é importante manter o desempenho em níveis ótimos para garantir os níveis de lucro expectáveis, através da implementação de métodos que permitem identificar falhas imediatas ou ocasionais, garantindo assim um funcionamento eficiente e confiável (Gradwohl et al., 2021).

Os autores Mellit et al., 2020 salientam a importância da previsão nos sistemas fotovoltaicos evidenciando que, ao longo do tempo muitos estudos abordam este tema, e vários métodos são identificados, em particular, métodos numéricos e probabilísticos, modelos físicos e técnicas de inteligência artificial (IA).

Assim sendo, fica evidente a importância dos métodos de previsão de falha nas mais diversas áreas, com especial interesse, na energia.

1.2. Questão e objetivos de investigação

O presente trabalho pretende responder às seguintes questões de investigação: *Como a manutenção preditiva pode ser integrada nuns sistemas de gestão de energia para melhorar a eficiência e a eficácia dos sistemas de painéis fotovoltaicos? Como a manutenção preditiva pode ser utilizada para reduzir os custos de manutenção dos painéis fotovoltaicos ao longo do tempo?*

Através do enquadramento do problema desenvolvido anteriormente, o principal objetivo do projeto é desenvolver modelos de previsão de valores futuros de determinadas variáveis com a finalidade de prever falhas nos equipamentos e perdas de eficiência num sistema fotovoltaico.

Como objetivos mais pormenorizados são enunciados os seguintes:

- Recolha e análise de dados;
- Análise e definição das variáveis importantes que podem indicar falhas e perdas de eficiência no sistema fotovoltaico;
- Análise através da literatura de modelos de previsão utilizados e perceber o mais adequado ao conjunto de dados e contexto do projeto para implementação;
- Construir e aplicar os modelos estipulados, inicialmente, com uma previsão para o conjunto de teste e posteriormente para datas futuras;
- Análise dos resultados por comparação dos valores de previsão com os valores reais do conjunto de teste;
- Análise dos valores de previsões futuras e comparação dos resultados dos modelos aplicados;
- Definir conclusões sobre os valores de previsão e mostrar como a ferramenta desenvolvida é importante para a organização;
- Estipular os desafios encontrados e possíveis melhorias futuras;

1.3. Opções metodológicas

De acordo com os objetivos estabelecidos anteriormente definiu-se que a metodologia mais apropriada seria *Action Research*.

De acordo com, (Kemmis, 2009), a metodologia *Action Research* tem por objetivo alterar os seguintes aspetos: a prática dos profissionais, compreensão da prática por parte dos profissionais que a implementam e as condições em que é implementada.

Existem várias formas de investigação associadas a esta metodologia (Kemmis, 2009).

Nesta abordagem os investigadores e participantes da situação cooperam entre si para a resolução de um problema ou ação (da Silva, 2011). Desta forma, (Ventura, 2012), afirma que deve existir um envolvimento ativo e ponderado no contexto da investigação do problema tanto por parte das pessoas envolvidas no problema como dos investigadores.

O objetivo do método *Action Research* é promover a ligação entre a ciência e a prática (Ventura, 2012). Esta metodologia possui uma característica única que envolve uma combinação de dois objetivos principais:

- Objetivo Prático: Este objetivo visa ajudar a resolver o problema em questão, fornecendo suporte aos envolvidos no processo para transformar a situação;

- **Objetivo de Conhecimento:** Aqui, o foco está em adquirir informações que seriam difíceis de obter por outros meios, a menos que houvesse interação direta com as pessoas envolvidas na situação investigada.

O método Action Research é atualmente muito utilizado nas mais diversas áreas, Educação, política e tecnologia, mas em particular, na área de sistemas de informação, inclusive em situações de apoio à decisão (Ventura, 2012).

1.4. Apresentação da empresa

O estágio foi desenvolvido na organização SENS – Iqony Solar Energy Solutions, com sede Würzburg (Baviera) e tem como principal objetivo fornecer serviços no setor das energias renováveis.

A empresa atua na construção de grandes parques solares, mantendo a operação contínua dos mesmos e ações de manutenção, e também oferece soluções abrangentes de eficiência energética adaptadas às necessidades de diversas empresas, com foco em aplicações industriais. Com presença em mais de 40 países e mais de 8000 instalações, a SENS possui décadas de experiência na geração de energia.

Além disso, a empresa acumula vários prêmios ao longo dos anos, incluindo o selo de melhores práticas em projetos de eficiência energética e o prêmio "Intersolar".

1.5. Estrutura do trabalho

O presente projeto encontra-se dividido em cinco capítulos principais. O presente capítulo é abordado o contexto geral do tema proposto, bem como o problema específico. Além disso, são indicados os objetivos que a pesquisa visa atingir, a opção metodológica implementada e é ainda desenvolvida uma pequena apresentação da empresa onde foi desenvolvido o projeto.

O capítulo 2, dedicado à revisão bibliográfica, tem como objetivo fornecer uma base teórica para os temas mais relevantes e significativos do projeto. Este capítulo desempenha um papel fundamental, pois fornece informações essenciais para a correta execução do projeto. É importante destacar que ao longo do desenvolvimento do projeto, este capítulo passou por modificações para garantir que todos os temas fossem incluídos de maneira informativa e precisa.

No capítulo 3 é apresentada a metodologia abordada e as técnicas aplicadas. Inicialmente é desenvolvida uma pequena apresentação da unidade de produção de autoconsumo analisada para o projeto, de seguida, são mencionadas todas as etapas importantes para o desenvolvimento do projeto, desde a análise exploratória dos dados à implementação dos métodos de previsão.

No capítulo 4 são mostrados os resultados das diferentes técnicas e metodologias apresentadas anteriormente. Em concreto, são evidenciados os resultados da análise exploratória desenvolvida, de seguida, são apresentados os valores obtidos de cada modelo de previsão implementada para as variáveis de interesse. No fim deste capítulo, é ainda desenvolvida uma discussão dos resultados.

Por fim, no capítulo 5, referente à conclusão do projeto, são evidenciados os principais desafios encontrados no desenvolvimento do projeto, bem como passos futuros de melhoria da ferramenta.

2. REVISÃO BIBLIOGRÁFICA

Com o objetivo de contextualizar a nível teórico o projeto, desenvolveu-se uma análise e pesquisa por temas relevantes.

O principal tema do projeto envolvia a análise de falhas em sistemas fotovoltaicos, o que, implica uma análise dos dados de forma consistente. E, portanto, o primeiro tema abordado foi a situação de *Big Data* com o desenvolvimento da análise de dados no âmbito da engenharia e gestão industrial.

Atualmente, a energia e a sustentabilidade energética assumem um papel de destaque, conduzindo à realização de pesquisas significativas nesses campos. Vale ressaltar a abordagem abrangente sobre sistemas fotovoltaicos, incluindo sua estrutura e funcionamento, como tópicos de relevância central no desenvolvimento do presente trabalho.

A manutenção desempenha um papel crucial no desenvolvimento eficiente e produtivo de uma organização. Para além disso, permite o diagnóstico de falhas atempadamente, sem inferir grandes prejuízos, portanto, foi também um dos temas abordados.

Por fim, para a construção de previsões, inicialmente, iniciou-se um estudo introdutório do tema, de seguida, de forma mais pormenorizada, abordou-se as redes neuronais, visto que seria um método amplamente praticado para o diagnóstico de falhas. No entanto, os resultados obtidos não foram satisfatórios, e por meio de uma análise exploratório dos dados e da literatura, aprofundou-se a análise de séries temporais. De seguida, destacou-se o modelo SARIMA, um método de previsão de análises temporais bastante utilizado tendo em conta as características dos dados. Por fim, foram abordadas metodologias de avaliação dos métodos de previsão.

2.1. *Big Data* e análise de dados na gestão industrial

Atualmente, a análise de dados é um tema crucial para as empresas desenvolverem vantagens competitivas. Efetivamente, a nova era industrial destaca-se pela inovação, quer em termos de mudança do modo de produção tradicional, como pelo desenvolvimento de técnicas de “reindustrialização”, nomeadamente a Indústria 4.0” e a “Internet industrial” (Yang & Ge, 2022).

Cada vez mais, é abordado o tema da análise dos dados industriais tanto a nível académico como industrial, visto que, se traduz numa fonte de informação necessária e importante para organizações se desenvolverem de uma forma eficaz. Desta maneira, a análise industrial de Big data é uma área em constante desenvolvimento, com grandes vantagens a serem trabalhadas (Yang & Ge, 2022).

Segundo (Addo-Tenkorang & Helo, 2016), *Big data* pode ser definido como um vasto conjunto de dados que não é possível explorar, adquirir, armazenar e analisar por sistemas de tecnologias de informação e software/hardware num limite admissível. Para além disso, de acordo com (Kwon et al., 2014), os profissionais deste ramo consideram que esta técnica de análise de dados será o próximo “oceano azul” na formação de oportunidades de negócio.

2.1.1. Características de Big data

A exploração e análise dos dados pelos métodos baseados em *Big data* é uma forma de auxiliar na toma de decisões bem como na garantia da qualidade do sistema. As características de *Big data* podem ser resumidas e definidas por 5 V (Addo-Tenkorang & Helo, 2016), nomeadamente, grande volume: grande quantidade de dados; alta velocidade: dados gerados e atualizados ao momento; alta diversidade: dados gerados por várias fontes e com diferentes formas; alta precisão e alto valor;

As características deste sistema são em tempo real, dinâmicas e adaptativas. Com a evolução do sistema de exploração e armazenamento de dados, as técnicas de tomada de decisão tendo por base a experiência do gestor são substituídas pela análise de *Big data* de forma a criar mais valor (Li et al., 2022).

2.1.2. Vantagens de Big data

Segundo (Li et al., 2022), a análise orientada por *Big data* confere várias vantagens para as organizações, nomeadamente, melhoria da competitividade do mercado da indústria, exploração dos dados de modo a extrair todo o conhecimento de valor para os gestores, e consequentemente, auxílio na tomada de decisões complexas. Efetivamente, informações sobre degradações nas máquinas, ineficiências nos recursos e processos são valorizadas e através de uma adequada análise de todo o conjunto de dados é possível obter informações para ajudar o gestor a desempenhar a sua função da melhor forma, quer seja para atuar na manutenção preventiva ou na gestão dos recursos (Wang et al., n.d.).

Desta forma, quando apoiada por outras tecnologias de ponta, esta técnica torna-se ainda mais competente.

Para além do que já foi mencionado, a tecnologia baseada por *Big data* permite a transformação da indústria tradicional como a conhecemos para a chamada “indústria inteligente” que tem como base a indústria 4.0. Desta forma, esta técnica tem se mostrado fundamental na criação de valor para a indústria (Li et al., 2022) .

Identificar os serviços e produtos que o cliente mais valoriza é uma informação muito importante quer seja para conseguir satisfazer as necessidades dos clientes que já existem, quer seja para aumentar o conjunto de consumidores. Desta forma, estudos evidenciam que a análise de *Big data* quando desenvolvida de forma adequada, para além de todas as vantagens enunciadas anteriormente, é capaz de diminuir os custos dos negócios, criar insights de negócio e demonstrar informações estratégicas para alcançar novos clientes (Kwon et al., 2014).

2.1.3. O impacto da análise de Big Data nas empresas

A utilização de análise de *Big data* numa empresa é influenciada pela capacidade das tecnologias de informação, quer seja em termos de gestão da qualidade dos dados, quer pela perceção dos benefícios da sua utilização. De uma forma geral, a análise de *Big data* nas empresas é fundamental e deve ser incorporada tanto na gestão da qualidade dos dados, como no desenvolvimento de novas oportunidades acompanhadas por novas tecnologias de informação. Por outro lado, a análise de *Big data* exige investimentos caros na sua implementação e são necessários recursos (humanos,

software, etc) para a sua aprendizagem. Assim, estas empresas podem tentar resistir às inovações tecnológicas deste tipo (Kwon et al., 2014).

A análise indica que as empresas que são bem-sucedidas na utilização de dados de fontes externas estão mais predispostas a explorar a análise de *Big data*.

2.1.4. Limitações de Big data

Como já mencionado, algumas características da análise de *Big data* são o volume de dados para explorar, a velocidade a que chegam estes dados e a sua diversidade. Estas propriedades por um lado traduzem-se em vantagens pela quantidade e precisão de informações que transmitem aos órgãos de gestão, por outro lado são necessários recursos tecnológicos e humanos bem desenvolvidos para trabalhar todos os dados.

Reunir e explorar dados desta magnitude não está ao alcance de muitas organizações, também pelo facto dos dados não serem bem preparados.

Desta forma, como é normal, segundo (Wang et al., n.d.) esta metodologia apresenta limitações:

- Falta de representação espaço temporal dos dados em larga escala: A aquisição de dados na indústria é desenvolvida por meio de vários dispositivos (sensores, por exemplo) que são colocados numa localização específica a uma data e hora correspondente. A localização e o espaço temporal dos dados são dimensões importantes para a posterior análise de dados.
- Falta de aprendizagem das máquinas online (*machine learning*) em larga escala de forma eficaz e eficiente: A análise de *Big data* industrial caracteriza-se pela diversidade dos dados, “ruído” dos mesmos, recursos não estruturados e muitas vezes, são necessárias respostas rápidas para que um problema não avance. Portanto, para além de aumentar os recursos tecnológicos para conseguir produzir respostas rápidas, é necessário aplicar algoritmos de *machine learning* em grande escala numa estrutura de análise de dados para desenvolver conhecimentos de uma forma eficaz e eficiente.
- Falta de gestão do ciclo de vida dos processos: Um grande problema da aplicação desta metodologia é que o sistema de armazenamento de dados não é capaz de admitir tantos dados como os que são gerados. E, portanto, uma análise da qualidade dos dados deve ser desenvolvida à priori de modo a identificar, os dados que oferecem informações de valor e os que devem ser armazenados dos que devem ser rejeitados.
- Reduzida visualização dos dados: Como a quantidade de dados a analisar é enorme, as informações desenvolvidas também são muitas e, portanto, deve existir uma boa representação das mesmas. Uma boa visualização dos conhecimentos, consegue expor correlações entre fatores, tornar a tomada de decisão mais imediata, expressar diferenças em informações difíceis de resumir.
- Falha na confidencialidade dos dados industriais: são necessários recursos humanos e ferramentas adicionais para conseguir analisar *Big data* de uma forma eficiente, visto que a capacidade dos serviços é limitada. De uma certa forma, a segurança dos dados fica comprometida. Portanto, medidas de prevenção para proteção dos dados devem ser delimitadas.

2.2. Panorama nacional da energia e sustentabilidade energética

Os recursos limitados do planeta é um tema crucial no mundo dos negócios levando as empresas a explorar formas de manter e/ou aumentar a sua eficiência de produção de forma sustentável. Portanto, os cargos de gestão das organizações cada vez mais tendem a optar por alternativas mais sustentáveis, como a energia solar, devido à sua viabilidade económica em comparação com a energia elétrica (Macedo et al., 2021).

Os combustíveis fósseis, como o carvão, o petróleo e o gás natural, compõem mais de 80% do consumo global de energia. O elevado consumo deste tipo de combustível deve-se principalmente aos avanços tecnológicos durante o século XX, que implicaram o aumento de consumo de energia. No entanto, esse consumo excessivo de combustíveis fósseis também acarreta vários problemas, nomeadamente, o esgotamento das reservas desses recursos, distribuição desigual das mesmas, causando problemas geopolíticos e a emissão de poluentes decorrente da queima desses combustíveis (Goldemberg, 2015).

2.2.1. Sustentabilidade energética

As fontes de energia podem ser renováveis e não renováveis. As fontes de energia renováveis têm origem em recursos naturais, como a água, vento e o sol, e podem ser renovadas, como exemplos destas energias temos, a energia solar, eólica e hidráulica. Por outro lado, as energias não renováveis, como o petróleo, o carvão e o gás natural têm origem na combustão de recursos fósseis, que demoram milhões de anos para se formar e são muito poluentes. A energia solar, em particular, destaca-se como uma opção viável, reduzindo o consumo de energia convencional em até 70%. São várias as vantagens, nomeadamente, é uma fonte inesgotável de energia, não poluente, exige pouca manutenção nas centrais e cada vez mais os painéis solares admitem custos menores e são mais eficientes. Por outro lado, como é normal também são identificadas algumas desvantagens, como o facto da produção de energia depender das condições atmosféricas, durante a noite não há produção de energia, a produção de energia depende da localização geográfica, (Ramalho et al., 2022) .

O termo sustentabilidade, é uma forma de resposta para todas as desvantagens enumeradas do consumo excessivo de combustíveis fósseis ao longo dos anos. A procura pela sustentabilidade assenta na utilização de energias renováveis em vez dos combustíveis fósseis (Goldemberg, 2015) .

No contexto empresarial, a sustentabilidade, pode ser definida como uma abordagem que equilibra os aspetos ambientais, sociais, e económicos, tendo por objetivo criar vantagens a longo prazo para as gerações futuras e partes interessadas. Para que as empresas sejam sustentáveis e competitivas nos dias de hoje, é imperativo que produzam mais e melhor, reduzindo a poluição, diminuindo o uso de recursos naturais e adotando uma postura socialmente responsável. A utilização de práticas sustentáveis nas organizações favorece as mesmas em vários pontos, nomeadamente, redução de custos, melhoria na qualidade e visibilidade do negócio, reconhecimento da organização e pode traduzir-se numa vantagem competitiva no mercado que a empresa opera, despesas de manutenção mais baixas, e conseqüentemente, os equipamentos admitem maior durabilidade, (Macedo et al., 2021) .

2.2.2. Panorama energético em Portugal

De seguida, é abordado o panorama energético de Portugal com base num relatório para o ano de 2019.

O estudo é desenvolvido por meio de comparação do consumo energético final nos anos de 1999 e 2019, de modo, a perceber a evolução do consumo energético do país (da Energia, 2021). Os resultados demonstram que o petróleo continua a ser a principal fonte de energia, no entanto tem vindo a diminuir a utilização deste combustível, ao contrário do peso da eletricidade, gás natural e energias renováveis que registaram um aumento. Todas estas conclusões mencionadas podem ser comprovadas pela Figura 1 que reflete o consumo total de energia final em 1999 e 2019.

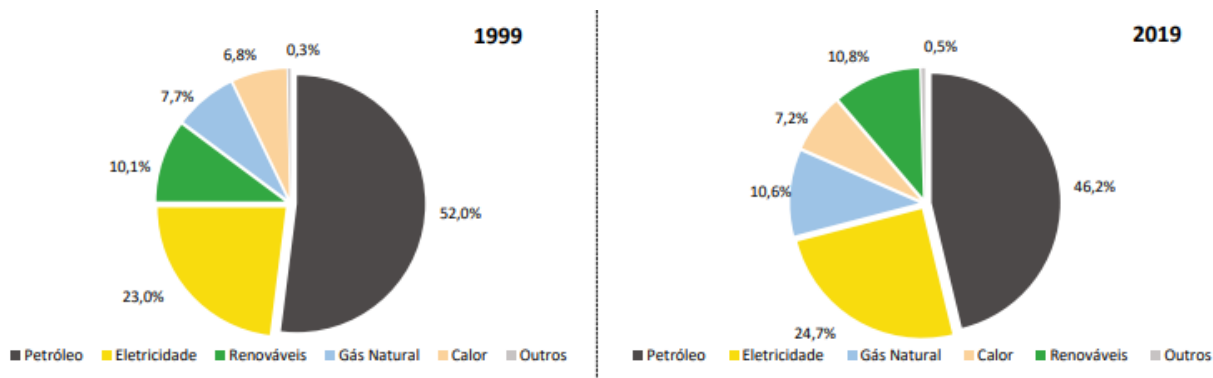


Figura 1 : Comparação de consumos energéticos em Portugal entre 1999 e 2019 (da Energia, 2021)

A aposta nas energias renováveis é evidente, os resultados transparecem isso mesmo. Para além disso, a dependência energética do país tem vindo a diminuir também.

É visível pela análise da Figura 2 que os setores que mais consomem energia são o setor do transporte e o da indústria. As percentagens de cada setor são semelhantes, no entanto, o setor da indústria revelou uma diminuição mais acentuada no consumo de energia entre 1999 e 2019.

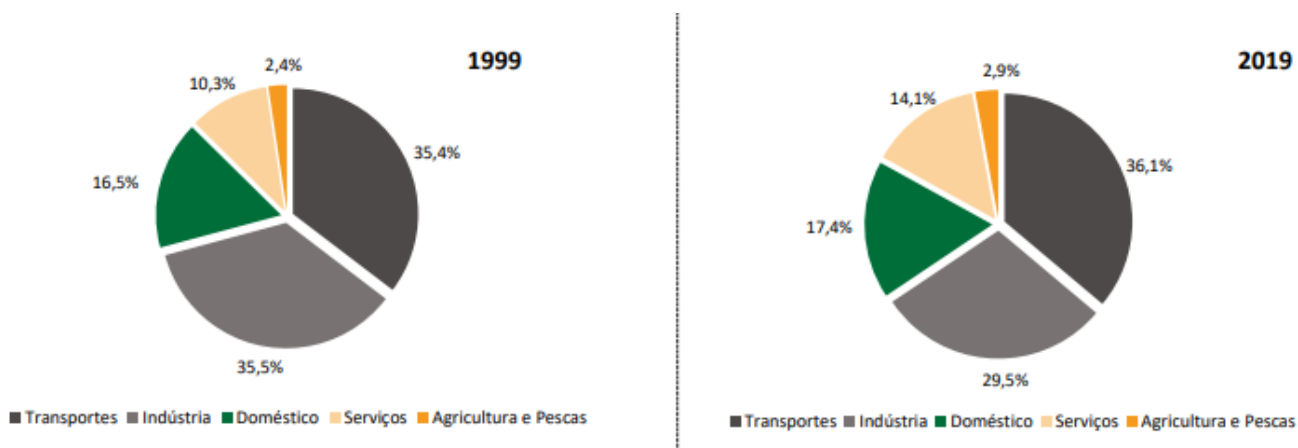


Figura 2: Consumo de energia final por setor de atividade em 1999 e 2019 (da Energia, 2021)

No mesmo estudo é revelado que em 2019, Portugal foi o 5º país com menor consumo de energia final por habitante em relação aos restantes países da união europeia, Figura 3 .

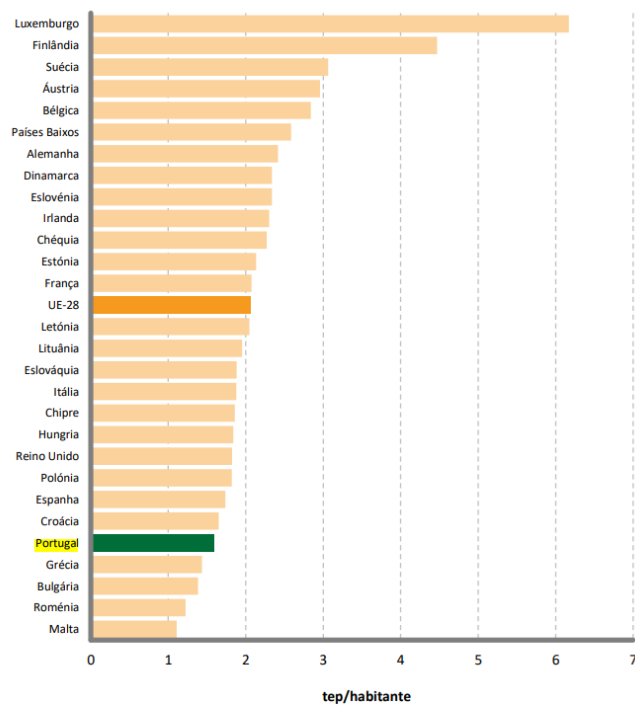


Figura 3: Consumo de energia final por habitante na união europeia em 2019 (da Energia, 2021)

Quanto à emissão de gases de efeito de estufa (GEE), em relação aos restantes países da união europeia, Portugal ocupa uma das posições mais baixas, cerca de 22% abaixo do valor médio registado na união europeia, Figura 4. De ressaltar que os dados da Figura 4 são em toneladas de dióxido de carbono equivalente por habitante, indica que a quantidade de emissões de GEE não é apenas de dióxido de carbono (CO₂), mas também inclui outros gases de efeito estufa.

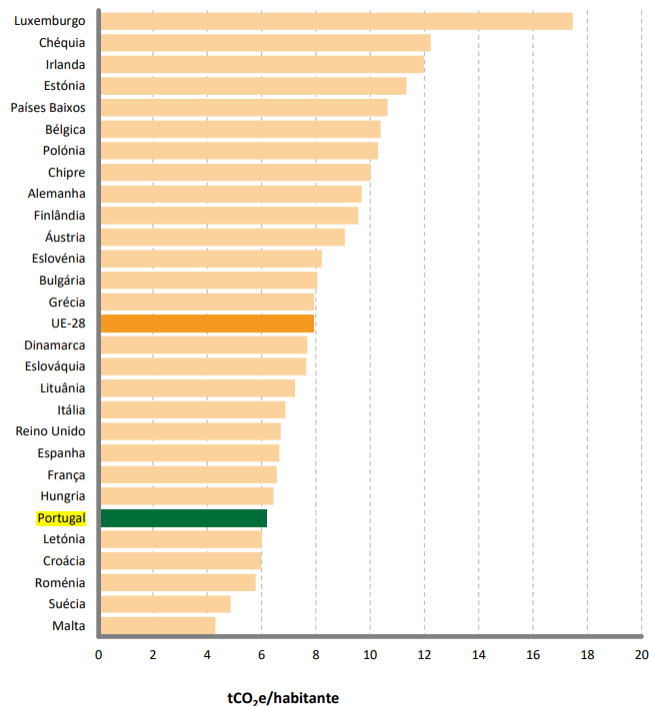


Figura 4: Emissões de GEE por habitante em 2019 na união europeia (da Energia, 2021)

2.3. Sistemas fotovoltaicos

À medida que procuramos avançar nas tecnologias de energia renovável, os sistemas fotovoltaicos estão em constante atualização e são amplamente utilizados (Almeida et al., n.d.). Também conhecidos como painéis solares, estes convertem a energia solar em eletricidade através de células fotovoltaicas. Essa tecnologia limpa e sustentável é amplamente utilizada para gerar eletricidade em residências, empresas e instalações industriais, contribuindo para a redução das emissões de carbono e a transição para uma energia mais verde.

2.3.1. Tipos de sistemas fotovoltaicos

De acordo com, (Francisco et al., n.d.) e (Moreira, 2016), os sistemas fotovoltaicos podem ser classificados em dois grupos principais, de acordo com o seu funcionamento:

- Com ligação à rede elétrica de serviço público (RESP):
 - UPAC (Unidade de Produção para Autoconsumo);
 - UPP (Unidade de Pequena Produção);
- Sem ligação à rede elétrica de serviço público (RESP):
 - Sistemas isolados/autónomos;

Dentro dos sistemas fotovoltaicos vinculados à rede elétrica de serviço público (RESP), há duas categorias específicas, dependendo de como o excedente de energia produzida é tratado em

relação à rede elétrica. Isso afeta a forma como esses sistemas são remunerados: o excesso de energia produzida pode ser injetado na rede elétrica ou o excesso de energia produzida não é injetado na rede elétrica.

Sistemas com ligação à rede elétrica de serviço público

Tendo por base, (Francisco et al., n.d.) e (Moreira, 2016), vão ser explicadas com mais algum detalhe cada tipo de sistema.

Uma UPAC, engloba uma ou mais unidades de geração de eletricidade que têm como objetivo principal atender às necessidades energéticas de uma instalação específica. O foco principal é reduzir a necessidade de comprar energia da Rede Elétrica de Serviço Público (RESP), tornando-se uma opção extremamente vantajosa para o consumidor. O cenário ideal seria gerar energia suficiente para o consumo próprio necessário. Nesse contexto, a energia gerada pela UPAC é destinada prioritariamente ao consumo interno da instalação em questão. Qualquer excesso de energia produzida além das necessidades da instalação pode ser vendido e injetado na Rede Elétrica de Serviço Público.

As UPAC's podem ser equipadas por equipamentos de armazenamento de energia, para situações de excedente. O que acontece na realidade, é que existem momentos em que a energia produzida é superior à necessária para consumo, e neste caso, existem duas hipóteses, vender e injetar o excedente na RESP, ou então, armazenar em baterias para uso posterior. Por outro lado, há momentos em que a produção de energia não é suficiente para atender ao consumo, sendo necessário adquirir eletricidade da rede elétrica de serviço público.

- UPAC sem sistema de armazenamento: Neste caso, as UPAC's não admitem baterias como sistema de armazenamento. E assim, toda a energia produzida é utilizada no momento e caso exista excedentes é injetada na RESP. Neste momento, dado o elevado custo de aquisição de baterias é a decisão mais rentável.
- UPAC com sistema de armazenamento: A principal vantagem deste sistema é que a energia pode ser utilizada posteriormente, quando não é possível de gerar energia suficiente para o consumo da própria instalação. Para além disso, diminui a necessidade de compra de energia à rede. Por outro lado, é um sistema bastante caro. Na Figura 5 está representada de forma direta o funcionamento principal de uma UPAC.

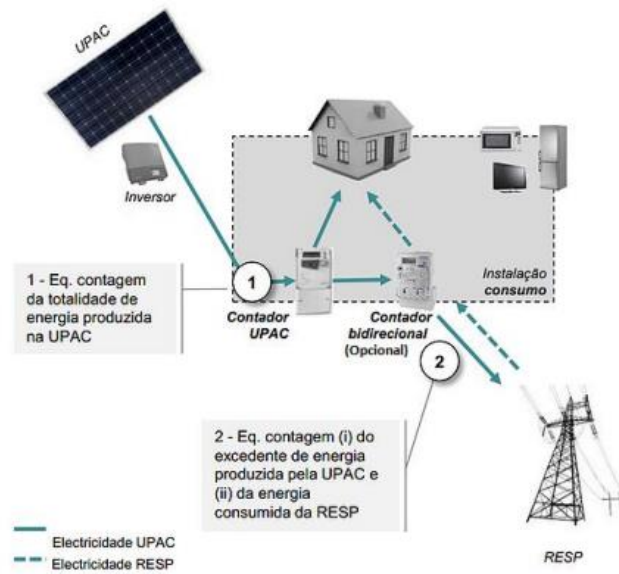


Figura 5 : Funcionamento de uma UPAC (Moreira, 2016)

- UPP – Unidade de Pequena Produção: Toda a energia produzida é vendida e injetada na RESP, a uma taxa fixada em leilão ou no mercado de eletricidade. Este modelo. Tem vindo a perder interesse visto, que as tarifas, têm vindo a descer. Na Figura 6 é apresentado o modo de funcionamento de uma UPP.

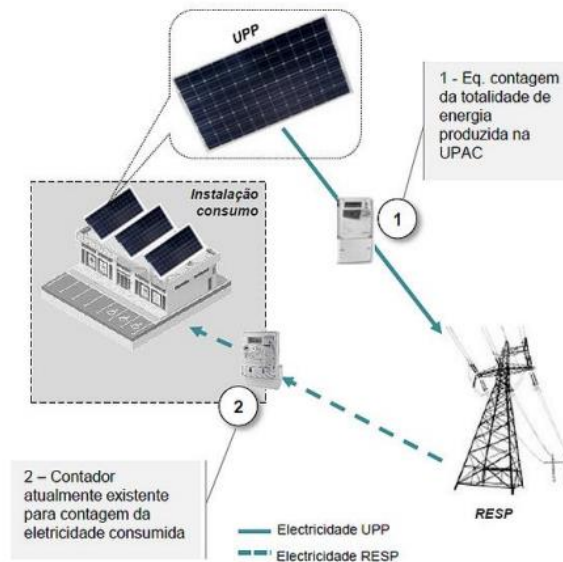


Figura 6: Modo de funcionamento de uma UPP (Moreira, 2016)

Sistemas sem ligação à rede elétrica de serviço público

- Sistemas isolados: Nestes sistemas, a energia produzida é consumida instantaneamente e localmente na instalação de utilização, Figura 7. Pode ou não ser necessário a instalação de um sistema de armazenamento de energia, de acordo com as necessidades energéticas do consumidor.

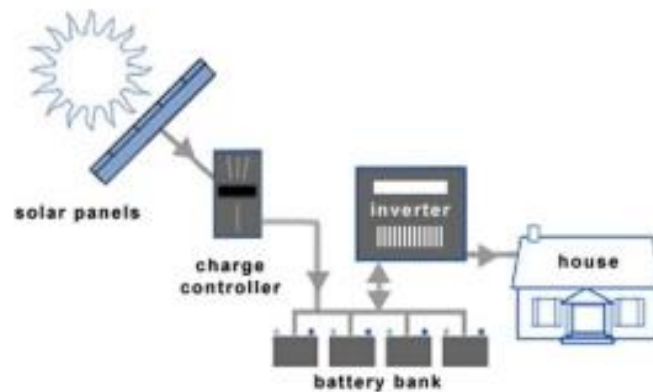


Figura 7 : Modo de funcionamento de um sistema isolado (Francisco et al., n.d.)

2.3.2. Constituição de um sistema fotovoltaico

De acordo com (Francisco et al., n.d.), um sistema fotovoltaico destinado a autoconsumo é constituído por:

- Módulos ou painéis fotovoltaicos: desempenham o papel central na produção de energia elétrica, pois convertem a energia proveniente do sol em energia elétrica.
- Estruturas de suporte.
- Inversores: responsáveis por converter a corrente contínua gerada em corrente alternada para ser possível utilizar.
- Cablagem e estruturas de proteção.
- Contador bidirecional: este equipamento não é obrigatório (Moreira, 2016) e tem como função contabilizar a energia trocada com a rede.
- Contador da energia produzida pelo sistema fotovoltaico da UPAC.
- Sistema de armazenamento (Baterias) para armazenar energia excedente. Consequentemente, será necessário a instalação de um regulador de carga (preservar o máximo da vida útil das baterias) e ainda, um controlador para gerir o armazenamento de acordo com a produção.

2.3.3. Funcionamento dos sistemas fotovoltaicos

Energia fotovoltaica é o termo usado para descrever o processo de captura de luz solar e a sua conversão em energia que podem ser utilizadas nas tarefas diárias (Moreira, 2016).

Desta forma, a energia elétrica é gerada diretamente a partir da luz solar, sendo transformada em eletricidade por meio de um fenômeno conhecido como "Efeito Fotovoltaico". De uma forma geral é apresentado o princípio de funcionamento de um sistema fotovoltaico.

O módulo ou painel fotovoltaico é constituído por uma matriz de células fotovoltaicas, ligadas entre si para obter os valores elétricos. As células fotovoltaicas, permitem a conversão direta da energia luminosa em energia elétrica, são constituídas por duas camadas de silício altamente purificados, (Moreira, 2016) . Segundo, (Miguel et al., 2023), um sistema fotovoltaico é constituído por vários módulos ligados em série, chamados de *strings*, as várias *strings* estão ligadas em paralelo entre si. A corrente elétrica gerada pelos módulos solares é convertida de corrente contínua (CC) em corrente alternada (CA) através do inversor. Desta forma, é transformada a energia solar captada pelos módulos em energia elétrica, pronta para ser utilizada. Na Figura 8 é apresentado uma configuração de um sistema fotovoltaico.

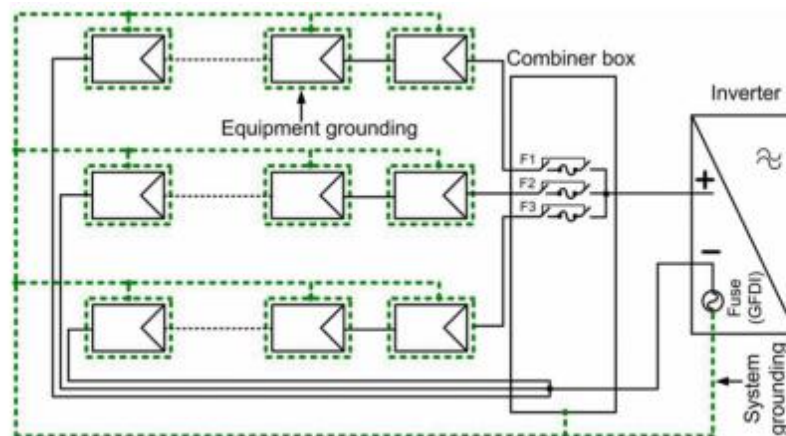


Figura 8 : Configuração de um sistema fotovoltaico (Miguel et al., 2023)

2.3.4. Origem e classificação das falhas nos sistemas fotovoltaicos

Como refere em (Lindig et al., 2018), “Uma falha é definida como a necessidade de substituir um componente FV, devido à sua falha final, econômica ou relacionada à segurança”.

É necessário controlar regularmente os sistemas fotovoltaicos para garantir o seu funcionamento adequado. Isso envolve ações preventivas para evitar e identificar possíveis falhas que possam causar uma diminuição na eficiência do sistema (Miguel et al., 2023).

Numa fase inicial, os fornecedores de painéis fotovoltaicos asseguram determinados níveis de rendimento, desta forma quando a potência de saída real do painel não coincide com os valores estabelecidos é sinónimo que existe alguma falha que está a impedir o funcionamento normal do dispositivo. Qualquer tipo de falha que ocorra num sistema fotovoltaico, que não é diagnosticada atempadamente, pode afetar seriamente a eficiência e segurança de todo o sistema fotovoltaico (Mellit et al., 2018). De acordo com, (Peinado Gonzalo et al., 2020a) , o desempenho dos módulos fotovoltaicos é influenciado pela luz recebida pelos painéis, estado das células e as ligações elétricas.

De uma forma geral, de acordo com, (Mellit et al., 2018) , a causa da falha pode ser a nível interno ou externo. Falhas internas estão relacionadas com os componentes do sistema, quanto às externas, são fatores externos ao sistema, como condições climáticas. O estudo realça também que as falhas podem ser classificadas como permanentes ou temporárias. Uma falha temporária pode ser resolvida dentro de um período específico, exemplificando, as sombras, acumular de poeira ou neve na superfície do painel (Madeti & Singh, 2017) . Por outro lado, as falhas permanentes só ficam resolvidas com a substituição do módulo responsável pela falha, nomeadamente, células queimadas, delaminação, superfície do painel com bolhas.

As falhas nos painéis fotovoltaicos podem ter origem no próprio painel ou conjunto de painéis (como a sujidade, degradação e corrosão) ou então, por ligações elétricas, curto-circuito (Abdulmawjood et al., 2018).

2.3.5. Exemplos de falhas

De seguida, com base na literatura são enunciadas as formas de falha mais evidentes e com mais ocorrência:

- Sujidade: (Peinado Gonzalo et al., 2020a), evidencia que um dos principais modos de falha é a sujidade e entenda-se por sujidade o acumular de poeira, pólen, entre outros constituintes do ar atmosférico. Locais desérticos são mais afetados por este tipo de falha, apesar de todos os inconvenientes que causa, é uma falha que facilmente pode ser resolvida, com uma limpeza adequada da superfície do painel.
- Acumulação de neve e gelo: A acumulação de neve ou gelo também condiciona o normal funcionamento e rendimento do painel, segundo (Peinado Gonzalo et al., 2020a), com bloqueio parcial ou por vezes total da luz solar, entrada de umidade para o sistema, verifica-se principalmente em países frios.
- Falha por sombreamento parcial e permanente: (Miguel et al., 2023) mostra que falhas de sombreamento parcial do sistema, podem ser provocadas pela passagem de nuvens ou então pela sombra das estruturas, árvores. Por outro lado, o sombreamento permanente pode provocar falhas por hot spots e conseqüentemente a degradação do módulo.
- Exposição ao ar: (Peinado Gonzalo et al., 2020a) A exposição ao ar livre também provoca danos no sistema, como a corrosão, delaminação e ainda degradação do sistema fotovoltaico. Os três modos de falha estão relacionados entre si, uma vez que, tanto a corrosão como a delaminação ocorrem pela entrada de umidade nos módulos, que potencializam o seu desgaste no longo prazo. De acordo com (Abdulmawjood et al., 2018), a delaminação resulta de falhas/fissuras entre as diversas camadas consequentes do módulo, que acabam por perder a aderência entre si. No estudo (Peinado Gonzalo et al., 2020a) evidencia que a delaminação potencia a entrada de umidade e facilita a corrosão do módulo, que se traduz em níveis de desempenho menores. Trata-se de efeitos irreversíveis e contribuem para a degradação dos módulos a longo prazo.
- Rachaduras: Quebras na estrutura do módulo fotovoltaico contribuem em perdas no desempenho do sistema se não forem verificadas logo no início (Peinado Gonzalo et al., 2020a).

- Falhas de ponto quente (hot spot): (Peinado Gonzalo et al., 2020a) e (Mellit et al., 2018) evidenciam este tipo de falha, como o aumento anormal da temperatura nas células fotovoltaicas, que pode ocorrer por diferentes motivos, nomeadamente, ao sombreamento parcial do módulo, pontos de solda fracos, sujidade acumulada, que causam a dissipação de energia na forma de calor em vez de eletricidade.
- Falhas de ligação à terra: Este tipo de falha pode ser considerado como curto-circuito elétrico inesperado entre os condutores de corrente do sistema fotovoltaico e o solo (Abdulmawjood et al., 2018) e (Mellit et al., 2018). É a falha mais comum nos sistemas fotovoltaicos e podem resultar em choques elétricos e risco de incêndio (Abdulmawjood et al., 2018). Por fim, esta falha pode acontecer por uma quebra ou falha no isolamento dos cabos devido a defeitos de fabrico ou superaquecimento e cabos envelhecidos.
- Falha linha a linha: Nos sistemas fotovoltaicos esta falha acontece quando existe um curto-circuito entre painéis solares ou cabos que admitem diferenciais elétricos diferentes (Mellit et al., 2018). (Abdulmawjood et al., 2018) mostra que este tipo de falha pode ocorrer entre módulos da mesma *string*, ou entre duas *strings* paralelas. A causa desta falha pode ser, mau isolamento dos cabos, curto-circuito acidental entre condutores de corrente (Mellit et al., 2018). Este tipo de falha está representado na Figura 9.

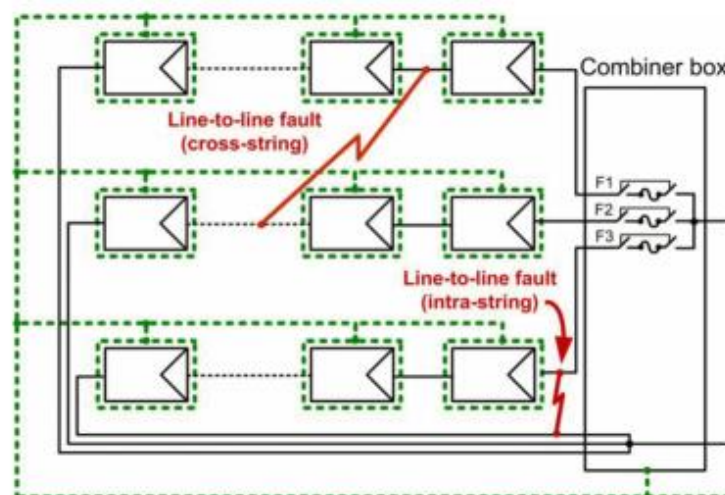


Figura 9 : Exemplo de uma falha entre linhas num sistema Fotovoltaico (Miguel et al., 2023)

Na Figura 10 é apresentado um esquema pormenorizado sobre os diferentes tipos de falha.

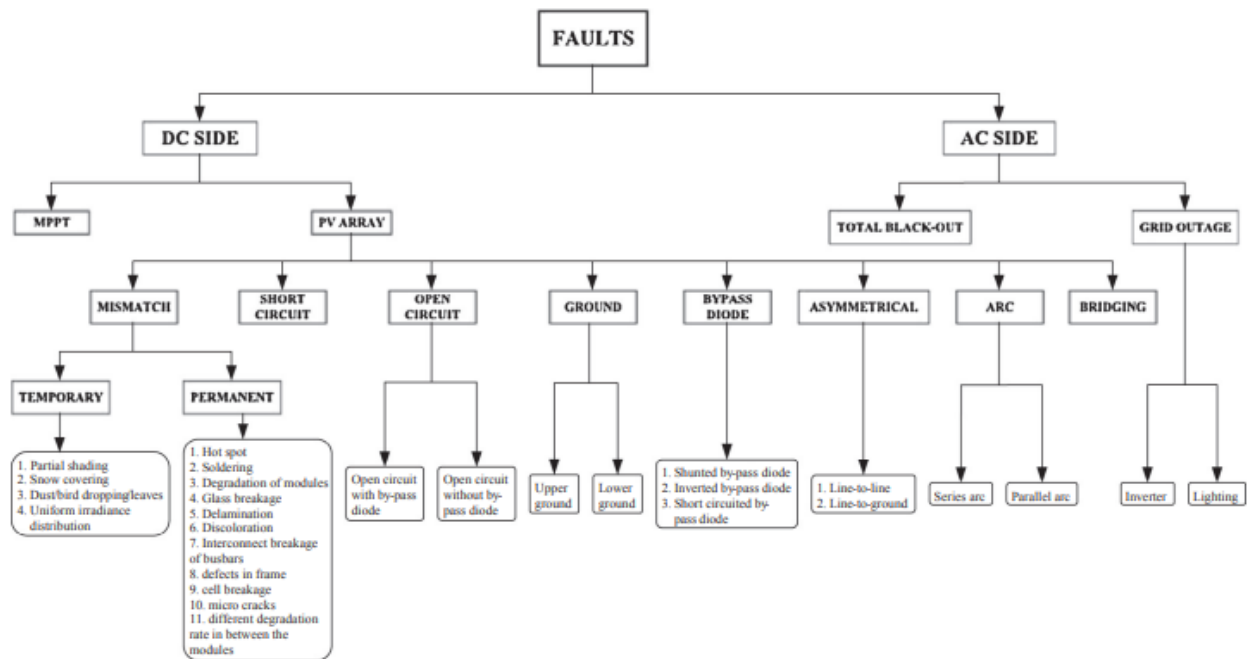


Figura 10 : Diferentes tipos de falha (Madeti & Singh, 2017)

2.3.6. Modelos de detecção e diagnóstico de falhas nos sistemas fotovoltaicos

A inovação tecnológica, os incentivos governamentais têm sido impulsionadores significativos do notável crescimento dos sistemas fotovoltaicos. No entanto, a imprevisibilidade e a variabilidade da energia solar são fatores que podem criar instabilidade e comprometerem a confiabilidade dos sistemas de energia. Desta forma, a previsão da produção fotovoltaica desempenha um papel fundamental para as empresas, permitindo a adaptação das operações das centrais fotovoltaicas e garantindo a estabilidade e eficiência do sistema, (Ramsami & Oree, 2015).

Os modelos de diagnóstico de falha auxiliam no cálculo da perda de desempenho dos sistemas fotovoltaicos durante as operações ao longo do tempo (Lindig et al., 2018).

O controlo do funcionamento dos sistemas fotovoltaicos, identificação de possíveis falhas podem ser desenvolvidos por meio de diferentes abordagens, em particular, verificação das mudanças de tensão nos inversores, comparar a produção do sistema fotovoltaico com outros sistemas futuros ou utilizar técnicas de *Machine Learning*, (Miguel et al., 2023).

(Mellit et al., 2018) é mais específico e distingue dois grupos principais para os métodos de detecção e diagnóstico de falhas, nomeadamente, métodos visuais e térmicos que são utilizados para detetar anomalias visuais (sujidade, descoloração, quebras), por outro lado, os métodos elétricos que podem ser utilizados para descobrir estragos nas *strings*, por exemplo.

2.3.7. Manutenção e monitorização de PF

De acordo com, (J. D. F. Alves, 2021), “A manutenção baseada na monitorização do estado é definida como um conjunto de tarefas a executar para a deteção de uma falha incipiente da máquina para evitar qualquer falha ou avaria importante que impeça a operação normal”.

Para se manter competitiva e conquistar novos mercados, a organização deve focar-se na qualidade e na produtividade. Assume um papel muito importante tanto na qualidade dos produtos desenvolvidos, como na produtividade da empresa (Marcorin et al., 2004).

Portanto, a manutenção deve ser entendida como uma estratégia para alcançar os resultados desejados e resolver problemas que possam surgir na produção. Tem a responsabilidade de apoiar a gestão da empresa e garantir que a produção funcione de forma eficiente, o que, por sua vez, coloca a empresa numa posição competitiva em termos de qualidade e produtividade. Os objetivos da empresa devem ser alinhados com a política de manutenção, por forma a diminuir os custos globais da organização. É essencial considerar a importância da função de manutenção e tomar decisões conscientes sobre a sua estratégia, considerando não apenas os custos isolados da aplicação de ações de manutenção, mas também o impacto positivo que uma manutenção eficaz pode ter no desempenho global da empresa (Marcorin et al., 2004).

Tipos de manutenção

A manutenção pode ser dividida em 3 tipos principais: corretiva, preventiva e preditiva, de uma forma direta, (Trojan & Marçal, 2017) , evidencia o conceito geral de cada um destes tipos de manutenção:

- Manutenção corretiva: resume-se a resolver os problemas depois de causarem perdas. Ou seja, este tipo de manutenção apenas ocorre após a falha do equipamento ou após a falha de desempenho do mesmo.
- Manutenção preditiva: É como se fosse uma forma avançada de manutenção preventiva, que tem em conta a condição atual do equipamento. A manutenção preditiva antecipa potenciais falhas do equipamento e auxilia a determinar o momento adequado para realizar intervenções de reparo. Envolve o controlo permanente do equipamento durante o seu normal funcionamento.
- Manutenção preventiva: O objetivo é prevenir as falhas nos equipamentos, ao contrário da manutenção corretiva. Neste caso, a manutenção realiza-se quando os equipamentos ainda estão a funcionar, mesmo que em condições mínimas. Pode acontecer em duas situações diferentes: é realizada a manutenção antes de ser necessário, e, portanto, o equipamento é desligado para manutenção mais cedo do que o devido, por outro lado, a segunda situação ocorre quando o equipamento falha devido a uma previsão incorreta do momento adequado para a manutenção.

Custos e manutenção

(Marcorin et al., 2004) , salienta que muitas vezes, a estratégia de manutenção passa para segundo plano, pois, as organizações analisam os custos da implementação desta estratégia de forma isolada, ou seja, apenas têm em conta dos custos imediatos de implementação da manutenção. No

entanto, a não manutenção dos equipamentos acarreta custos às empresas que muitas vezes passam despercebidos e não são contabilizados.

Em suma, os autores demonstram que a inexistência de uma política de manutenção eficaz provoca vários custos, nomeadamente, custos de mão de obra, materiais e os custos relacionados com a indisponibilidade do equipamento que são os mais significativos (perda de produção, aumento do tempo de produção, produtos de baixa qualidade, imagem da organização). Por fim, é relevante mencionar os custos que não conseguem ser contabilizados, relacionados com a imagem da empresa.

Ainda de salientar, que estudos realizados evidenciam que as implementações de ações de manutenção preventivas admitem uma relação custo-benefício mais favorável do que quando a manutenção não é abordada, culminando no descontrolo do processo produtivo.

Manutenção em sistemas fotovoltaicos

A instalação de sistemas fotovoltaicos tem vindo a aumentar, e, portanto, é importante focar na operação e na manutenção destes sistemas, para garantir que continuam a funcionar de maneira eficiente e produtiva ao longo da sua vida útil, que geralmente são de aproximadamente 25 anos. Efetivamente, a falta de manutenção num sistema fotovoltaico pode provocar: redução da produção de energia, superaquecimento das ligações, danos na estrutura de suporte das placas solares, entre muitos outros problemas. Os intervalos de manutenção também são um fator importante e depende de cada equipamento. É normal, que quanto mais frequente for a manutenção dos equipamentos, melhor será o acompanhamento dos equipamentos e, conseqüentemente, a probabilidade de prevenir as falhas será muito maior. Caso o controlo de um sistema fotovoltaico for constante e exigente, a produtividade mantém-se entre 90% a 100% da sua produtividade original, (Conrado, 2021).

A produção de energia e a eficiência de um sistema fotovoltaico são influenciadas pelas condições adversas a que estão sujeitos. Desta forma, a gestão de manutenção e mecanismos de prevenção de falha são temas cada vez mais essenciais para aumentar o desempenho, rentabilidade e eficiência destes sistemas (Peinado Gonzalo et al., 2020b).

Os proprietários de sistemas fotovoltaicos devem implementar estratégias de manutenção semestralmente. É crucial considerar tanto o tipo de manutenção quanto os intervalos de manutenção. Um estudo, mostra um caso de análise de 16 centrais fotovoltaicas, onde observou-se que a taxa de degradação de sistemas submetidos a manutenção preventiva foi inferior aos que foram sujeitos à manutenção corretiva. Estes dois fatores impactaram negativamente, a potência de saída dos sistemas e a vida útil dos sistemas (Aboagye et al., 2022).

De acordo com, (Haney & Burstein, 2013), analisar o local de forma pormenorizada é uma das abordagens mais eficazes de detetar problemas, para tal, deve ser realizada uma inspeção visual prática e abrangente dos elementos, periodicamente.

A maioria dos problemas em sistemas fotovoltaicos acontece nos inversores, de seguida nos módulos, por fim, nos restantes elementos do sistema (Conrado, 2021). Desta forma, a Figura 11 mostra a periodicidade de manutenção recomendada para cada elemento de um sistema fotovoltaico.

Periodicidade	Equipamento	Verificação
Diário	Inversor	Se está em operação ou se existe mensagem de erro
Mensal	Verificação da produção de energia	Registro das medições de energia e se estão de acordo com o esperado
	Superfície do painel fotovoltaico	Se há acúmulo de sujeira, como folhas ou dejetos de pássaros. Removê-los com água.
A cada 6 meses	Caixa de junção	Se há insetos. Se há corrosão/umidade excessiva. Testar a continuidade dos fusíveis.
	Proteção contra descargas atmosféricas	Se está operacional após tempestades com descargas
	Cablagem	Se há isolamentos danificados, marcas chamuscadas provenientes de arcos elétricos etc. Se as conexões estão boas.

Figura 11: Periodicidade de manutenção dos elementos de um sistema fotovoltaico (Conrado, 2021)

2.4. Métodos de previsão

Os métodos de previsão servem como metodologias de precaução para qualquer área, por exemplo, na economia para prever riscos de investimento, por outro lado, no caso da indústria, estes métodos são fundamentais para delimitar conclusões sobre os processos e operações. De salientar ainda, a importância destes métodos para prever o desenvolvimento socioeconómico da sociedade, bem como antecipar as consequências de determinadas decisões políticas, desta forma, a principal vantagem dos métodos de previsão, é apoiar a tomada de decisão. A literatura define um método de previsão como um conjunto de etapas teóricas e práticas com o objetivo de desenvolver previsões, que podem variar desde cálculos simples até procedimentos mais complexos, (Buchatskaya et al., 2015).

(Vieira da Silva & Carolina Nicolosi da Rocha Gracioso, 2021) estabelece que o principal objetivo da previsão é fornecer informações sobre a procura futura, e assim, projetar valores futuros, antecipar variações e facilitar o desenvolvimento de planos a longo prazo.

Pela análise dos artigos realizados por (Buchatskaya et al., 2015), (Gradwohl et al., 2021) e (Lindig et al., 2018), fica evidente que os métodos de previsão podem ser categorizados em duas abordagens principais: métodos físicos e métodos estatísticos.

Métodos físicos: envolvem a caracterização do sistema fotovoltaico em função de algumas variáveis independentes. Baseiam-se em teorias físicas de um modo de falha específico (Lindig et al., 2018).

Métodos estatísticos: são utilizados para prever parâmetros por meios de técnicas de previsão estatísticas, técnicas como análise de regressão, análise de séries temporais e inteligência artificial (técnicas de Machine Learning), (Buchatskaya et al., 2015), (Gradwohl et al., 2021). São necessários dados históricos para a previsão. A implementação de modelos estatísticos para prever determinados parâmetros tem sido uma etapa essencial na identificação de falhas em sistemas fotovoltaicos (Gradwohl et al., 2021).

2.4.1. Análise de séries temporais

(Latorre & Cardoso, 2001), define uma série temporal como uma sequência de dados, relativos a uma variável, em intervalos de tempo regulares durante um período específico.

A variável de controlo é o tempo, as séries temporais são ordenadas cronologicamente, no qual variar a ordem pode modificar a informação contida na série. O intervalo de tempo de recolha dos dados tem de ser regular, podem ser observações diárias, mensais, trimestrais, anuais, entre outros. Além disso, a principal característica que uma série deve possuir é ser estacionária, quer isto dizer que “além de ser estocástico, o processo deve estar em equilíbrio em relação a uma média e com variância constante”, (MARTIN et al., 2016).

No caso de, (Latorre & Cardoso, 2001), defende que as séries temporais podem ser compostas por três componentes: tendência, sazonalidade e a variação aleatória denominada de ruído branco. Já (MARTIN et al., 2016), admite que os principais de uma série temporal são: tendência, sazonalidade e ciclo. De seguida, é apresentada uma breve definição dos componentes destacados segundo os dois autores mencionados:

- Tendência: Os dados admitem um crescimento ou diminuição ao longo do tempo. Também é possível que a tendência mude de direção, passar de uma tendência crescente para decrescente.
- Sazonalidade: quando fatores sazonais influenciam a serie. Normalmente, um fenómeno sazonal ocorre regularmente em períodos fixo.
- Variação aleatória / ruído branco: são variações não explicáveis por tendência ou ciclos. Contribuem para aumentos ou diminuição de valores da série.
- Ciclo: as variações cíclicas diferem da sazonalidade, na medida em que ocorrem em períodos não conhecidos.

A previsão com base em séries tem como objetivo analisar e estudar valores históricos para criar e implementar modelos apropriados, capazes de caracterizar a estrutura e particularidades da série e prever valores futuros (P. Chen et al., 2018).

2.4.2. Redes neuronais

O sistema de construção de uma rede neuronal artificial é um processo inspirado nos neurónios interligados do sistema biológico. Uma rede neuronal é constituída por um conjunto de neurónios interligados, em que cada nó desenvolve um cálculo independente e estão conectados por ligações com determinados pesos que determinam o grau de conectividade entre estes. Durante a fase de treino da rede, de acordo com a apresentação dos dados, os pesos das ligações são ajustados (Furtado, 2019) e (Runge & Zmeureanu, 2019).

Na Figura 12 está representado a estrutura geral de uma rede neuronal.

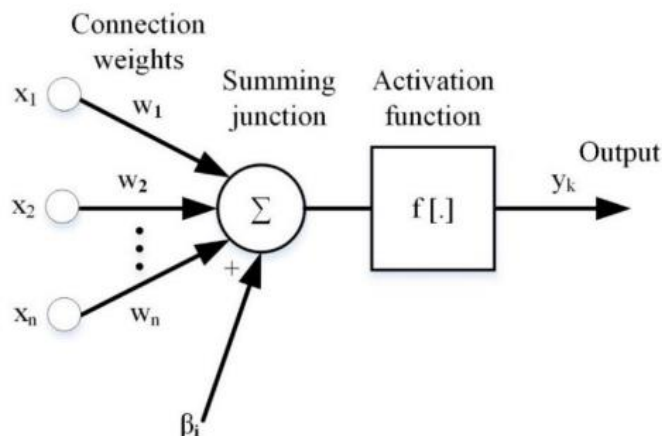


Figura 12 : Estrutura geral de uma rede neuronal (Runge & Zmeureanu, 2019)

Pela análise da Figura 12, os termos x_1, x_2, \dots, x_n correspondem aos valores de entrada; w_1, w_2, \dots, w_n são os pesos de cada ligação, β_i é o valor de polarização (bias); Σ corresponde à função de somatório, para realizar a soma da multiplicação dos neurónios de entrada pelos seus pesos, f corresponde à função de ativação e por fim, y_k são os valores de saída.

As redes neurais desempenham um papel proeminente no domínio da previsão, principalmente devido à sua capacidade específica de aprendizagem e aos resultados eficazes que podem alcançar em problemas complexos, permitindo que sejam identificados padrões complexos diretamente dos dados, sem a necessidade de um estudo prévio dos dados (M. F. Alves et al., 2014).

Situações onde os conjuntos de dados apresentam variações acentuadas, as redes neuronais demonstram ser uma escolha promissora, no entanto, vale a pena ressaltar que a dimensão do conjunto de dados pode representar um desafio para obter bons resultados de previsões (Mitrea et al., 2009).

Um aspeto relevante a ser considerado é as características dos dados, que exercem uma influência significativa no desempenho das técnicas de previsão. Efetivamente, uma análise comparativa entre técnicas estatísticas e a aplicação de redes neuronais na previsão de dados com tendência e/ou sazonalidade, conclui que as técnicas estatísticas evidenciam resultados mais eficientes. Nesse contexto, é interessante notar que diferentes tipos de redes neuronais possuem desempenhos distintos, como as redes LSTM apresentando bons resultados em relação a outro tipo de redes. Portanto, é possível concluir que para séries temporais com características de tendência e / ou sazonalidade as redes neuronais apontam determinadas limitações e as técnicas estatísticas sobressaem-se. Por fim, um último apontamento referido pelo mesmo autor é a dimensão dos conjuntos de dados das séries que também exerce um impacto direto nos resultados obtidos, para obter melhores resultados são necessários conjuntos de dados robustos (Moreno, 2021).

(Data Science Academy, 2022) e (Sharkawy, 2020) identificam as principais arquiteturas de redes neuronais:

- Redes *multilayer perceptrons* (MLP): são frequentemente utilizados para problemas de aprendizagem supervisionados. (Ferreira et al., 2012) e (Data Science Academy, 2022) realçam, que este é o modelo utilizado mais frequentemente.

- Redes neurais convulsionais (CNN): estes tipos de redes são mais utilizados para processamentos de imagem e visão computacional.
- Redes neurais recorrentes (RNN): são uma poderosa ferramenta para o processamento de dados sequenciais. Os LSTM surgiram como uma solução para os defeitos das redes neurais recorrentes, é o mais indicado para análise de séries temporais.
- Rede de função de base radial: este tipo de redes utiliza funções de base radial como funções de ativação. São utilizadas para problemas de classificação, previsões de séries temporais.

2.4.3. SARIMA

De acordo com (P. Chen et al., 2018), o modelo de previsão de séries temporais mais utilizado e reconhecido é o Autoregressive Integrated Moving Average (ARIMA).

De uma forma muito breve, o modelo ARIMA, pode ser dividido em 3 termos, autoregressão pelo AR, média móvel pelo MA e ainda admite a possibilidade de converter as séries temporais não estacionárias em estacionárias pelo termo I (Musbah & El-Hawary, n.d.).

Desta forma, quando a série sofrer pelo menos uma diferenciação para se tornar estacionária e combinada com os modelos AR e MA, é gerado um modelo ARIMA não sazonal. O método ARIMA apresenta melhores resultados para dados estacionários, ou seja, dados que admitem um padrão consistente ao longo do tempo (Aditya Satrio et al., 2021). Para ter em conta o efeito sazonal na série temporal, utiliza-se uma variação do modelo ARIMA clássico, o modelo ARIMA sazonal, denominado por SARIMA (Kardakos et al., 2013). De acordo com a literatura, inicialmente, os modelos de previsão ARIMA e SARIMA foram desenvolvidos, pelos engenheiros *Norbert Wiener*, posteriormente, *Box e Jenkins* modificaram-no (Rabbani et al., 2021). Em 1970, *Box e Jenkins*, propuseram uma nova abordagem capaz de prever diversos tipos de séries temporais (Musbah & El-Hawary, n.d.).

De uma forma geral, com a adição dos termos sazonais ao modelo ARIMA, obtemos o SARIMA. Desta forma, os modelos SARIMA podem ser escritos desta forma (Equação 1):

$$SARIMA(p, d, q)(P, D, Q)_m \quad \text{Equação 1}$$

Desta forma, pela equação (1) é possível comprovar que a metodologia SARIMA é constituída pelos termos p , d e q representam a parte não sazonal e P , D , Q representam a parte sazonal do modelo (P. Chen et al., 2018). Em particular:

- p e P representam as ordens dos parâmetros autorregressivos (AR) não sazonais e sazonais respetivamente;
- q e Q representam as ordens dos parâmetros de média móvel (MA) não sazonal e sazonal respetivamente;
- d e D representam os valores de diferenciação, na parte não sazonal e sazonal, respetivamente;
- M define o período de sazonalidade (K. Y. Chen & Wang, 2007).

A literatura é capaz de indicar valores ou intervalos de valores adequados para determinados parâmetros que seguidamente devem ser ajustados. De seguida, serão apresentadas as equações matemáticas que representam a modelo matemático do SARIMA, tendo por base (Narasimha Murthy et al., 2018), (Deretić et al., 2022).

Assim, os polinómios autoregressivos (AR) não sazonais e sazonais, de ordem p e P , respetivamente, são representados pela Equação 2 e Equação 3:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p (B)^p \quad \text{Equação 2}$$

$$\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P (B)^{PS} \quad \text{Equação 3}$$

Os polinómios que representam a média móvel não sazonais e sazonais, de ordem q e Q , respetivamente, podem ser demonstrados pela Equação 4 e Equação 5:

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad \text{Equação 4}$$

$$\Theta_Q(B^S) = 1 + \Theta_1 B^S + \Theta_2 B^{2S} + \dots + \Theta_Q B^{QS} \quad \text{Equação 5}$$

O operador de diferenciação não sazonal de ordem d , utilizado para eliminar tendências polinomiais, está representado na Equação 6:

$$Z(1 - B)^d \quad \text{Equação 6}$$

Da mesma forma, através da Equação 7 pode ser representado o operador de diferenciação sazonal de ordem D , utilizado para eliminar padrões sazonais.

$$(1 - B^S)^D \quad \text{Equação 7}$$

Por último, de salientar o parâmetro B que representa o desfasamento. O efeito deste parâmetro numa série sazonal Y_t é demonstrado na Equação 8:

$$B^d Y_t = Y_{t-d} \quad \text{Equação 8}$$

Por fim, a forma geral do modelo *SARIMA* $(p, d, q)(P, D, Q)m$ para uma série Y_t é evidenciada na Equação 9:

$$\phi_p(B)\Phi_P(B^S)(1 - B)^d(1 - B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad \text{Equação 9}$$

no qual, o termo s representa o período de sazonalidade da série e ε_t é a sequência de ruído branco. Através de uma vasta pesquisa na literatura (Musbah & El-Hawary, n.d.) , (Kardakos et al., 2013) ,(Bouzerdoum et al., 2013), foi possível constatar que o desenvolvimento do modelo *SARIMA* é frequentemente baseado no método iterativo de *Box e Jenkins*. Este método é constituído por quatro etapas principais, em particular, a identificação do modelo, atribuir os parâmetros do modelo, avaliação do mesmo e por fim a previsão dos valores futuros (Bouzerdoum et al., 2013) .

Método de previsão de Box & Jenkins

Efetivamente, a literatura retrata que este é um dos métodos mais utilizado e aplicado na análise de previsão de séries temporais (Albertt Santos Carvalho, 2020) , na Figura 13 é apresentado um fluxograma com as principais etapas de desenvolvimento do método.

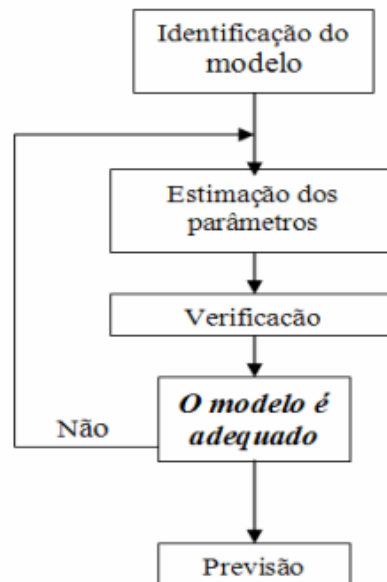


Figura 13 : Principais etapas de desenvolvimento do método de Box & Jenkins (Kirchner, 2006)

A primeira fase consiste em identificar a variável a estudar, e, simultaneamente, determinar o conjunto de dados que deve servir de input para o modelo. Para além disso, é analisada a estacionariedade da série temporal, caso a série não seja estacionária deve ser transformada (Bouzerdoum et al., 2013) . Com recurso a ferramentas como as Funções de Autocorrelação (ACF) e Autocorrelação Parcial (PACF) é possível, determinar valores de hipótese para os parâmetros p e q do modelo e assim identificar o modelo que melhor se adapta aos dados (Albertt Santos Carvalho, 2020).

De seguida, na segunda fase, são estudados os modelos “hipótese” estabelecidos na primeira fase e seleciona-se o mais eficiente (Bouzerdoum et al., 2013) . Para esta fase, o cálculo dos parâmetros pode ser desenvolvido com recursos a diferentes métodos e critérios, em particular, (Albertt Santos Carvalho, 2020) , realça o método da máxima verosimilhança, mas outros critérios como o, critério de informação de Akaike (AIC), critério de informação de Hannon-Quinn (HIC) e também o critério de informação Bayesiano (BIC), são válidos para examinar a complexidade do modelo, bem como a qualidade de ajuste dos parâmetros aos dados da amostra.

Segue-se a fase de diagnóstico do modelo na penúltima etapa. Efetivamente, após identificação do modelo e cálculo dos parâmetros, é necessário verificar se o modelo está bem ajustado à série temporal (Albertt Santos Carvalho, 2020). Nesta etapa o modelo é escolhido tendo por base dois fatores, as características dos resíduos e os parâmetros estimados. De salientar que o resíduo é a diferença entre os valores observados e os valores previstos pelo modelo. Desta forma,

(Bouzerdoum et al., 2013) (Albertt Santos Carvalho, 2020), com o objetivo de corroborar a qualidade do ajuste do modelo é necessário que os resíduos apresentem propriedades de ruído branco, ou seja, devem apresentar média zero, variância constante, devem ser independentes entre si (não existir qualquer padrão de autocorrelação) e assumir uma distribuição semelhante à distribuição normal. Nesta fase podem ainda ser implementadas melhorias (Bouzerdoum et al., 2013), (Musbah & El-Hawary, n.d.).

Por fim, nesta última fase é possível desenvolver as previsões e avaliar o modelo. Existem vários métodos para comparar modelos bem como para avaliar a qualidade das previsões, nomeadamente, calcular o Erro Quadrático Médio (*Root Mean Squared Error* – RMSE) através da diferença quadrática entre os valores reais e os previstos (Campana BARBOSA et al., 2015). Para além disso, (Bouzerdoum et al., 2013), evidencia outros métodos, o cálculo do coeficiente de correlação, cálculo do erro de viés normalizado (*Normalized Mean Bias Error* - NMBE). Por fim, é possível avaliar a qualidade das previsões de forma gráfica, através da comparação dos valores do conjunto de teste (reais) com os valores previstos pelo modelo, esta forma pode servir de suporte para os outros métodos já mencionados (P. Chen et al., 2018).

Características dos dados

Segundo (Rabbani et al., 2021), a estacionariedade da série é um requisito para que o SARIMA seja capaz de descrever previsões de qualidade. Dados cuja média, variância e autocorrelação permanecem constantes ao longo do tempo são denominados por dados estacionários. Desta forma, métricas como a diferenciação devem ser implementadas caso necessário para tornar a série temporal estacionária.

No caso do modelo SARIMA, apesar de ter sido muito bem-sucedido nas últimas décadas, quer em pesquisas académicas quer em aplicações reais a nível industrial, este método apresenta uma limitação, a sua forma linear pré-assumida (K. Y. Chen & Wang, 2007).

É importante realçar que quanto maior for o histórico de dados, melhor é o ajuste do modelo aos dados, e conseqüentemente são desenvolvidas previsões com maior qualidade.

Estacionariedade

A primeira etapa de construção do modelo SARIMA é analisar os dados quanto à estacionariedade. Como já foi referido, para a série ser estacionária, os dados devem apresentar um aspeto constante ao longo do tempo, a nível de média e desvio padrão (Rabbani et al., 2021). Desta forma, os gráficos da média móvel e do desvio padrão são uma forma de identificar tendências e padrões sazonais da série, o que auxilia no estudo da estacionariedade, apesar de não ser um teste formal (Brendan Artley, 2022).

Durante o estudo na literatura, é possível afirmar que os testes mais relevantes e aplicados para o estudo da estacionariedade da série são o teste de Dickey-Fuller e Kwiatkowski–Phillips–Schmidt–Shin (KPSS) (Rabbani et al., 2021), (Musbah & El-Hawary, n.d.).

Teste de *Dickey-Fuller* Aumentado

O teste de Dickey-Fuller é uma ferramenta clássica para avaliar a estacionariedade de séries temporais. Para tornar o teste de Dickey-Fuller mais robusto, a versão original foi melhorada e a nova versão ficou denominada por teste de Dickey-Fuller aumentado (ADF). Esta nova versão inclui novas “desfasagens”, quer isto dizer, novos parâmetros, permitindo assim, que o teste seja capaz de lidar com séries temporais de comportamentos mais complexos, e possa ser aplicado a outros tipos de modelos de séries temporais (Campana BARBOSA et al., 2015).

De acordo com, (Rabbani et al., 2021), o teste ADF trata-se de um teste estatístico para estudar a estacionariedade da série, pela existência ou não de uma raiz unitária. O objetivo é testar a hipótese de $p=1$, por isso o nome de raiz unitária e detetar tendências na série temporal. A presença de uma raiz unitária implica que a série não é estacionária (Lopes, 2001) .

Para aplicar o teste de Dickey-Fuller Aumentado, inicialmente, é necessário definir um nível de significância antes do teste e as conclusões sobre as hipóteses vão de acordo com p_value associado ao teste. A hipótese nula do teste ADF afirma que a série não é estacionária, enquanto a hipótese alternativa admite que a série é estacionária. Assim sendo, caso o p_value associado ao teste for maior que o nível de significância definido (por exemplo 5 %), é possível concluir que os dados não são estacionários (Albertt Santos Carvalho, 2020). Assim, a não rejeição da hipótese nula implica que a série admite uma raiz unitária e, portanto, não é estacionária, pelo que devem ser utilizados procedimentos matemáticos para tornar a série estacionária (Campana BARBOSA et al., 2015) .

Identificação dos parâmetros e seleção do modelo

Antes de mais a literatura evidencia que todos os parâmetros do modelo SARIMA devem ser inteiros e positivos. Como já mencionado anteriormente, o parâmetro d corresponde ao número de diferenças não sazonais, o valor deste parâmetro deve ser menor ou igual a dois. Para além disso, o parâmetro D , deve admitir o valor de um, se a série temporal admitir sazonalidade, caso contrário deve ser zero (Bouzerdoum et al., 2013) .

A atribuição de valores para determinados parâmetros pode ser feita pela análise do gráfico da função de autocorrelação (ACF) e do gráfico da função de autocorrelação parcial (PACF) (P. Chen et al., 2018) , (Albertt Santos Carvalho, 2020). Assim, é construído um conjunto de modelos candidatos, para serem analisados e avaliados, através de critérios como AIC, BIC, HIC com o objetivo de selecionar o mais eficiente.

ACF e PACF

No contexto da análise de séries temporais, tanto a função de autocorrelação (Autocorrelation function – ACF), como a função de autocorrelação parcial (Partial Autocorrelation Function - PACF) assumem um papel importante.

A autocorrelação explora a relação que existe entre os valores de uma série temporal em momentos diferentes, o que fornece informações importantes como padrões e tendências dos dados ao longo do tempo. É muito importante para perceber como os valores passados influenciam os valores presentes e futuros. Por outro lado, mediante a análise da PAFC é possível identificar em que momentos podem existir termos autoregressivos que influenciam de forma significativa

(Albertt Santos Carvalho, 2020) .ACF auxilia na escolha dos parâmetros da média móvel (q e Q), enquanto que a PACF é usada para determinar os parâmetros autoregressivos (p e P), (Narasimha Murthy et al., 2018).

De acordo com as análises dos gráficos ACF e PACF observadas na literatura (Campana BARBOSA et al., 2015), (Deretić et al., 2022) , é importante observar os padrões e as quedas abruptas nestes gráficos para auxiliar na escolha de valores para os parâmetros da média móvel e dos parâmetros autoregressivos, de modo a realizar uma seleção adequada dos parâmetros, observar pontos significativos em defasagens sazonais e não sazonais.

AIC e BIC

A seleção do modelo mais apropriado, é um tema relevante, dois dos critérios mais utilizados para selecionar o “melhor” modelo são o critério de informação de *Akaike* (AIC) e o critério de informação bayesiano de Schwartz (BIC) (Narasimha Murthy et al., 2018), (Akaike, 1974).

De uma forma geral, o critério AIC quantifica a informação que um modelo em específico perde. Desta forma, esta metodologia baseia-se no princípio de que quanto menos informação um modelo perde, maior será a sua qualidade (Rossi et al., 2020). Por outro lado, o critério BIC, baseia-se na teoria de bayes, tenta encontrar um equilíbrio entre um bom ajuste do modelo e a complexidade do mesmo. O modelo ótimo será selecionado com base no mínimo de AIC e BIC (Mohammed et al., 2015). Efetivamente, conforme a complexidade do modelo aumenta o valor BIC aumenta e conforme a probabilidade de observar os dados dado um conjunto de dados aumenta o valor BIC diminui (Deretić et al., 2022) .

Por fim, os dois critérios tratam-se de funções de custo, no qual, favorecem modelos com alta verosimilhança (bom ajuste) e atribuem penalizações conforme aumenta a complexidade do modelo (Rossi et al., 2020).

De seguida, é apresentada a Equação 10 e Equação 11 que representam a formulação matemática dos critérios AIC (Akaike, 1974) e BIC, respetivamente.

$$AIC = -2 * \ln(L) + 2k \quad \text{Equação 10}$$

$$BIC = -2 \ln(L) + \ln(n) * k \quad \text{Equação 11}$$

Na Equação 10, o termo L representa a verosimilhança do modelo, o termo k representa a quantidade de parâmetros a ajustar. Por seu turno, na Equação 11, o termo n evidencia o número de observações registadas.

2.5. Métricas de avaliação dos modelos de previsão

O princípio da avaliação dos modelos de previsão assenta na diferença observada entre os valores das previsões e os valores observados reais. Uma metodologia normalmente utilizada é dividir o conjunto de dados em conjunto de treino e teste e as previsões são desenvolvidas para o conjunto

teste, assim, é possível comparar os valores das previsões com valores reais observados (P. Chen et al., 2018).

Metodologias como o cálculo da raiz do erro quadrático médio (*Root Mean Squared Error* – RMSE), Erro absoluto médio (*Mean Absolute Error* - MAE), Erro médio de viés (*Mean Bias errors* - MBE) são amplamente aplicadas para avaliar a exatidão os modelos de previsão (Chai & Draxler, 2014).

O MBE indica se os dados calculados são menores ou maiores que os valores reais observados, desta forma, se os dados calculados forem menores que os valores observados o MBE apresenta um valor negativo, caso contrário um valor positivo. O MAE representa a média dos erros absolutos, e consegue assim, avaliar o quão perto os valores previstos pelo modelo estão dos valores reais observados. Por outro lado, o RMSE avalia a magnitude do erro e admite sempre um valor positivo. Assim sendo, quanto mais próximo de zero o RMSE estiver, maior será a acurácia do modelo (Nogueira & Moreira, 2015).

Para além disso, (Chai & Draxler, 2014), evidencia uma comparação entre os dois tipos de erros mencionados anteriormente, e revela que o RMSE apresenta maior sensibilidade com valores discrepantes (outliers), e, portanto, torna-se uma vantagem em relação ao MAE.

Por fim, o MAE e o RMSE podem ser calculados de acordo com a Equação 12 e Equação 13, respetivamente (Hodson, 2022):

$$MAE = \frac{1}{n} * \sum_{i=1}^n |y_i - \hat{y}_i| \quad \text{Equação 12}$$

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{Equação 13}$$

No qual, o termo n representa a quantidade de observações, o termo y_i indica os valores observados, e por fim \hat{y}_i evidencia os valores previstos.

3. METODOLOGIA E TÉCNICAS APLICADAS

Ao longo deste capítulo vai ser abordado toda a lógica de implementação do projeto, desde o software de programação utilizado, mecanismos de otimização de código e fácil interpretação utilizados. Os dados são descritos, para posteriormente, mencionar o procedimento utilizado para a análise exploratória dos dados. Por fim, é descrito o pré-processamento dos dados necessário para implementar cada um dos métodos de previsão.

O propósito fundamental deste projeto centra-se na previsão antecipada de falhas em equipamentos, particularmente aqueles que apresentam maior propensão a problemas recorrentes. Para alcançar esse objetivo, o projeto é constituído por três etapas essenciais. Em primeiro lugar, foi conduzida uma análise cuidadosa para identificar quais equipamentos eram mais suscetíveis a falhas e se havia dados disponíveis para fundamentar essa investigação. Esta fase inicial distribuiu as bases para a seleção dos equipamentos de interesse. Em seguida, procedeu-se à análise exploratória dos dados, com o intuito de obter insights valiosos para o desenvolvimento do modelo de previsão. Das principais falhas apresentadas, foram destacadas as relacionadas aos sensores de temperatura e irradiância, cujos valores médios foram minuciosamente avaliados, bem como as falhas nos inversores, considerando os valores de potencia do lado DC e AC dos mesmos. A abordagem de previsão baseou-se na ideia de que, se os valores previstos futuros não alterarem o seu comportamento em relação aos dados históricos, não há expectativas de falha. No entanto, um desvio significativo em relação aos dados históricos pode indicar uma possível falha. Essa estratégia permite antecipar problemas, proporcionando vantagens como a redução de custos operacionais, a prevenção do aumento de falhas para problemas mais graves e uma gestão mais eficaz dos ativos.

3.1. Descrição da unidade de produção de autoconsumo

A unidade de produção de autoconsumo em análise é um empreendimento comercial associado a uma empresa situado na zona do Porto. O sistema fotovoltaico admite no total, 324 painéis. Com o objetivo de manter a produção sempre no seu máximo e tirar proveito do maior rendimento possível do sistema, os sistemas fotovoltaicos foram instalados no telhado do edifício. Para além disso, para manter a produção de energia equilibrada ao longo do dia, os 324 painéis foram instalados com orientações específicas. Portanto, 162 painéis foram instalados com orientação para o leste (nascente), enquanto os outros 162 foram direcionados para o oeste (poente). Desta forma, a produção de energia durante a manhã (painéis localizados a nascente) é equilibrada e estável em relação à produção de energia durante a tarde (painéis localizados a poente). O sistema fotovoltaico, é constituído por dois inversores, um inversor para os 162 painéis localizados a nascente e outro para os 162 painéis localizados a poente.

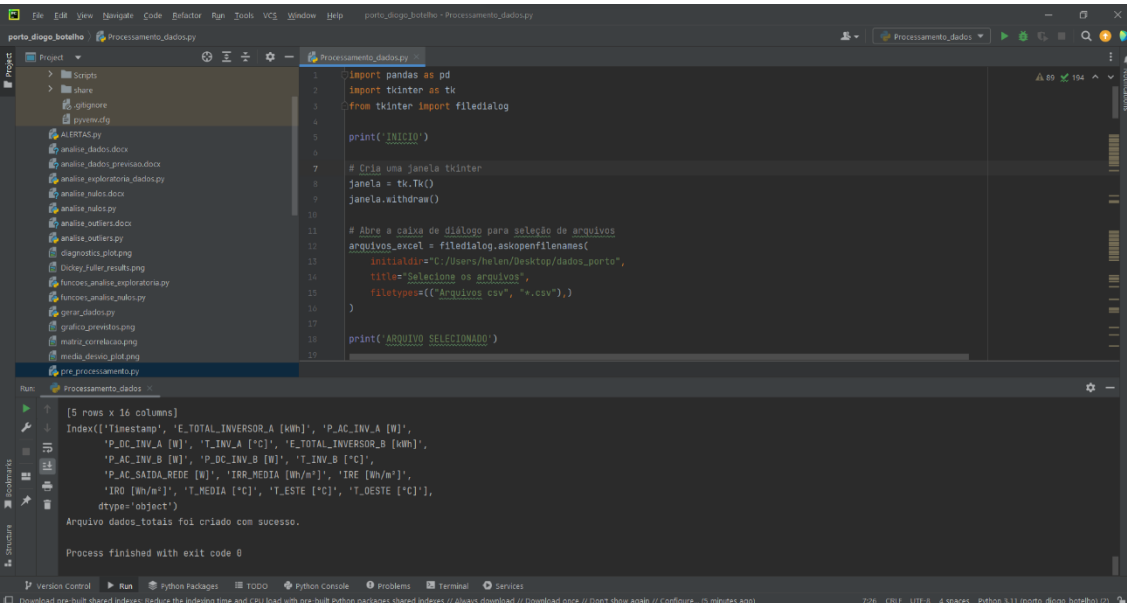
Por fim de salientar que, os 324 painéis estão distribuídos por 20 *strings*, no qual 10 estão ligadas a um inversor e dizem respeito aos painéis a nascente, e as outras 10 *strings* estão ligadas ao outro inversor e dizem respeito aos painéis ligados a poente.

3.2. Software utilizado

Para o desenvolvimento completo do projeto, foi utilizada a linguagem de programação *Python*, que atualmente é uma das mais populares e versáteis em diversas áreas. É a linguagem de programação mais utilizada em aplicações em ciência de dados e inteligência artificial. Para além disso é amplamente utilizada no desenvolvimento de software, análise de dados e *machine learning*, pois oferece uma vasta gama de bibliotecas e que facilitam a implementação de projetos complexos. Para além do que já foi referido, esta linguagem de programação conta com muita informação pormenorizada de todo o tipo de funções e outros temas acedendo à página oficial da linguagem (Python Software Foundation, 2022) .

Vale ressaltar que, ao enfrentar o desafio da análise de conjuntos de dados muito extensos, conhecido como *Big Data*, o *Python* destaca-se como uma solução viável, devido à sua capacidade de lidar eficientemente com volumes consideráveis de informações. Destaca-se também a sua ampla variedade de bibliotecas disponíveis, quer para tratamento dos dados, visualizações gráficas e aplicação de métodos de *machine learning*.

No decorrer deste projeto, o IDE (Ambiente de Desenvolvimento Integrado) escolhido para programar em *Python* foi o *PyCharm*, cujo layout de programação é apresentado na Figura 14.



```

1 import pandas as pd
2 import tkinter as tk
3 from tkinter import filedialog
4
5 print(' INICIO ')
6
7 # Cria uma janela tkinter
8 janela = tk.Tk()
9 janela.withdraw()
10
11 # Abre a caixa de diálogo para seleção de arquivos
12 arquivos_excel = filedialog.askopenfilenames(
13     initialdir='C:/Users/foleiv/Desktop/dados_porto',
14     title='Selecione os arquivos',
15     filetypes=[('Arquivos csv', '*.csv')])
16
17 print(' ARQUIVO SELECIONADO ')

```

```

[5 rows x 16 columns]
Index(['Timestamp', 'E_TOTAL_INVERSOR_A [kW]', 'P_AC_INV_A [W]',
      'P_DC_INV_A [W]', 'T_INV_A [°C]', 'E_TOTAL_INVERSOR_B [kW]',
      'P_AC_INV_B [W]', 'P_DC_INV_B [W]', 'T_INV_B [°C]',
      'P_AC_SAIDA_REDE [W]', 'IRR_MEDIA [W/m²]', 'IRE [W/m²]',
      'IRO [W/m²]', 'T_MEDIA [°C]', 'T_ESTE [°C]', 'T_OESTE [°C]'],
      dtype='object')
Arquivo dados_totais foi criado com sucesso.
Process finished with exit code 0

```

Figura 14 : Ambiente de programação python no IDE pycharm

Bibliotecas utilizadas

Normalmente, cada arquivo *Python* começa com a importação das bibliotecas permitidas, como uma forma de organizar o código. As bibliotecas desempenham papéis importantes na importação, manipulação e apresentação de dados, bem como na construção de modelos de previsão, como no caso deste projeto. Seguidamente, são apresentadas as bibliotecas incluídas neste projeto:

- Manipulação dos dados: *pandas*, *calendar*, *numpy*, *joblib*

- Apresentação gráfica dos dados: *matplotlib.pyplot*
- Criar documentos e interfaces: *Tkinter, Docx*
- Modelos de previsão: *Seaborn, pmdarima, statsmodels.tsa.stattools, sklearn, keras, tensorflow*

3.3. Descrição e tratamento dos dados

De seguida, é apresentada uma breve descrição dos dados, para que fique mais clara a abordagem deste projeto.

3.3.1. Recolha e descrição dos dados

Os dados analisados neste projeto foram gentilmente fornecidos pela empresa, sob a condição de que eles não sejam compartilhados com indivíduos não envolvidos diretamente no projeto. Esses dados foram fornecidos em formato CSV e abrangem os anos de 2020, 2021 e 2022 e são referentes a uma instalação situada na zona do Porto.

Mais especificamente, os dados estão estruturados da seguinte forma:

- Para o ano de 2020, temos informações referentes aos meses de outubro, novembro e dezembro.
- No ano de 2021, os dados abrangem todos os meses, de janeiro a dezembro.
- O ano de 2022 também engloba todos os meses, de janeiro a dezembro.

Cada conjunto de dados está registado em intervalos de 5 minutos, fornecendo dados com grande detalhe.

Em termos de variáveis foram fornecidas cerca de 93 variáveis relacionadas com várias características do sistema fotovoltaico. Posteriormente, será retratado o processamento destes dados.

3.3.2. Tratamento dos dados

Seguidamente, são apresentadas em detalhe todas as fases específicas do processamento dos dados, desde sua importação inicial até a formação do conjunto de dados final, pronto para análise exploratória.

Para uma análise mais detalhada, vão ser pormenorizadas todas as etapas desenvolvidas nos diferentes ficheiros *Python*.

Inicialmente vão ser retratadas as etapas incluídas no ficheiro *Python* denominado como "*Processamento_dados*".

Importação dos dados

Os dados foram importados usando a biblioteca *Tkinter*, que admite como parâmetro o caminho dos arquivos e o tipo de arquivos e, desta forma, o utilizador tem a liberdade de escolher os dados a ser incorporados na ferramenta. De seguida, são desenvolvidas várias fases de manipulação dos dados.

Seleção e renomeação das variáveis

Os dados a analisar foram retirados diretamente do sistema de armazenamento de dados da empresa e, portanto, contavam com 93 variáveis, no entanto, muitas das variáveis não seriam importantes para o tema em questão, e como os dados já admitiam grande granularidade, as variáveis foram selecionadas.

Como já evidenciado o principal objetivo do projeto é encontrar um mecanismo de previsão de falhas dos sistemas fotovoltaicos. Para tal, através da literatura e sobretudo, através da comunicação com a empresa, verificou-se que os equipamentos mais propensos a falha e passíveis de encontrar falhas pela análise e previsão dos dados seriam: sensor de irradiância, sensor de temperatura, inversores e por fim, ainda de realçar as falhas de comunicação através da omissão de dados como já mencionado. Desta forma, a seleção das variáveis foi orientada pelos possíveis modos de falha, e assim direcionar a criação do conjunto de dados de acordo com os requisitos do projeto.

De salientar também que o nome das variáveis dos ficheiros fornecidos não era óbvio e, portanto, procedeu-se à renomeação das mesmas.

Antes de apresentar cada variável em pormenor, é importante desde já salientar que foram analisados dados de dois inversores, nomeadamente, *SunGrow SG50CX A2003160753* e *SunGrow SG50CX A2003160763*, no qual foram abreviados por INVERSOR_A e INVERSOR_B, respetivamente.

De seguida, são apresentadas todas as variáveis selecionadas, seguidas da sua definição e por fim, é apresentada a abreviatura utilizada, Tabela 1.

De salientar que o nome das variáveis não é igual ao evidenciado no conjunto de dados, por motivos de confidencialidade.

Tabela 1: Variáveis utilizadas, sua definição e abreviatura

Variável	Definição	Abreviatura
Timestamp	corresponde à data de registo dos dados (dia, mês, ano, hora, minuto)	Timestamp
(...) SunGrow SG50CX A2003160753 - Total Energy [kWh] - E_TOTAL	indica a energia total DC [kWh] convertida pelo inversor A	E_TOTAL_INVERSOR_A[kWh]
(...) SunGrow SG50CX A2003160753 - AC power [W] - P_AC:	indica a potência [W] alternada que o inversor A converteu, diz respeito à potencia de saída	P_AC_INV_A [W]
(...) SunGrow SG50CX A2003160753 - Power DC [W] - P_DC:	indica a potência [W] total DC que chega ao inversor A	P_DC_INV_A [W]
(...) SunGrow SG50CX A2003160753 - Device Temperature [°C] - T_WR	indica a temperatura [°C] do inversor A	T_INV_A [°C]
(...) SunGrow SG50CX A2003160763 - Total Energy [kWh] - E_TOTAL	indica a energia total DC [kWh] convertida pelo inversor B	E_TOTAL_INVERSOR_B [kWh]

(...) SunGrow SG50CX A2003160763 - AC power [W] - P_AC	indica a potência [W] alternada que o inversor B converteu, diz respeito à potencia de saída	P_AC_INV_B [W]
(...) SunGrow SG50CX A2003160763 - Power DC [W] - P_DC	indica a potência [W] total DC que chega ao inversor B	P_DC_INV_B [W]
(...) SunGrow SG50CX A2003160763 - Device Temperature [°C] - T_WR	indica a temperatura [°C] do inversor B	T_INV_B [°C]
(...) Real Energy Systems PRISMA310A 2170001742 2 - Power [W] - M_AC_P	indica a potência alternada de saída da rede	P_AC_SAIDA_REDE [W]
(...) Basics - Irradiation Sensor Average [W/m ²] - G_M0	indica o valor da irradiância média mostrada pelo sensor	IRR_MEDIA [W/m ²]
(...) Basics - Irradiation Sensor East [W/m ²] - G_M1	indica o valor da irradiância do sensor a este	IRE [W/m ²]
(...) Basics - Irradiation Sensor West [W/m ²] - G_M2	indica o valor da irradiância do sensor a oeste	IRO [W/m ²]
(...) Basics - Module Temperature Average [°C] - T_M0	indica o valor da temperatura média do módulo	T_MEDIA [°C]
(...) Basics - Module Temperature Sensor East [°C] - T_M1	indica o valor da temperatura do módulo a este	T_ESTE [°C]
(...) Basics - Module Temperature Sensor West [°C] - T_M2	indica o valor da temperatura do módulo a oeste	T_OESTE [°C]

Transformações desenvolvidas nos dados

De seguida, são analisadas algumas modificações desenvolvidas no conjunto de dados.

– Transformação do Timestamp:

Inicialmente, os dados foram fornecidos com uma resolução temporal de 5 minutos, capturando informações em intervalos bastante curtos. No entanto, considerando a complexidade das análises a serem realizadas e a praticidade no processamento dos dados, houve uma decisão estratégica de ajustar a frequência para intervalos de 15 minutos.

– Dataset nulo:

Durante uma análise visual preliminar dos dados, foi observado que o conjunto de dados correspondente ao mês de maio de 2021 possuía apenas a coluna "Timestamp" com valores, as restantes colunas referentes às variáveis em estudo apresentavam-se vazias. Tal facto, pode ser devido a uma falha de comunicação, bem como algum erro associado ao armazenamento dos dados.

Considerando que os dados fornecidos abrangem um período de dois anos (2021 e 2022), a remoção completa do mês de maio de 2021 da análise reduziria significativamente a quantidade de informações disponíveis sobre aquele mês, restando apenas um mês de maio do ano de 2022.

No entanto, a literatura especializada ressalta que a constituição adequada de um conjunto de dados é um fator crucial para obter resultados precisos em análises e previsões. Diante disso, optou-se por tentar encontrar uma solução que não a de eliminar aquele mês.

Efetivamente, para solucionar o desafio apresentado pelo ficheiro de maio de 2021, os valores correspondentes ao mês de abril de 2021 foram copiados e utilizados para preencher o mês de maio de 2021, de modo a evitar a perda de informação.

Cálculo dos valores de eficiência para cada inversor

O principal objetivo do projeto é tentar prever falhas em sistemas fotovoltaicos. Para atingir esse objetivo, é crucial avaliar a eficiência dos equipamentos ao longo do tempo. Portanto, após a manipulação dos dados mencionados anteriormente, procedeu-se ao cálculo da eficiência de cada inversor.

A função `calcular_eficiencia_inversor` desempenha um papel essencial nesse processo. Os inversores são dispositivos responsáveis por converter corrente contínua em alternada para consumo pessoal. Com base nas variáveis relacionadas aos valores de potência dos lados DC e AC de cada inversor, a função calcula a eficiência do equipamento expressa em percentagem.

É importante mencionar que, em alguns casos, as variáveis podem ser iguais a zero, o que indica uma situação em que não há produção de energia. Nesses casos, a função substitui os valores de eficiência por 0 para evitar distorções nos resultados. De salientar, que as equações de cálculo foram fornecidas por membros da organização. Na Figura 15 é apresentada a função capaz de calcular as novas variáveis.

```
def calcular_eficiencia_inversor(df, variavel_DC_A, variavel_AC_A, variavel_DC_B, variavel_AC_B):
    df['Eficiencia_INV_A'] = (df[variavel_AC_A] / df[variavel_DC_A]) * 100
    df['Eficiencia_INV_B'] = (df[variavel_AC_B] / df[variavel_DC_B]) * 100

    # Substituir eficiencia por 0 se pelo menos uma das variáveis for zero
    df.loc[(df[variavel_DC_A] == 0) | (df[variavel_AC_A] == 0), 'Eficiencia_INV_A'] = 0
    df.loc[(df[variavel_DC_B] == 0) | (df[variavel_AC_B] == 0), 'Eficiencia_INV_B'] = 0

    return df

# calcular a eficiencia para os dois inversores
variavel_DC_A = 'P_DC_INV_A [W]'
variavel_AC_A = 'P_AC_INV_A [W]'
variavel_DC_B = 'P_DC_INV_B [W]'
variavel_AC_B = 'P_AC_INV_B [W]'
calcular_eficiencia_inversor(df_dados, variavel_DC_A, variavel_AC_A, variavel_DC_B, variavel_AC_B)
```

Figura 15 : Função `calcular_eficiencia_inversor` em Python e aplicação

Tratamento dos dados nulos

Uma etapa fundamental do tratamento de dados para análise posterior e previsão dos mesmos, é analisar os valores nulos do conjunto de dados original.

Para além disso, os valores nulos são uma fonte de informação no que toca a falhas, efetivamente, uma quantidade definida de valores nulos consecutivos indica uma falha de comunicação com o sistema de armazenamento de dados. Assim sendo, a análise dos valores tornou-se um ponto importante deste projeto, e para tal, foram criados dois ficheiros *Python* para proceder à sua análise que serão apresentados de seguida.

De seguida, são abordadas todas as funções desenvolvidas no Ficheiro `funcoes_analise_nulos`, para analisar e substituir os valores nulos. As informações específicas sobre essas funções podem ser encontradas no Anexo A.

– Função `contar_nulos`

Uma parte essencial da análise de dados é a verificação e tratamento de valores nulos, uma vez que esses valores podem impactar significativamente a qualidade das análises subsequentes. Neste contexto, a função `contar_nulos` foi desenvolvida para avaliar a presença de valores nulos num conjunto de dados e identificar possíveis falhas de comunicação.

Esta função é capaz de contar a quantidade de valores nulos em cada ano para cada variável e a quantidade de valores nulos nos dois anos para cada variável. Para além disso, foi também desenvolvida para emitir alertas sobre falhas de comunicação. Uma falha de comunicação acontece quando a quantidade de valores nulos num período excede o limite mínimo definido.

Desta forma, estabeleceu-se uma variável para armazenar a quantidade de nulos consecutivos que deviam existir para ser lançado o alerta de falha de comunicação. Por outro lado, se nenhuma exceção for encontrada, é apresentada uma notificação a indicar que não existem falhas de comunicação. Por fim, todos os resultados são apresentados de forma clara e organizados em uma tabela que é guardada num documento word.

– Função `get_daytime`

Esta função foi criada, como suporte para a função que será apresentada de seguida. Basicamente, retorna se é “dia” ou “noite” mediante a hora de registo da observação.

– Função `substituir_nulos`

A função anterior começa por criar uma coluna denominada 'Daytime', que categoriza cada observação como "Dia" ou "Noite" com base na hora registada na coluna 'Timestamp', que é fundamental para a etapa subsequente de substituição de valores nulos.

De seguida, é calculada a média mensal da variável objetivo, para o ano de análise, sem ter em conta os valores nulos, com base no período do dia. Ou seja, é capaz de calcular a média de cada mês para cada ano para cada período do dia.

Cada observação é analisada e se encontrar um valor nulo, é substituído pela média do ano correspondente, com base no mês e no período do dia (diurno ou noturno).

Ao substituir os valores nulos com as médias específicas para o período do dia, é preservada a coerência das informações, considerando as variações naturais que podem ocorrer durante o ciclo de 24 horas. Esta função foi desenvolvida com este rigor, visto que o conjunto de dados fornecido reflete observações de variáveis sensíveis à luz solar.

Por fim, destaca-se o ficheiro `analise_nulos`, que desempenha um papel fundamental na análise de dados. É o ponto central onde as funções desenvolvidas são utilizadas para analisar e substituir os valores nulos no nosso conjunto de dados.

Além disso, vale ressaltar que, após a substituição dos valores nulos, os valores de eficiência para cada inversor são calculados novamente. Isso ocorre porque os valores nulos não podem ser incluídos no cálculo dessas variáveis. Desta forma, os dados foram tratados, e desta forma não foram perdidas informações importantes para análises futuras.

Tratamento de valores *outliers*

Inicialmente, foi ponderada a possibilidade de examinar e tratar valores *outliers* do conjunto de dados original.

No entanto, os valores *outliers* podem ser uma fonte potencial de falhas ou anomalias. A presença de valores discrepantes não implica apenas que os dados estejam incorretos, podem indicar a existência de condições especiais ou eventos fora do comum que devem ser considerados no contexto da análise. Portanto, esta foi a abordagem seguida.

Alarmes sobre alertas de falhas de comunicação

Com o objetivo de salientar a existência de uma falha de comunicação, foi incorporada uma mensagem de alerta (caixa de mensagem no monitor) revelando-se uma abordagem altamente eficaz e prática. Esta metodologia oferece uma série de vantagens em relação aos métodos mais tradicionais, como a simples impressão dos resultados. Uma mensagem de alerta chama imediatamente a atenção do utilizador, garantindo que qualquer informação crítica não seja facilmente ignorada ou negligenciada.

A utilização de uma mensagem de alerta como parte do processo de análise de dados é uma metodologia que melhora significativamente a eficácia do nosso sistema de deteção de problemas. Garantindo que nenhum problema crítico passe despercebido e promove uma abordagem mais proativa e eficiente na gestão de nossa análise de dados. De seguida, são apresentados os exemplos das mensagens de alerta de falha de comunicação e de alerta de correção na Figura 16.

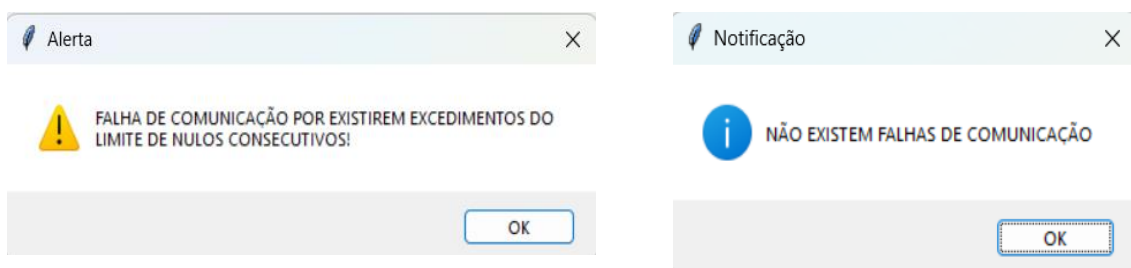


Figura 16 : Do lado esquerdo vemos um exemplo de alerta de identificação de falha de comunicação, do lado direito um exemplo de alerta da não existência de falhas

3.4. Análise exploratória

Neste capítulo, serão descritas as etapas realizadas durante a análise exploratória de dados. Antes de embarcarmos na fase de previsão, é crucial entender as características dos dados, os comportamentos das séries temporais e outros fatores relevantes. Um conhecimento sólido dos dados auxilia na seleção de um modelo de previsão adequado, para além do que é relatado na literatura.

Portanto, foi desenvolvida uma análise exploratória dos dados, embora não tenha sido uma análise detalhada. O objetivo principal era obter insights preliminares que seriam usados na fase de previsão. Vale ressaltar que os dados originais foram registados a cada 5 minutos, mas realizamos um pré-processamento para transformar essas observações em intervalos de 15 minutos. Para a análise dos dados, optamos por uma abordagem mensal, uma vez que as previsões seriam mensais. Esta escolha será destacada ao longo deste capítulo.

Foram criados dois ficheiros python, um para a construção das funções necessárias para a análise exploratória, outro ficheiro, para aplicar as funções e obter os resultados, na Figura 17 é apresentado uma síntese dos ficheiros.

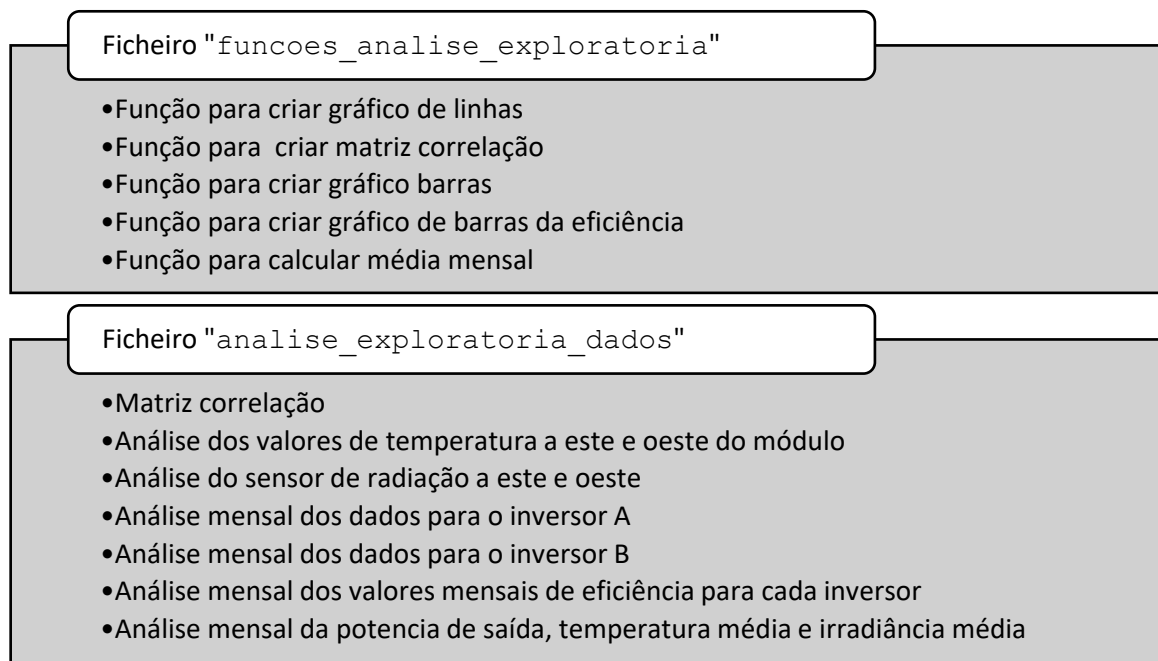


Figura 17 : Síntese dos Ficheiros criados para desenvolver a análise exploratória dos dados

Posteriormente, é apresentada uma análise mais detalhada do Ficheiro `funcoes_analise_exploratoria`, abordando individualmente cada função criada. As informações específicas sobre essas funções podem ser encontradas no Anexo B.

– Função `criar_grafico_linhas`

Tem como objetivo gerar gráficos de linha a partir de um determinado conjunto de dados contendo dados temporais. Admite a flexibilidade de conseguir apresentar graficamente uma ou duas

variáveis em relação ao tempo, permitindo uma análise visual das tendências e padrões ao longo de um período específico.

É importante notar que, inicialmente, os dados da variável em análise são filtrados com base no ano fornecido como entrada, e em seguida, é calculada a média mensal dessa variável para o ano em específico. Portanto, os valores usados para construir o gráfico de linhas representa a média mensal da variável de entrada. Para facilitar a análise, o gráfico é salvo em um documento Word, caso um documento seja fornecido como entrada.

– Função `criar_matriz_correlacao`

Esta função desempenha um papel importante na análise de dados, permitindo a criação de uma matriz de correlação a partir de um conjunto de dados, de acordo com as colunas selecionadas inicialmente. Esta função é extremamente útil para resumir as relações entre as variáveis de um conjunto de dados e apresentá-las de forma visual num documento Word, tornando a sua análise mais acessível e informativa.

– Função `criar_grafico_barras`

O objetivo desta função é criar um gráfico de barras para uma variável. Efetivamente, começa por filtrar os dados de acordo com o ano específico, de seguida, é calculada a média mensal dessa variável para esse ano. Para além disso, é calculada a quantidade de nulos para cada variável em cada mês. À partida, serão nulos porque já foi desenvolvida a análise e substituição dos nulos. Por fim, é estabelecido o gráfico de barras usando a média mensal calculada, onde cada barra representa um mês.

– Função `criar_grafico_barras_eficiencia_word`

A finalidade desta função é criar um gráfico de barras para cada mês tendo em conta a média mensal como a função anterior, mas neste caso para a variável eficiência. Efetivamente, para o cálculo da média mensal da eficiência não podem existir valores nulos. Visto que, a eficiência mede a capacidade do sistema fotovoltaico converter corrente DC em AC, ou seja, se não existir produção de energia, por várias razões, nomeadamente, a falta de luz solar, o valor de potencia do lado DC e AC do inversor será zero, mas não é por isso que o sistema fotovoltaico não foi eficiente, portanto, para o caso da eficiência teve que ser uma análise mais pormenorizada e para tal foi desenvolvida esta função, para que os valores desta variável correspondessem à realidade. A metodologia é em tudo semelhante à função referida anteriormente, apenas com este detalhe.

– Função `media_mensal`

Tem como objetivo calcular a média mensal de uma variável específica em um conjunto de dados e, opcionalmente, criar uma tabela para apresentar esses valores.

3.5. Pré-processamento

O pré-processamento dos dados é uma etapa muito importante da análise de dados, antes de implementar modelos de previsão. Efetivamente, o objetivo é preparar os dados para serem aplicados os modelos. Uma etapa muito importante do pré-processamento dos dados é a normalização dos mesmos, quer isto dizer, transformar os dados todos na mesma escala.

No estudo desenvolvido por (Bernardino Da Cunha et al., 2010) antes de aplicar uma rede neuronal como método de previsão, procedeu-se à normalização dos dados que seriam utilizados para aplicar o modelo. Ainda no mesmo estudo, é referida a importância deste passo, e que a normalização utilizada foi no intervalo entre 0 e 1.

Portanto, para este projeto segue-se a mesma ideia, e procedeu-se à normalização dos dados numa escala de 0 e 1, para todo o conjunto. O pré-processamento dos dados foi desenvolvido no ficheiro `pre-processamento`, no qual, são evidenciadas todas as etapas desenvolvidas.

Inicialmente, são filtradas as colunas, o conjunto de dados é constituído apenas por variáveis de interesse. De seguida, são calculadas as médias mensais das variáveis para cada ano, é importante destacar que a eficiência é calculada excluindo os valores zero, pois esses valores não são importantes para a análise da eficiência dos inversores, por fim todos os dados são armazenados num novo conjunto de dados, denominado *df_media*.

Assim sendo, este novo conjunto de dados, é constituído apenas por 24 observações para cada variável, correspondentes à média mensal dos dois anos. Para que, a previsão seja desenvolvida mensalmente.

Por fim, são normalizados os dados para garantir que todas as variáveis tenham a mesma escala, o que é fundamental para muitos algoritmos de aprendizagem de máquina e análise estatística. O código utiliza a função `MinMaxScaler` da biblioteca `Scikit-Learn` para realizar a normalização das variáveis. Cada variável é normalizada separadamente, garantindo que os valores estejam no intervalo [0, 1].

As linhas de código deste ficheiro são apresentadas no Anexo C.

3.6. Redes Neurais

Neste ponto, será retratada de uma forma geral a metodologia implementada para o modelo MLP (Multilayer Perceptron) de rede neuronal. Um MLP é uma arquitetura de rede neuronal artificial que consiste em pelo menos três camadas de nós: uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída.

Bibliotecas utilizadas para a implementação do modelo de redes neuronais

Antes de abordar as funções e a lógica do modelo, é relevante realçar o vasto conjunto de bibliotecas que o software dispõe. As bibliotecas utilizadas desempenham papéis importantes em diferentes aspetos, na análise de dados, na construção de modelos de previsão.

Deste modo, é relevante indicar as bibliotecas utilizadas para a construção da ferramenta, em especial, para a construção do modelo MLP de rede neuronal:

- *sklearn*: utilizada para treinar modelos de aprendizado de máquina, fazer avaliação de desempenho de modelos, realizar seleção de recursos e pré-processamento de dados.
- *matplotlib.pyplot* : Usada para criar gráficos que auxiliam na análise exploratória de dados e na apresentação dos resultados.
- *docx*: usada para criar relatórios, neste caso, para guardar informações sobre cada variável.
- *numpy*: realizar cálculos e operações matemáticas eficientes nos dados
- *keras* e *tensorflow*: Utilizados para criar e treinar redes neurais profundas

Metodologia implementada

De seguida, são apresentadas as etapas de construção do modelo de acordo com os passos utilizados para a programação do mesmo em *Python*. As linhas de código responsáveis pela implementação deste método encontram-se no Anexo D.

Importação do ficheiro com os dados

Numa fase inicial procedeu-se à importação dos dados. De salientar que os dados devem estar normalizados para a construção de qualquer modelo de previsão. Desta forma, na etapa anterior procedeu-se ao pré-processamento dos dados para que de forma prática, a implementação dos modelos inicie com a importação dos dados já com o formato devido.

Seleção da variável a prever

Neste passo, é definida a variável a prever e guardada numa variável, designada por *variavel_x*. Assim, apenas é necessário definir a variável a prever uma vez, e previne erros por parte do utilizador e torna o código mais simples, mais fácil de alterar.

Seleção de variáveis

Inicialmente, foram selecionadas algumas variáveis a partir do conjunto de dados original, focando apenas as essenciais para a identificação de falhas em sistemas fotovoltaicos. No entanto, ao implementar o modelo, tornou-se evidente a necessidade de incluir apenas as variáveis que fossem relacionadas com a variável a prever.

Para realizar a seleção de variáveis, foi construída uma matriz de correlação entre as diversas variáveis. Foi estabelecido um limite correlação mínimo a ser alcançado com a variável a prever. Apenas as variáveis que atingiram ou ultrapassaram o valor mínimo de correlação em relação à variável a prever foram incluídas no conjunto de dados.

Preparação dos dados de entrada e saída

Neste ponto foram criadas duas variáveis:

- “x”: dados de entrada, inclui todas as variáveis exceto a variável a prever;
- “y”: dados de saída, inclui a variável a prever;

Semente de inicialização

Após preparar os dados é importante definir a semente de início do número aleatório.

Muitas bibliotecas de *machine learning* utilizam números aleatórios para iniciar uma solução, se não for definida uma semente, quer isto dizer, se não for estabelecido um ponto de partida consistente para a geração de números aleatórios, de cada vez que o código é executado os resultados podem ser diferentes. A consistência dos resultados é importante, e, portanto, é relevante atribuir um valor de início à nossa semente.

Na procura pela reprodutibilidade e consistência dos resultados, a semente de inicialização, desempenhou um papel crucial. Ao utilizar a biblioteca `TensorFlow`, foi incorporada a semente de inicialização para controlar a aleatoriedade relativa a diversas transações que envolvem números aleatórios, incluindo a inicialização de pesos em redes neuronais.

O valor da semente de inicialização, neste caso, foi dois. No entanto, a escolha do valor é arbitrária, o que interessa é que seja a mesma de cada vez que se execute o código, para garantir que os resultados não alteram muito significativamente.

Em suma, a semente de inicialização, ao estabelecer um ponto de partida consistente para a geração de números aleatórios, desempenhou um papel fundamental na procura por resultados precisos neste estudo.

Divisão do conjunto de dados em treino e teste

Com recurso à biblioteca `sklearn`, em particular, recorrendo à função `train_test_split` procedeu-se à divisão do conjunto de dados em treino e teste. O conjunto de treino é utilizado para treinar a rede, enquanto o de teste é utilizado para posteriormente avaliar as previsões desenvolvidas pelo modelo.

Um parâmetro importante a salientar, é partição dos dados, ou seja, definir a quantidade de dados para o conjunto de treino e para o conjunto de teste, este é um parâmetro que afeta a qualidade dos dados. Efetivamente, quanto maior for o conjunto de treino maior será a aprendizagem do modelo sobre o comportamento dos dados, no entanto, deve se ter em conta também o sobreajuste do modelo aos dados históricos.

(Antoniolli et al., n.d.) menciona que o conjunto de treino geralmente deve abranger de 60% a 90% dos dados, enquanto o conjunto de teste deve corresponder a 10% a 40% dos dados disponíveis.

No caso deste projeto, os dados analisados eram constituídos por observações bastante pormenorizadas, que foram transformadas numa base de mensal através da média, portanto, o conjunto final dos dados era constituído por 24 observações. Deste modo, optou-se por um valor intermédio ao que é mencionado pela revisão literatura, ou seja 25% dos dados para teste e 75% para treino da rede. Na abordagem implementada, não foram alterados estes valores por se tratar de um conjunto de dados reduzido.

De seguida, as linhas de código responsáveis pela divisão do conjunto de dados são apresentadas na Figura 18.

```

print('-----CONJ. TREINO E TESTE-----')
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5) #, random_state=40
print(X_train)
print(X_test)
print(y_train)
print(y_test)

```

Figura 18 : Divisão do conjunto de dados em treino e teste da variável X e Y

De salientar, que esta primeira metodologia admite um erro, pois desta forma, o conjunto de dados seria dividido de forma aleatória, ou seja, a sequência temporal não era atendida. De acordo com este método, apenas seria possível avaliar a qualidade de previsão pela comparação entre os valores reais e os previstos, e não seria possível concluir sobre o comportamento da série. Assim sendo, após uma pesquisa, foi visível em alguns exemplos de redes neurais a utilização do comando *shuffle* para travar a aleatoriedade na divisão do conjunto de dados, ou seja, no momento da divisão os dados não são sorteados para treino e teste de forma aleatória, são divididos sequencialmente. Portanto, para o presente projeto foi atualizado o código apresentado na Figura 18 para o código revelado na Figura 19, com a introdução do comando, desta forma a aleatoriedade na divisão do conjunto de dados já não é um problema. Visto que se trata de uma série temporal, a divisão do conjunto em treino e teste de forma cronológica é um passo muito importante, para representar o comportamento da série.

```

print('-----CONJ. TREINO E TESTE-----')
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, shuffle=False)
print(X_train)
print(X_test)
print(y_train)
print(y_test)

```

Figura 19 : Divisão do conjunto de dados em treino e teste da variável X e Y com comando shuffle

Definição da estrutura da rede MLP

Neste projeto, optou-se por implementar um algoritmo *Multilayer Perception* (MLP) por apresentarem bons resultados e serem frequentemente utilizados, segundo a literatura. De seguida, é abordada a definição da estrutura da rede.

Utilizando a biblioteca *Keras*, parte-se para a definição da estrutura da rede, usando a classe *Sequential* da biblioteca *Keras*. Para definir a estrutura da rede é necessário atender a diversos parâmetros. A definição de cada parâmetro é evidenciada na Figura 20.

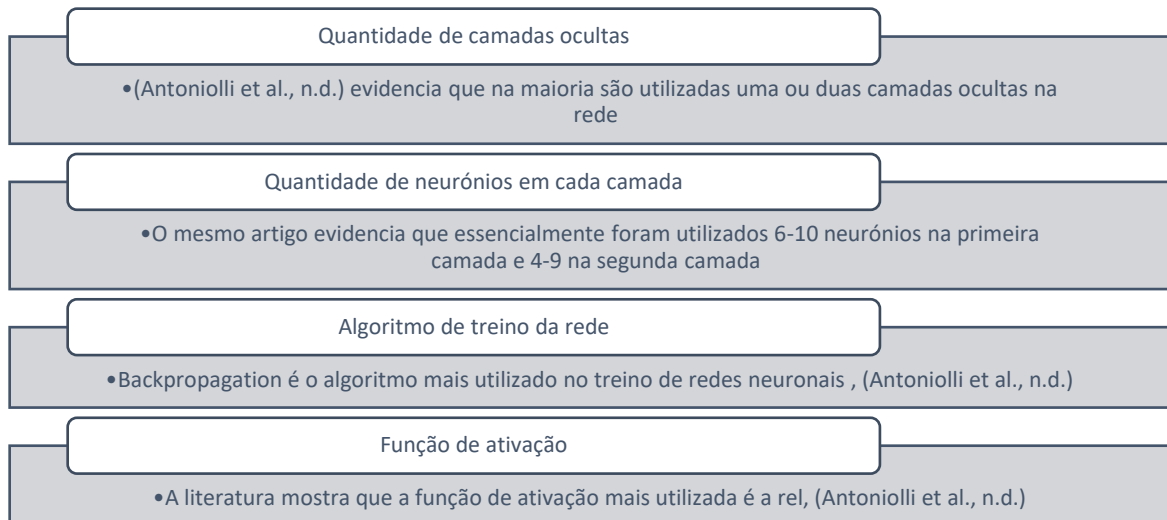


Figura 20 : Parâmetros de análise da estrutura da rede neuronal

Conforme é mostrado na Figura 20 os parâmetros de estrutura da rede, foram estabelecidos com base no que é referido na literatura. No entanto, no que toca à quantidade de camadas ocultas e quantidade de neurónios em cada camada não existe uma forma automática de os calcular, nem nenhum estudo que evidencie os valores exatos para estes parâmetros, pelo que foram selecionados numa abordagem de tentativa erro.

Para determinar o modelo mais adequado, foi desenvolvida uma análise comparativa entre os valores reais (obtidos a partir do conjunto de teste) e as previsões do modelo, de forma visual. Além disso, foram calculados os valores do Erro Absoluto Médio (MAE) e Erro Quadrático Médio (MSE) para cada modelo. A escolha do modelo final foi baseada no modelo que apresentou os menores valores de MAE e MSE.

Processo de compilação do modelo

Após a definição da estrutura da rede, segue-se a fase de compilar a rede definindo uma função perda e o otimizador. No exemplo que serviu de base, na fase de compilar o modelo também foi definida uma métrica de avaliação de desempenho do modelo, que também foi definida no presente projeto.

Numa primeira fase, foi definido o otimizador do modelo com *optimizer='adam'*. O otimizador é responsável por ajustar os pesos da rede neuronal durante o treinamento de modo a minimizar a função de perda (Secreto, 2020). *Adam* é um dos algoritmos de otimização mais populares e eficazes, e, portanto, foi o utilizado neste projeto.

Em seguida, foi identificada a função de perda que o modelo deve minimizar durante o treinamento. (Ferreira et al., 2012) evidencia que uma das funções perdas mais utilizada é Erro Quadrático Médio (MSE), assim foi a utilizada. O objetivo é minimizar essa função, tornando as previsões do modelo o mais próximo possível dos valores reais.

Por fim, como já mencionado, foi adicionada uma métrica para avaliar o desempenho da rede, neste caso, definiu-se o Erro Absoluto Médio (MAE) a ser calculado.

Processo de treino da rede

A fase de treino da rede é uma etapa fundamental, para tal foi utilizada a função `fit` para treinar o modelo com os dados de treino fornecidos anteriormente.

Para tal, é necessário definir valores para os parâmetros `batch_size` e `epochs`.

O parâmetro `batch_size` define o tamanho dos lotes de dados que são apresentados ao modelo em cada iteração do treino. Neste projeto, utilizou-se um tamanho de lote de 128. Isso significa que serão apresentados ao modelo 128 amostras de o conjunto de treino antes dos pesos serem atualizados. Basicamente define a quantidade de dados processados em cada época. Este parâmetro que permite acelerar o processo de treino do modelo.

O parâmetro `epochs` define a quantidade de épocas de treino. Uma época é uma passagem completa por todo o conjunto de dados de treino. Neste exemplo, foi utilizado 80 épocas de treino, o que significa que o modelo passará pelo conjunto de treino 80 vezes para aprender a partir dos dados. Os valores destes parâmetros são encontrados através de testes.

As linhas de código das fases de compilar o modelo e treinar a rede são evidenciadas de seguida Figura 21.

```
print('-----COMPILAR O MODELO-----')
model.compile(optimizer='adam', loss='mean_squared_error', metrics=['mean_absolute_error'])

print('-----TREINAR O MODELO-----')
model.fit(X_train, y_train, batch_size=128, epochs=80, verbose=False)
#batch_size = 32 e 64
# epochs = 50
```

Figura 21 : Fases de compilar o modelo e treinar da rede

É importante referir que quando foi utilizado o `model.compile()` e posteriormente treino do modelo através do `model.fit()`, está implícito a utilização do algoritmo *backpropagation* para treinar o modelo. Efetivamente, quando são utilizadas as bibliotecas *Keras* em conjunto com o *TensorFlow*, o algoritmo de *backpropagation* é implementado automaticamente.

Desenvolvimento das previsões

Por fim, são construídas as previsões de acordo com a parametrização do modelo evidenciada anteriormente. De salientar, que este é um processo de teste dos parâmetros. Efetivamente, alguns parâmetros já se encontram estabelecidos, mas outros são estabelecidos por tentativa erro e, portanto, através de testes é que são identificados.

Nesta fase, são então desenvolvidos os testes das previsões, de seguida, são avaliadas e representadas graficamente. Na Figura 22, são evidenciados os parâmetros para os quais foram desenvolvidos testes, com o objetivo de obter a rede que produzia previsões de melhor qualidade.

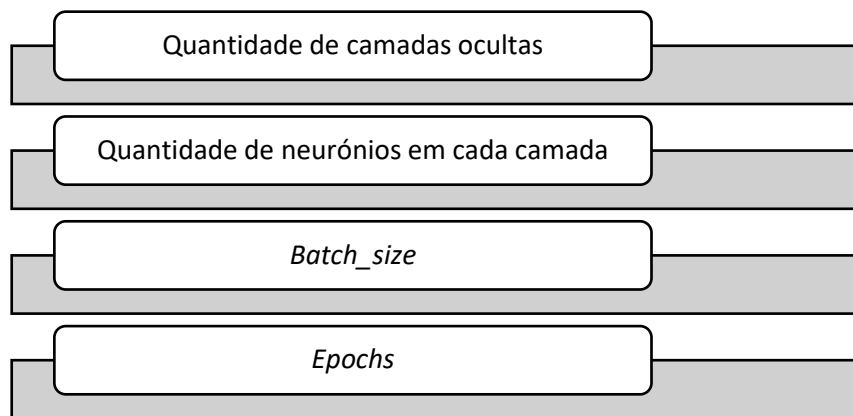


Figura 22 : Parâmetros selecionados através de testes

A avaliação das previsões ocorreu por meio do cálculo do Erro Absoluto Médio (MAE) e do Erro Quadrático Médio (MSE). O teste que evidenciasse um menor valor destas métricas, admitia o melhor modelo e, conseqüentemente, seria o modelo selecionado.

Por fim, para realçar a qualidade das previsões foi desenvolvido um gráfico de comparação dos valores reais com os valores obtidos pelo modelo. Desta forma, a avaliação do modelo foi desenvolvida em duas formas, a nível de cálculo e graficamente.

3.7. SARIMA

De seguida, é descrita toda a lógica fundamentada para o desenvolvimento do método de previsão SARIMA.

Bibliotecas utilizadas para a implementação do SARIMA

Antes de abordar as funções e a lógica do modelo, é importante referir que o *Python* se torna uma linguagem de programação muito interessante, dado o vasto conjunto de bibliotecas que dispõe. Cada uma dessas bibliotecas desempenha um papel específico na análise de séries temporais, tornando possível realizar várias tarefas, desde a preparação dos dados até a avaliação do modelo de uma forma muito mais rápida do que escrita do código de forma manual.

Deste modo, é relevante indicar as bibliotecas utilizadas para a construção da ferramenta, em especial, para a construção do modelo SARIMA:

- *pandas*: usada para carregar, estruturar e preparar os dados temporais;
- *numpy*: realizar cálculos e operações matemáticas eficientes nos dados;
- *matplotlib.pyplot*: no código, foi usada para criar gráficos que mostram as tendências e variabilidades dos dados;
- *docx*: usada para criar relatórios, neste caso, para guardar informações sobre cada variável;
- *pmdarima*: biblioteca específica para que de forma automática seja selecionado o melhor modelo ARIMA e SARIMA. No código, foi usada para ajustar um modelo SARIMA aos dados de treino;

- *statsmodels.tsa.stattools*: Parte do *Statsmodels*, uma biblioteca para análise estatística em *Python*. Utilizada para realizar o teste de *Dickey-Fuller*, que verifica a estacionariedade dos dados;
- *sklearn.metrics*: a biblioteca *Scikit-learn*, usada para calcular métricas de avaliação, como o Erro Médio Absoluto (MAE) e a Raiz do Erro Médio Quadrático (RMSE), para avaliar a qualidade das previsões do modelo SARIMA.

Metodologia implementada

A metodologia de implementação baseou-se nas ideias apresentadas na revisão da literatura. Nesse sentido, utilizou-se o método de *Box & Jenkins*, (Albertt Santos Carvalho, 2020), que compreende quatro etapas fundamentais: identificação, cálculo dos parâmetros, diagnóstico do modelo e, por fim, avaliação. De seguida, Figura 23 é apresentada a divisão das funções criadas, seguindo a linha de pensamento do método analisado na revisão literatura.

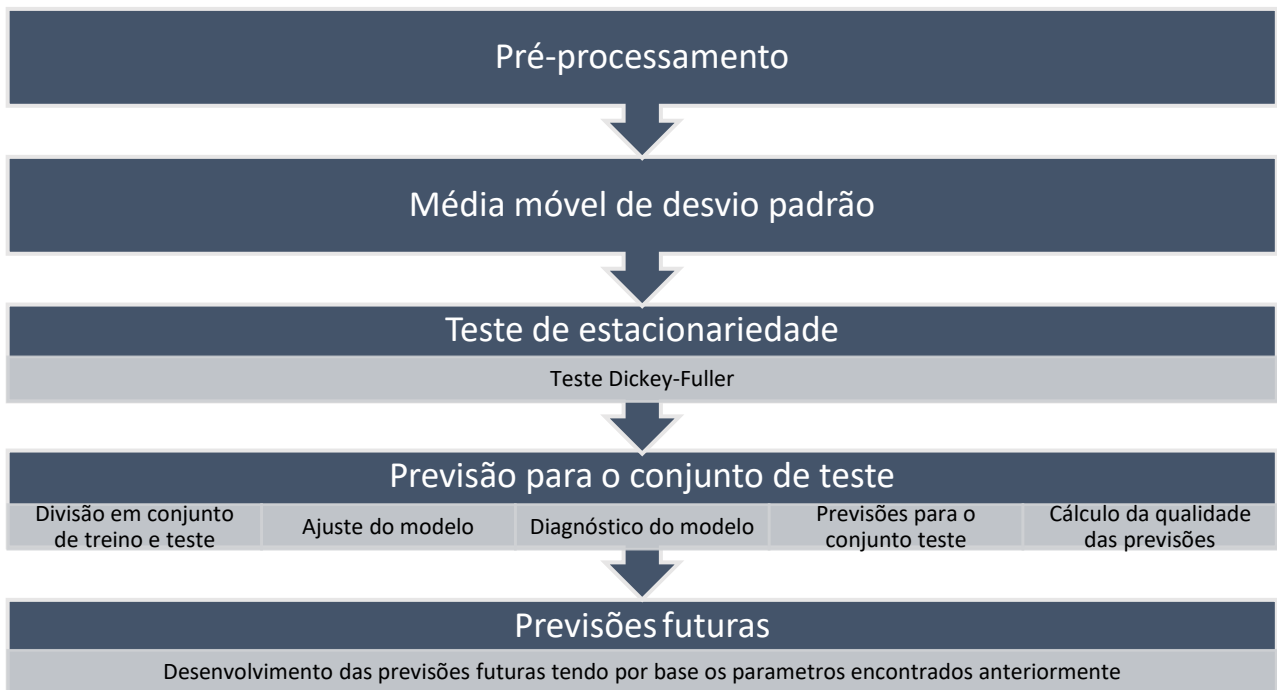


Figura 23 : Metodologia aplicada

Por fim, é apresentada a estrutura de implementação do método seguida para qualquer variável analisada, Figura 24.

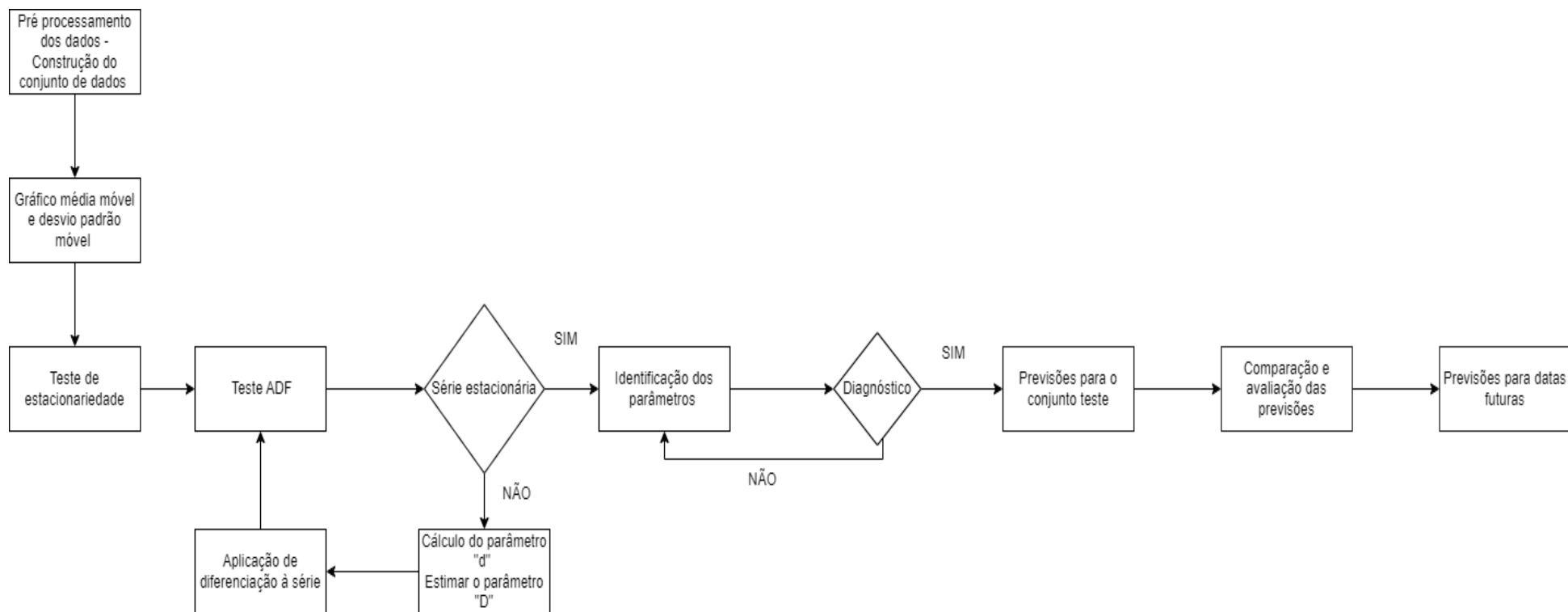


Figura 24: Estrutura de implementação do método

Variáveis estudadas e organização dos ficheiros

O objetivo do projeto era construir um modelo de previsão de falha, dos equipamentos mais propensos a paragens ou distúrbios recorrentes.

Com base nas comunicações da empresa que destacaram os equipamentos mais expostos às falhas e após realizar uma análise exploratória dos dados, identificamos as variáveis de estudo que foram representadas e catalogadas na Tabela 2 para aplicar o modelo SARIMA para prever valores futuros de cada variável. Na Tabela 2 são também referidos o tipo de falha e o equipamento associado e por fim, é evidenciada a abreviatura da variável.

Tabela 2 : Identificação da variável em estudo, tipo de falha e abreviatura da variável

Variável em estudo	Tipo de falha	Abreviatura
Temperatura média	Falha no sensor de temperatura (falha do equipamento, falha de comunicação com o sistema de informação, etc)	T_MEDIA [°C]
Irradiância média	Falha no sensor de radiação solar (falha do equipamento, falha de comunicação com o sistema de informação, etc)	IRR_MEDIA [W/m ²]
Potência contínua	Falha no sistema fotovoltaico do lado DC (diminuição da energia gerada em DC)	P_DC_INV_A [W] P_DC_INV_B [W]
Potência alternada	Falha no inversor (falha no equipamento de conversão DC/AC)	P_AC_INV_A [W] P_AC_INV_B [W]
Temperatura do inversor	Falha no inversor (perdas de eficiência por Temperatura demasiado elevada)	T_INV_A [°C] T_INV_B [°C]

Inicialmente, foi criado o ficheiro *Python* SARIIMA_FUNCOES, no qual foram desenvolvidas todas as funções necessárias para a construção do modelo. Como a análise de previsão foi desenvolvida para várias variáveis, para cada variável foi criado um ficheiro para aplicar o modelo e, de acordo com, a variável em estudo são introduzidos os inputs necessários para cada função. As vantagens da técnica de modularização já foram evidenciadas anteriormente. De seguida, para tornar mais fácil a compreensão do projeto, na Tabela 3 está representado o ficheiro de construção das funções para o modelo, enquanto, na Tabela 4 os ficheiros para cada variável que utilizam as funções representadas na Tabela 3.

Tabela 3 : Desenvolvimento de funções método SARIMA no Ficheiro SARIIMA_FUNCOES

Ficheiro Python	Funções
SARIIMA_FUNCOES	pre_processamento_SARIMA
	calcular_media_desvio_padrao
	teste_dickey_fuller
	plot_acf_pacf_to_word
	Previsao_Sarima_treino_teste
	fazer_previsoes_futuras

Tabela 4 : Ficheiros Python para cada variável para aplicação do método SARIMA

Ficheiro Python
SARIMA_T_MEDIA
SARIMA_IRR_MEDIA
SARIMA_T_INV_A
SARIMA_T_INV_B
SARIMA_P_AC_INV_A
SARIMA_P_DC_INV_A
SARIMA_P_AC_INV_B
SARIMA_P_DC_INV_B

Etapas de construção

De seguida, é apresentada cada etapa em particular, tendo em conta a Figura 24. De notar ainda que, as etapas de desenvolvimento foram baseadas no que foi retratado na literatura, e ainda, em (Brendan Artley, 2022).

Pré-processamento

O processo inicia-se com a construção do conjunto de dados para ser aplicado o modelo, e, para tal foi desenvolvida a função `pre_processamento_SARIMA`, apresentada no Anexo E.

De acordo com os exemplos demonstrados na literatura, o conjunto de dados deve ser constituído, pela variável de tempo que representa a série temporal ao longo do tempo, e pela variável a ser prevista. Para serem introduzidas outras variáveis independentes que influenciam a variável dependente (variável a prever) deve ser aplicada uma extensão do modelo SARIMA, denominada de SARIMAX (Vagropoulos et al., 2016) . Desta forma, como o objetivo é aplicar o modelo SARIMA, apenas serão consideradas as duas variáveis já descritas.

O conjunto de dados será formado pelas variáveis *Timestamp* (variável de tempo) e por uma *variável_x* (variável a estudar – input).

De seguida, para avaliar de uma forma geral a nível estatístico os dados, foram desenvolvidos os gráficos da média móvel e do desvio padrão móvel.

Média móvel e desvio padrão

Com o objetivo de estudar visualmente a variabilidade dos dados e conseqüentemente, a sua estacionariedade foram desenvolvidos os gráficos de média móvel e do desvio padrão móvel. Efetivamente, através do gráfico da média móvel é possível analisar se a série temporal admite algum tipo de tendência. Quanto ao gráfico do desvio padrão móvel, é utilizado para verificar a variabilidade dos dados ao longo do tempo. Por fim de salientar que, se os gráficos de média móvel e desvio padrão móvel não apresentarem grandes oscilações, pode ser um indicador de que se trata de uma série estacionária, mas serão necessários outros mecanismos que comprovem esta conclusão.

De salientar que, que as representações destes gráficos foram desenvolvidas através da função `calcular_media_desvio_padrao`, demonstrada no Anexo E.

Teste de estacionariedade: Teste ADF

De acordo com a literatura, a estacionariedade é uma condição fundamental para a aplicação eficaz dos modelos de séries temporais. Efetivamente, dados não estacionários não podem ser utilizados, devido à sua natureza imprevisível, pois nestes casos, os dados variam de forma inesperada. De ressaltar, que uma série é estacionária se admitir, média e variância constante ao longo do tempo. Se a série não for estacionária devem ser aplicadas diferenciações para tornar a série estacionária.

Para avaliar a estacionariedade da série temporal, foi utilizado o Teste de Dickey-Fuller Aumentado (ADF). Se o resultado deste teste indicar que a série é estacionária (ou seja, os valores dos parâmetros d e D são ambos iguais a zero), avança-se para a próxima fase. No entanto, se o teste indicar que a série não é estacionária, é necessário recorrer a diferenciações, e para tal, como já foi referido, os termos relacionados com a diferenciação no modelo SARIMA são d e D . O parâmetro d refere-se à ordem de diferenciação não sazonal, enquanto, D refere-se à ordem de diferenciação sazonal. Assim, caso se verifique a não estacionariedade da série, estes parâmetros devem ser atualizados.

Caso a série admita sazonalidade, podemos começar por atualizar $D = 1$. No caso do parâmetro d , pode ser calculado, através do código da Figura 25. O código foi baseado em exemplos encontrados na internet (pmdarima, 2017).

```
print('Cálculo parametro "d"')
### calcular parametro "d"
d = pm.arima.ndiffs(df_preprocessado[variavel_x], alpha=0.05, test='kpss', max_d=2)
print(f'valor do parametro "d" é :{d}')
```

Figura 25: Linhas de código para calcular parâmetro “d”

De salientar, pela Figura 25 que o cálculo do parâmetro “ d ” foi desenvolvido pelo teste “kpss”, visto que, é o teste de estacionariedade padrão associado.

O cálculo do parâmetro d pela Figura 25, nem sempre se mostrou eficiente, portanto, dependendo do caso, foram elaborados alguns testes de valores para este parâmetro, mediante o caso serão apresentados.

Com o cálculo do valor d é importante criar testes para os restantes parâmetros. É importante, que a série seja estacionária para continuar o modelo, portanto, pelas linhas de código evidenciadas na Figura 26, é possível aplicar diferenciações sazonais, D , e não sazonais, d , separadamente ou em combinação, com o objetivo de atingir a estacionariedade da série.

Após cada aplicação de diferenciação, é realizado novamente o teste ADF. Esta etapa é crucial para verificar se a série se tornou estacionária após as diferenciações aplicadas. Caso a série seja considerada estacionária, pode-se avançar para a próxima fase do método SARIMA. Caso contrário, devem ser realizados novos testes e os dados devem ser analisados de forma mais pormenorizada.

É um processo de análise adaptativo a cada variável em questão, onde o objetivo principal é garantir que a série é estacionária, o que é fundamental para a construção de modelos de séries temporais, eficazes e confiáveis.

É importante, ainda ressaltar, que os testes dos parâmetros d e D foram realizados sobre uma cópia do conjunto de testes originais, pois o objetivo é estabelecer o valor destes parâmetros.

Na Figura 26, é apresentado então as linhas de código de aplicação das diferenciações, bem como o teste ADF pela segunda vez.

```
print('Aplicação das diferenciações e teste ADF 2ª vez')
# segunda avaliação de estacionariedade - TESTAR PARAMETROS "d" E "D"
# Criar uma cópia do DataFrame original para a diferenciação
df_preprocessado_2 = df_preprocessado.copy()

# Aplicar a primeira diferenciação à cópia do DataFrame com D=1 se quiser alterar o "d" tenho que tornar a diferenciar
# SAZONAL ORDEM 12
df_preprocessado_2['Dif1'] = df_preprocessado_2[variavel_x].diff(12)
# NÃO SAZONAL ORDEM 1
df_preprocessado_2['Dif1'] = df_preprocessado_2[variavel_x].diff(1)
# DIFERENCIAÇÃO SAZONAL E NÃO SAZONAL em simultâneo
df_preprocessado_2['Dif1'] = df_preprocessado_2[variavel_x].diff(1)
df_preprocessado_2['Dif2'] = df_preprocessado_2[variavel_x].diff(12)

# Remover os valores NaN resultantes da diferenciação
df_preprocessado_2 = df_preprocessado_2.dropna()
# Chamar a função de teste Dickey-Fuller à série diferenciada
output_path2 = "dickey_fuller_results2.png"
dfoutput2 = teste_dickey_fuller(df_preprocessado_2, 'Dif1', output_path2)
```

Figura 26 : Exemplo de aplicação das diferenciações e teste ADF pela 2ª vez.

A Figura 27 oferece um resumo do processo de teste de estacionariedade no projeto.

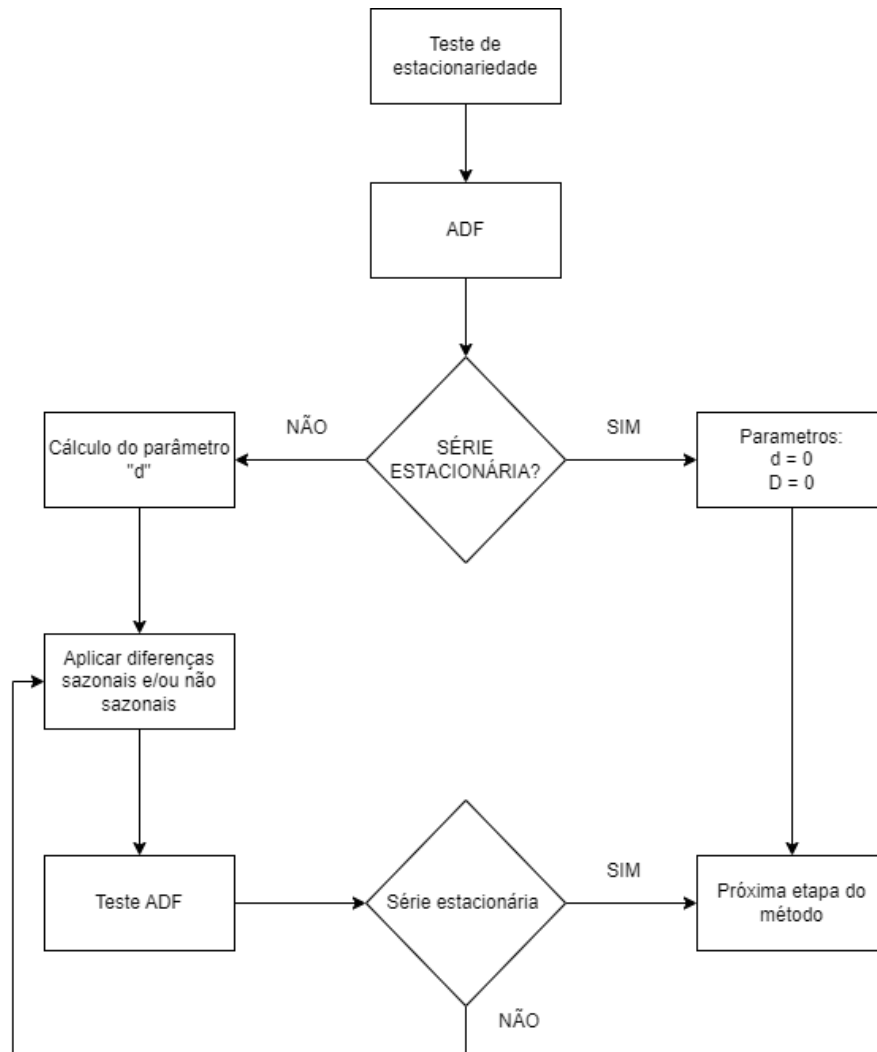


Figura 27 : Fluxograma do teste de estacionariedade da série

De seguida, são apresentadas as funções responsáveis pela aplicação do teste de ADF e desenvolvimento dos gráficos ACF e PACF.

– Função `teste_dickey_fuller`

Esta função é capaz de aplicar o teste ADF através da biblioteca `statsmodels.tsa.stattools` e está representada no Anexo E.

O teste de *Dickey Fuller Aumentado* já foi apresentado e esclarecido na revisão literatura, no entanto, é apresentada de uma forma mais intuitiva, o seu funcionamento, Figura 28.

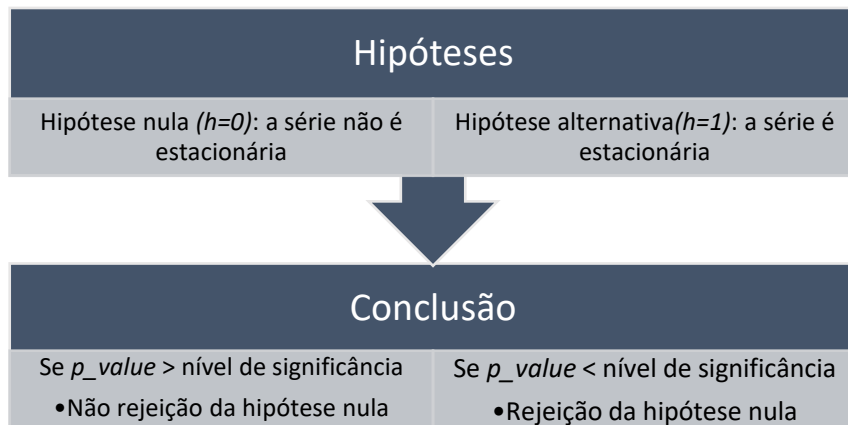


Figura 28 : Método ADF

Desenvolvimento das previsões

No contexto da análise de séries temporais e da aplicação do modelo SARIMA, é comum usar os gráficos ACF e PACF como ferramentas visuais para ajudar na escolha dos parâmetros do modelo. No entanto, no presente projeto, para não depender exclusivamente desses gráficos para determinar os valores dos parâmetros, optou-se por utilizar uma abordagem mais automatizada que será abordada na próxima etapa.

Neste ponto, o teste de estacionariedade já foi aplicado e os valores dos parâmetros d e D já foram definidos. É importante continuar com as etapas do método de *Box&Jenkins* evidenciadas anteriormente, nomeadamente, calcular os valores dos parâmetros, desenvolver o diagnóstico do modelo, e por fim desenvolver as previsões. Portanto, neste caso, os gráficos de ACF e PACF serviram como um mecanismo inicial para determinar o intervalo de valores dos parâmetros, assim, foram utilizados para o cálculo dos parâmetros como a literatura indica, mas em conjunto com uma técnica automatizada. De seguida, é apresentada a função capaz de gerar os gráficos.

– Função `plot_acf_pacf_to_word`

A função foi desenvolvida através da biblioteca `statsmodels.graphics.tsaplots`, esta função é capaz de representar os gráficos de ACF e PACF de uma determinada variável e está representada no Anexo E. Os gráficos ACF e PACF auxiliam no estudo da estacionariedade e ainda a refletir sobre os parâmetros: p, P, q, Q .

De acordo com (Hyndman, 2018), “para uma série temporal estacionária, o ACF tende para zero rapidamente, enquanto o ACF de dados não estacionários diminuirá lentamente.”

A avaliação visual das funções de ACF e PACF é uma maneira útil de atribuir inicialmente valores aos parâmetros relativos à média móvel e autoregressão, respetivamente, Figura 29.

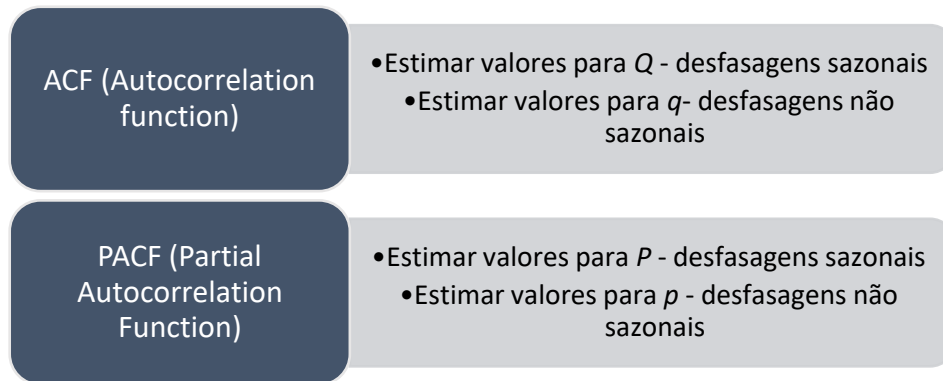


Figura 29: Parâmetros a analisar para ACF e PACF

Durante o estudo do método pela revisão da literatura, foi visível a utilização de uma abordagem comum para a previsão de valores futuros. Inicialmente envolvia a divisão do conjunto de dados em dois subconjuntos: treino e teste. O conjunto de treino é utilizado para aprender o modelo e ajustar-se aos dados disponíveis. O conjunto de teste é utilizado para o cálculo das previsões no período correspondente ao do conjunto de teste, e, como o nome sugere, é usado para testar o desempenho do modelo, comparando os valores previstos pelo modelo com os valores reais apresentados no conjunto de teste. Finalmente, foram aplicados os métodos de avaliação incluídos na literatura para avaliar a qualidade das previsões desenvolvidas. Por fim, através do modelo ajustado para o conjunto de teste foram elaboradas as previsões para datas futuras ao conjunto de dados.

Portanto, este foi o procedimento utilizado para a construção do projeto. Inicialmente, foram desenvolvidas as previsões para o conjunto de teste, de seguida, mediante a qualidade das previsões obtidas, foram utilizados os parâmetros encontrados para desenvolver as previsões para os valores futuros. Com esse objetivo foram criadas duas funções, representadas na Figura 30.

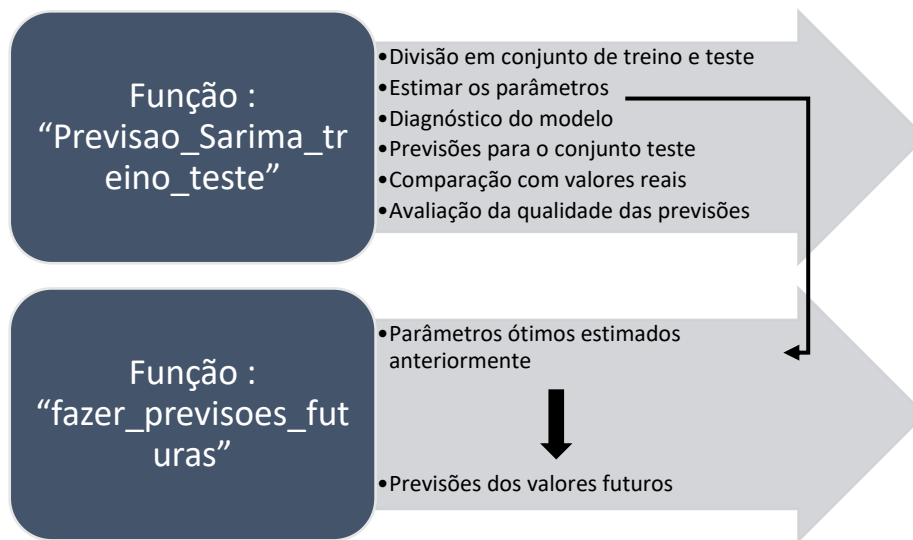


Figura 30 : Funções para o desenvolvimento das previsões

De seguida, são abordadas as duas funções.

Previsão para o conjunto de teste

– Função `Previsao_Sarima_treino_teste`

É responsável por desenvolver as previsões para o conjunto de teste. Seguidamente, são apresentadas cada uma das etapas que constitui esta função. Para além disso, esta função está representada no Anexo E.

Divisão do conjunto de dados em treino e teste

Para iniciar o projeto, procedeu-se à divisão do conjunto de dados em conjuntos de treino e teste. É importante destacar que a escolha dessa divisão pode afetar a qualidade das previsões. No entanto, no contexto deste estudo, optou-se por seguir uma abordagem fixa com base nas diretrizes da literatura.

De facto, a revisão da literatura, especificamente a partir do (Antoniolli et al., n.d.), mostra que o conjunto de treino geralmente deve abranger de 60% a 90% dos dados, enquanto o conjunto de teste deve corresponder a 10% a 40% dos dados disponíveis.

Neste caso, os dados foram armazenados ao longo de dois anos e foram analisados numa base mensal, resultando num conjunto de dados com 24 meses. Portanto, decidiu-se que o conjunto de treino seria composto pelos dados do período de janeiro de 2021 a junho de 2022, inclusive, enquanto o conjunto de teste seria composto pelos dados restantes, ou seja, de julho de 2022 a dezembro de 2022. Desta forma, optou-se por um valor intermédio ao que a revisão literatura introduz, em particular, 75% dos dados para o conjunto de treino, e os restantes, 25% para o conjunto de teste.

Calcular os parâmetros ótimos

A construção do modelo SARIMA envolve o cálculo de vários parâmetros cruciais, tais como p , d , q , P , D , Q e s . Os parâmetros d e D já foram calculados, portanto resta determinar os restantes parâmetros.

Uma abordagem comum é a análise visual dos gráficos de ACF e PACF. No entanto, estes gráficos apenas serviram de ponto de partida para a parametrização do modelo, pois, em conjunto, foi utilizada uma metodologia mais eficiente, aproveitando as bibliotecas disponíveis no *Python*, especialmente a *pmdarima*.

Com a biblioteca *pmdarima*, é possível automatizar a seleção dos melhores parâmetros de forma eficaz. Através da função *autoarima*, o melhor modelo é selecionado com base nas classificações do critério de informação de *Akaike* (AIC), o modelo escolhe automaticamente as configurações que resultam no menor valor de AIC, como mostrado na revisão literatura deste método.

É relevante indicar que esta função já foi utilizada na literatura, por (Deretić et al., 2022) e (Brendan Artley, 2022), tornando-a uma metodologia confiável. Além disso, o facto de ser automática a seleção do melhor modelo simplifica significativamente o processo de construção do modelo SARIMA, economizando tempo e recursos.

Seguidamente, é apresentado na Figura 31, a aplicação da função *auto_arima* para escolha do melhor modelo tendo por base o critério AIC.

```
print('-----MODELO SARIMA-----')
# Ajustar o modelo SARIMA com os parâmetros selecionados usando os dados de treino
SARIMA_model = pm.auto_arima(train_data[variavel_x],
                             start_p=0, start_q=0,
                             #test='adf',
                             max_p=3, max_q=3,
                             m=m, # Frequência do ciclo (sazonalidade anual)
                             start_P=0, start_Q=0,
                             max_P=3, max_Q=3,
                             seasonal=Seasonal, # Sazonalidade definida como True
                             d=d, # Diferenciação não sazonal de primeira ordem
                             D=D, # Diferenciação sazonal de primeira ordem
                             trace=False,
                             error_action='ignore',
                             suppress_warnings=True,
                             stepwise=True,
                             #
                             information_criterion='aic')

# Visualizar os parâmetros ótimos encontrados
print("Parâmetros Ótimos do Modelo SARIMA:")
print(SARIMA_model.order) # (p, d, q) do modelo
print(SARIMA_model.seasonal_order) # (P, D, Q, s) do modelo (caso sazonal)

# Valor do AIC
print("Valor do AIC:")
print(SARIMA_model.aic())
```

Figura 31: Utilização da biblioteca *auto_arima* para seleção do modelo

Pela Figura 31, é visível a utilização da função `auto_arima`, para selecionar o melhor modelo.

A literatura evidencia que os parâmetros com valores elevados, aumentam a complexidade do modelo. O limite dos parâmetros é estudado mais à frente, porém, é relevante indicar que, recorreu-se aos gráficos ACF e PACF para verificar a complexidade dos dados, ou seja, estudar a quantidade de pontos significativos, para estabelecer um limite razoável dos parâmetros.

Diagnóstico do modelo

A fase de diagnóstico do modelo foca-se em perceber a qualidade do ajuste do modelo, ou seja, perceber se são necessárias alterações nos valores dos parâmetros encontrados nas fases anteriores.

Desta forma, para a etapa de diagnóstico do modelo recorreu-se à função `plot_diagnostics` disponível no *Python*. Mais uma vez, esta técnica de desenvolver o diagnóstico do modelo foi também utilizada por (Brendan Artley, 2022) , (P. Chen et al., 2018).

Esta função retorna quatro gráficos, nomeadamente, gráfico dos resíduos ao longo do tempo, histograma dos resíduos e gráfico da distribuição KDE, gráfico normal QQ dos resíduos e gráfico de autocorrelação dos resíduos.

Efetivamente, (Thomaz & Favarini, 2018), menciona que um bom modelo de previsão deve ser capaz de gerar resíduos com características de ruído branco. Portanto, a fase de diagnóstico avalia o comportamento dos resíduos para verificar o ajuste do modelo. É importante, ainda evidenciar características em cada gráfico que mostrem que os resíduos admitem comportamento de ruído branco, e para tal, recorreu-se à literatura. De uma forma breve, na Figura 32 é evidenciado a estrutura que cada gráfico deve admitir para se tratar de resíduos com características de ruído branco, de acordo com (P. Chen et al., 2018) E (Brendan Artley, 2022).

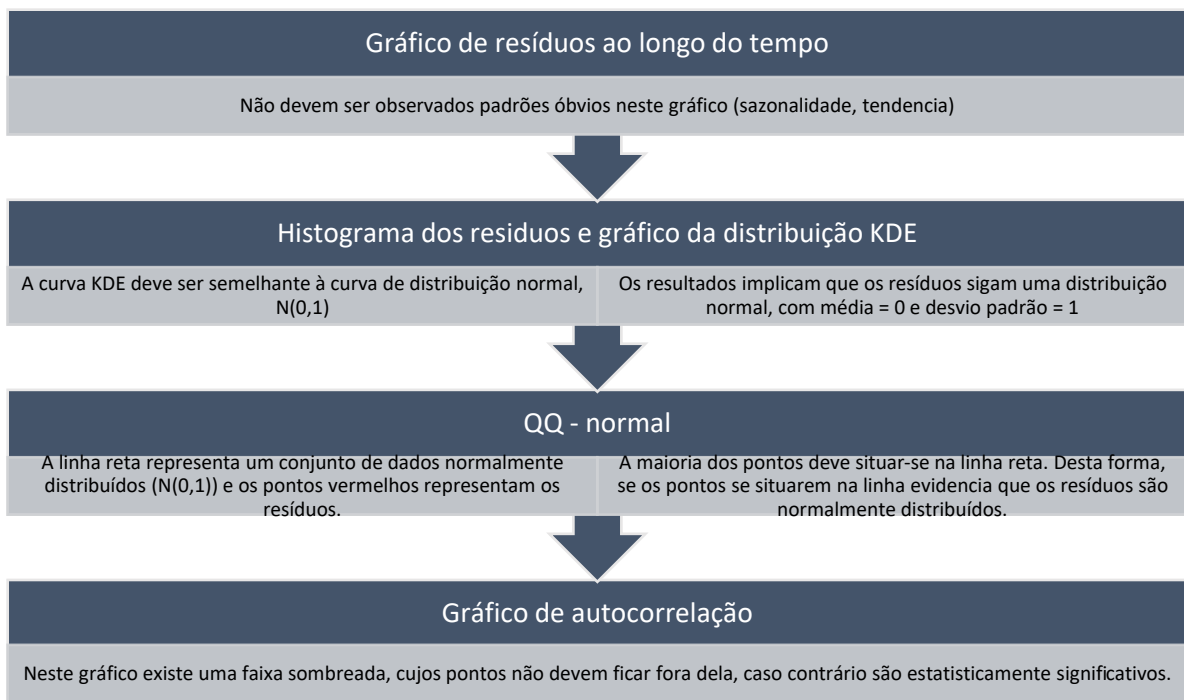


Figura 32: Estrutura dos gráficos de diagnóstico para admitir características de ruído branco

Previsões para o conjunto teste e avaliação

Caso os resíduos gerados pelo modelo admitam características de ruído branco, o modelo está apto para calcular as previsões para o conjunto de teste.

Assim, através da biblioteca *pandas* é calculada a quantidade de períodos de previsão e armazenada no variável *n_periods*. Neste caso, o intervalo admite a dimensão do conjunto de teste. O objetivo é preparar um período de previsão que inclua todos os meses a partir do primeiro mês do conjunto de teste.

De seguida, utiliza-se o modelo SARIMA previamente ajustado para calcular as previsões para o número de períodos que a variável *n_periods* admite. Também é calculado um intervalo de confiança para as previsões, admitindo assim um intervalo de variação.

Seguidamente, são comparados os valores de previsão para o conjunto de teste pelo modelo SARIMA ajustado, com os valores reais do conjunto de teste, através de um gráfico de linhas.

Para uma fácil verificação e análise dos gráficos, definiu-se um limite vertical para o gráfico entre -2 e 2, e ainda, tem-se que:

- Conjunto de treino: linha azul
- Conjunto de teste: linha verde
- Conjunto de previsões: linha vermelha
- Área sombreada: sombreada a rosa

O gráfico é guardado segundo uma referência para posterior utilização. Por fim, o modelo é avaliado de acordo com as métricas evidenciadas na revisão literatura:

- Erro Médio Absoluto (MAE): quanto menor melhor

- Raiz do erro Médio Quadrático (RMSE): quanto mais próximo de zero melhor

De salientar, um pequeno pormenor, os valores de previsão estão com um período de avanço. O cálculo de MAE e RMSE são calculados sobre os valores do conjunto de teste e os valores previstos pelo modelo, portanto, estes dois conjuntos encontram-se com um período de diferença. Assim sendo, foram calculadas as datas em comum dos dois conjuntos para posteriormente serem calculados MAE e RMSE. A lógica está evidenciada na função `Previsao_Sarima_treino_teste` apresentada no Anexo E.

Neste ponto do projeto, foram construídas as previsões para o conjunto de teste utilizando o modelo SARIMA previamente ajustado. Essas previsões foram comparadas com os valores reais do conjunto de teste usando métricas de avaliação, a fim de avaliar a precisão e a qualidade dos valores previstos pelo modelo. Além disso, as projeções foram visualizadas num gráfico.

Previsões futuras

Como último passo na análise e previsão de séries temporais, foi crucial desenvolver estratégias para dados futuros com base nos dados fornecidos. Para atingir esse objetivo, foi desenvolvida a função `fazer_previsoes_futuras`, que utiliza os parâmetros previamente ajustadas pelo modelo SARIMA criado anteriormente. De notar, que a função `fazer_previsoes_futuras` está evidenciada no Anexo E.

É fundamental ressaltar que, durante todo o projeto, foi dedicada uma atenção especial à qualidade dos parâmetros do modelo SARIMA. Inicialmente, foram ajustados com base em dados históricos, em particular, o conjunto de treino e teste. Essa etapa de ajuste incluiu uma escolha de critérios de parâmetros não sazonais (p, d, q) e sazonais (P, D, Q, s) para garantir que o modelo representasse com precisão os padrões subjacentes nos dados.

Além disso, o modelo foi submetido a testes rigorosos de diagnóstico e avaliação. Esse processo garantiu que o modelo selecionado fosse ótimo e adequado para o conjunto de dados específicos. Qualquer ajuste ou refinamento necessário foi realizado nessa fase, garantindo que o modelo estivesse na sua melhor forma. Por fim, com recurso aos parâmetros encontrados e à função mencionada anteriormente, foram desenvolvidas as previsões futuras das variáveis necessárias, com o objetivo de identificar falhas.

4. RESULTADOS E ANÁLISES

Ao longo deste capítulo serão apresentados os resultados das análises desenvolvidas, nomeadamente, a análise dos valores nulos, resultados da análise exploratória, resultados de cada um dos modelos e por fim é apresentada uma discussão geral dos mesmos.

Por fim, de salientar que a Tabela 2 revela as variáveis de maior interesse para o projeto. A análise exploratória recaiu sobre as diversas variáveis, mas a implementação dos modelos apenas se justificava para as variáveis de relvo para a questão do projeto evidenciadas na Tabela 2. Apesar de que, no caso das redes neuronais não foi possível aplicar a todas as variáveis importantes.

4.1. Resultados da análise nulos

Uma análise visual dos dados permitiu aferir que o conjunto de dados original fornecido admitia muitos valores nulos, portanto, antes da análise exploratória das variáveis, foram analisados os valores nulos, através da função `contar_nulos`. Desta forma, a análise de nulos para além de servir para contabilizar os valores nulos que cada variável admite, também foi utilizada para emitir alarmes de falhas de comunicação como já explicado.

Desta forma, a Tabela 5 indica a quantidade de nulos que cada variável admite no conjunto total dos dados para cada ano observado e ainda é apresentado o valor total.

Tabela 5 : Análise da quantidade de valores nulos do conjunto original de dados por variável

Variável	Quantidade de Nulos totais	Nulos - 2021	Nulos - 2022
E_TOTAL_INVERSOR_A [kWh]	9142	7637	1505
P_AC_INV_A [W]	9142	7637	1505
P_DC_INV_A [W]	9142	7637	1505
T_INV_A [°C]	9142	7637	1505
E_TOTAL_INVERSOR_B [kWh]	9272	7623	1649
P_AC_INV_B [W]	9272	7623	1649
P_DC_INV_B [W]	9272	7623	1649
T_INV_B [°C]	9272	7623	1649
P_AC_SAIDA_REDE [W]	8938	7488	1450
IRR_MEDIA [Wh/m ²]	2871	2701	170
IRE [Wh/m ²]	2871	2701	170
IRO [Wh/m ²]	2871	2701	170
T_MEDIA [°C]	2871	2701	170
T_ESTE [°C]	2871	2701	170
T_OESTE [°C]	2871	2701	170
Eficiencia_INV_A	9142	7637	1505
Eficiencia_INV_B	9272	7623	1649

4.2. Resultados da análise exploratória dos dados

Neste capítulo são apresentados os dados da análise exploratória desenvolvida para as diversas variáveis em estudo.

Análise dos valores de temperatura do módulo a este e oeste

De seguida, foram analisados os valores da temperatura dos módulos a este e temperatura dos módulos a oeste, para os anos de 2021 e 2022 em separado, através das variáveis T_ESTE [°C], T_OESTE [°C], respetivamente.

Inicialmente são apresentados os resultados para a temperatura dos módulos a Este para os dois anos referidos ao mesmo tempo, Figura 33. Foi também desenvolvido um gráfico de barras para cada variável relativa à temperatura média mensal do módulo a Este para o ano de 2021 e 2022, representados na Figura 130 e apresentada no Anexo F.

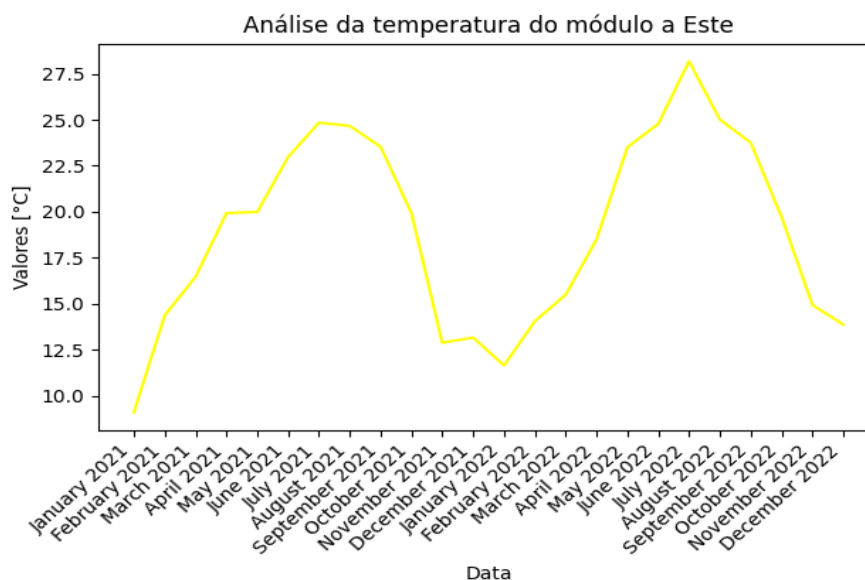


Figura 33 : Análise dos valores médios mensais da temperatura dos módulos a Este para os dois anos

Pela análise da Figura 33 e Figura 130, facilmente se percebe que os valores de temperatura média mensais, aumentam de janeiro a julho, e até dezembro diminuem. Este padrão reflete-se nos dois anos de análise, e, portanto, pode-se concluir que os dados admitem sazonalidade anual, visto que, o padrão de comportamento repetiu-se nos dois anos de forma semelhante.

A análise das figuras permite também comprovar que a temperatura máxima, registada em julho, aumentou de 2021 para 2022, para além disso a temperatura de janeiro e dezembro de 2021 (representam as temperaturas mais baixas no ano) também aumentaram em 2022.

A Figura 33 auxilia também na identificação de tendências na série temporal. A análise do gráfico de linhas, Figura 33, permite, efetivamente, perceber que no geral, os valores aumentaram de 2021 em relação a 2022, e, portanto, podemos estar perante uma tendência crescente. No entanto, é

bom lembrar que os dados passaram por um processo de transformação, com substituição de nulos que pode ter influenciado as características dos dados. Para além disso, o conjunto de dados é muito limitado, apenas são analisados dois anos de observações, assim podemos concluir que admitem sazonalidade anual e possivelmente uma tendência crescente.

De seguida, foram analisados os valores de temperatura média mensal do módulo a oeste para 2021 e 2022, representados no gráfico de linhas, Figura 34. Para além disso, foram construídos gráfico de barras com análise da mesma variável, para cada ano, representados na Figura 131 presente no Anexo F.

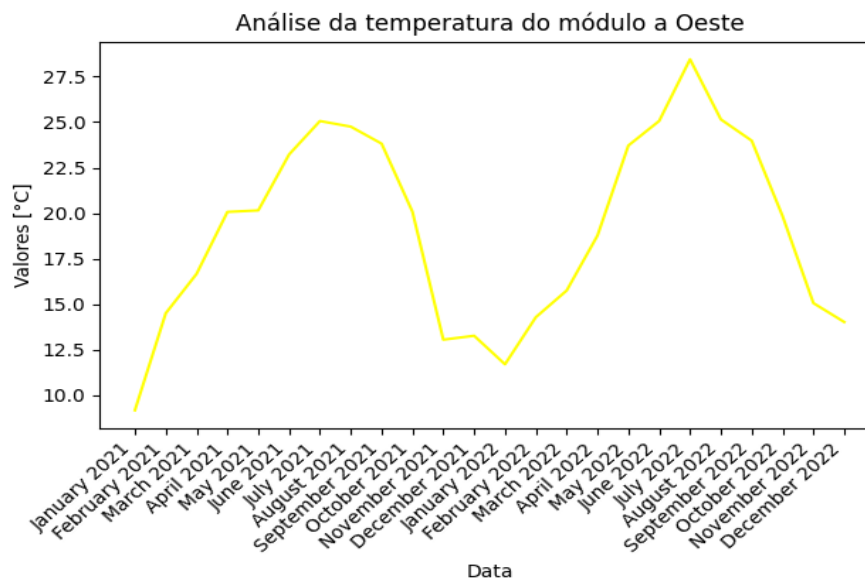


Figura 34 : Análise dos valores médios mensais da temperatura dos módulos a Oeste para os dois anos

Pela observação da Figura 34 e da Figura 131, é possível admitir que o comportamento dos valores de temperatura médios mensais do módulo a oeste é o mesmo que o comportamento dos valores de temperatura médios mensais do módulo a este. Assim, pela Figura 34 verificamos que a série temporal admite sazonalidade anual, e o valor máximo e mínimo de temperatura aumentaram de 2021 para 2022, representados em julho e janeiro, respetivamente.

A Figura 34, transmite a ideia de uma tendência crescente dos valores de temperatura a oeste, no entanto, é importante ter em conta que o conjunto de dados é limitado e os dados forma transformados.

Análise dos valores do sensor de radiação a este e oeste

De seguida, são abordados os valores do sensor de radiação a este e oeste para o ano de 2021 e 2022 (Figura 35). Inicialmente são abordados os valores do sensor de radiação a Este na Figura 132 apresentada no Anexo D.

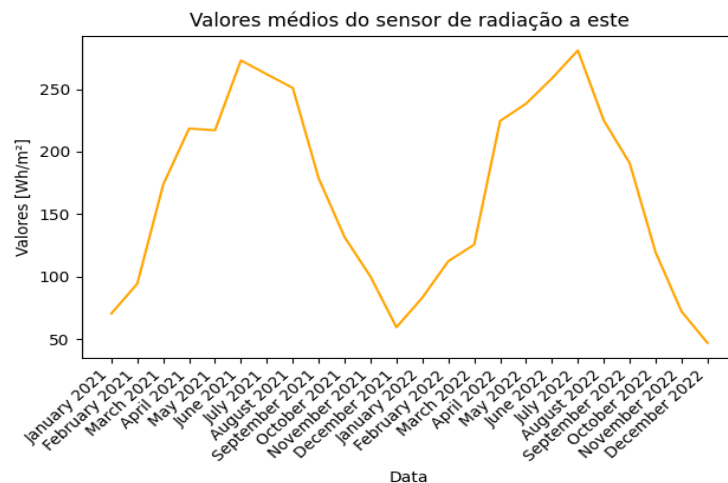


Figura 35 : Análise dos valores médios mensais da radiação a Este para os dois anos

O comportamento dos valores médios mensais do sensor de radiação a este em 2021 e 2022 são semelhantes, observando a Figura 35 . Pela análise da Figura 132 , percebe-se que os valores desta variável aumentam junho/julho e a partir desta data começa a decrescer, o valor máximo nos dois anos foi atingido em meses diferentes. Pela Figura 35 o comportamento dos dados nos dois anos de análise parece semelhante, pelo que para já, a série não apresenta tendência.

De seguida, são analisados os valores do sensor de radiação a Oeste, representados por um gráfico de linhas na Figura 36 e pelos gráficos de barras desenvolvidos para o ano de 2021 e 2022, na figura Figura 133 no Anexo F.

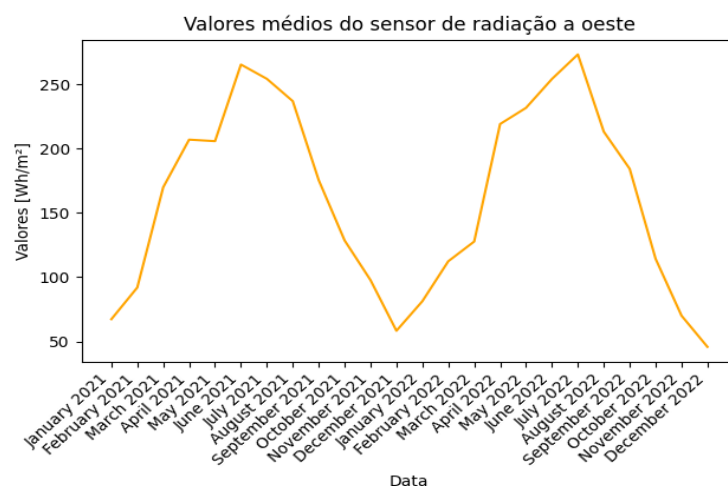


Figura 36 : Análise dos valores médios mensais da radiação a Oeste para os dois anos

A Figura 36 e a Figura 133, mostram que os valores médios mensais do sensor de radiação a oeste admitem comportamentos semelhantes para os dois anos de estudo, e portanto, a série admite

sazonalidade anual. Por fim a Figura 36, revela que para já a série não admite tendência e comprova o efeito de sazonalidade anual.

Análise mensal dos dados para o inversor A

De seguida é apresentada uma análise geral de variáveis relativas ao inversor A.

Inicialmente são abordados os valores da energia acumulada convertida pelo inversor A nos dois anos de análise, Figura 37.

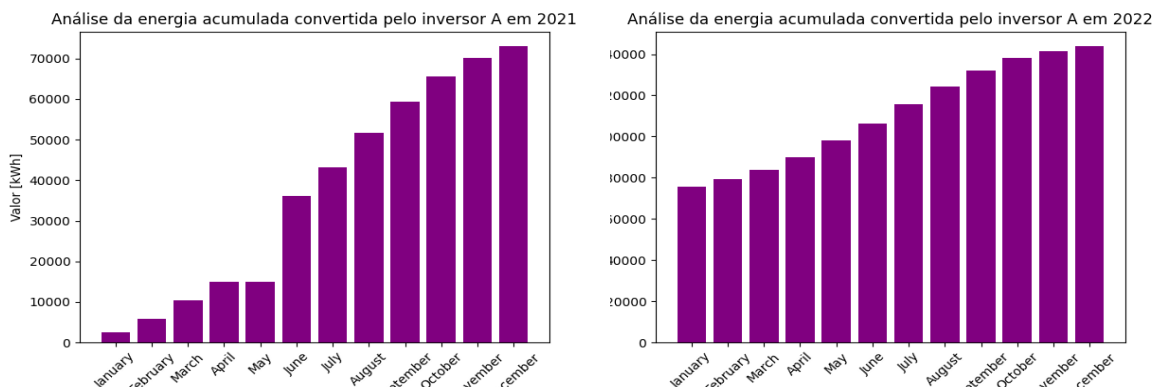


Figura 37 : Análise dos valores médios mensais da energia acumulada convertida pelo inversor A em 2021 (lado esquerdo) e 2022 (lado direito)

A análise desta variável destaca uma possível falha de comunicação que ocorreu no ano de 2021, especificamente no mês de janeiro. Esta suspeita é sustentada pela seguinte observação: a energia acumulada convertida pelo inversor A em janeiro de 2021, Figura 37, foi significativamente diferente daquela observada no mesmo período em 2022, Figura 37. Esta discrepância chama a atenção para a possibilidade de um problema de comunicação de dados.

A Tabela 5 também fornece insights importantes para a análise. Pois indica que houve uma quantidade significativamente maior de valores nulos nos dados de 2021 em comparação com 2022, em particular, em janeiro. No entanto, é importante notar que esses valores nulos já foram tratados. Como mencionado anteriormente, os valores nulos foram substituídos pelas médias mensais diurnas ou noturnas. Devido à grande quantidade de valores nulos, é normal que essas médias resultem em valores relativamente baixos

De seguida são analisados e apresentados os gráficos dos valores de potência do lado AC e DC do inversor A, representado na Figura 38.

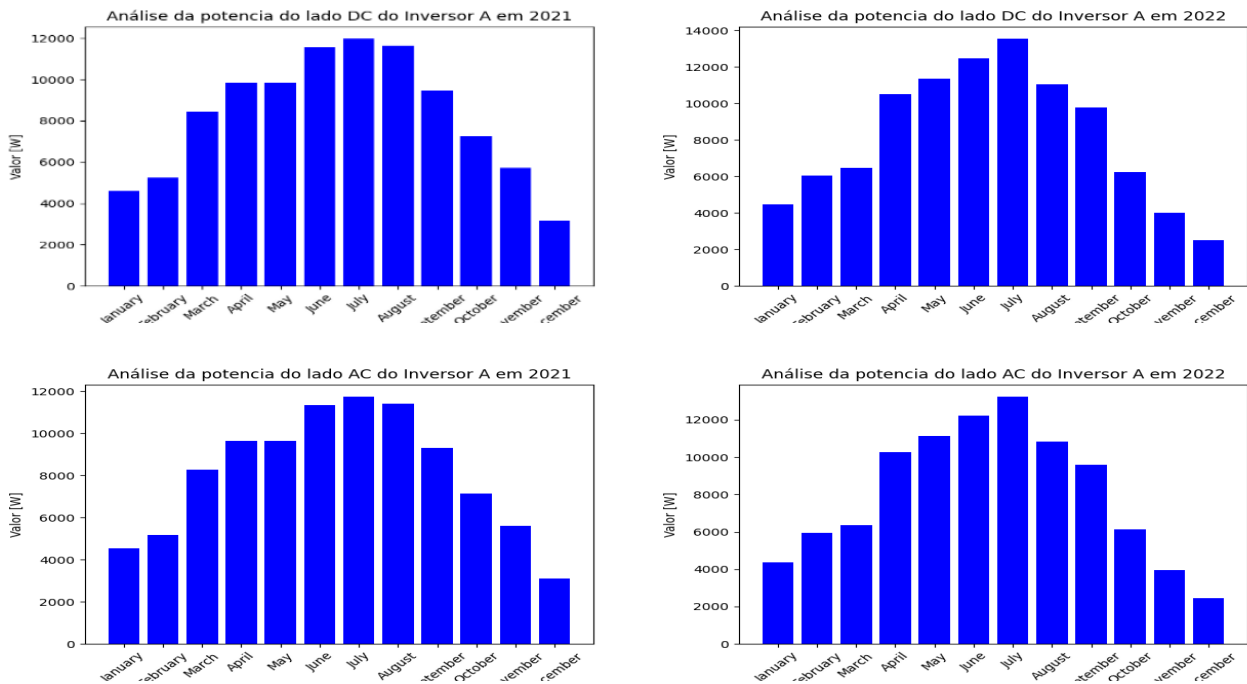


Figura 38 : Valores médios mensais da potência do lado DC em 2021 (canto superior esquerdo), em 2022 (canto superior direito) , Valores médios mensais da potência do lado AC em 2021 (canto inferior esquerdo), em 2022 (canto inferior direito) para o inversor A

Pela observação das figuras é perceptível que ambas as variáveis admitem sazonalidade anual. De salientar também que no ano de 2022 a potência máxima do lado DC foi maior do que no ano de 2021, e conseqüentemente, o mesmo aconteceu para a variável relativa à potência do lado AC do inversor, como seria de esperar. Caso não acontecesse a eficiência do inversor não seria a expectável. Por fim, é importante evidenciar a análise relativa à temperatura do inversor A, é uma característica que pode evidenciar perdas de eficiência no equipamento, Figura 39 .

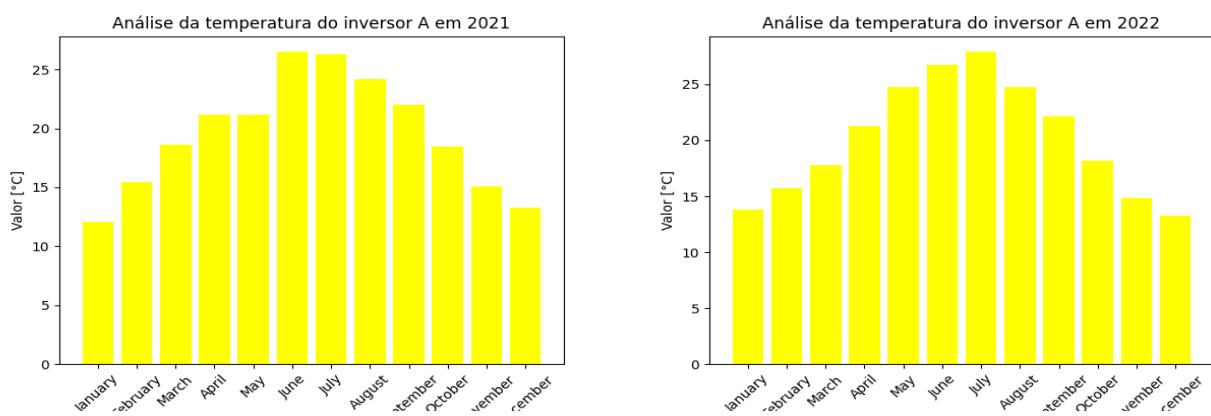


Figura 39 : Análise dos valores de temperatura médios mensais do Inversor A em 2021 (lado esquerdo) em 2022 (lado direito)

Esta variável admite sazonalidade anual, como seria de esperar. Pois é normal que a temperatura seja mais elevada em meses mais quentes, e mais baixa em meses mais frios.

Análise mensal dos dados para o inversor B

Inicialmente procedeu-se à análise da potência lado DC e AC do inversor B para os dois anos de análise, representados na Figura 40.

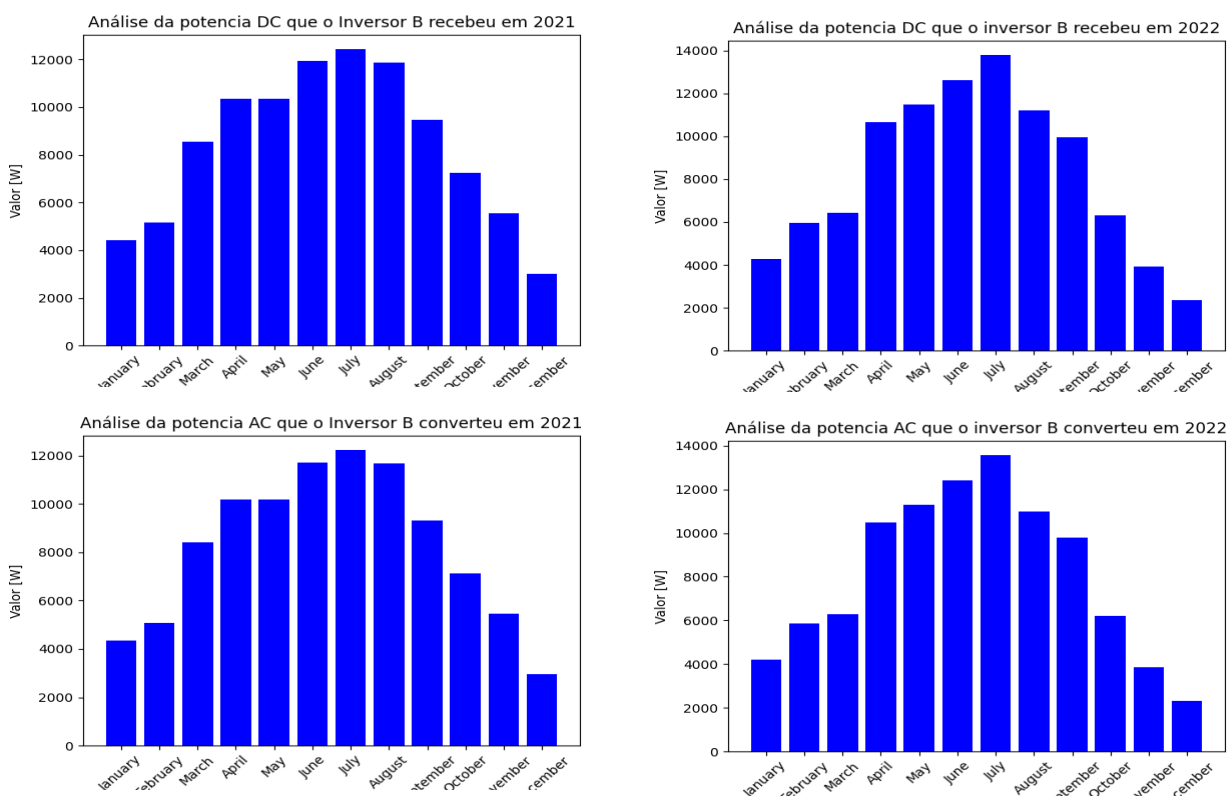


Figura 40 : Valores médios mensais da potência do lado DC em 2021 (canto superior esquerdo), em 2022 (canto superior direito), Valores médios mensais da potência do lado AC em 2021 (canto inferior esquerdo), em 2022 (canto inferior direito) do Inversor B

Pela análise da Figura 40, é perceptível que as duas variáveis também admitem sazonalidade anual como seria esperado, visto que, apresentam comportamentos semelhantes entre o ano de 2021 e 2022, com subidas dos valores quando esperado, e depois, decréscimo dos mesmos. De salientar, também, que pela análise da Figura 40 que não existiram perdas de eficiência aparentemente, visto que o gráfico dos valores do lado DC é semelhante ao gráfico de valores do lado AC para os dois anos (do lado esquerdo a figura de cima é semelhante à figura de baixo e o mesmo para o lado direito) e o processo implícito que distingue as duas variáveis é a conversão do lado DC em AC através do inversor B, neste caso.

Por fim, é avaliada a temperatura do inversor B, Figura 41, pelo que mais uma vez, podemos aferir sazonalidade anual da série.

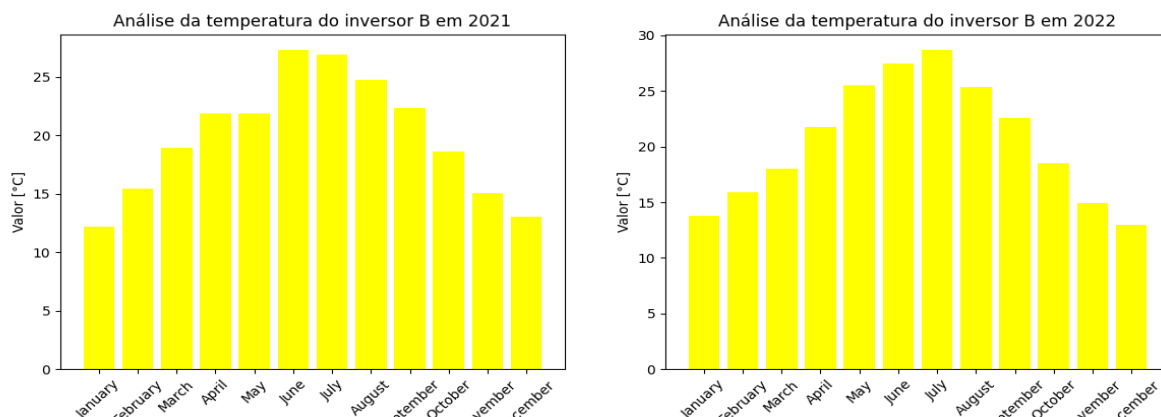


Figura 41 : Análise dos valores de temperatura médios mensais do Inversor B em 2021 (lado esquerdo) em 2022 (lado direito)

Com o objetivo de enfatizar as conclusões relativas com a eficiência dos dois inversores, foi construído o gráfico que mostra a média da eficiência mensal para cada inversor, em cada ano. Pela análise da Figura 134 e Figura 135 é possível verificar que para qualquer ano, o valor desta variável permanece nos valores esperados, nos dois inversores. As figuras relativas a esta conclusão encontram-se no Anexo F.

Análise mensal da irradiância média, temperatura média e potência de saída

Na Figura 42 é representado o comportamento dos valores de radiação média para os dois anos de observação.

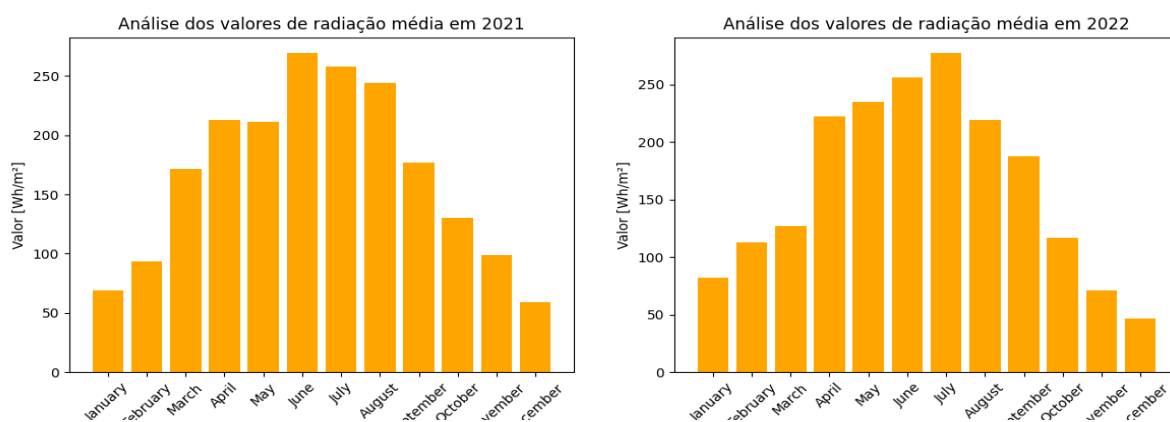


Figura 42 : Análise dos valores médios mensais de irradiância média em 2021 (lado esquerdo) e 2022 (lado direito)

Numa análise geral das figuras, facilmente se percebe que nos dois anos a temperatura média registada aumenta até aos meses de verão, depois decresce até dezembro, e, portanto, a série temporal relativa à variável da radiação média admite sazonalidade anual. Para além disso, salientar que os valores da radiação média para o ano de 2022, decresceram de forma mais rápida do que no mesmo período para o ano de 2021. No entanto, seriam necessários mais anos de análise para uma conclusão sustentada.

De seguida, são apresentados graficamente os valores de temperatura média registados para 2021 e 2022, Figura 43 .

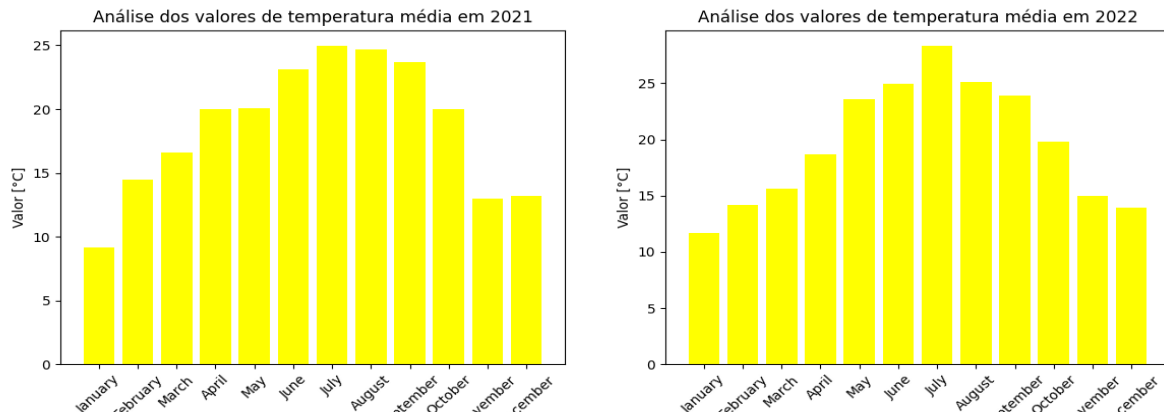


Figura 43 : Análise dos valores médios mensais de temperatura média em 2021 (lado esquerdo) e 2022 (lado direito)

Os gráficos ilustrados na Figura 43 indicam sazonalidade anual, como as variáveis anteriormente abordadas. Em 2022 (lado direito da Figura 43) é evidente um aumento da temperatura média seguido, de um decréscimo de forma mais acentuada, em comparação com 2021(lado esquerdo da Figura 43). Para além disso, os valores de temperatura em 2022 são geralmente mais altos do que em comparação com 2021 em cada mês.

Por fim, foi desenvolvido um estudo entre as variáveis relativas à temperatura média, radiação média e potência do lado AC do inversor A para os anos de 2021 e 2022.

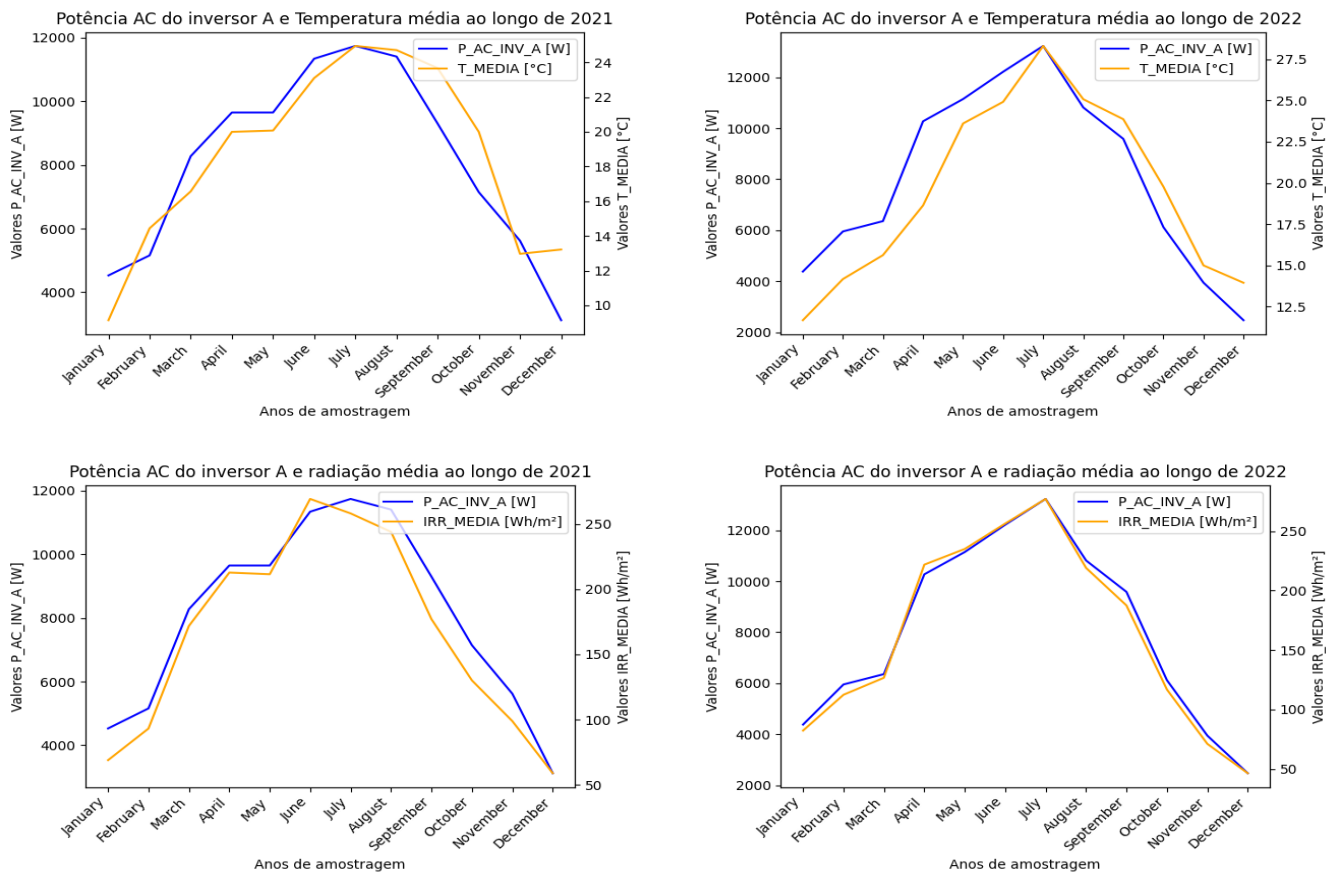


Figura 44 : Influência da temperatura na potência do lado AC ao longo de 2021 (em cima lado esquerdo) e 2022 (em cima lado direito) e influência da irradiância na potência do lado AC ao longo de 2021 (em cima baixo esquerdo) e 2022 (em baixo lado direito)

Pela análise da Figura 44 , verificamos que tanto a temperatura média como a irradiância media acompanham o comportamento da potência do lado AC do inversor A durante os dois anos de observação.

Ao analisar os gráficos, é possível observar que tanto a temperatura média quanto a radiação média seguem o comportamento da potência no lado AC do inversor A durante os dois anos de observação, o que seria expectável. De facto, normalmente, à medida que a temperatura aumente e a radiação também, é expectável que a produção de energia aumente se em nenhum dos equipamentos existir perdas de eficiência.

O estudo desenvolveu-se sobre os valores de potência no lado AC do inversor, uma vez que esses dados já foram convertidos e estão prontos para uso pessoal. Também permitiu confirmar que as transformações aplicadas aos dados não afetaram a lógica dos mesmos, como a análise de nulos.

4.3. Resultados Redes neuronais

A rede MLP (Perceptron de Múltiplas Camadas) foi o primeiro modelo a ser implementado nesta pesquisa. No entanto, devido à falta de conhecimento inicial, o desempenho do modelo MLP não

atingiu os resultados desejados. Isso levou à decisão de explorar alternativas, culminando na construção do modelo SARIMA, que será detalhado posteriormente neste trabalho.

Portanto, o modelo de redes neurais foi inicialmente implementado apenas para uma variável e os resultados obtidos não foram esmagadores. Consequentemente, a decisão de empregar o modelo SARIMA foi tomada.

À medida que a implementação do modelo SARIMA avançava, retornamos ao modelo de redes neurais e identificamos o erro na sua configuração original. Assim, ficou evidente que o desempenho do modelo de redes neurais foi melhorado significativamente. No entanto, devido a restrições de tempo, optamos por manter a implementação do modelo de redes neurais focadas na variável inicialmente selecionada para esta análise.

Essa abordagem incluiu uma avaliação mais aprofundada do desempenho do modelo SARIMA, enquanto ainda reconhecíamos o potencial das redes neurais como uma alternativa viável para futuras investigações.

Assim sendo, neste ponto serão apresentados os testes referentes à única variável testada deste modelo, que foi a variável relativa à temperatura do inversor A em graus celsius, com o objetivo de prever perdas de eficiência do equipamento.

Na Figura 45, são lembrados os passos implementados no desenvolvimento de uma rede neuronal MLP.

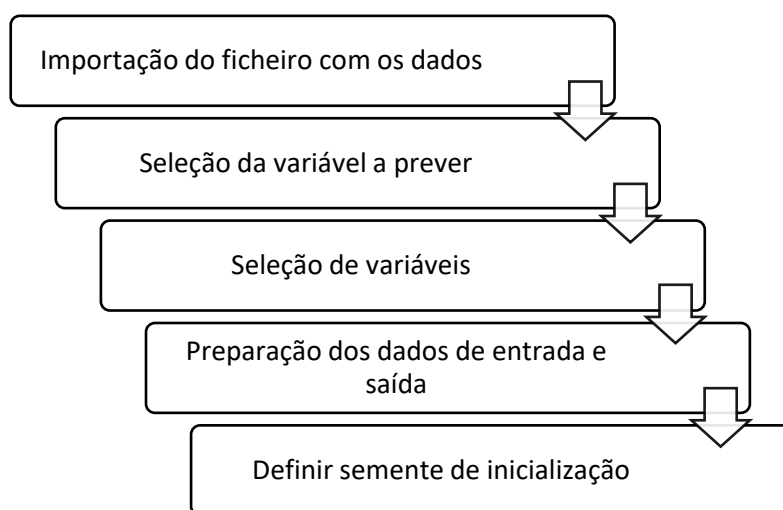


Figura 45 : Etapas de construção de uma rede neuronal multicamadas (MLP)

Pela análise da Figura 22, verificamos os parâmetros cujos valores foram encontrados através de teste. Na abordagem delimitada, inicialmente, foi testada a quantidade de neurónios em cada camada. De seguida, foram realizados testes para a quantidade de camadas e por fim para os parâmetros relativos à época de treino e ao tamanho do lote. Assim sendo, de seguida são apresentados os resultados dos testes para construir uma rede neuronal MLP capaz de prever valores da temperatura do inversor A de qualidade.

Primeiro teste

Na Tabela 6 são apresentados os valores dos parâmetros neste primeiro teste.

Tabela 6 : Parâmetros da rede do primeiro teste

Estrutura da rede	
Quantidade de camadas	Quantidade de neurónios
Camada 1	6
Camada 2	8
Camada 3	10
Camada 4	12
Camada 5	1
Outros parâmetros	
<i>Batch_size</i>	128
<i>Epochs</i>	80

Os valores do Erro Absoluto Médio (MAE) e do Erro Quadrático Médio (MSE) são apresentados de seguida, Tabela 7.

Tabela 7 : Valores de MAE e MSE para o primeiro teste

Avaliação do desempenho do modelo	
MAE	MSE
0.33862	0.178825

De seguida, é apresentada uma comparação gráfica entre os valores reais e os previstos, Figura 46.

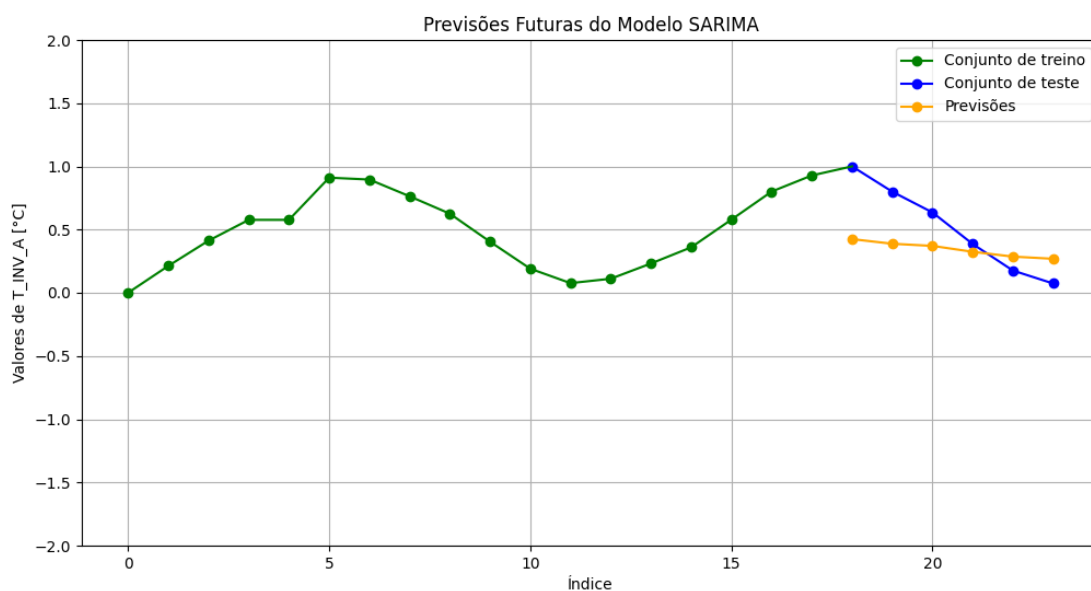


Figura 46 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do primeiro teste

Pela análise da Figura 46, verificamos que o modelo não apresenta resultados satisfatórios e, portanto, procedeu-se à alteração da quantidade de neurónios em cada camada. Na Tabela 8 são apresentados os parâmetros para um novo teste.

Tabela 8 : Parâmetros da rede do segundo teste

Estrutura da rede	
Quantidade de camadas	Quantidade de neurónios
Camada 1	18
Camada 2	16
Camada 3	20
Camada 4	12
Camada 5	1
Outros parâmetros	
<i>Batch_size</i>	128
<i>Epochs</i>	80

Os valores do Erro Absoluto Médio (MAE) e do Erro Quadrático Médio (MSE), para o segundo teste, são apresentados de seguida, Tabela 9.

Tabela 9 : Valores de MAE e MSE para o segundo teste

Avaliação do desempenho do modelo	
MAE	MSE
0.15030	0.02936

De seguida, é apresentada a comparação gráfica entre os valores reais e os previstos, Figura 47.

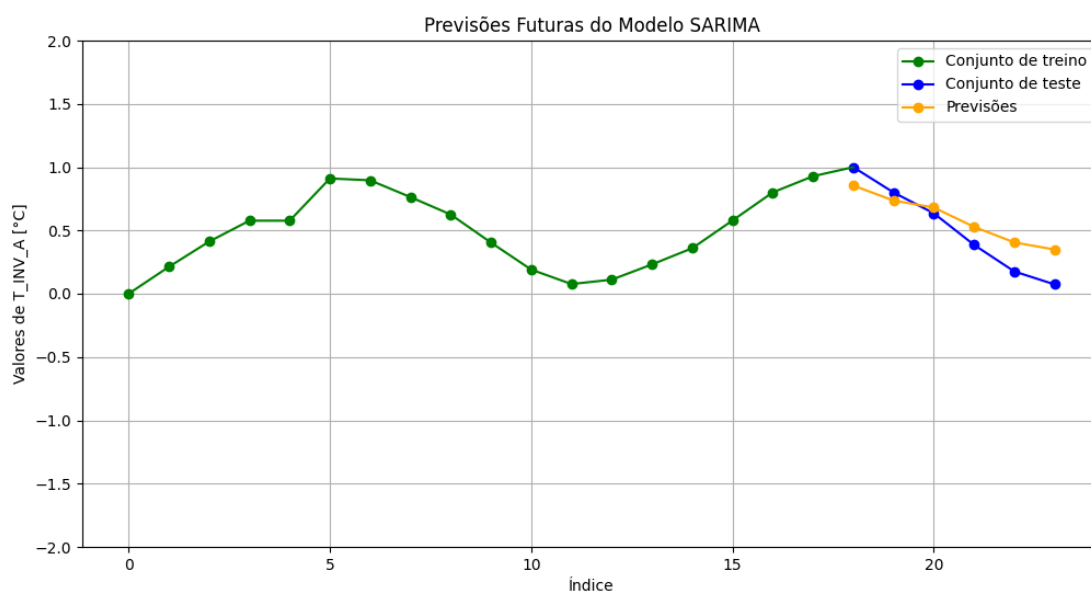


Figura 47 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do segundo teste

Comparando os valores de MAE e MSE entre o primeiro e o segundo teste, representados na Tabela 7 e Tabela 9, respetivamente, é perceptível que diminuíram. Assim sendo, podemos concluir que o aumento do número de neurónios em cada camada melhorou o desempenho do modelo. Assim sendo, no próximo teste estes valores vão ser novamente aumentados para perceber o comportamento do modelo e conseqüentemente, o comportamento dos valores previstos. Na Tabela 10 são retratados os parâmetros para o terceiro teste.

Tabela 10 : Parâmetros da rede do terceiro teste

Estrutura da rede	
Quantidade de camadas	Quantidade de neurónios
Camada 1	30
Camada 2	28
Camada 3	26
Camada 4	24
Camada 5	1
Outros parâmetros	
<i>Batch_size</i>	128
<i>Epochs</i>	80

Os valores do Erro Absoluto Médio (MAE) e do Erro Quadrático Médio (MSE), para o segundo teste, são apresentados de seguida, Tabela 11.

Tabela 11 : Valores de MAE e MSE para o terceiro teste

Avaliação do desempenho do modelo	
MAE	MSE
0.102615	0.013025

De seguida, é apresentada a comparação gráfica entre os valores reais e os previstos, Figura 48.

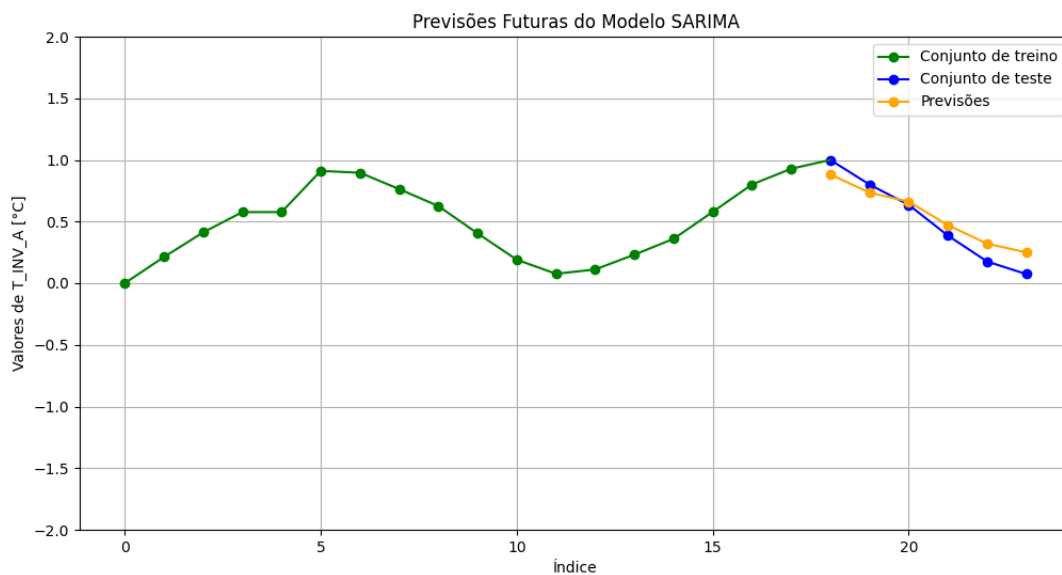


Figura 48 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do terceiro teste

Mais uma vez, por comparação dos resultados de MAE e MSE Obtidos no segundo e terceiro testes, representados na Tabela 9 e Tabela 11, percebe-se facilmente que o desempenho do modelo melhorou.

Tendo por base a ordem de ideias apresentada anteriormente, para o próximo teste torna-se a aumentar o número de neurónios para perceber se continua a melhorar o modelo, e mesmo, se o aumento de desempenho do modelo justifica a complexidade da rede. Os parâmetros são evidenciados na Tabela 12.

Tabela 12 : Parâmetros da rede do quarto teste

Estrutura da rede	
Quantidade de camadas	Quantidade de neurónios
Camada 1	40
Camada 2	38
Camada 3	36
Camada 4	34
Camada 5	1
Outros parâmetros	
<i>Batch_size</i>	128
<i>Epochs</i>	80

Os valores do Erro Absoluto Médio (MAE) e do Erro Quadrático Médio (MSE), para o quarto teste, são apresentados de seguida, Tabela 13.

Tabela 13 : Valores de MAE e MSE para o quarto teste

Avaliação do desempenho do modelo	
MAE	MSE
0.089667	0.010370

De seguida, é apresentada a comparação gráfica entre os valores reais e os previstos, Figura 49.

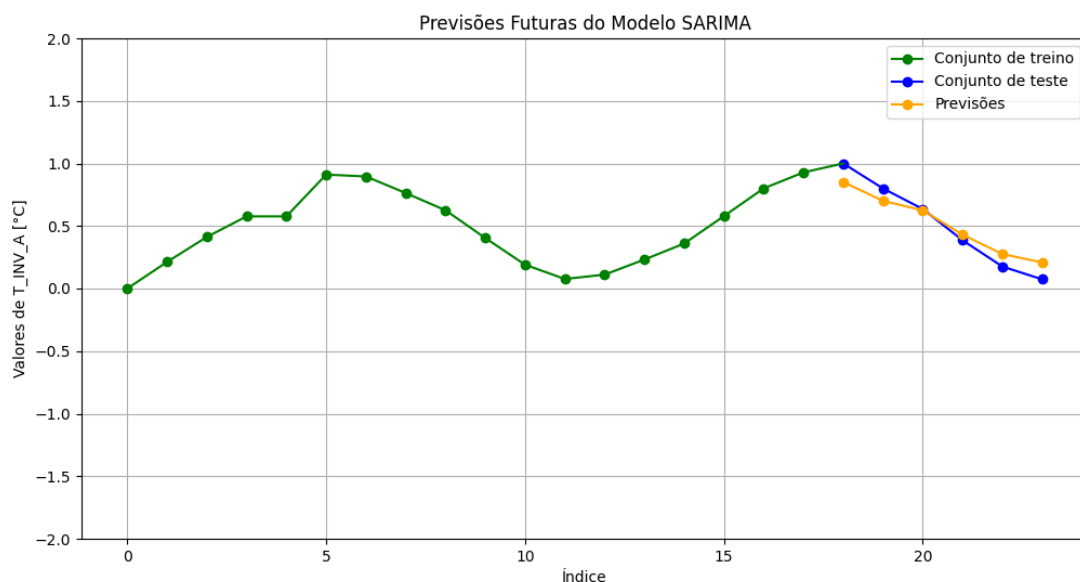


Figura 49 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do quarto teste

Pela análise dos valores de MAE e MSE do terceiro e quarto teste, representados, pelas Tabela 11 e Tabela 13, respetivamente, verificou-se efetivamente uma diminuição. No entanto, não tão impactante em relação às outras, em particular o valor de MSE, que se alterou muito pouco. Para além disso, a nível visual não se identificam diferenças significativas, sobre os gráficos da Figura 48 e Figura 49.

Portanto, testes relativos ao número de neurónios não são mais necessários, visto que, para melhorar as previsões será necessário um novo aumento da quantidade de neurónios, que aumenta a complexidade da rede. Desta forma, o próximo parâmetro a ser alterado, será a quantidade de camadas ocultas.

Para já, vai ser eliminada uma camada oculta, para perceber como o modelo se comporta, os parâmetros são apresentados na Tabela 14.

Tabela 14 : Parâmetros da rede do quinto teste

Estrutura da rede	
Quantidade de camadas	Quantidade de neurónios
Camada 1	40
Camada 2	38
Camada 3	36
Camada 4	1
Outros parâmetros	
<i>Batch_size</i>	128
<i>Epochs</i>	80

Os valores do Erro Absoluto Médio (MAE) e do Erro Quadrático Médio (MSE), para o quinto teste, são apresentados de seguida, Tabela 15.

Tabela 15 : Valores de MAE e MSE para o quinto teste

Avaliação do desempenho do modelo	
MAE	MSE
0.062594	0.005125

De seguida, é apresentada a comparação gráfica entre os valores reais e os previstos, Figura 50.

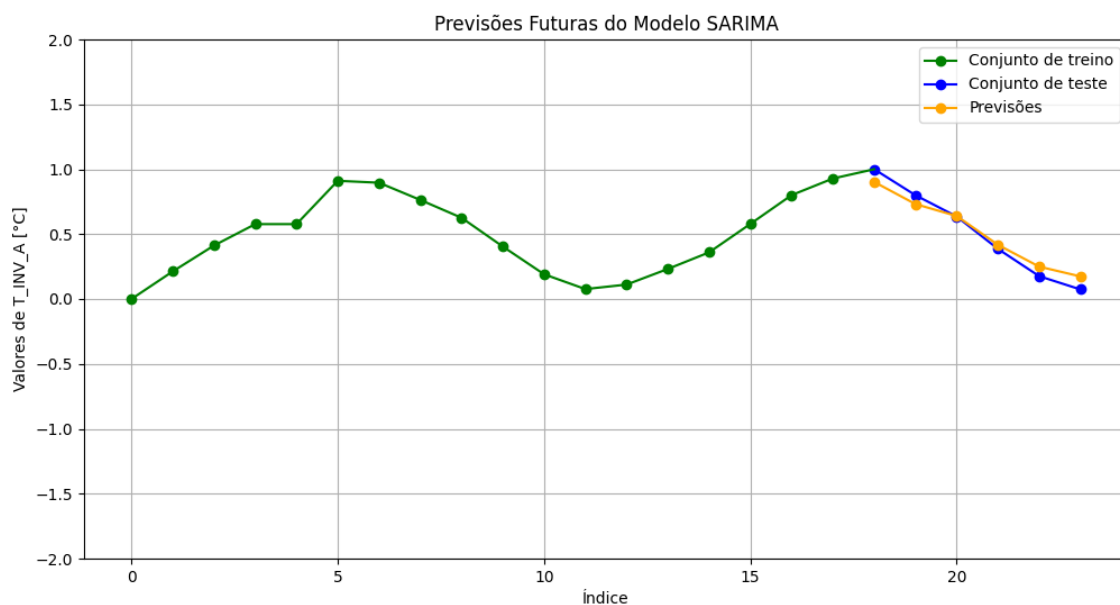


Figura 50 : Comparação entre os valores reais e os valores previstos segundo os parâmetros do quinto teste

Pela análise dos valores de MAE e MSE dos dois últimos testes, apresentados na Tabela 13 e Tabela 15, é evidente uma redução. Pelo que a redução do número de camadas apresentou efeitos

positivos no que toca ao desempenho do modelo. A Figura 50, mostra que a linha das previsões (laranja) por vezes se sobrepõe à linha do conjunto de teste, portanto, mais uma vez é realçada a qualidade do modelo.

Desta forma, como o quinto teste de parâmetros apresentou bons resultados, a quantidade de camadas e a quantidade de neurónios não foram mais alteradas. Por fim, procedeu-se à alteração dos restantes dois parâmetros.

Inicialmente, procedeu-se à redução do *batch_size*, por duas vezes, 64 e 32. Os parâmetros e resultados para o teste de *batch_size* assumindo o valor de 64 é apresentado na Tabela 16. Enquanto, a hipótese de *batch_size* assumindo o valor de 32 é apresentada na Tabela 17.

Tabela 16 : Alteração do parâmetro *batch_size* para 64

Estrutura da rede	
Quantidade de camadas	Quantidade de neurónios
Camada 1	40
Camada 2	38
Camada 3	36
Camada 4	1
Outros parâmetros	
<i>Batch_size</i>	64
<i>Epochs</i>	80
Resultados	
MAE	MSE
0.077073	0.007105

Tabela 17 : Alteração do parâmetro *batch_size* para 32

Estrutura da rede	
Quantidade de camadas	Quantidade de neurónios
Camada 1	40
Camada 2	38
Camada 3	36
Camada 4	1
Outros parâmetros	
<i>batch_size</i>	32
<i>Epochs</i>	80
Resultados	
MAE	MSE
0.077017	0.006973

A redução do *batch_size* para 64, provocou um pequeno aumento do valor de MAE e MSE, comparando os valores retratados na Tabela 15 e Tabela 16.

A alteração do mesmo parâmetro para 32, provocou novamente, um pequeno aumento do valor de MAE e MSE, comparando os valores retratados na Tabela 15 e Tabela 17.

A nível de *batch_size* utilizaram-se valores bastantes discrepantes do original, e o efeito foi um ligeiro aumento das métricas de qualidade. Portanto, o parâmetro *batch_size* não influenciou de forma significativa os resultados do modelo originalmente. No entanto, o teste de *batch_size* a assumir o valor de 128 foi o que apresentou melhores resultados, portanto, foi esse o valor definido para este parâmetro.

Por fim, abordou-se o parâmetro *epochs*, com uma redução para 50, os parâmetros e os resultados desta alteração são apresentados na Tabela 18.

Tabela 18 : Alteração do parâmetro epochs para 50

Estrutura da rede	
Quantidade de camadas	Quantidade de neurónios
Camada 1	40
Camada 2	38
Camada 3	36
Camada 4	1
Outros parâmetros	
<i>batch_size</i>	128
<i>epochs</i>	50
Resultados	
MAE	MSE
0.102999	0.013846

Os resultados apresentados na Tabela 18 revelam que uma redução do valor parâmetro em estudo, aumenta os valores de MAE e MSE, ou seja, a qualidade de previsão do modelo piorou.

Desta forma, após os vários testes retratados, é possível concluir que os parâmetros que mais influenciam a qualidade de previsão do modelo, são, a quantidade de neurónios em cada camada e a quantidade de camadas. É relevante, salientar que o modelo com os parâmetros apresentados, na Tabela 15, foi o que admitiu melhores resultados. Para além disso, as redes neuronais apresentaram resultados bastante promissores para esta variável em particular, apresentando valores de previsão muito próximos dos valores reais.

É notório o trabalho intrínseco na parametrização do modelo. No entanto, é importante notar que, embora as redes neuronais tenham demonstrado um desempenho promissor na previsão do conjunto de teste, devido a restrições de tempo, não foi possível explorar outras variáveis como já mencionado anteriormente.

4.4. Resultados SARIMA

O método SARIMA, é uma extensão do modelo ARIMA clássico projetado para lidar com séries temporais que exibem padrões sazonais.

Como já mencionada a implementação do método SARIMA foi desenvolvida através da linguagem *Python*. Cada uma das funções já foi devidamente abordada na secção anterior e demonstrada no respetivo apêndice.

Por fim, de salientar que para a construir este método é necessário atender à condição de estacionariedade da série, bem como a identificação dos parâmetros ótimos de ajuste do modelo e todos os passos realçados na Figura 24. De seguida, são abordados os resultados do teste SARIMA para as variáveis destacadas na Tabela 2.

4.4.1. Variável 1: Temperatura média ($T_MEDIA [^{\circ}C]$)

Para estudar a estacionariedade da série recorreu-se ao teste ADF. Inicialmente, recorreu-se ao estudo dos gráficos que mostram a média móvel e o desvio padrão móvel da série ao longo do tempo.

Neste caso, o conjunto de dados admitia contagens mensais, para além disso, admitiam sazonalidade anual (ou seja, 12 meses) e, portanto, foi estudada a média móvel com uma janela temporal de 12 observações. Desta forma, foram obtidos os gráficos da média móvel e desvio padrão móvel anual, Figura 51.

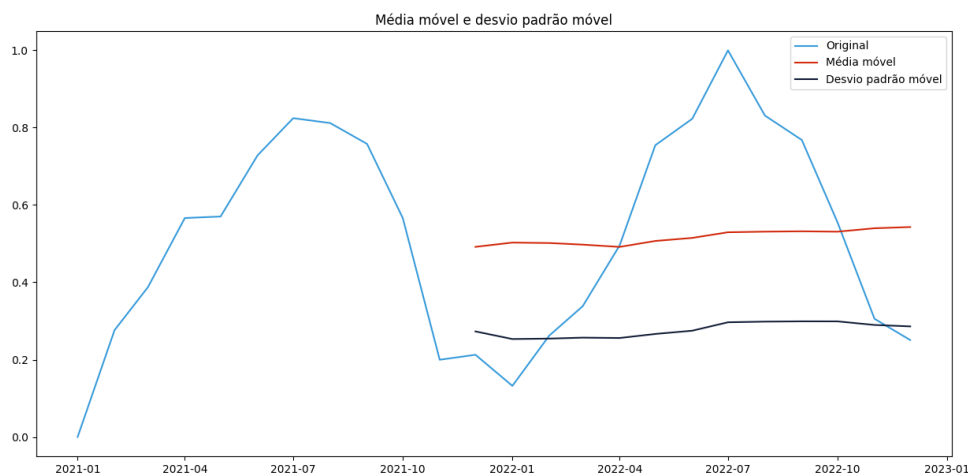


Figura 51 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável $T_MEDIA [^{\circ}C]$

Pela análise da Figura 51, é óbvio que tanto a linha da média móvel (em vermelho) quanto a linha do desvio padrão móvel (em azul-escuro) não mostram oscilações significativas. Isso pode indicar uma possível estacionariedade na série de dados. No entanto, é importante notar que o conjunto de dados é relativamente curto, e a janela da média móvel de 12 observações torna os dados ainda mais limitados para uma análise robusta. Portanto, não é possível para já constatar que a série é estacionária.

De salientar, que a estacionariedade implica que as propriedades estatísticas da série, como a média e a variância, não mudem ao longo do tempo, portanto, como a série em questão apresenta sazonalidade, à partida não deverá ser estacionária. Para tal, foi construído o teste ADF e os dados são apresentados na Figura 52.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-0.041005	0.955007	-4.012034	-3.104184	-2.690987

Figura 52 : Resultados teste ADF para a variável $T_MEDIA [^{\circ}C]$

Pela Figura 52, é possível concluir que a série não é estacionária, pois o valor da estatística de ADF é superior aos valores críticos apresentados, e p_value é superior ao nível de significância de 0,05.

Como a série não é estacionária, os valores dos parâmetros d e D devem ser estudados. Na literatura, é evidenciado que se os dados apresentarem sazonalidade o parâmetro D deve ser alterado para 1. Na análise exploratória da temperatura média foi possível concluir que a série admitia sazonalidade, portanto o parâmetro $D = 1$. Quanto ao d , foi calculado pela expressão apresentada na Figura 25 e obteve-se que $d=0$.

Desta forma, com os parâmetros estabelecidos aplicou-se diferenciação sazonal na série e foi aplicado novamente o teste de ADF à nova série diferenciada, como apresentado na Figura 53.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
- 5.930938	2.381566e-07	-4.938690	-3.477583	-2.843868

Figura 53 : Resultados teste ADF para a variável $T_MEDIA [^{\circ}C]$ após diferenciação sazonal da série

Pela Figura 53, é possível concluir que a série se tornou estacionária após a aplicação de uma diferenciação sazonal, apresentando o valor da estatística de ADF menor que os valores críticos, e um p_value menor que 0,05. Assim sendo, como a série se tornou estacionária para estes valores dos parâmetros d e D , não foi necessário realizar mais testes por alteração de d .

De seguida, procedeu-se à parametrização do modelo, sendo que, neste caso, os parâmetros d e D já estão definidos.

Na Figura 54 estão representados os gráficos de ACF e PACF. Desta forma, pelos gráficos constata-se que quer no gráfico PACF (indicador dos parâmetros p e P), como no ACF, não muitos existem pontos significativos quer em defasagens sazonais como não sazonais. Desta forma, os valores dos restantes parâmetros não devem ser muito elevados, pois iria tornar o modelo mais complexo sem evidencia de tal necessidade nos gráficos apresentados na Figura 54. Por fim salientar, que o valor máximo para os quatros parâmetros foi estabelecido de 3 e o mínimo de zero, pois a literatura indica que devem assumir valores inteiros positivos.

No entanto, é de ter em consideração que os dados fornecidos apenas são constituídos por 24 observações e o algoritmo apenas permite a computação de menos de metade das observações totais, neste caso, apenas 11 observações.

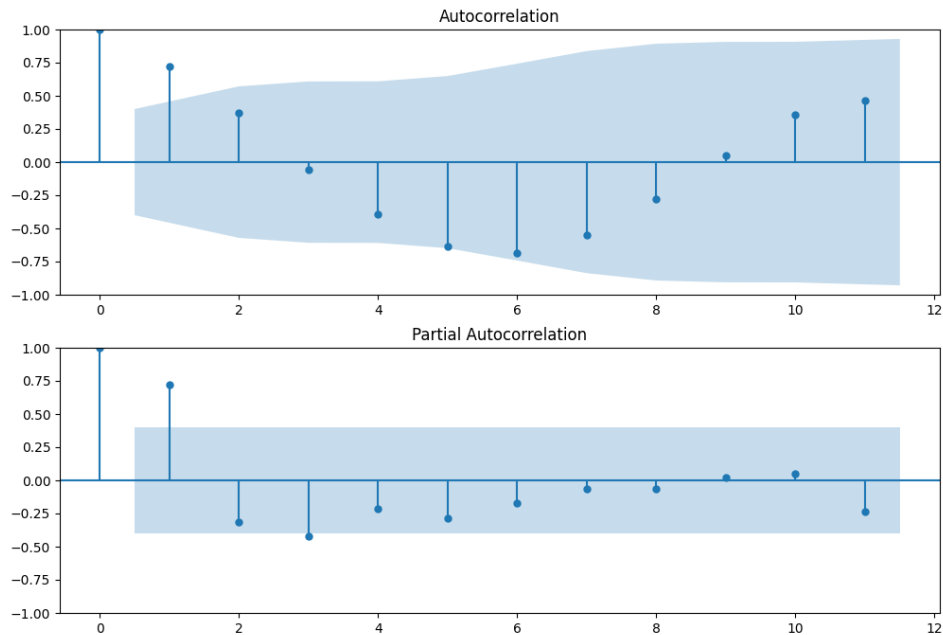


Figura 54: Gráficos ACF e PACF para a variável T_MEDIA [°C]

Os parâmetros foram selecionados com recurso à biblioteca *pmdarima*, em particular, à função *auto-arima*, segundo o critério de informação AIC, no qual, o melhor modelo é o que admite menor valor deste critério. Os resultados dos parâmetros e do valor AIC são demonstrados na Tabela 19.

Tabela 19 : Parâmetros modelo SARIMA e valor de AIC

(p, d, q) (P, D, Q, s)	AIC
(0,0,0) (0,1,0,12)	-8.403975

De seguida, é desenvolvido o diagnóstico do modelo e representado na Figura 55.

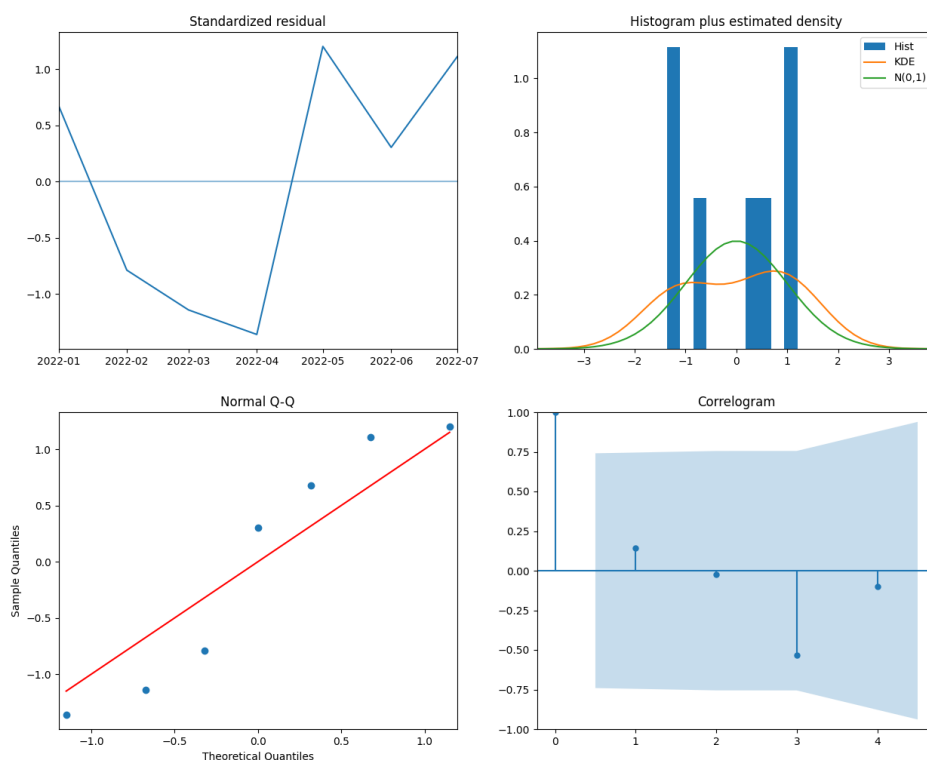


Figura 55 :Diagnóstico do modelo

Pela Figura 55 é possível constatar que:

- Gráfico dos resíduos ao longo do tempo não revela padrões óbvios;
- Histograma dos resíduos e gráfico da distribuição KDE aparenta semelhanças com a curva de distribuição normal;
- Pelo gráfico QQ dos resíduos mostra que os pontos seguem uma tendência linear, evidenciando assim que os resíduos são normalmente distribuídos;
- O gráfico de autocorrelação não apresenta nenhum ponto fora da zona sombreada;

Pela análise demonstrada, o modelo aparenta boa precisão e está apto para desenvolver as previsões. Seguidamente, são apresentadas as previsões para o conjunto teste na Tabela 20 e evidenciada uma comparação entre os valores reais e os previstos na Figura 56.

Tabela 20 : Valores das previsões para o conjunto de teste para a variável T_MEDIA [°C]

Data	Previsão
2022-08	0.876696
2022-09	0.822667
2022-10	0.630850
2022-11	0.264328
2022-12	0.277262

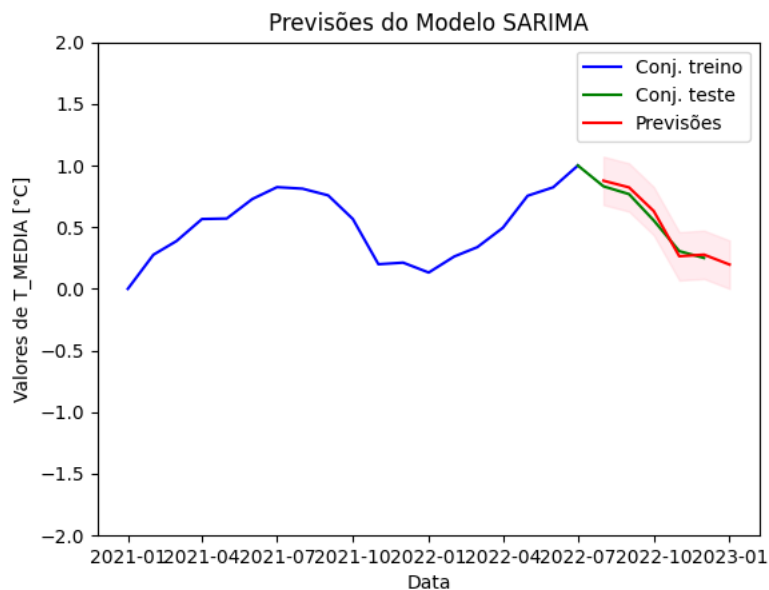


Figura 56: Comparação dados reais e previstos para a variável T_MEDIA [°C]

Pela análise do gráfico da Figura 56, é possível perceber que a linha vermelha, que representa os valores das previsões, acompanha a linha verde, do conjunto de teste. A Figura 56 no eixo y evidencia os valores para a temperatura média em graus celsius.

Foram ainda calculados os valores do erro absoluto médio (MAE) e da raiz do erro quadrático médio (RMSE) e estão apresentados na Tabela 21.

Tabela 21: Valores de MAE e RMSE para modelo SARIMA para a variável T_MEDIA [°C]

Modelos SARIMA	MAE	RMSE
(0,0,0) (0,1,0,12)	0.048727	0.051390

Tendo em conta a literatura os valores de MAE e RMSE devem estar próximos de zero, e perante os valores revelados na Tabela 21, é possível constatar que as previsões estão próximas dos valores reais, pois MAE e RMSE assumem valores muito próximos de zero. Desta forma, foram desenvolvidas as previsões para datas futuras com intervalo de confiança, representadas na Tabela 22 e graficamente apresentadas na Figura 57.

Tabela 22 : Valores previsões futuras com intervalos de confiança para a variável T_MEDIA [°C]

Data	Previsão	Limite inferior	Limite superior
2023-01	0,18342	0,022675	0,344165
2023-02	0,31358	0,152835	0,474326
2023-03	0,38932	0,228574	0,550065
2023-04	0,546472	0,385727	0,707218
2023-05	0,805934	0,645189	0,96668
2023-06	0,874287	0,713542	1,035032
2023-07	1,051256	0,890511	1,212001
2023-08	0,882656	0,72191	1,043401
2023-09	0,819398	0,658653	0,980144
2023-10	0,606162	0,445417	0,766907
2023-11	0,356908	0,196163	0,517653
2023-12	0,30197	0,141225	0,462715

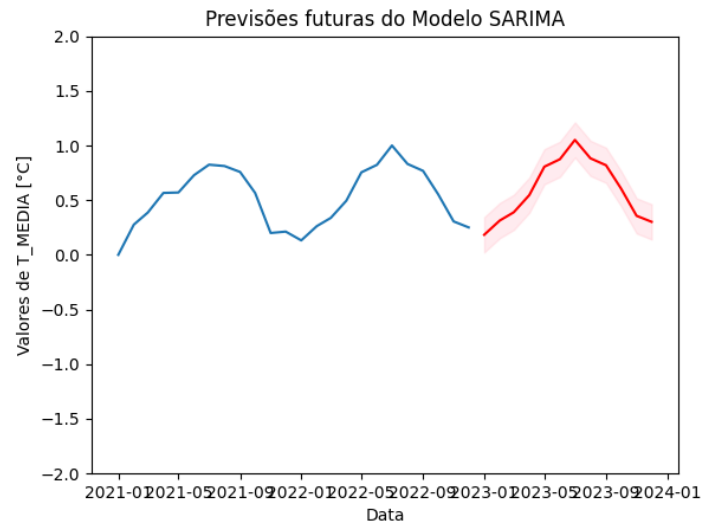


Figura 57: Previsões futuras da temperatura média para a variável T_MEDIA [°C]

Por fim, salientar pela Figura 57, que para o ano de 2023 não aparecem oscilações significativas em comparação com os anos anteriores. No entanto, não é demais lembrar que o conjunto de dados é limitado e, portanto, a previsão dos dados e conclusão sobre as mesmas é limitada.

4.4.2. Variável 2: Irradiância média ($IRR_MEDIA [Wh/m^2]$)

A metodologia aplicada para a irradiância média foi igual à da temperatura média mencionada anteriormente. Portanto, na Figura 58 está representado o gráfico que evidencia a média móvel e o desvio padrão móvel da série ao longo do tempo.

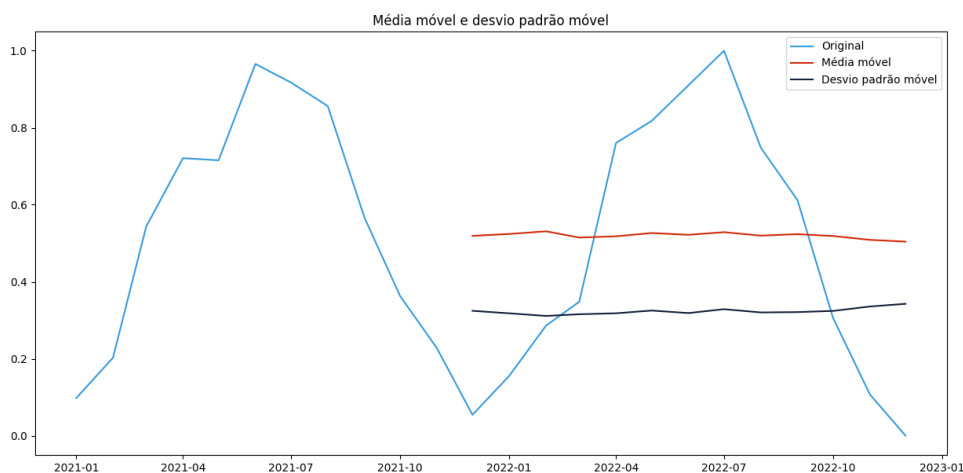


Figura 58 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo para a variável $IRR_MEDIA [Wh/m^2]$

Pela análise da Figura 58, é possível verificar que tanto a média como o desvio padrão não admitem grandes oscilações. No entanto, como já foi referido o conjunto de dados é de apenas dois anos, e utilizou-se uma janela de média móvel e desvio padrão móvel de 12 observações, ou seja, de um ano de observações, então torna os dados ainda mais limitados para uma análise rigorosa. Desta forma, implementou-se o teste ADF para verificar a estacionariedade da série e os resultados são demonstrados na Figura 59.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
1.302531	0.996623	-4.012034	-3.104184	-2.690987

Figura 59 : Resultados teste ADF para a variável $IRR_MEDIA [Wh/m^2]$

Pela análise da Figura 59, é possível constatar que a série não é estacionária, visto que, o valor da estatística de ADF é superior aos valores críticos apresentados, e p_value é superior ao nível de significância de 0,05.

Portanto, deve ser feita uma análise dos valores dos parâmetros d e D . Pela análise exploratória foi possível verificar que a irradiância média admite sazonalidade anual, desta forma, parâmetro $D = 1$. Quanto ao parâmetro d foi calculado o seu valor de acordo com a Figura 25, no qual obteve-se um valor de 0 para este parâmetro. Com os novos valores dos parâmetros aplicou-se diferenciação sazonal na série e realizou-se de novo o teste ADF à nova série diferenciada. Os resultados do teste ADF pela segunda vez são apresentados na Figura 60.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
1.739338	0.998221	-4.938690	-3.477583	-2.843868

Figura 60 : Resultados teste ADF para a variável IRR_MEDIA [Wh/m²] após diferenciação sazonal da série

Assim sendo, pela Figura 60, verifica-se que a série não se tornou estacionária após a diferenciação sazonal aplicada, pois p_value não é menor que 0,05 e o valor da estatística de ADF é superior aos valores críticos apresentados. Desta forma, o cálculo do parâmetro d não foi eficaz e, portanto, admitiu-se o valor de um para os parâmetros d e D , foi aplicada uma diferenciação sazonal e não sazonal com o objetivo de tornar a série estacionária e por fim, aplicado o teste ADF esta série. Os resultados são apresentados na Figura 61.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-2,920923	0.042968	-4.938690	-3.477583	-2.843868

Figura 61 : Resultados teste ADF para a variável IRR_MEDIA [Wh/m²] após diferenciação sazonal e não sazonal da série

Pela Figura 61, é possível indicar que p_value é menor que 0,05, mas o valor da estatística de teste ADF não é menor que todos os valores críticos. No entanto, a literatura evidencia que a decisão sobre a estacionariedade da série é baseada no p_value . Desta forma, optou-se por admitir o modelo em que $D=1$ e $d=1$, salientando desta forma, que a parametrização do modelo é realmente um desafio em que várias componentes devem ser atendidas, principalmente, que o objetivo seja cumprido, realizar previsões de qualidade.

De seguida, procedeu-se à parametrização do modelo para os restantes parâmetros em falta. Para tal, procedeu-se à elaboração dos gráficos ACF e PACF de modo a estudar o intervalo de valores para os restantes parâmetros, Figura 62.

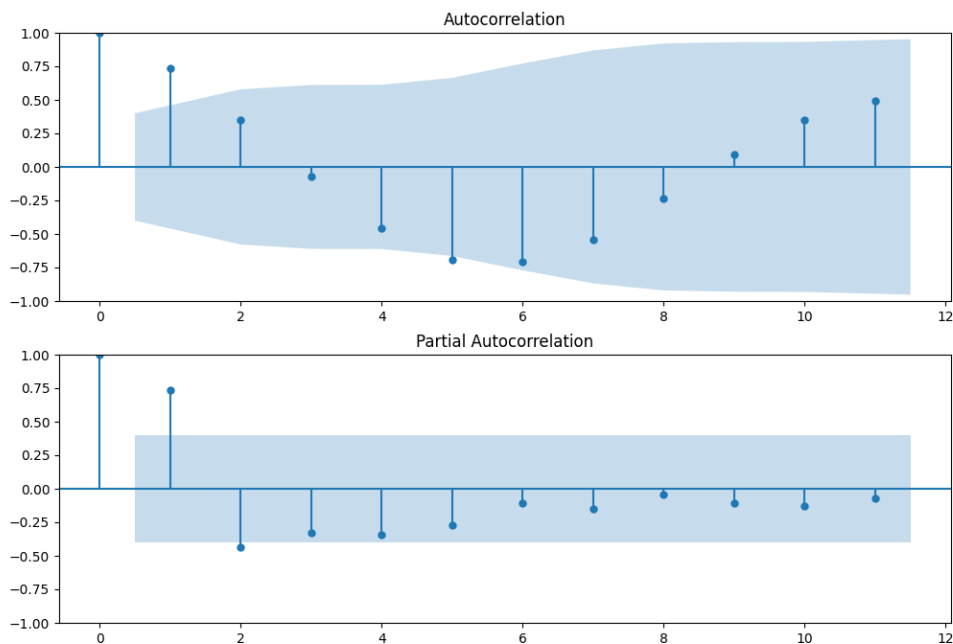


Figura 62 : Gráficos ACF e PACF para a variável IRR_MEDIA [Wh/m²]

Pela observação da Figura 62, é facilmente perceptível que tanto o gráfico ACF como PACF não admitem muitos pontos significativos, portanto, não há necessidade de altera de alterar valor máximo destes parâmetros, pelo que se mantém em 3 o mínimo em 0.

Desta forma, com os intervalos dos parâmetros já estabelecidos, os parâmetros do modelo foram estabelecidos com recurso à biblioteca `pmdarima`, em particular, à função `auto-arima`, segundo o critério de informação AIC, no qual, o melhor modelo é o que admite menor valor deste critério. Os resultados dos parâmetros e do valor AIC são demonstrados na Tabela 23.

Tabela 23 : Parâmetros modelo SARIMA e valor de AIC

(p, d, q) (P, D, Q, s)	AIC
(0,1,1) (0,1,0,12)	-3.858390

Seguidamente, foi desenvolvido o diagnóstico do modelo de forma a avaliar o valor dos parâmetros e se podem ser introduzidas melhorias, os gráficos estão apresentados na Figura 63.

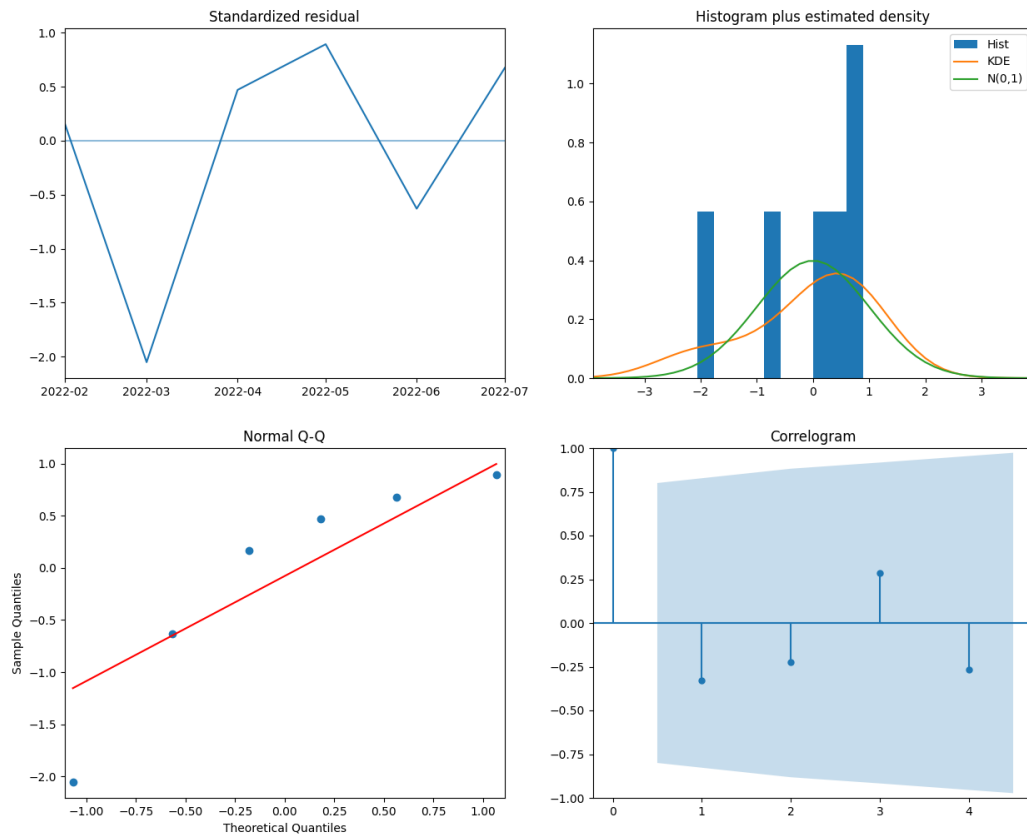


Figura 63 : Diagnóstico do modelo

Pela análise da Figura 63 é possível constatar que:

- Gráfico dos resíduos ao longo do tempo não revela padrões óbvios;
- Histograma dos resíduos e gráfico da distribuição KDE apresenta uma forma semelhante com a curva da distribuição normal;
- Pelo gráfico QQ dos resíduos mostra que os pontos seguem uma tendência linear, evidenciando assim que os resíduos são normalmente distribuídos;
- O gráfico de autocorrelação não apresenta nenhum ponto fora da zona sombreada;

Pela o estudo demonstrado, o modelo estará apto para desenvolver as previsões. Seguidamente, são apresentadas as previsões para o conjunto teste na Tabela 24 e evidenciada uma comparação entre os valores reais e os previstos na Figura 64.

Tabela 24 : Valores das previsões para o conjunto de teste para a variável IRR_MEDIA [Wh/m²]

Data	Previsão
2022-08-01	0.872602
2022-09-01	0.582625
2022-10-01	0.379166
2022-11-01	0.243717
2022-12-01	0.070525

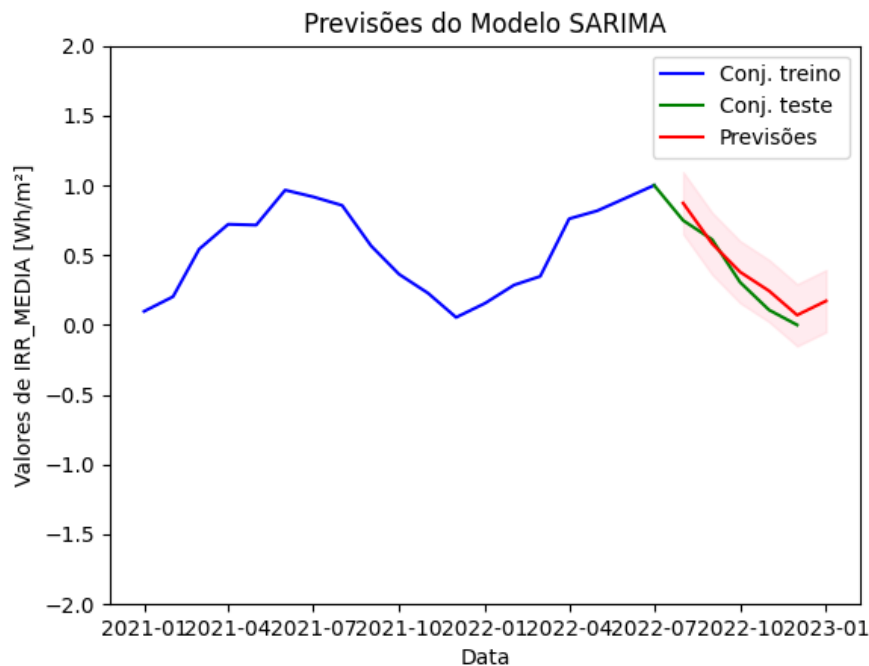


Figura 64 : Comparação dados reais e previstos para a variável IRR_MEDIA [Wh/m²]

Pela análise do gráfico da Figura 64, é possível perceber que a linha vermelha, que representa os valores das previsões, acompanha a linha verde, do conjunto de teste e, portanto, o modelo apresenta valores de previsão adequados. A Figura 64 no eixo y evidencia os valores para a temperatura média Wh/m^2 .

Foram ainda calculados os valores do erro absoluto médio (MAE) e da raiz do erro quadrático médio (RMSE) e estão apresentados na Tabela 25.

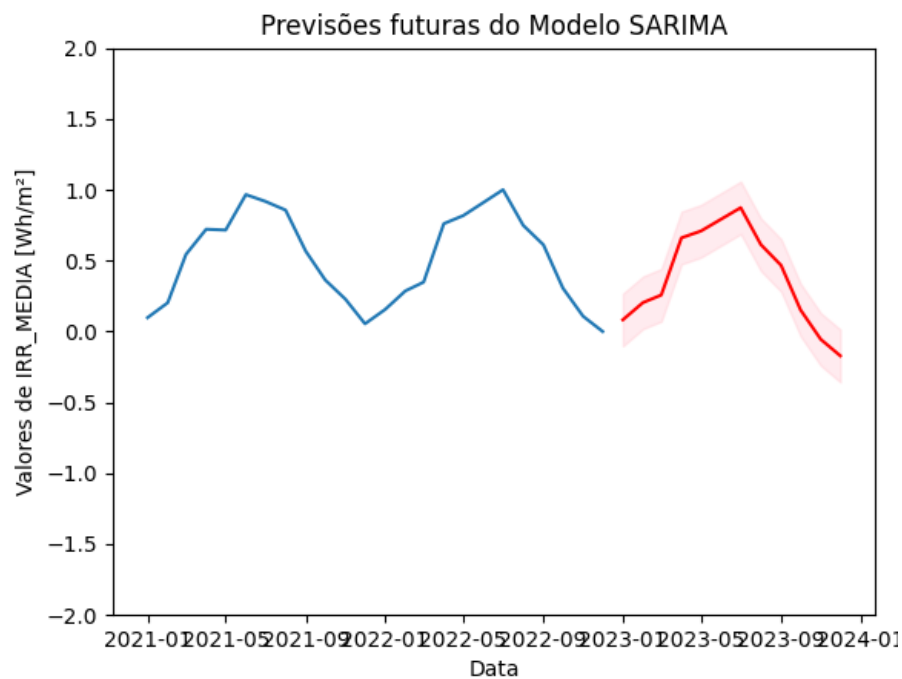
Tabela 25 :Valores de MAE e RMSE para modelo SARIMA para a variável IRR_MEDIA [Wh/m²]

Modelo SARIMA	MAE	RMSE
(0,1,1) (0,1,0,12)	0.086803	0.095197

Os valores apresentados na Tabela 25, indicam os parâmetros do modelo estão bem ajustados e os valores de previsão acompanham os valores reais, visto que as métricas de avaliação MAE e RMSE admitem valores próximos de zero. Desta forma, procedeu-se ao desenvolvimento de previsões para datas futuras com intervalo de confiança, representadas na Tabela 26 e graficamente apresentadas na Figura 65.

Tabela 26 : Valores previsões futuras com intervalos de confiança para a variável IRR_MEDIA [Wh/m²]

Data	Previsão	Limite inferior	Limite superior
2023-01	0,082281	-0,10281	0,267377
2023-02	0,203836	0,018715	0,388958
2023-03	0,256848	0,071701	0,441996
2023-04	0,660399	0,475226	0,845572
2023-05	0,70858	0,523382	0,893779
2023-06	0,792182	0,606958	0,977407
2023-07	0,873095	0,687844	1,058345
2023-08	0,612671	0,427395	0,797947
2023-09	0,466863	0,281561	0,652164
2023-10	0,151634	-0,03369	0,336961
2023-11	-0,0556	-0,24095	0,129757
2023-12	-0,17167	-0,35705	0,013712

Figura 65 : Previsões futuras da temperatura média para a variável IRR_MEDIA [Wh/m²]

Finalmente, realçar pela Figura 65, que para o ano de 2023 não aparecem oscilações significativas em comparação com os anos anteriores. No entanto, não é demais lembrar que o conjunto de dados é limitado e, portanto, a previsão dos dados e conclusão sobre as mesmas é limitada.

4.4.3. Variável 3: Potencia do lado DC inversor A ($P_{DC_INV_A}$ [W])

A metodologia utilizada para aplicação do modelo SARIMA para a variável da potência alternada convertida pelo inversor A seguiu os mesmos passos que a metodologia evidenciada para as variáveis estudadas até agora.

Inicialmente, procedeu-se à elaboração do gráfico da média móvel e desvio padrão móvel da série ao longo do tempo, utilizando as mesmas janelas temporais. O gráfico está representado na Figura 66.

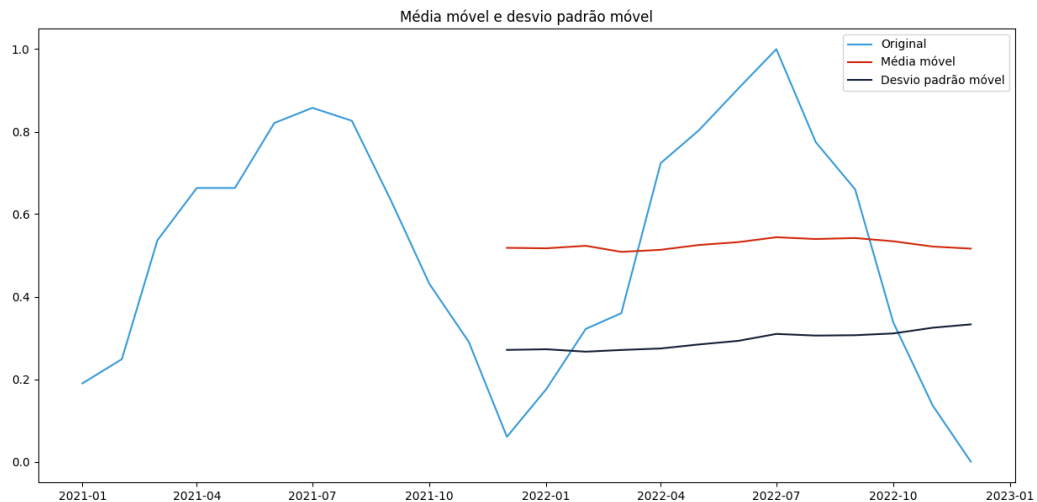


Figura 66 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável $P_{DC_INV_A}$ [W]

Pela observação da Figura 66, é facilmente visível que as linhas azul-escuro e vermelha, correspondente ao desvio padrão móvel e média móvel respetivamente, admitem-se mais ou menos constante ao longo do tempo, o que pode inferir uma certa estacionariedade da série. No entanto, este não é um teste formal de estacionariedade, apenas serve para de forma geral perceber como as características da série se comportam. De salientar também, pela análise exploratória que os dados referentes a esta variável apresentam uma certa sazonalidade o que realmente, já pode indicar que a série não será estacionária.

Deste modo, para estudar a estacionariedade da série foi implementado o teste ADF, cujos resultados são apresentados na Figura 67.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-1.688601	0.436995	-4.012034	-3.104184	-2.690987

Figura 67 : Resultados teste ADF para a variável $P_{DC_INV_A}$ [W]

A Figura 67 apresenta os resultados do teste ADF, em que p_value é maior que 0,05 e o valor da estatística do teste ADF é maior que qualquer valor crítico. Através desta análise, é possível concluir que a série não é estacionária e são necessárias diferenciações para a tornar estacionária.

Para tal, como na análise exploratória, esta variável apresentou sazonalidade anual, inicialmente, aplicou-se uma diferenciação sazonal ($D = 1$), quanto ao valor do parâmetro d foi calculado tendo em conta a Figura 25, cujo resultado foi $d = 0$, e de seguida, foi aplicado o teste ADF a esta nova série diferenciada, cujo resultados estão representados na Figura 68.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-1.589527	0.488852	-4.665186	-3.367187	-2.802961

Figura 68 : Resultados teste ADF para a variável P_DC_INV_A [W] após diferenciação sazonal

Analisando os resultados, da Figura 68 é visível que a diferenciação sazonal não tornou a série estacionária. Mais uma vez o cálculo do parâmetro d não se mostrou eficiente e para além, da diferenciação sazonal, passou a aplicar uma diferenciação não sazonal. Seguidamente foi aplicado novamente o teste ADF para $D=1$ e $d=1$, os resultados deste teste são apresentados na Figura 69.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-2.865253	0.049540	-4.938690	-3.477583	-2.843868

Figura 69 : Resultados teste ADF para a variável P_DC_INV_A [W] após diferenciação sazonal e não sazonal

Finalmente, pela Figura 69 comprovamos que a série se tornou estacionária, visto que, p_value é menor que 0,05 e o valor da estatística do teste ADF um menor valor que o valor crítico (10%). De salientar que, apesar dos outros dois valores críticos serem menores que o valor da estatística de teste ADF esta não é por vezes considerada como condição crucial para estudar a sazonalidade da série.

Com a estacionariedade da série verificada seguidamente, procedeu-se à parametrização dos restantes valores do modelo, sendo que, neste caso, os parâmetros d e D já estão definidos.

Foram desenvolvidos os gráficos ACF e PACF, Figura 70, para tentar perceber a complexidade dos dados e assim, perceber os valores máximos que devem ser atribuídos aos restantes parâmetros.

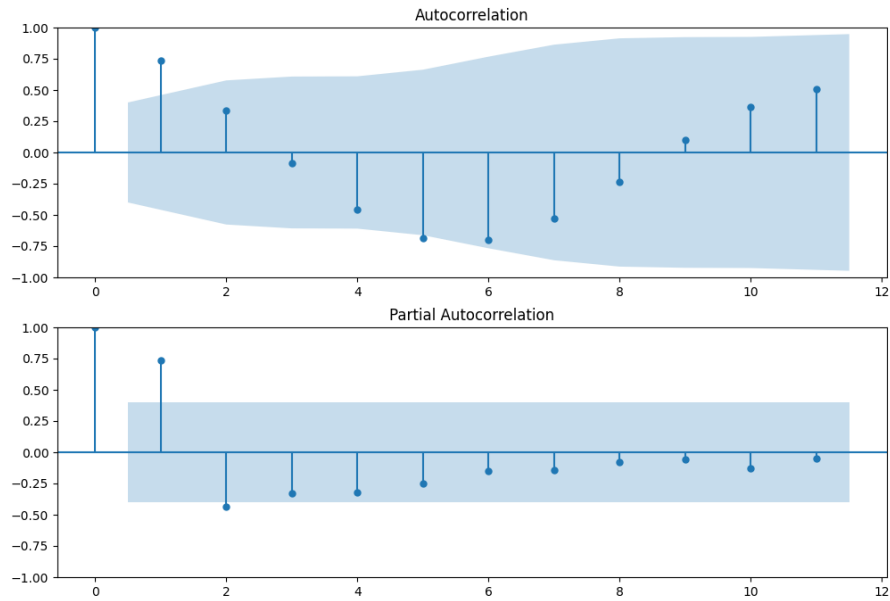


Figura 70 : Gráficos ACF e PACF para a variável P_DC_INV_A [W]

A Figura 70 evidencia que tanto no gráfico ACF como PACF existem poucos pontos significativos, quer isto dizer, fora da zona sombreada, e, portanto, o valor máximo dos parâmetros não foi alterado e manteve-se como 3 e os mínimos como 0, pois não podem admitir valores negativos.

Os intervalos dos parâmetros já estão estabelecidos, para encontrar o melhor modelo recorreu-se à biblioteca *pmdarima*, em particular, à função `auto-arima`, segundo o critério de informação AIC, no qual, o melhor modelo é o que admite menor valor deste critério. Os resultados dos parâmetros e do valor AIC são demonstrados na Tabela 27.

Tabela 27 : Parâmetros modelo SARIMA e valor de AIC

Modelo SARIMA	AIC
(0,1,1) (0,1,0,12)	-3.727893

Posteriormente, procedeu-se ao diagnostico modelo e representado na Figura 71.

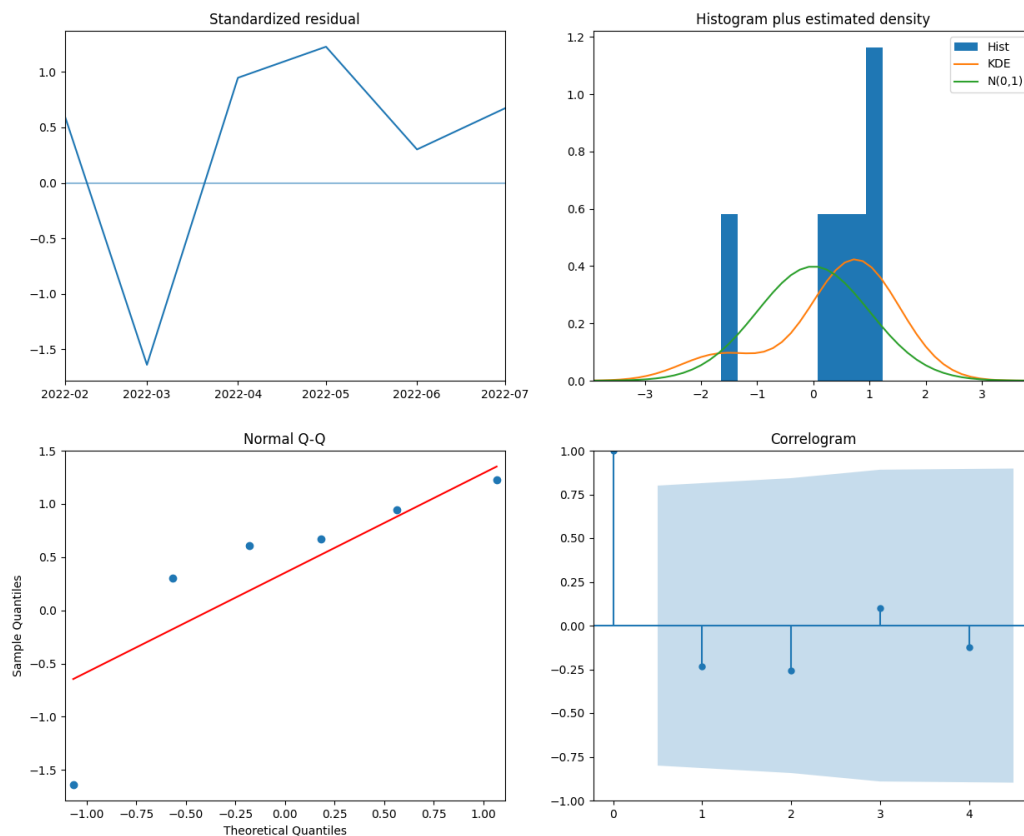


Figura 71 : Diagnóstico do modelo

Pela análise da Figura 71 é possível evidenciar que:

- Gráfico dos resíduos ao longo do tempo não revela padrões óbvios;
- Histograma dos resíduos e gráfico da distribuição KDE aparenta semelhanças com a curva de distribuição normal;
- Pelo gráfico QQ dos resíduos mostra que os pontos seguem uma tendência linear, evidenciando assim que os resíduos são normalmente distribuídos;
- O gráfico de autocorrelação não apresenta nenhum ponto fora da zona sombreada;

Pela análise demonstrada, o modelo aparenta boa precisão e está apto para desenvolver as previsões. Na Tabela 28, são apresentadas as previsões para o conjunto teste e na Figura 72 a representação gráfica com comparação com os valores reais.

Tabela 28 : Valores das previsões para o conjunto de teste para a variável P_DC_INV_A [W]

Data	Previsão
2022-08	0.917131
2022-09	0.722357
2022-10	0.521817
2022-11	0.381087
2022-12	0.151026

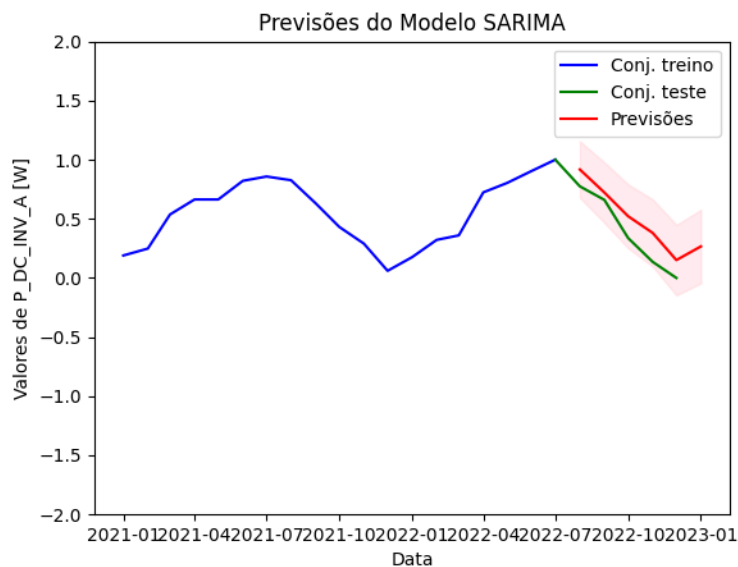


Figura 72 : Comparação dados reais e previstos para a variável P_DC_INV_A [W]

A linha que representa as previsões do modelo a vermelho, segue um comportamento semelhante à do conjunto de teste, apesar de que, para esta variável esta linha encontra-se ligeiramente mais distante da linha do conjunto de teste, em comparação com outras variáveis estudadas anteriormente. Para avaliar a precisão do modelo foram calculados os valores de MAE e RMSE, evidenciados na Tabela 29.

Tabela 29 : Valores de MAE e RMSE para modelo SARIMA para a variável P_DC_INV_A [W]

Modelo SARIMA	MAE	RMSE
(0,1,1) (0,1,0,12)	0.157002	0.167888

A Tabela 29, evidencia que os valores de MAE e RMSE são próximos de zero, que indica uma boa qualidade de previsão de valores desta variável por parte deste modelo. Portanto, procedeu-se ao desenvolvimento de previsões futuras, estabelecidas na Tabela 30 e representadas na Figura 73 com intervalo de confiança.

Tabela 30 : Valores previsões futuras com intervalos de confiança para a variável P_DC_INV_A [W]

Data	Previsão	Limite inferior	Limite superior
2023-01	0,113572	-0,09487	0,322011
2023-02	0,25011	0,041655	0,458565
2023-03	0,279119	0,070648	0,487589
2023-04	0,633537	0,425051	0,842023
2023-05	0,703752	0,495251	0,912254
2023-06	0,79557	0,587053	1,004088
2023-07	0,882051	0,673518	1,090585
2023-08	0,647358	0,438809	0,855907
2023-09	0,523562	0,314997	0,732127
2023-10	0,191796	-0,01678	0,400377
2023-11	-0,01888	-0,22748	0,189712
2023-12	-0,16434	-0,37295	0,044271

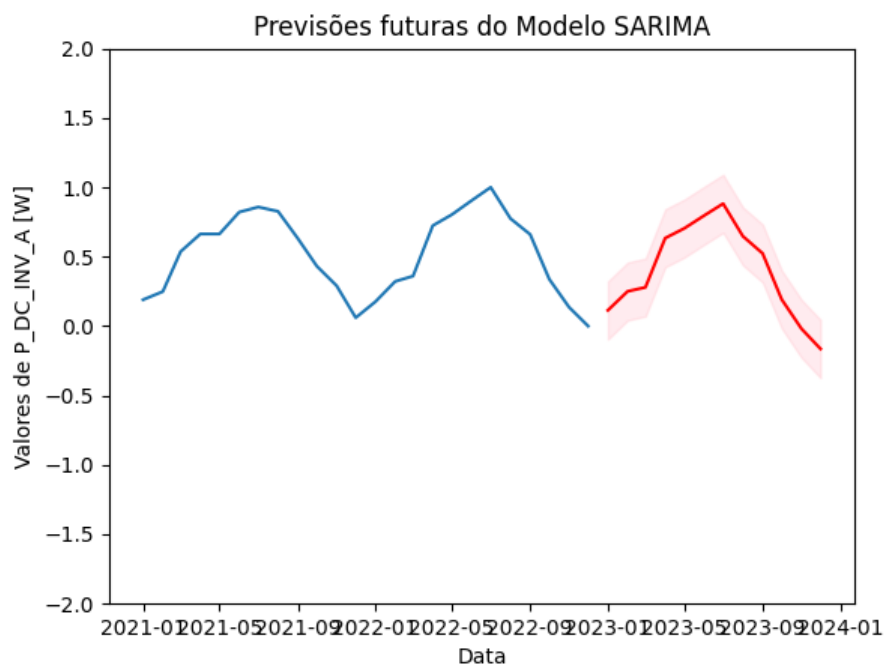


Figura 73 : Previsões futuras da temperatura média para a variável P_DC_INV_A [W]

Para concluir, uma breve análise da Figura 73, mostra que os valores de previsão assumem um comportamento semelhante ao histórico, e, portanto, nenhuma falha é para já evidenciada como seria de esperar, visto que o sistema fotovoltaico ainda é recente. Um histórico de dados mais robusto, com mais dados de amostragem seria pertinente, para perceber a evolução do comportamento dos mesmos.

4.4.4. Variável 4: Potencia do lado AC inversor A ($P_{AC_INV_A}$ [W])

Com o objetivo de construir previsões futuras sobre valores relativos à potência do lado AC do inversor A, através do método SARIMA, foram abordadas várias etapas.

Inicialmente, foi elaborado o gráfico da média móvel e desvio padrão móvel da série ao longo do tempo para perceber o seu comportamento, Figura 74. A janela temporal utilizada foi a mesma que a utilizada para outras variáveis, 12 observações.

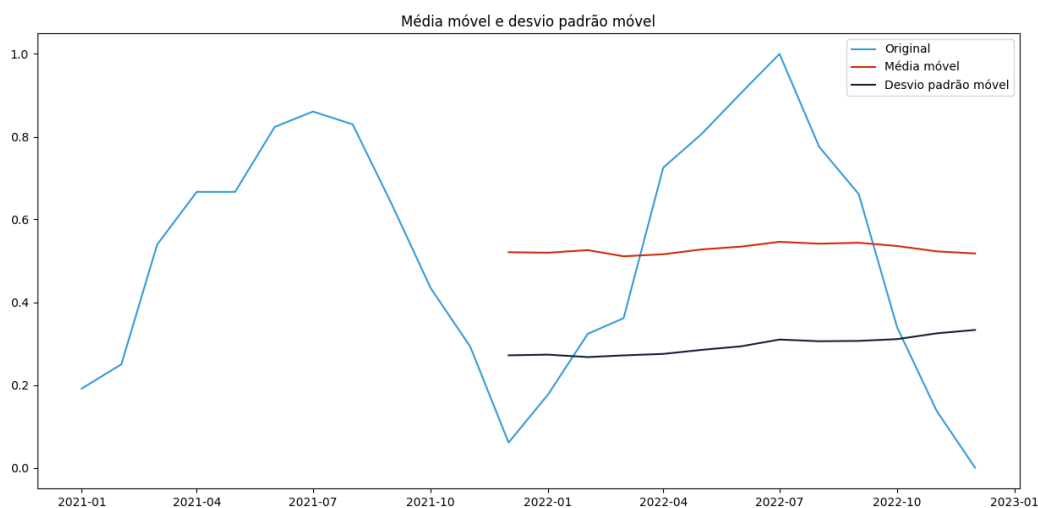


Figura 74 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável $P_{AC_INV_A}$ [W]

Pela Figura 74, é possível observar que tanto a média móvel como o desvio padrão móvel não apresentam oscilações ao longo do tempo. Porém, o facto de o conjunto de dados ser reduzido não permite aferir conclusões robustas sobre os dados, nomeadamente, aferir quanto à estacionariedade da série. Para avaliar a estacionariedade da série aplicou-se o teste ADF, os resultados são evidenciados na Figura 75.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-1.398721	0.582889	-4.012034	-3.104184	-2.690987

Figura 75 : Resultados teste ADF para a variável $P_{AC_INV_A}$ [W]

A Figura 75 evidencia que a série não é estacionária, pelo que é necessário aplicar diferenciações para tentar torná-la. Como a variável apresentou sazonalidade durante a sua análise, inicialmente, foi aplicada uma diferenciação sazonal aos dados, de seguida, foi novamente aplicado o teste ADF para avaliar a estacionariedade da série diferenciada, os resultados são demonstrados na Figura 76.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-1.556145	0.505577	-4.665186	-3.367187	-2.802961

Figura 76 : Resultados teste ADF para a variável P_AC_INV_A [W] após diferenciação sazonal

Pela análise da Figura 76, é possível verificar que p_value continua acima de 0,05, então a série não se tornou estacionária. Desta forma, para além da diferenciação sazonal aplicada, foi também aplicada uma diferenciação não sazonal, tendo os parâmetros d e D assumido o valor 1. Após as diferenciações foi realizado o teste ADF e demonstrado os resultados na Figura 77.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-2.809749	0.056895	-4.938690	-3.477583	-2.843868

Figura 77 : Resultados teste ADF para a variável P_AC_INV_A [W] após diferenciação sazonal e não sazonal

Após a aplicação de uma diferenciação sazonal e não sazonal na série é possível constatar pela Figura 77, que a série não se tornou estacionária. Quando aplicadas as duas diferenciações, verificamos que p_value encontra-se muito próximo de 0,05, isso sugere que a série pode estar a aproximar-se da estacionariedade, mas ainda não é completamente estacionária. Portanto, provavelmente, com ordens de diferenciação mais altas a série tornava-se estacionária. No entanto, perante o reduzido conjunto de dados não foi possível aplicar ordens de diferenciação mais altas, assim sendo não foi possível tornar a série estacionária.

Assim sendo, pela literatura (MARTIN et al., 2016), as séries temporais que não admitam estacionariedade não são de grande valor para o campo da previsão, dado a sua natureza incerta. E para além disso, a estacionariedade é uma condição importante para aplicar o SARIMA. No entanto, pelos testes ADF desenvolvidos observa-se que a série tende a tornar-se estacionária e como p_value está muito próximo do valor limite, decidiu-se aplicar o modelo e analisar as previsões do mesmo. Desta forma, os parâmetros d e D admitem o valor 1 que foi o teste ADF com melhor resultado, Figura 77.

Prosseguindo, foram desenvolvidos os gráficos ACF e PACF, Figura 78, para avaliar a complexidade os dados e os possíveis intervalos de valores para os restantes parâmetros do modelo.

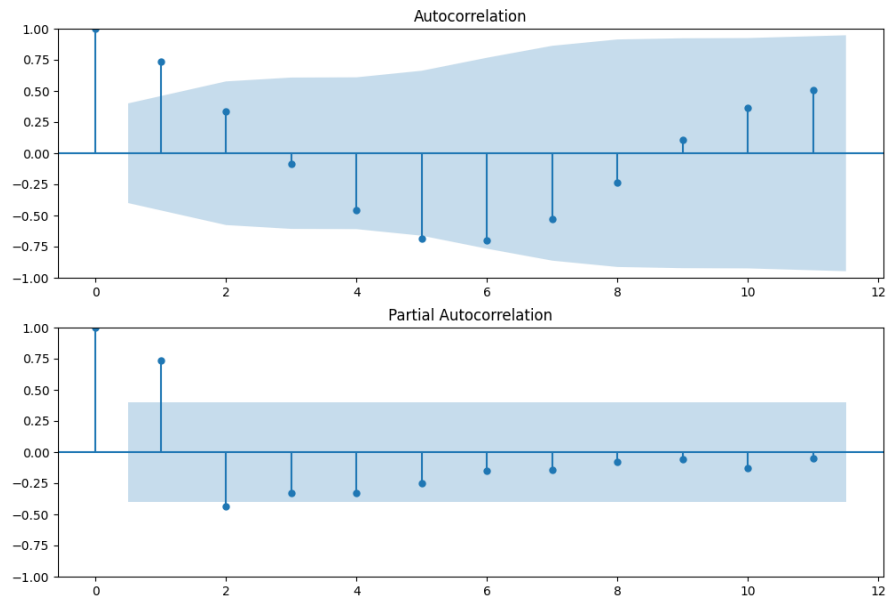


Figura 78 : Gráficos ACF e PACF para a variável P_AC_INV_A [W]

Os gráficos ACF e PACF não demonstram muitos pontos significativos, pela Figura 78, então o valor máximo e mínimo dos restantes parâmetros não foram alterados e mantiveram-se como 3 e 0, respetivamente.

De seguida, procedeu-se à seleção de valores dos restantes parâmetros, com recurso à biblioteca *pmdarima*, em particular, à função *auto-arma*, segundo o critério de informação AIC, no qual, o melhor modelo é o que admite menor valor deste critério. Os resultados dos parâmetros e do valor AIC são demonstrados na Tabela 31.

Tabela 31 : Parâmetros modelo SARIMA e valor de AIC

Modelo SARIMA	AIC
(0,1,1) (0,1,0,12)	-3.713535

De seguida, procedeu-se à construção do diagnostico do modelo, representado na Figura 79.

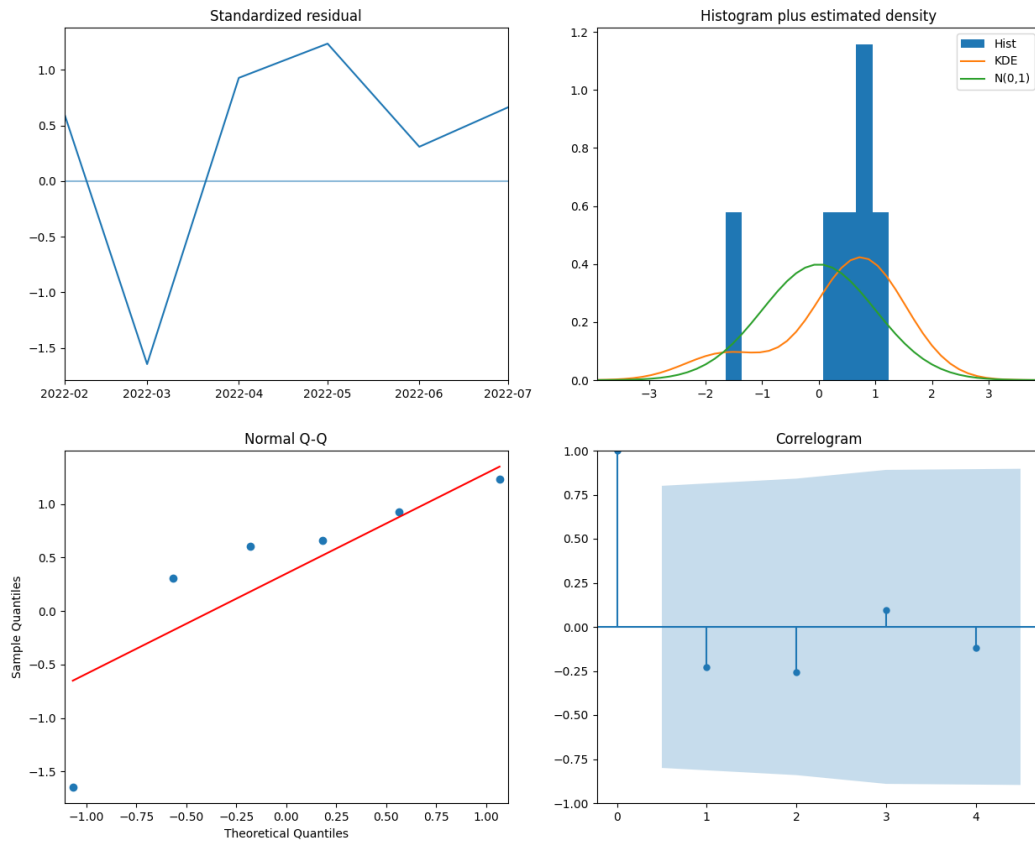


Figura 79 : Diagnóstico do modelo

Analisando a Figura 79 é possível afirmar o seguinte:

- Gráfico dos resíduos ao longo do tempo não apresenta padrões;
- A curva da distribuição KDE aparenta semelhanças com a curva de distribuição normal a nível de forma apesar de se apresentar um pouco desviada;
- Pelo gráfico QQ dos resíduos mostra que os pontos seguem uma tendência linear, evidenciando assim que os resíduos são normalmente distribuídos;
- O gráfico de autocorrelação não apresenta nenhum ponto fora da zona sombreada;

Assim, o modelo estará apto para desenvolver previsões. Inicialmente, foram construídas as previsões para os valores do conjunto de teste, no qual, são apresentados os valores previstos na Tabela 32, e posteriormente, são evidenciadas visualmente com uma comparação com os valores reais, Figura 80.

Tabela 32 : Valores das previsões para o conjunto de teste para a variável P_AC_INV_A [W]

Data	Previsão
2022-08	0.917497
2022-09	0.721809
2022-10	0.521993
2022-11	0.380347
2022-12	0.148479

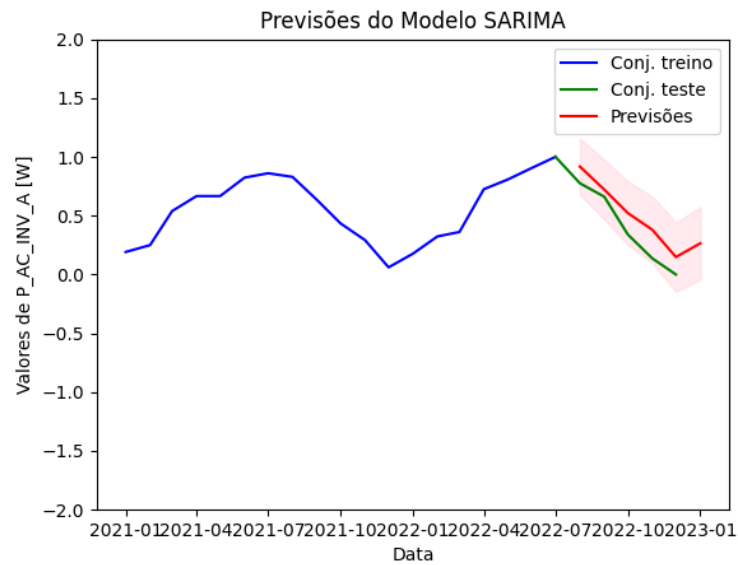


Figura 80 : Comparação dados reais e previstos para a variável P_AC_INV_A [W]

Ao analisar a Figura 80 percebe-se que a linha a vermelha, representa as previsões do modelo, apresenta um comportamento semelhante ao do conjunto de teste e até bastante próxima da linha verde, representa o conjunto de teste. Assim sendo, aplicar o modelo mesmo sem a estacionariedade da série estar completamente assegurada foi uma decisão, porque o modelo foi capaz de estudar os dados históricos e formar previsões semelhantes aos dados históricos.

Foram ainda calculados os valores do erro absoluto médio (MAE) e da raiz do erro quadrático médio (RMSE) e estão apresentados na Tabela 33, que avaliam a qualidade das previsões do modelo.

Tabela 33 : Valores de MAE e RMSE para modelo SARIMA para a variável P_AC_INV_A [W]

Modelo SARIMA	MAE	RMSE
(0,1,1) (0,1,0,12)	0.155350	0.166325

Tanto MAE como RMSE apresentam valores próximos de zero, apesar de mais altos quando comparados com o desempenho de outros modelos para outras variáveis. Desta forma, foram desenvolvidas as previsões para períodos futuros com intervalo de confiança, representadas na Tabela 34 e graficamente apresentadas na Figura 81.

Tabela 34 : Valores previsões futuras com intervalos de confiança para a variável P_AC_INV_A [W]

Data	Previsão	Limite inferior	Limite superior
2023-01	0,113326	-0,09466	0,321312
2023-02	0,250235	0,042208	0,458262
2023-03	0,278523	0,070457	0,48659
2023-04	0,633044	0,424938	0,84115
2023-05	0,704866	0,49672	0,913012
2023-06	0,794935	0,586749	1,00312
2023-07	0,879768	0,671542	1,087994
2023-08	0,646031	0,437765	0,854296
2023-09	0,522172	0,313867	0,730477
2023-10	0,191141	-0,0172	0,399486
2023-11	-0,02057	-0,22896	0,187811
2023-12	-0,16707	-0,3755	0,041353

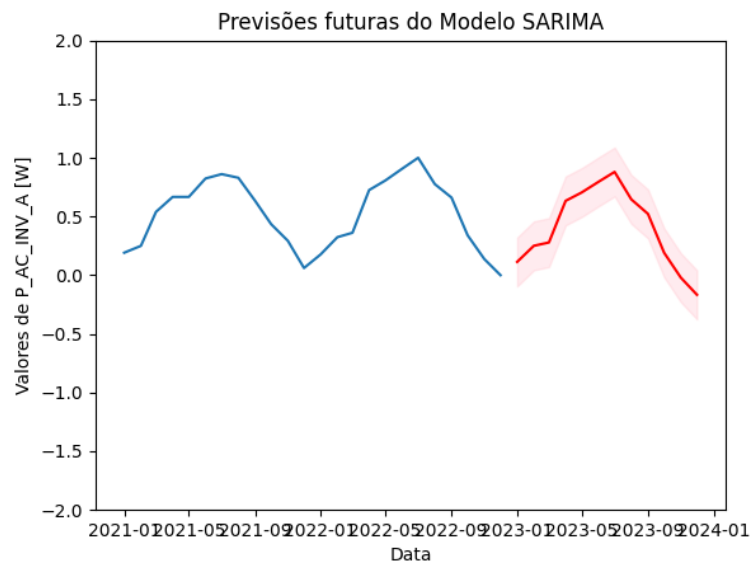


Figura 81 : Previsões futuras da temperatura média para a variável P_AC_INV_A [W]

As previsões para datas futuras revelam um comportamento semelhante ao dos dados históricos. Quer isto dizer que, o modelo foi capaz de analisar os dados históricos e realizar previsões adequadas. De salientar, uma vez mais que observações de apenas dois anos não permite pelo menor para já averiguar falhas que supostamente acontecem em anos mais tardios do sistema fotovoltaico. No entanto, pela análise dos dados previstos e comparando com os dados reais é possível perceber falhas que possam estar a acontecer no momento, e desta forma, prevenir estas situações.

Em conclusão, notar que, apesar da não estacionariedade da série temporal o modelo apresentou um bom desempenho, também pela análise a série tenderia para estacionaria, ou seja, não admitia completamente a não estacionariedade.

4.4.5. Variável 5: Temperatura do inversor A (T_INV_A [°C])

Na mesma linha de pensamento das variáveis estudadas até ao momento, as etapas de implementação da variável relativa à temperatura do inversor A são as mesmas que já evidenciadas para as outras variáveis.

Deste modo, com a mesma janela de tempo foi desenvolvido o gráfico que mostra a média móvel e o desvio padrão móvel da série ao longo do tempo, Figura 82, com o objetivo de estudar o comportamento da série, ou seja, se as duas características se manterem constantes, indica que os valores da série não se alteram muito, e pode indicar uma possível estacionariedade da série, no entanto, para o formal da estacionariedade é desenvolvido posteriormente.

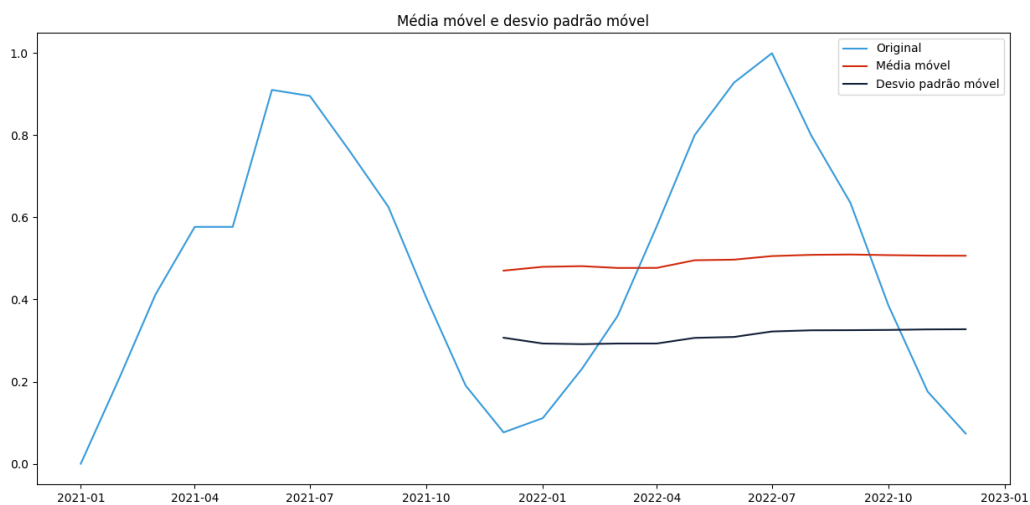


Figura 82 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável T_INV_A [°C]

Como demonstrado nas variáveis anteriores, a linha azul-escuro e a vermelha, correspondentes, ao desvio padrão móvel e à média móvel, respetivamente, permanecem constante ao longo do tempo. Mas como já foi referido, tal facto pode se dever ao conjunto de dados de apenas dois anos fornecido.

Para analisar a estacionariedade da série foi implementado o teste ADF, os resultados estão apresentados na Figura 83.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-0.695006	0.848039	-4.012034	-3.104184	-2.690987

Figura 83 : Resultados teste ADF para a variável T_INV_A [°C]

De acordo com a Figura 83, é possível concluir que a série não é estacionária, pois p_value é maior que o teste de significância. Portanto, devem ser aplicadas diferenciações, e, portanto, alterados os valores dos parâmetros referentes à diferenciação (d e D). Tal como as outras variáveis, a temperatura do inversor A também admite sazonalidade, portanto, inicialmente aplicou-se uma diferenciação sazonal, $D=1$. O parâmetro referente à diferenciação não sazonal, foi calculado a partir da Figura 25 e o resultado foi $d=0$. Desta forma, inicialmente, com vista a tornar a série estacionária apenas se procedeu à diferenciação sazonal da série, de seguida, foi aplicado o teste ADF à série diferenciada, cujos resultados são apresentados na Figura 84.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-3.851707	0.002420	-4.938690	-3.477583	-2.843868

Figura 84 : Resultados teste ADF para a variável T_INV_A [°C] após diferenciação sazonal da série

Os resultados apresentados na Figura 84 demonstram que a série se tornou estacionária, visto que, p_value é menor que 0,05, e neste caso, o valor da estatística do teste ADF é menor que o valor crítico (1%) e valor crítico (10%).

Assim sendo, o parâmetro D , referente à diferenciação sazonal assume o valor de 1, enquanto, d assume zero, visto que, não foi necessário aplicar diferenciação não sazonal à série.

O próximo passo foi proceder à restante parametrização do modelo. Para tal, mais uma vez, para estudar a possível complexidade dos dados e assim determinar os intervalos de valores dos parâmetros, foram desenvolvidos os gráficos ACF e PACF, apresentados na Figura 85.

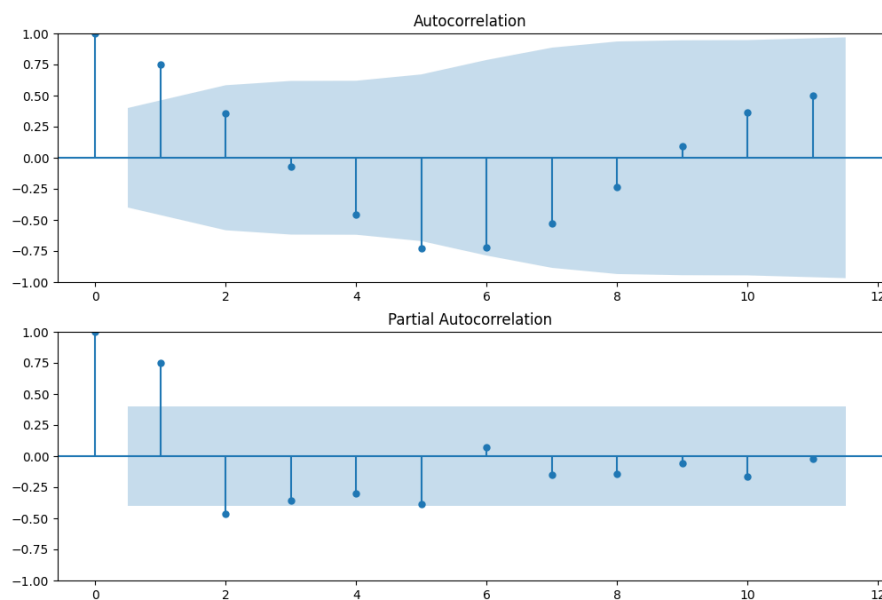


Figura 85 : Gráficos ACF e PACF para a variável T_INV_A [°C]

A Figura 85 demonstra que os dados não admitem, muitos pontos fora da área sombreada, ou seja, pontos significativos tanto no gráfico ACF como PACF, e, portanto, os parâmetros p , P , Q e q não devem admitir valores máximos muito elevados, porque a complexidade dos dados assim não o exige, desta forma, o valor máximo e mínimo para cada um deles não foi alterado e mantiveram-se como 3 e 0, respetivamente.

Com os intervalos dos parâmetros já definidos, posteriormente, com objetivo de encontrar o melhor modelo recorreu-se à biblioteca `pmdarima`, em particular, à função `auto-arima`, segundo o critério de informação AIC, no qual, o melhor modelo é o que admite menor valor deste critério. Os resultados dos parâmetros e do valor AIC são demonstrados na Tabela 35.

Tabela 35 : Parâmetros modelo SARIMA e valor de AIC

(p, d, q) (P, D, Q, s)	AIC
(0,0,0) (0,1,0,12)	-10.603272

Após a identificação dos parâmetros passou-se à etapa de diagnóstico do modelo, cujo resultado é demonstrado na Figura 86.

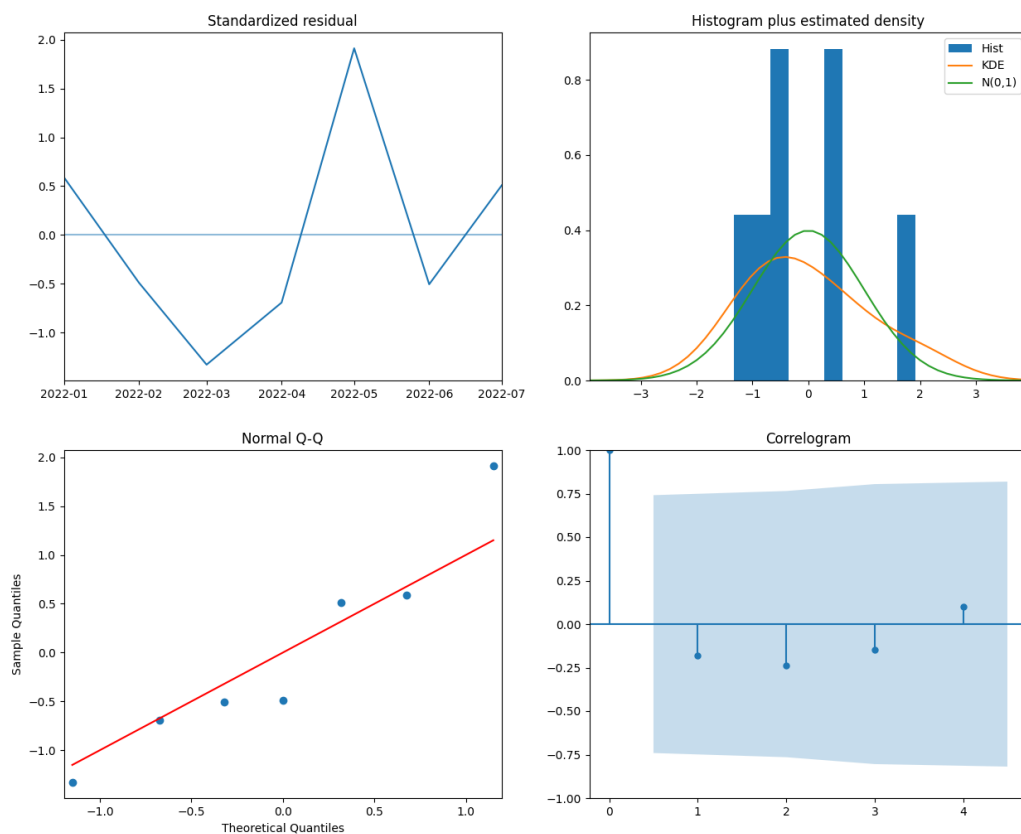


Figura 86 : Diagnóstico do modelo

Analisando os gráficos de diagnóstico da Figura 86:

- Gráfico dos resíduos ao longo do tempo não apresenta padrões;
- Histograma dos resíduos e gráfico da distribuição KDE aparenta uma forma muito semelhante com a curva de distribuição normal;
- Pelo gráfico QQ dos resíduos mostra que os pontos seguem uma tendência linear, com os pontos seguindo maioritariamente a linha vermelha, evidenciando assim que os resíduos são normalmente distribuídos;
- O gráfico de autocorrelação não apresenta nenhum ponto significativo;

Assim sendo, o modelo está adequado para desenvolver as previsões. De seguida, são apresentadas as previsões para o conjunto teste na Tabela 36 e evidenciada uma comparação entre os valores reais e os previstos na Figura 87.

Tabela 36 : Valores das previsões para o conjunto de teste para a variável T_INV_A [°C]

Data	Previsão
2022-08	0.824099
2022-09	0.686448
2022-10	0.464786
2022-11	0.251247
2022-12	0.137128

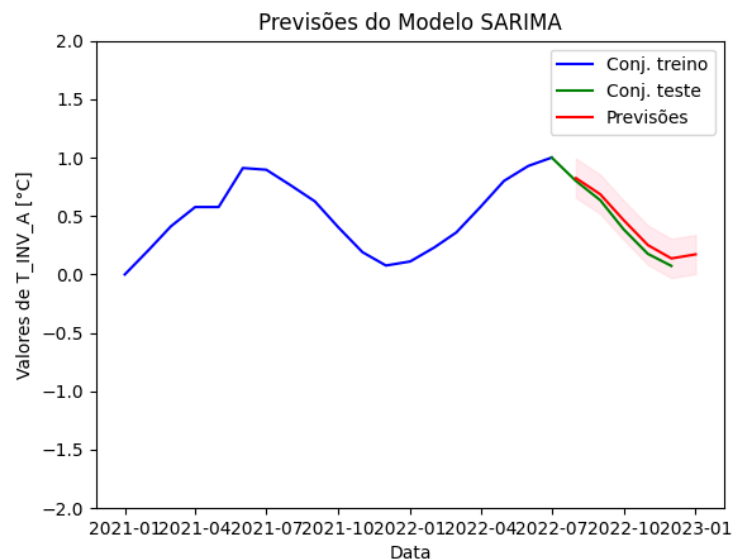


Figura 87 : Comparação dados reais e previstos para a variável T_INV_A [°C]

É interessante analisar a Figura 87 e perceber que as previsões desenvolvidas pelo modelo acabam por retratar quase por completo os dados reais do conjunto de teste. No entanto, é importante analisar, analisar o mesmo período do conjunto de teste no ano de 2021, pois é possível verificar

que admitem comportamentos muito semelhantes. Assim, fica evidenciada a importância dos dados históricos na implementação deste modelo.

Para constatar o que já foi referido, são apresentados os valores de MAE e RMSE como métricas de avaliação do modelo, na Tabela 37.

Tabela 37 : Valores de MAE e RMSE para modelo SARIMA para a T_INV_A [°C]

Modelo SARIMA	MAE	RMSE
(0,1,1) (0,1,0,12)	0.058633	0.06188

A Tabela 37 apresenta valores muito próximos de zero para os dois tipos de erro, quer isto dizer, que os valores previstos são muito semelhantes com os valores reais, que acaba por afirmar o que já concluído anteriormente.

Por fim, são criadas as previsões para datas futuros com um intervalo de confiança e apresentadas os valores das mesmas, na Tabela 38, e apresentado o seu comportamento gráfico na Figura 88.

Tabela 38 : Valores previsões futuras com intervalos de confiança para a variável T_INV_A [°C]

Data	Previsão	Limite inferior	Limite superior
2023-01	0,14747	0,005603	0,289336
2023-02	0,267972	0,126105	0,409838
2023-03	0,395871	0,254005	0,537738
2023-04	0,614925	0,473058	0,756791
2023-05	0,836968	0,695101	0,978835
2023-06	0,964616	0,822749	1,106483
2023-07	1,03628	0,894413	1,178147
2023-08	0,836086	0,69422	0,977953
2023-09	0,67175	0,529883	0,813617
2023-10	0,422107	0,28024	0,563974
2023-11	0,212114	0,070247	0,35398
2023-12	0,109886	-0,03198	0,251753

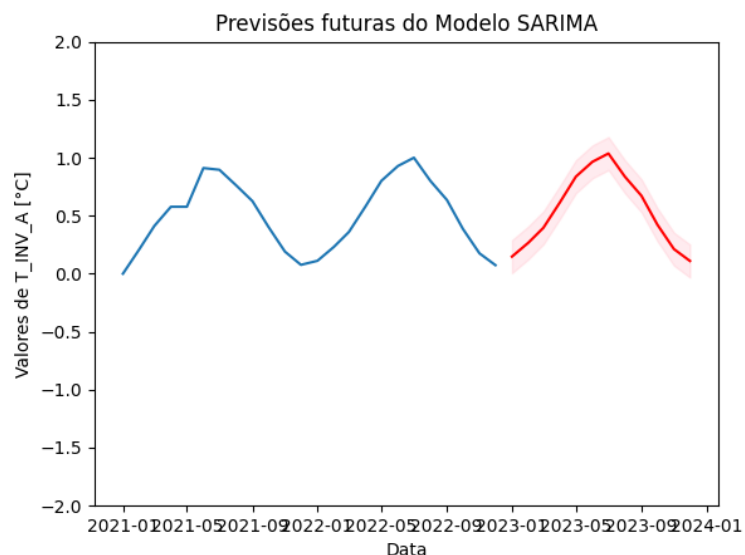


Figura 88 : Previsões futuras da temperatura média para a variável T_MEDIA [°C]

Mais uma vez, o grande entrave para um melhor ajuste e até mesmo conclusões mais realistas é o reduzido número de anos analisados. No entanto, o modelo está apto a carregar mais ficheiros de análise à medida que existam mais dados, e desta forma, prever as falhas que supostamente ocorrerão em períodos mais tardios. No entanto, realçar o comportamento das previsões muito semelhante ao dos dados históricos.

4.4.6. Variável 6: Potencia do lado DC do inversor B ($P_{DC_INV_B}$ [W])

A metodologia de implementação do modelo SARIMA para a variável relacionada com a potência do lado DC do inversor B admite os mesmos passos da implementação do mesmo modelo para as variáveis já retratadas. Portanto, inicialmente, desenvolveu-se o gráfico de média móvel e desvio padrão móvel da série ao longo do tempo, com a mesma janela de tempo das outras variáveis, e o gráfico está representado na Figura 89.

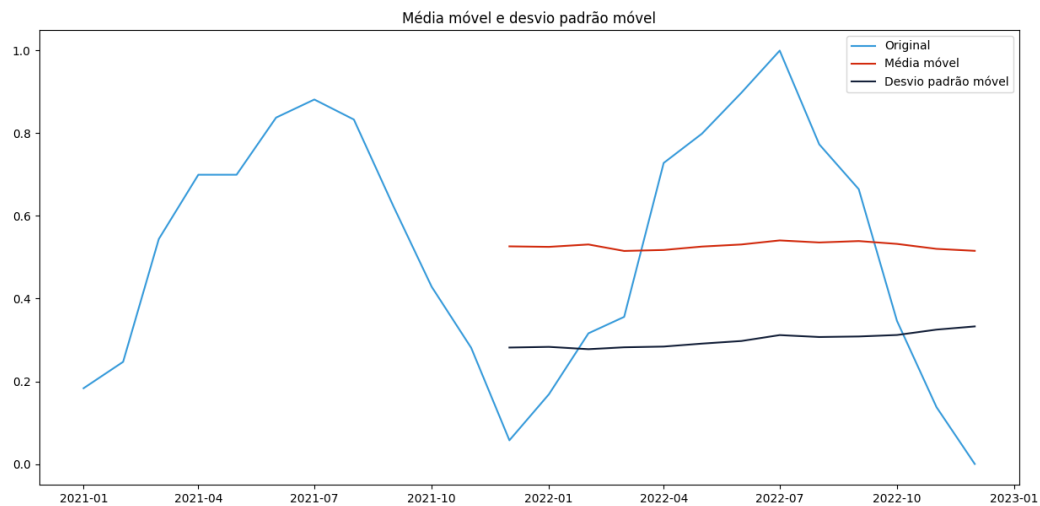


Figura 89: Gráfico da média móvel e desvio padrão móvel ao longo do tempo para a variável P_DC_INV_B [W]

Analogamente com o que já foi retratado com as outras variáveis, tanto, a média móvel como o desvio padrão móvel mantém-se mais ou menos constantes ao longo do tempo, pela Figura 89. No entanto, a quantidade de observações dos dados é limitada e, portanto, nada pode ser afirmado com robustez. Para avaliar a estacionariedade da série aplicou-se o teste ADF, os resultados são apresentados na Figura 90.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-4.006206	0.001378	-3.924019	-3.068498	-2.673893

Figura 90 : Resultados teste ADF para a variável P_DC_INV_B [W]

O teste ADF indica que a série é estacionária sem a necessidade de diferenciação quer sazonal, quer não sazonal, apesar da sazonalidade mostrada durante a análise exploratória. No entanto, o teste ADF é um teste estatístico capaz de determinar a estacionariedade da série e neste caso, p_value admite um valor menor que 0,05 e o valor do estatístico do teste ADF é menor do que qualquer valor crítico, então a série original é estacionária. Assim sendo, é possível passar de imediato para a próxima fase de parametrização dos restantes valores do modelo. Para tal, foram construídos os gráficos ACF e PACF, com vista a estudar possíveis intervalos de valores para os parâmetros p , P , q , Q de acordo com a quantidade de valores significativos evidenciados nos gráficos, representados na Figura 91.

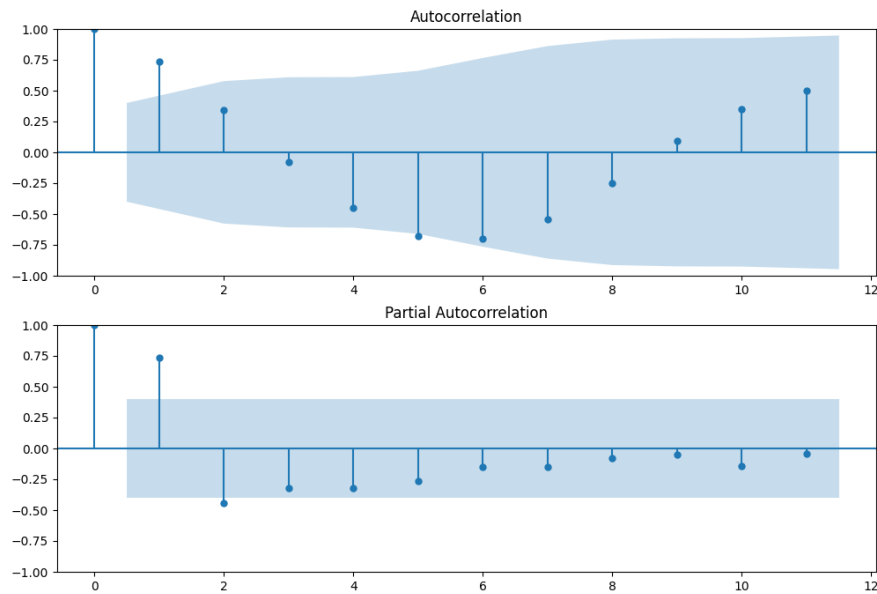


Figura 91 : Gráficos ACF e PACF para a variável P_DC_INV_B [W]

Pela Figura 91, são evidenciados alguns pontos, nas defasagens não sazonais, fora da área sombreada, no entanto, não são muito representativos, e assim, não vai ser alterado o valor máximo proposto para os parâmetros p , P , q e Q .

Para encontrar o melhor modelo, utilizou-se uma forma mais automática, com recurso à biblioteca `pmdarima`, em particular, à função `auto-arima`, segundo o critério de informação AIC, no qual, o melhor modelo é o que admite menor valor deste critério. Os resultados dos parâmetros e do valor AIC são demonstrados na Tabela 39.

Tabela 39 : Parâmetros modelo SARIMA e valor de AIC

(p, d, q) (P, D, Q, s)	AIC
(3,0,1) (0,0,0,12)	-24.767683

Posto isto, procedeu-se ao diagnóstico do modelo, representado na Figura 92.

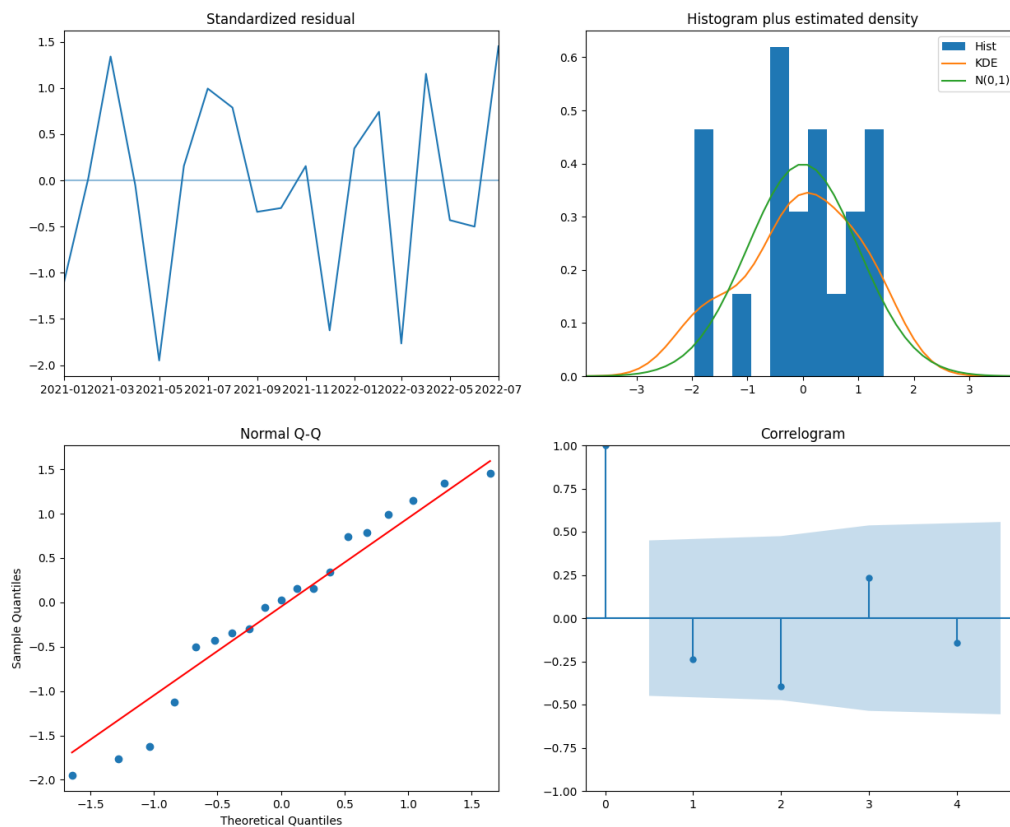


Figura 92 : Diagnóstico do modelo

Pela Figura 92 é possível delimitar a seguinte análise:

- Gráfico dos resíduos ao longo do tempo não apresenta padrões;
- A curva do gráfico da distribuição KDE apresenta uma forma semelhante à curva de distribuição normal;
- Pelo gráfico QQ dos resíduos mostra que os pontos se situam praticamente sobre a linha vermelha, seguem uma tendência linear, evidenciando assim que os resíduos são normalmente distribuídos;
- O gráfico de autocorrelação não apresenta nenhum ponto fora da zona sombreada;

Pela análise realizada, parece que o modelo apresenta uma boa capacidade de previsão e pode ser utilizado para fazer previsões. Em seguida, são apresentadas as previsões para o conjunto de teste na Tabela 40 e reveladas graficamente, na Figura 93, possibilitando uma comparação entre os valores reais e os valores previstos.

Tabela 40 : Valores das previsões para o conjunto de teste para a variável P_DC_INV_B [W]

Data	Previsão
2022-08	0.826103
2022-09	0.643089
2022-10	0.403373
2022-11	0.214814
2022-12	0.107532

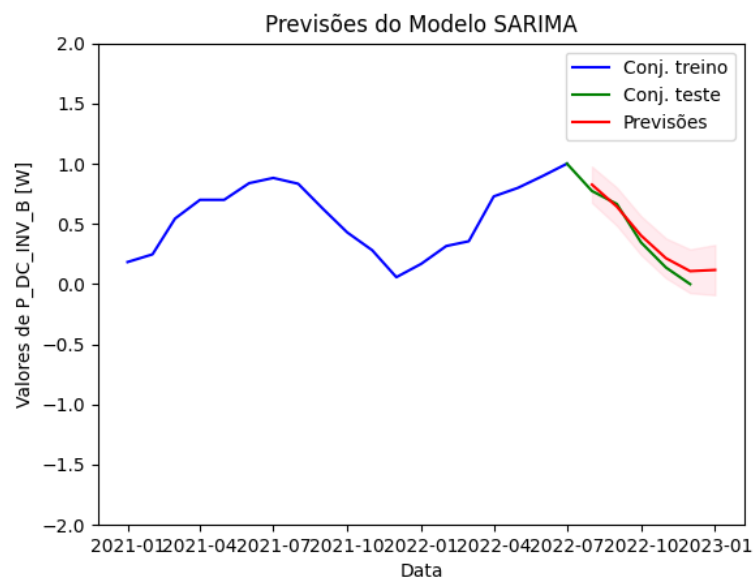


Figura 93 : Comparação dados reais e previstos para a variável P_DC_INV_B [W]

A Figura 93 apresenta uma boa aproximação das previsões desenvolvidas pelo modelo ao conjunto de teste, a linha vermelha que representa as previsões do modelo encontra-se em parte sobreposta à linha verde, representante do conjunto de teste. A nível visual, o modelo aparenta um bom desempenho, para confirmar esta conclusão, foram calculados os valores de MAE e RMSE e representados na Tabela 41.

Tabela 41 : Valores de MAE e RMSE para modelo SARIMA para a variável P_DC_INV_B [W]

Modelo SARIMA	MAE	RMSE
(3,0,1) (0,0,0,12)	0.063321	0.069395

O modelo admite valores de MAE e RMSE, Tabela 41, muito próximos de zero, o que indica uma boa precisão e qualidade de previsão.

Desta forma, foram desenvolvidas as previsões para datas futuras com intervalo de confiança, representadas na Tabela 42 e apresentadas graficamente na Figura 94.

Tabela 42 : Valores previsões futuras com intervalos de confiança para a variável P_DC_INV_B [W]

Data	Previsão	Limite inferior	Limite superior
2023-01	0,081797	-0,05685	0,220444
2023-02	0,247111	0,104174	0,390048
2023-03	0,505051	0,360113	0,649988
2023-04	0,755944	0,605016	0,906871
2023-05	0,945129	0,777771	1,112487
2023-06	1,014574	0,822232	1,206916
2023-07	0,950316	0,740272	1,160361
2023-08	0,769911	0,553038	0,986784
2023-09	0,52515	0,308254	0,742046
2023-10	0,283457	0,063197	0,503718
2023-11	0,111214	-0,12147	0,343895
2023-12	0,054592	-0,19491	0,304093

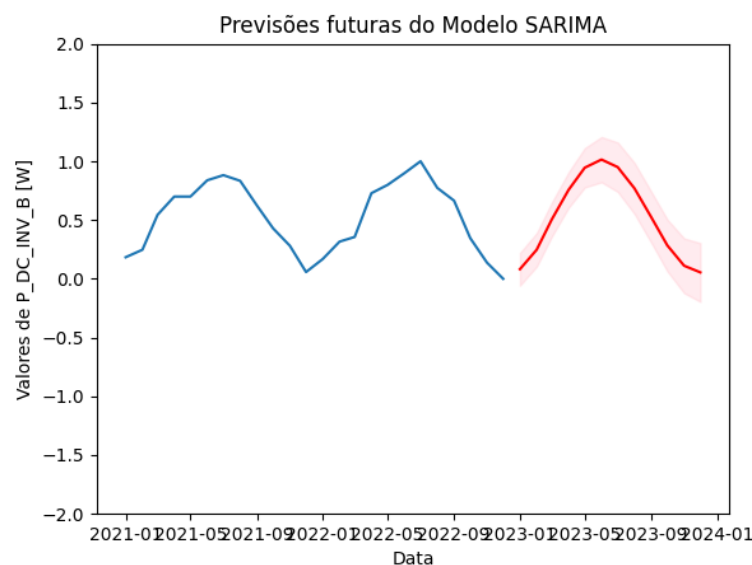


Figura 94 : Previsões futuras da temperatura média para a variável P_DC_INV_B [W]

Em conclusão, a Figura 94, apresenta as previsões para os valores da potência do lado DC do inversor B para o ano seguinte, e facilmente, percebe-se que o comportamento dos dados não será alterado de forma significativa, o que faz sentido, visto que, o sistema fotovoltaico ainda é recente.

4.4.7. Variável 7: Potencia do lado AC do inversor B ($P_{AC_INV_B}$ [W])

De seguida, é implementado o modelo SARIMA para estudar a previsão de valores para a variável relacionada com a potência do lado AC do inversor B, neste caso. A metodologia aplicada passo pelos mesmos passos das variáveis já estudadas.

Inicialmente, procedeu-se à elaboração do gráfico que evidencia a média móvel e o desvio padrão móvel da série ao longo do tempo. De notar que, a janela de tempo utilizada foi a mesma e mais, uma vez, a quantidade de observações limitou este estudo. O gráfico é apresentado na Figura 95.

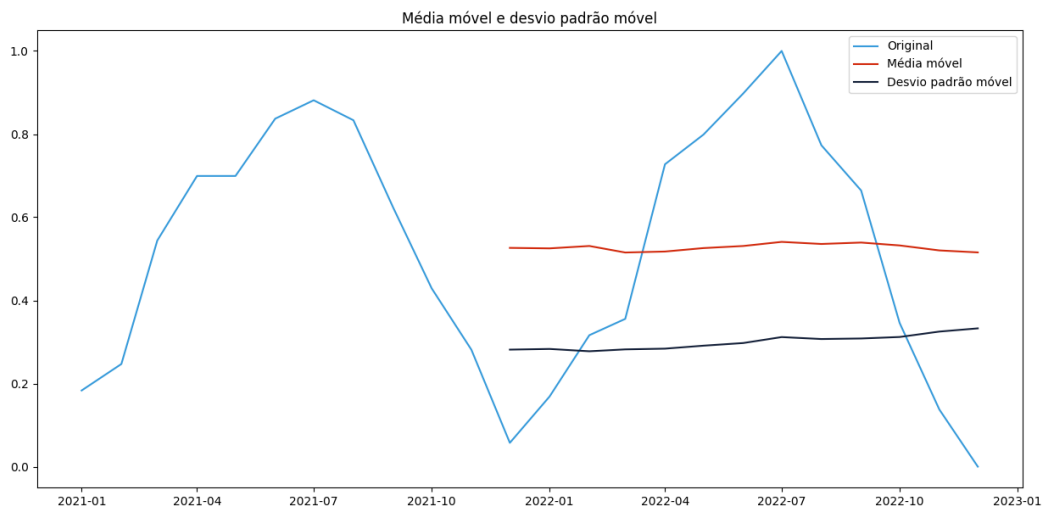


Figura 95 : Gráfico da média móvel e desvio padrão móvel da série ao longo do tempo da variável P_AC_INV_B [W]

Pela análise da Figura 95, é possível verificar que tanto a média móvel como o desvio padrão móvel da série não apresentam oscilações significativas. Como não é possível, apenas pela figura 30x aferir quanto à estacionariedade da série, de seguida, foi aplicado o teste ADF, cujos resultados são apresentados na Figura 96.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-4.010331	0.001357	-3.924019	-3.068498	-2.673893

Figura 96 : Resultados teste ADF para a variável P_AC_INV_B [W]

Pela análise da Figura 96, é possível afirmar que a série é estacionária, visto que p_value é menor que 0,05 e, ainda, o valor da estatística do teste ADF é menor do que qualquer valor crítico. Apesar da variável apresentar sazonalidade (pela análise exploratória dos dados) não são necessárias diferenciações para a tornar estacionária, pelo que $d=0$ e $D=0$.

De seguida, procedeu-se à restante parametrização do modelo. Inicialmente, para traçar valores máximos dos restantes parâmetros, desenvolveram-se o gráfico ACF e PACF, com o objetivo de avaliar a complexidade da série, a quantidade de números significativos, representados na Figura 97.

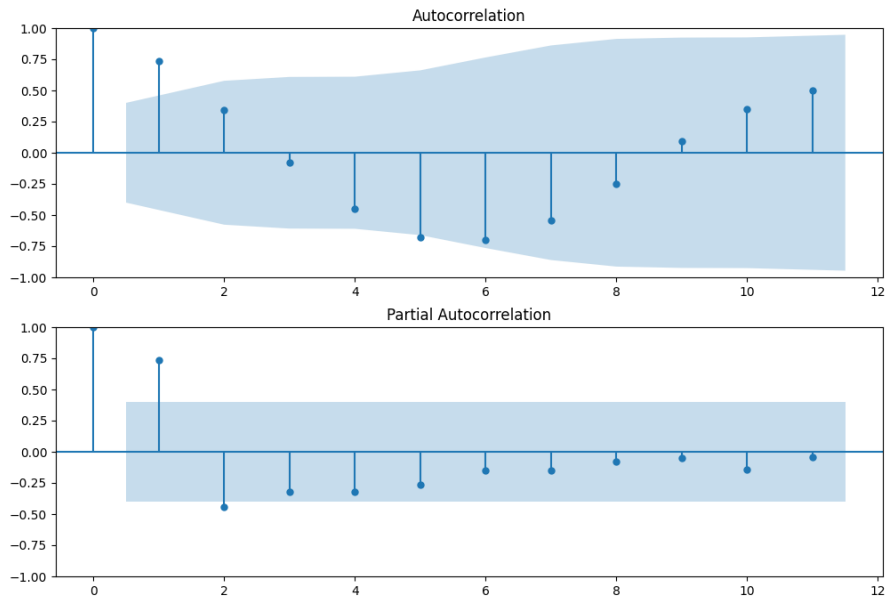


Figura 97 : Gráficos ACF e PACF para a variável P_AC_INV_B [W]

Os gráficos apresentados na Figura 97 não apresentam muitos pontos significativos, quer isto dizer, pontos fora da área sombreada, pelo que mais uma vez não foi alterado o valor máximo de qualquer parâmetro.

Recorrendo às bibliotecas do *Python* para selecionar os melhores parâmetros, em particular, à biblioteca *pmdarima*, especificamente, à função *auto-arima*, segundo o critério de informação AIC, no qual, o melhor modelo é o que admite menor valor deste critério. Os resultados dos parâmetros e do valor AIC são demonstrados na Tabela 43.

Tabela 43 : Parâmetros modelo SARIMA e valor de AIC

(p, d, q) (P, D, Q, s)	AIC
$(3,0,1)$ $(0,0,0,12)$	-24.730860

De seguida, procedeu-se ao diagnóstico do modelo, Figura 98.

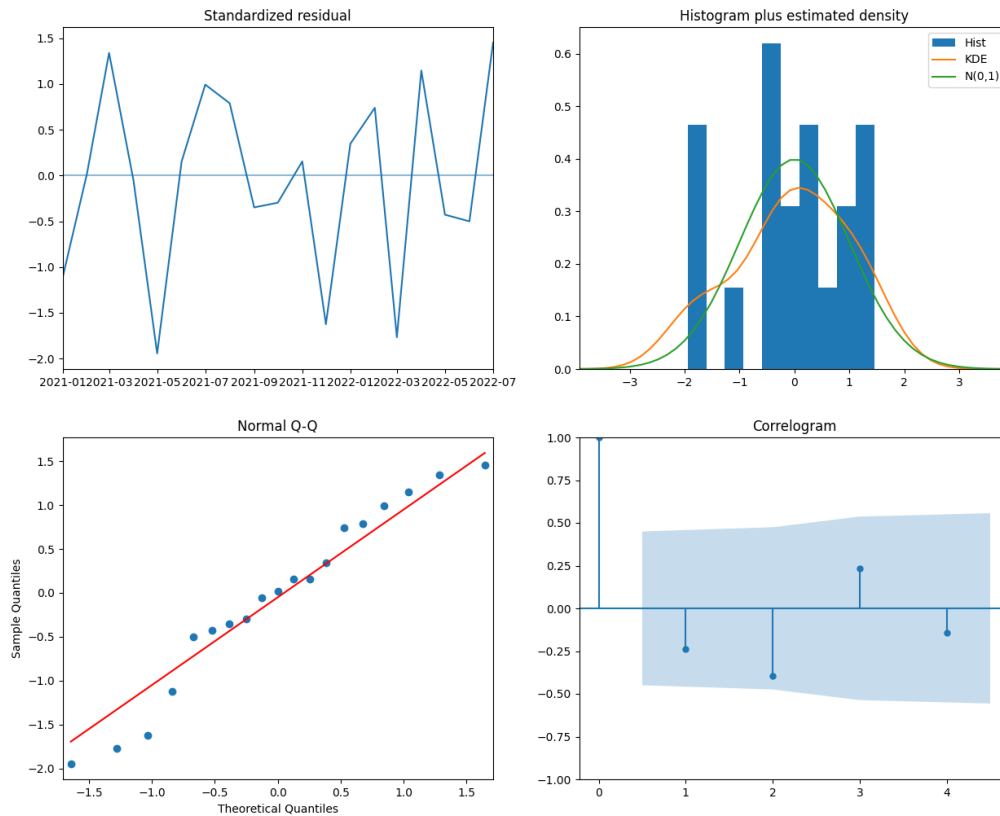


Figura 98 : Diagnóstico do modelo

Pela Figura 98 é possível constatar que:

- Gráfico dos resíduos ao longo do tempo não admite padrões;
- A curva do gráfico da distribuição KDE apresenta uma forma semelhante à curva da distribuição normal;
- Pelo gráfico QQ dos resíduos mostra que, muitos pontos se sobrepõem à reta vermelha, seguindo uma tendência linear, evidenciando assim que os resíduos são normalmente distribuídos;
- O gráfico de autocorrelação não apresenta nenhum ponto significativo;

Com base na análise realizada, parece que o modelo possui uma boa precisão e estará pronto para fazer previsões. A seguir, são apresentadas as previsões para o conjunto de teste na Tabela 44 e mostramos uma comparação entre os valores reais e os valores previstos na Figura 99.

Tabela 44 : Valores das previsões para o conjunto de teste para a variável P_AC_INV_B [W]

Data	Previsão
2022-08	0.826233
2022-09	0.643271
2022-10	0.403618
2022-11	0.214936
2022-12	0.107520

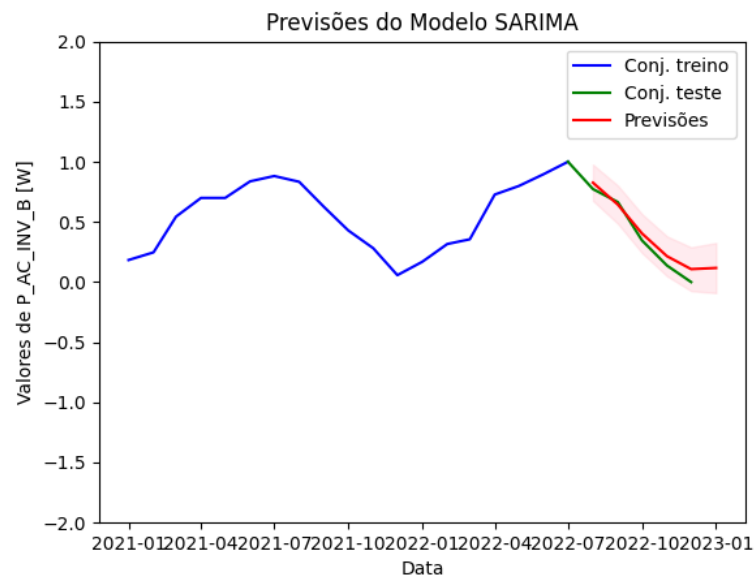


Figura 99 : Comparação dados reais e previstos para a variável P_AC_INV_B [W]

A Figura 99, mostra uma boa aproximação do modelo ao conjunto de dados reais, com momentos em que por vezes as duas linhas se sobrepõem, linha verde (conjunto de teste) e a linha vermelha (previsões). Desta forma, é desde já possível concluir que o modelo acompanha o comportamento dos dados históricos e consegue prever valores com qualidade, para confirmar são apresentados os valores de MAE e RMSE como metodologias de avaliação do modelo, Tabela 45.

Tabela 45 : Valores de MAE e RMSE para modelo SARIMA para a variável P_AC_INV_B [W]

Modelo SARIMA	MAE	RMSE
(3,0,1) (0,0,0,12)	0.063321	0.069395

Os valores de MAE e RMSE, mostrados na Tabela 45, enfatiza o que já concluído anteriormente, o modelo apresenta boa capacidade de previsão, e, portanto, procedeu-se ao desenvolvimento das previsões para datas futuras, cujos valores são apresentados na Tabela 46 e são representados graficamente na Figura 100.

Tabela 46 : Valores previsões futuras com intervalos de confiança para a variável P_AC_INV_B [W]

Data	Previsão	Limite inferior	Limite superior
2023-01	0,082023	-0,05618	0,220222
2023-02	0,246624	0,104562	0,388686
2023-03	0,505243	0,361088	0,649398
2023-04	0,756201	0,605886	0,906515
2023-05	0,94602	0,77961	1,11243
2023-06	1,015508	0,82427	1,206745
2023-07	0,951076	0,742445	1,159708
2023-08	0,769898	0,554552	0,985245
2023-09	0,524119	0,308765	0,739472
2023-10	0,281357	0,062603	0,500111
2023-11	0,108442	-0,12274	0,339622
2023-12	0,051824	-0,19609	0,299741

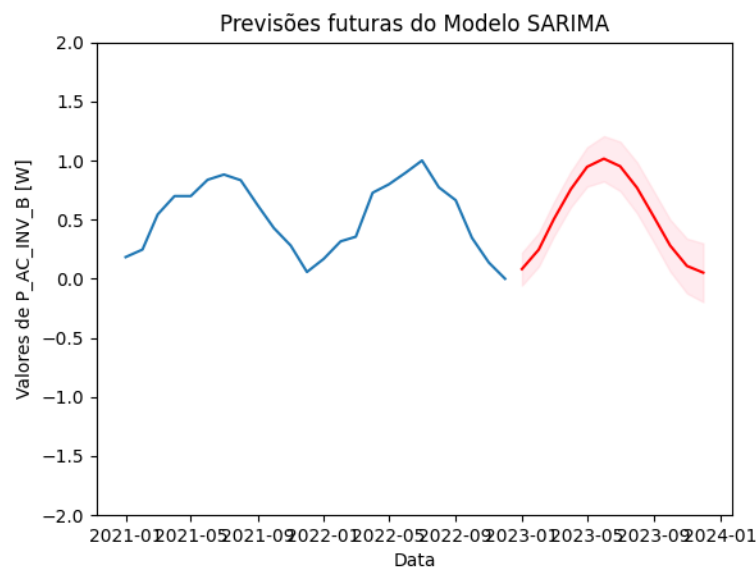


Figura 100 : Previsões futuras da temperatura média para a variável P_AC_INV_B [W]

Em conclusão, como seria de esperar, o modelo apresenta previsões para o ano seguinte expectáveis, visto que, não é normal, ocorrerem falhas nos primeiros anos do sistema fotovoltaico. Para além disso, lembrar o facto de apenas existirem observações para dois anos o que dificulta a análise de falhas, no entanto, o ficheiro fica capaz de ser alimentado por novos dados.

4.4.8. Variável 8: Temperatura do inversor B (T_INV_B [°C])

A temperatura do inversor B também é um variável relevante, para estudar perdas de eficiência neste equipamento. Neste sentido, é aplicado o modelo SARIMA com o foco de apresentar valores de previsão para datas futuras. Desta forma, as etapas de construção foram as mesmas aplicadas às variáveis que têm vindo a ser estudadas.

Inicialmente, desenvolveu-se o gráfico que se ilustra o comportamento da média móvel e desvio padrão móvel da serie ao longo do tempo, representado na Figura 101. A janela temporal de desenvolvimento do gráfico foi de 12 observações, tal como nas outras variáveis.

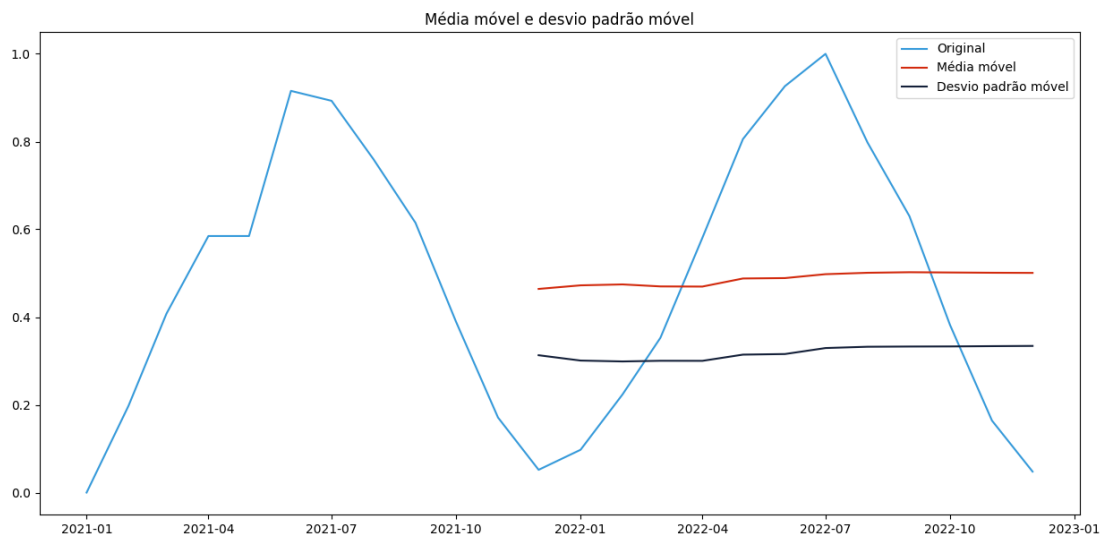


Figura 101 : Gráfico da média móvel e desvio padrão móvel ao longo do tempo da variável T_INV_B [°C]

A Figura 101 mostra uma média móvel e desvio padrão móvel praticamente constantes, mas não pode ser concluído nada sobre a estacionariedade da série, dado o número reduzido de observações e também o facto de apenas o teste ADF permitir conclusões de valor sobre este tema. Desta forma, para estudar a estacionariedade da série foi aplicado o método ADF e os resultados são apresentados na Figura 102.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-0.765005	0.829186	-4.012034	-3.104184	-2.690987

Figura 102 : Resultados teste ADF para a variável T_INV_B [°C]

Pela Figura 102, é possível concluir que a série não é estacionária pois p_value é maior que 0.05 e, portanto, fica violada a condição fulcral para admitir a série estacionária. Desta forma é necessário aplicar diferenciações à série para a tornar estacionária. O parâmetro d foi calculado tendo em conta a Figura 25 e o resultado foi que $d=0$. No entanto, na análise exploratória a série admitia sazonalidade, pelo que o parâmetro relativo às diferenciações sazonais deve ser um, $D=1$, inicialmente. Assim, foi aplicada uma diferenciação sazonal à série e logo de seguida, aplicado teste

ADF para testar a estacionariedade da série diferenciada. Os valores do resultado do teste ADF à série diferenciada são apresentados na Figura 103.

Test Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-5.247767	0.000007	-4.938690	-3.477583	-2.843868

Figura 103 :Resultados teste ADF para a variável T_INV_B [°C] após diferenciação sazonal da série

Após a diferenciação sazonal da série, os parâmetros do teste ADF alteram-se drasticamente, e p_value é menor que 0,05, para além disso, o valor da estatística do teste ADF é menor do que qualquer valor critico. Portanto, a diferenciação sazonal tornou a série estacionária e é possível avançar para a próxima etapa.

Os gráficos ACF e PACF forma utilizados para analisar a complexidade dos dados, analisar a quantidade de dados significativos que cada um apresentava, Figura 104.

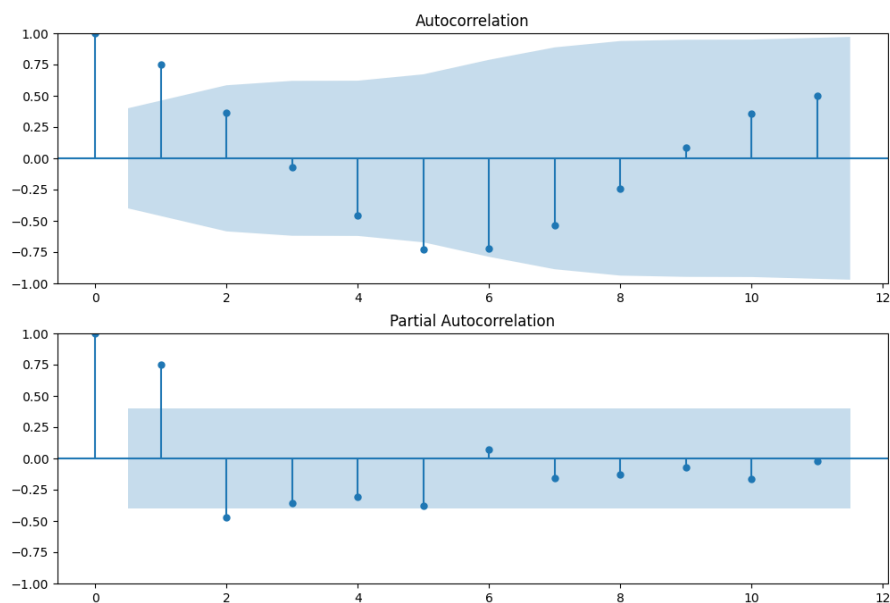


Figura 104 : Gráficos ACF e PACF para a variável T_INV_B [°C]

Tanto no gráfico ACF como PACF não são evidentes muitos pontos significativos, portanto, não se deve optar por valores de parâmetros muito elevados porque torna o modelo mais complexo sem necessidade. Deste modo, o valor máximo para qualquer parâmetro permaneceu em 3 e o mínimo em 0.

Os intervalos de valores a serem estudados para os restantes parâmetros já foram estudados, assim com recurso à biblioteca `pmdarima`, foram selecionados os melhores parâmetros, em particular, com recurso à função `auto-arima`, segundo o critério de informação AIC, no qual, o melhor

modelo é o que admite menor valor deste critério. Os resultados dos parâmetros e do valor AIC são demonstrados na Tabela 47.

Tabela 47 : Parâmetros modelo SARIMA e valor de AIC

(p, d, q) (P, D, Q, s)	AIC
$(0,0,0)$ $(0,1,0,12)$	-10.660558

Posteriormente, é desenvolvido o diagnóstico do modelo, Figura 105.

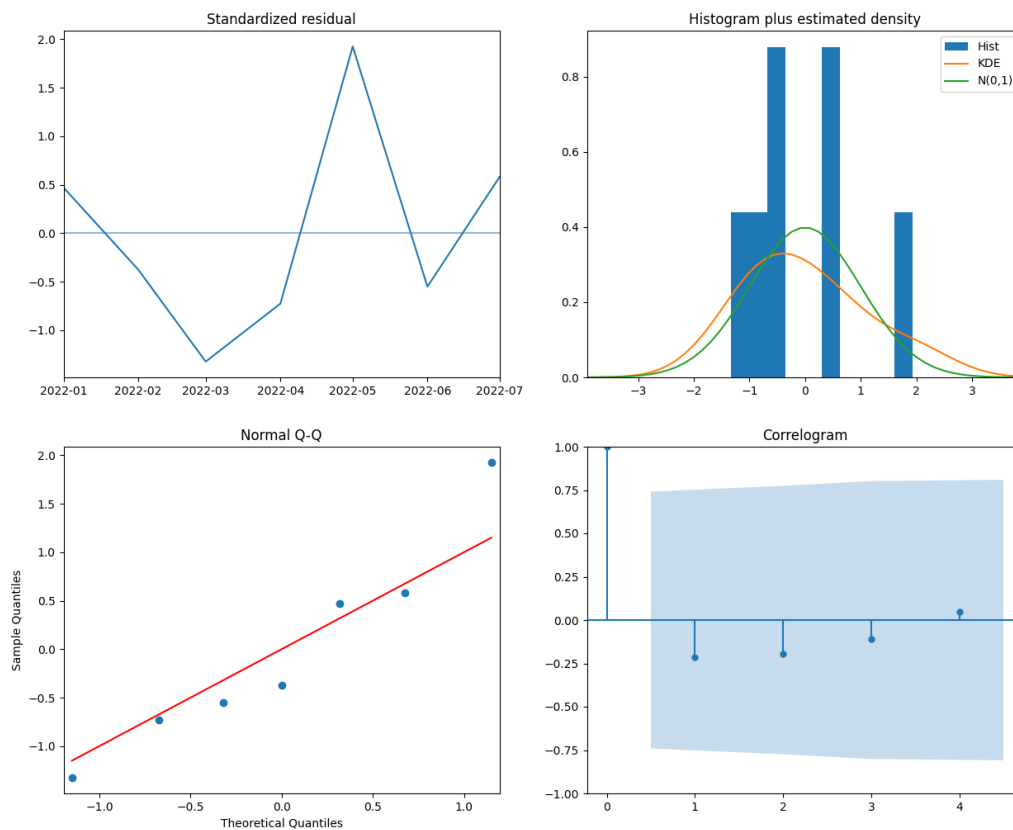


Figura 105 : Diagnóstico do modelo

Pela Figura 105 pode-se fazer a seguinte análise:

- Gráfico dos resíduos ao longo do tempo não mostra padrões;
- A curva da distribuição KDE aparenta semelhanças com a curva de distribuição normal;
- Pelo gráfico QQ dos resíduos verifica-se que os pontos azuis seguem uma tendência linear, evidenciando assim que os resíduos são normalmente distribuídos;
- O gráfico de autocorrelação não apresenta nenhum ponto significativo;

De acordo com a análise anterior, é possível concluir que o modelo está apto para desenvolver previsões. Seguidamente, são apresentadas as previsões para o conjunto teste na Tabela 48 e evidenciada uma comparação entre os valores reais e os previstos na Figura 106.

Tabela 48 : Valores das previsões para o conjunto de teste para a variável T_INV_B [°C]

Data	Previsão
2022-08	0.817236
2022-09	0.672493
2022-10	0.446726
2022-11	0.228589
2022-12	0.109417

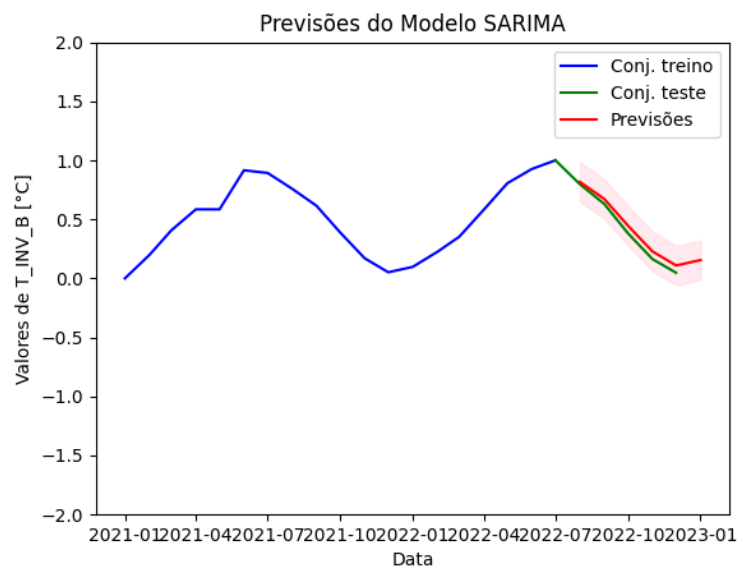


Figura 106 : Comparação dados reais e previstos para a variável T_INV_B [°C]

A Figura 106 mostra que o modelo SARIMA foi capaz de prever valores para o período relativo ao conjunto de teste, idênticos aos valores reais. Efetivamente, a linha vermelha, que representa as previsões, adquire um comportamento muito semelhante, em certos momentos quase que se sobrepõe à linha verde. A Figura 106 no eixo y evidencia os valores para a temperatura média do inversor B em graus celsius.

Além de avaliar o modelo visualmente, foram ainda calculados os valores do erro absoluto médio (MAE) e da raiz do erro quadrático médio (RMSE) os resultados são apresentados na Tabela 49.

Tabela 49 : Valores de MAE e RMSE para modelo SARIMA para a variável T_INV_B [°C]

Modelo SARIMA	MAE	RMSE
(0,0,0) (0,1,0,12)	0.050499	0.053517

Os valores apresentados de MAE e RMSE são muito próximos de zero, o que confirma a qualidade do modelo já mencionada anteriormente pela análise visual. Deste modo, foram desenvolvidas as previsões para datas futuras com intervalo de confiança, representadas na Tabela 50 e graficamente apresentadas na Figura 107.

Tabela 50 : Valores previsões futuras com intervalos de confiança para a variável T_INV_B [°C]

Data	Previsão	Limite inferior	Limite superior
2023-01	0,134068	-0,00392	0,272056
2023-02	0,260193	0,122205	0,398181
2023-03	0,389322	0,251334	0,52731
2023-04	0,617116	0,479128	0,755105
2023-05	0,842381	0,704393	0,98037
2023-06	0,96285	0,824861	1,100838
2023-07	1,036525	0,898537	1,174513
2023-08	0,834582	0,696594	0,972571
2023-09	0,666594	0,528605	0,804582
2023-10	0,418531	0,280543	0,556519
2023-11	0,200561	0,062573	0,338549
2023-12	0,084322	-0,05367	0,22231

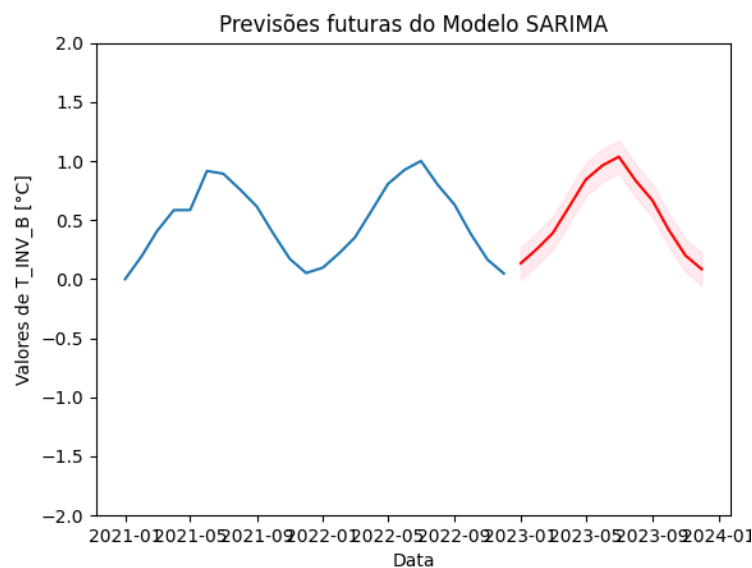


Figura 107 : Previsões futuras da temperatura média para a variável T_INV_B [°C]

As previsões futuras apresentam um comportamento semelhante aos dados históricos, e portanto, para já não são expectáveis falhas no período previsto Tal facto, seria expectável porque mais uma vez o sistema fotovoltaico é recente.

4.5. Discussão de resultados

Em geral todas as séries temporais das diversas variáveis de interesse apresentavam sazonalidade anual e, portanto, foi possível aplicar o modelo SARIMA. Uma variável importante de implementar seria a eficiência de cada inversor, no entanto, não admitiam sazonalidade e, portanto, não foram estudadas as previsões destas variáveis. No entanto, a sua análise desempenha um papel importante na identificação de perdas no sistema.

Com a análise global dos dois métodos aplicados, é possível concluir que ambos apresentam um bom desempenho. Para além disso, é importante relatar mais uma vez que, o principal fator limitante desta abordagem foi um conjunto de dados curto.

Os sistemas fotovoltaicos normalmente têm uma vida útil estimada de 25 anos, mas os dados utilizados na análise referem-se apenas aos primeiros dois anos da instalação. Como resultado, era natural que nenhuma falha fosse prevista no período considerado. À medida que mais dados são armazenados ao longo do tempo e fornecidos a ferramenta torna-se mais capaz de identificar falhas e perdas de eficiência à medida que ocorrem.

Comparação entre os dois modelos

Os modelos SARIMA e redes neurais MLP foram aplicados à variável que representa a temperatura do inversor A em graus Celsius. Esta abordagem possibilita a comparação dos valores de MAE (Erro Absoluto Médio) entre os dois modelos, permitindo avaliar a qualidade das previsões de cada um. O MAE foi uma métrica de desempenho comum calculada, e os resultados estão apresentados na Tabela 51.

Tabela 51: Comparação do valor de MAE entre os dois algoritmos para a variável relativa à temperatura do inversor A

SARIMA - MAE	Redes neurais MLP - MAE
0.058633	0.062594

Os valores de MAE obtidos para ambos os algoritmos, apresentados na Tabela 51 indicam uma alta qualidade de previsão, com valores previstos muito próximos dos valores reais do conjunto de teste. Ambos os modelos apresentaram desempenho promissor na previsão desta variável.

É importante notar que as redes neurais MLP são mais complexas e implicam uma maior capacidade computacional, embora neste caso não tenha apresentado uma diferença significativa em relação ao SARIMA devido ao tamanho reduzido do conjunto de dados. No entanto, as redes neurais têm a vantagem de identificar padrões de maneira mais sofisticada, conforme sugerido pela literatura. Em resumo, ambos os algoritmos produziram variações de alta qualidade para essa variável.

Vantagens da utilização de um intervalo de confiança

A análise em concreto dos resultados do modelo SARIMA, para qualquer variável mostram um bom desempenho e por parte do modelo e previsões de qualidade, quando comparamos os valores previstos com os valores reais do conjunto de teste.

Para além das previsões das datas específicas, foi também considerado um intervalo de confiança, representado por uma área sombreada nas figuras de previsão para datas futuras. Esta implementação tem como objetivo estabelecer limites mínimos e máximos para os valores de cada variável.

Após a conclusão dos modelos de previsão, podem ser gerados dois tipos de alertas com base na comparação com os valores reais. Emitir alerta quando os valores reais estão dentro dos valores estipulados pela área sombreada. Em segundo lugar, emitir um alerta que destaca a possibilidade de uma falha no sistema quando os valores reais estão fora da área sombreada.

Esta abordagem evita o envio de alertas de falhas desnecessárias que ocorreriam se fossem apenas considerados os valores exatos de previsão. Por exemplo, em variáveis como temperatura e radiação, onde as condições climáticas podem ser imprevisíveis e extremas, o sistema não deve emitir um alerta de falha apenas porque a temperatura real não coincide exatamente com a previsão para o respetivo mês, pois pode ser apenas um dia fora do normal, como cada vez mais acontece. O intervalo de confiança oferece uma margem para acomodação nessas previsões. Depois todas as variáveis são afetadas, e por isso esta abordagem foi inserida em todas as variáveis.

É importante ressaltar que o limite de confiança pode ser ajustado de acordo com as preferências dos utilizadores da ferramenta. É recomendável, uma pesquisa na literatura para determinar a melhor forma de adaptar esse parâmetro.

Além de indicar possíveis falhas, o modelo de previsão pode ser utilizado para comparar os valores reais com os previstos. Para tal, era importante, que os dados fossem recolhidos em tempo real, ou então, num período de 3 meses, por exemplo, porque neste caso, o modelo está habilitado para previsões mensais. Efetivamente, mesmo quando os valores reais estão dentro dos limites estabelecidos pelo intervalo de confiança, é possível identificar tendências e comportamentos nos dados. Por exemplo, se a cada 3 meses os dados forem fornecidos, é possível avaliar valores e comportamentos trimestralmente, os dados podem indicar uma tendência crescente ou decrescente nos seus valores, assim a equipa de análise dados pode tomar medidas de manutenção preventiva nesse sistema fotovoltaico específico. Pois apesar dos valores estarem dentro dos limites, pode indicar uma possível falha futuramente.

Essas medidas preventivas são essenciais para evitar problemas, danos nos equipamentos e, conseqüentemente, diminuição da geração de energia ou até mesmo a interrupção completa da geração de energia elétrica. Para além disso, estas medidas são bastante vantajosas para a economia da organização, evitando custos elevados de reparo ou até substituição de equipamentos, bem como os prejuízos associados às quedas na produção de energia.

A utilização de intervalos de confiança não apenas para detetar falhas, mas também para identificar tendências e comportamentos precocemente, é uma vantagem fundamental desta abordagem, que contribui para a eficiência operacional e a confiabilidade dos sistemas fotovoltaicos.

Valores de previsão ligeiramente acima dos valores de conjunto de teste

O desenvolvimento de previsões para o conjunto de teste com recurso ao modelo SARIMA permite-nos concluir, que o modelo admite previsões com valores ligeiramente acima dos valores reais. Esta abordagem, merece uma análise cuidadosa, das suas vantagens e desvantagens.

Uma das principais vantagens dessa abordagem é a capacidade de gerar alertas atempadamente de possíveis falhas em equipamentos ou processos. Os alertas precoces permitem agendar uma ação de manutenção preventiva, antes que uma falha real ocorra, economizando tempo e recursos.

A antecipação de potenciais problemas, permite evitar quebras inesperadas e interrupções no processo de produção de energia elétrica.

A principal desvantagem dessa abordagem é a possibilidade de gerar falsos alertas ou alertas antes do tempo. Como o modelo SARIMA prevê ligeiramente acima do valor pode levar à programação de manutenção desnecessária. Isso pode resultar em custos extras, tempo perdido, alocação de recursos, como mão de obra e peças de reposição, quando não são realmente necessários. Isso pode sobrecarregar os departamentos de manutenção e aumentar os custos operacionais.

No entanto, embora haja o risco de agendar ações de manutenção antes do que realmente é necessário, isso pode ser compensado pela redução de custos associados a reparos de danos causados por falhas. A manutenção preventiva tende a ser menos dispendiosa do que a correção de danos graves, especialmente em equipamentos dispendiosos.

5. CONCLUSÃO

No presente capítulo, será analisado de uma forma geral o projeto desenvolvido com o objetivo de abordar as questões mencionadas no início do projeto. Além disso, serão evidenciados os principais desafios encontrados na construção da ferramenta, e por fim, são propostas direções futuras de pesquisa.

5.1. Principais conclusões

O projeto apresentado procurou criar uma ferramenta, capaz de analisar dados fornecidos pela SENS e seguidamente, desenvolver modelos capazes de identificar e prever precocemente as falhas e problemas em painéis fotovoltaicos no sentido de evitar prejuízos adicionais e prolongar a vida útil dos painéis. Para tal, procurou-se responder às seguintes questões de investigação: Como a manutenção preditiva pode ser integrada nuns sistemas de gestão de energia para melhorar a eficiência e a eficácia dos sistemas de painéis fotovoltaicos? Como a manutenção preditiva pode ser utilizada para reduzir os custos de manutenção dos painéis fotovoltaicos ao longo do tempo?

Para abordar as questões delineadas e alcançar os objetivos propostos anteriormente, foi utilizada a linguagem de programação *Python* para desenvolver métodos de previsão, nomeadamente, redes neuronais e SARIMA. Os métodos de previsão implementados permitiram não apenas a análise e previsão de valores das variáveis essenciais para o bom funcionamento do sistema, mas também a avaliação de possíveis falhas e o comportamento futuro dessas variáveis em relação aos equipamentos em estudo.

Como resposta às questões evidenciadas revela-se que através da análise de dados desenvolvida e da construção dos modelos de previsão, demonstrou-se como a manutenção preditiva pode ser eficazmente integrada em sistemas de gestão de energia para melhorar a eficiência e o desempenho dos sistemas de painéis fotovoltaicos. Os resultados obtidos, para já não evidenciaram problemas ou falhas num futuro próximo nos equipamentos estudados, no entanto, estes resultados eram previsíveis, visto que, nos primeiros anos de vida útil dos sistemas, é garantido um bom funcionamento dos mesmos e para além disso, são garantidos valores de desempenho elevados. Porém, a ferramenta desenvolvida, é capaz de incorporar novos dados e estudar novamente valores de previsão futuros ao longo da vida útil dos sistemas. Assim, os resultados dos modelos aplicados permitem a identificação precoce de falhas nos sistemas fotovoltaicos, desencadeando ações de manutenção preventiva de forma oportuna. Não contribuindo apenas para a maximização do desempenho do sistema fotovoltaico, mas também para a redução dos custos operacionais e a otimização do mesmo. Desta forma, o projeto desenvolvido oferece uma base sólida para a integração bem-sucedida da manutenção preditiva nos sistemas de gestão de energia, melhorias na eficiência e melhoria dos sistemas de painéis fotovoltaicos.

5.2. Desafios encontrados

No desenvolvimento deste projeto foram encontrados alguns desafios, que tornaram a análise de dados e a implementação de métodos de previsão de falha num sistema fotovoltaico um processo mais desafiador e complexo.

O conjunto de dados fornecidos é o principal motor para todo o desenvolvimento do projeto, e, portanto, deve ser o mais adequado possível. Desta forma de seguida, são apresentados os principais desafios encontrados e de que forma influenciou o desenvolvimento do projeto.

– Conjunto de dados limitado

Um dos principais desafios enfrentados no desenvolvimento deste projeto foi a disponibilidade de dados com um período de observação relativamente curto. A limitação temporal dificulta a identificação de tendências de longo prazo e padrões sazonais.

Os modelos de previsão baseiam-se em dados históricos para prever valores futuros, e, portanto, o conjunto de dados revelou-se um fator limitante para o estudo. De facto, quanto mais completo for o conjunto de dados históricos melhor o modelo consegue captar variações e características de determinada variável ao longo do tempo.

– Confusão na Interpretação das Variáveis

A segunda dificuldade enfrentada durante o desenvolvimento do projeto foi a dificuldade na interpretação das variáveis. Algumas variáveis careciam de designações mais informativas, tornando difícil entender seu significado exato. Isso complicou a análise e a interpretação dos resultados.

– Dados de Variáveis Sem Sentido

Outro desafio crítico foi a presença de valores de variáveis que não faziam sentido ou continham valores irrealistas. Esses valores sem sentido prejudicaram a capacidade de identificar falhas de forma precisa e detalhada. Em específico, este desafio mostrou-se mais frequentemente em variáveis relacionadas com valores relativos aos *MPPT's* e *string*, desta forma, foi impossível analisar estes valores, e conseqüentemente, não permitiu o desenvolvimento de previsões de falha de forma tão pormenorizada.

Para abordar a falta de dados, foi considerada a geração de dados históricos através de pequenas variações dos dados existentes. No entanto, não se revelou uma abordagem eficaz. Dado que o conjunto de dados já tinha sido modificado para eliminar nulos, a geração de dados com pequenas variações poderia levar a conclusões falsas, ao forçar a similaridade entre os valores passados e presentes.

Na tentativa de abordar esses desafios, foi feita a troca do conjunto de dados original por um novo conjunto. Embora o novo conjunto tenha atenuado alguns problemas, os desafios fundamentais persistiram. Mesmo com menos problemas de qualidade de dados, ainda era necessário realizar um tratamento significativo dos dados para torná-los adequados para análise.

Apesar dos desafios mencionados, ambos o modelos apresentaram boa qualidade de previsão como evidenciado anteriormente.

5.3. Trabalhos futuros

É possível delinear várias propostas de trabalhos futuros, nomeadamente, aplicar o algoritmo de redes neuronais às restantes variáveis do estudo, e conseqüentemente, desenvolver uma análise comparativa entre os dois modelos implementados para as restantes variáveis. Para além disso, aplicar novos modelos, com o objetivo de escolher o melhor modelo, apesar dos dois modelos já implementados. Em específico, o modelo *LSTM*, que é uma variação das redes neuronais que consegue captar comportamentos de series temporais.

Outro ponto a considerar para investigação futura seria construir alertas de possíveis falhas de acordo com valores fora dos limites do intervalo de confiança, os valores de previsão para qualquer variável, são guardados num Excel, para que, posteriormente fosse mais fácil de desenvolver este ponto.

Por fim, integrar a ferramenta com estudos em *Power BI*, permitindo a apresentação de dados de forma atraente e informativa.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abdulmawjood, K., Refaat, S. S., & Morsi, W. G. (2018). Detection and prediction of faults in photovoltaic arrays: A review. *Proceedings - 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering, CPE-POWERENG 2018*, 1–8. <https://doi.org/10.1109/CPE.2018.8372609>
- Aboagye, B., Gyamfi, S., Ofosu, E. A., & Djordjevic, S. (2022). Investigation into the impacts of design, installation, operation and maintenance issues on performance and degradation of installed solar photovoltaic (PV) systems. *Energy for Sustainable Development*, 66, 165–176. <https://doi.org/10.1016/j.esd.2021.12.003>
- Addo-Tenkorang, R., & Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literature review. *Computers and Industrial Engineering*, 101, 528–543. <https://doi.org/10.1016/j.cie.2016.09.023>
- Aditya Satrio, C. B., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, 179, 524–532. <https://doi.org/10.1016/j.procs.2021.01.036>
- Akalke, H. (1974). A New Look at the Statistical Model Identification. In *Aulomatic Control* (Vol. 19).
- Albertt Santos Carvalho, G. (2020). USO DE SÉRIES TEMPORAIS PARA PREVISÃO DE IRRADIAÇÃO SOLAR.
- Almeida, E., Clara ROSA, A., Cristina Lima Sales DIAS, F., Thais Mariotto BRAZ, K., Teixeira Costa LANA, L., Castro do Espírito SANTO, O., & Cristina Bajur SACRAMENTO, T. (n.d.). ENERGIA SOLAR FOTOVOLTAICA: REVISÃO BIBLIOGRÁFICA.
- Alves, M. F., Lotufo, A. D. P., & Lopes, M. L. M. (2014, July 30). Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas. <https://doi.org/10.5540/03.2013.001.01.0144>
- Antonioli, E. A., Adriana, C., & Schmidt, P. (n.d.). REVISÃO SISTEMÁTICA: APLICAÇÃO DE REDES NEURAS PARA PREVISÃO DO CONSUMO DE ENERGIA ELÉTRICA CONSUMPTION.
- Bernardino Da Cunha, G., Luitgards-Moura, J. F., Lázaro, E., Naves, M., Andrade, A. O., Pereira, A. A., & Milagre, S. T. (2010). A utilização de uma rede neural artificial para previsão da incidência da malária no Município de Cantá, Estado de Roraima. In *Revista da Sociedade Brasileira de Medicina Tropical* (Vol. 43, Issue 5).
- Bouzerdoum, M., Mellit, A., & Massi Pavan, A. (2013). A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*, 98(PC), 226–235. <https://doi.org/10.1016/j.solener.2013.10.002>
- Brendan Artley. (2022, April 26). Time Series Forecasting with ARIMA , SARIMA and SARIMAX. *Towards Data Science*. <https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6>

- Buchatskaya, V., Buchatsky, P., & Teploukhov, S. (2015). Forecasting Methods Classification and its Applicability. *Indian Journal of Science and Technology*, 8(30).
<https://doi.org/10.17485/ijst/2015/v8i1/84224>
- Campana BARBOSA, E., Sáfadi, T., NASCIMENTO, M., NASCIMENTO, A. C. C., SILVA, C. H. O., & MANULI, R. C. (2015). METODOLOGIA BOX & JENKINS PARA PREVISÃO DE TEMPERATURA MÉDIA MENSAL DA CIDADE DE BAURU (SP) (Issue 1).
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chen, K. Y., & Wang, C. H. (2007). A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications*, 32(1), 254–264. <https://doi.org/10.1016/j.eswa.2005.11.027>
- Chen, P., Niu, A., Liu, D., Jiang, W., & Ma, B. (2018). Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing. *IOP Conference Series: Materials Science and Engineering*, 394(5). <https://doi.org/10.1088/1757-899X/394/5/052024>
- Conrado, D. M. (2021). Estudo dos principais aspectos de manutenção em sistemas fotovoltaicos on-grid. <https://repository.ufrpe.br/handle/123456789/2705>
- da Energia, O. (2021). *Energia em Números - Edição 2021*.
- da Silva, P. G. (2011). Processo de renegociação de créditos individuais - Análise e Implementação utilizando action research.
- Alves, J. D. F. (2021). Análise de potenciais causas raiz de danos nos rolamentos principais de aerogeradores e análise da manutenção preventiva, corretiva e preditiva aplicável (Master's thesis).
- Data Science Academy. (2022). *Deep Learning Book - Capítulo 10 – As 10 Principais Arquiteturas de Redes Neurais*. Deep Learning Book.
<https://www.deeplearningbook.com.br/as-10-principais-arquiteturas-de-redes-neurais/>
- Deretić, N., Stanimirović, D., Al Awadh, M., Vujanović, N., & Djukić, A. (2022). SARIMA Modelling Approach for Forecasting of Traffic Accidents. *Sustainability (Switzerland)*, 14(8). <https://doi.org/10.3390/su14084403>
- Ferreira, L., Moura, G. L. de, Borenstein, D., & Fischmann, A. A. (2012). THE USE OF ARTIFICIAL NEURAL NETWORKS AS A STRATEGY FOR FORECASTING PRICES IN THE CONTEXT OF AGRIBUSINESS. *Review of Administration and Innovation - RAI*, 8(4).
<https://doi.org/10.5773/rai.v8i4.475>
- Francisco, D., Pinheiro, S., Rui, J., Rocha, D., & Ferreira, P. (n.d.). Avaliação e instalação de uma Unidade de Produção de Autoconsumo (UPAC) com armazenamento de energia.
- Furtado, M. I. V. (2019). *Redes Neurais Artificiais: Uma Abordagem Para Sala de Aula*. In A. C. Oliveira (Ed.), *Atena (Editora Atena)*. <https://doi.org/10.22533/at.ed.262191504>
- Goldemberg, J. (2015). Energia e Sustentabilidade. *Revista de Cultura e Extensão USP*, 14, 33. <https://doi.org/10.11606/issn.2316-9060.v14i0p33-43>

- Gradwohl, C., Dimitrievska, V., Pittino, F., Muehleisen, W., Montvay, A., Langmayr, F., & Kienberger, T. (2021). A combined approach for model-based PV power plant failure detection and diagnostic. *Energies*, 14(5). <https://doi.org/10.3390/en14051261>
- Haney, J., & Burstein, A. (2013). PV System Operations and Maintenance Fundamentals Solar America Board for Codes and Standards. www.solarabcs.org
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. In *Geoscientific Model Development* (Vol. 15, Issue 14, pp. 5481–5487). Copernicus GmbH. <https://doi.org/10.5194/gmd-15-5481-2022>
- Hyndman, R. J. Athanasopoulos, G. (2018). *Previsão: princípios e prática*, 2a edição. OTexts: Melbourne, Austrália. <https://otexts.com/fpp2/>
- Kardakos, E. G., Alexiadis, M. C., Vagropoulos, S. I., Simoglou, C. K., Biskas, P. N., & Bakirtzis, A. G. (2013). Application of Time Series and Artificial Neural Network Models in Short-term Forecasting of PV Power Generation. <https://doi.org/10.1109/UPEC.2013.6714975>
- Kemmis, S. (2009). Action research as a practice-based practice. *Educational Action Research*, 17(3), 463–474. <https://doi.org/10.1080/09650790903093284>
- Kirchner, R. M. (2006). *MODELAGEM E PREVISÃO POR MEIO DE METODOLOGIA BOX & JENKINS: UMA FERRAMENTA DE GESTÃO*.
- Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387–394. <https://doi.org/10.1016/j.ijinfomgt.2014.02.002>
- Latorre, M. do R. D. de O., & Cardoso, M. R. A. (2001). *Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos*.
- Li, C., Chen, Y., & Shang, Y. (2022). A review of industrial big data for decision making in intelligent manufacturing. In *Engineering Science and Technology, an International Journal* (Vol. 29). Elsevier B.V. <https://doi.org/10.1016/j.jestch.2021.06.001>
- Lindig, S., Kaaya, I., Weis, K. A., Moser, D., & Topic, M. (2018). Review of statistical and analytical degradation models for photovoltaic modules and systems as well as related improvements. In *IEEE Journal of Photovoltaics* (Vol. 8, Issue 6, pp. 1773–1786). IEEE Electron Devices Society. <https://doi.org/10.1109/JPHOTOV.2018.2870532>
- Lopes, A. C. B. da silva. (2001). *Modelização Univariada de Séries Temporais: uma Introdução*. <http://gsbwww.uchicago.edu/fac/john.co->
- Macedo, M. A., Melo, M. de S., Lima, S. E. da S., & Silva, T. S. (2021). ENERGIA RENOVÁVEL E SUSTENTABILIDADE: a utilização da energia solar e os seus impactos na gestão das organizações. *Mundo Econômico*, 52–64. <https://doi.org/10.47592/mundec07042021>
- Madeti, S. R., & Singh, S. N. (2017). A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. In *Solar Energy* (Vol. 158, pp. 161–185). Elsevier Ltd. <https://doi.org/10.1016/j.solener.2017.08.069>

- Marcorin, W. R., Roberto, C., & Lima, C. (2004). Análise dos Custos de Manutenção e de Não-manutenção de Equipamentos Produtivos Cost Analysis of Maintenance and Non-Maintenance Policies for Productive Equipments.
- MARTIN, A. C., HENNING, E., WALTER, O. M. F. C., & KONRATH, A. C. (2016, November 22). Análise de séries temporais para previsão da evolução do número de automóveis no Município de Joinville. *Revista ESPACIOS*.
<https://www.revistaespacios.com/a16v37n06/16370629.html>
- Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., & Lughì, V. (2020). Advanced Methods for Photovoltaic Output Power Forecasting: A Review. *Applied Sciences*, 10(2), 487.
<https://doi.org/10.3390/app10020487>
- Mellit, A., Tina, G. M., & Kalogirou, S. A. (2018). Fault detection and diagnosis methods for photovoltaic systems: A review. In *Renewable and Sustainable Energy Reviews* (Vol. 91, pp. 1–17). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2018.03.062>
- Miguel, D., Batista, S., & Brito, M. C. (2023). Modelação de sistemas fotovoltaicos.
- Mitreá, C. A., Lee, C. K. M., & Wu, Z. (2009). A Comparison between Neural Networks and Traditional Forecasting Methods: A Case Study. In *International Journal of Engineering Business Management* (Vol. 1, Issue 2).
- Mohammed, E. A., Naugler, C., & Far, B. H. (2015). Emerging Business Intelligence Framework for a Clinical Laboratory Through Big Data Analytics. In *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology: Algorithms and Software Tools* (pp. 577–602). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-802508-6.00032-6>
- Moreira, F. M. B. (2016). Sistemas Fotovoltaicos. <http://hdl.handle.net/10400.26/18617>
- Moreno, F. J. T. dos S. (2021). Estudo comparativo de métodos de análise e previsão de séries temporais: métodos estatísticos versus redes neuronais.
<https://doi.org/http://hdl.handle.net/10174/29054>
- Musbah, H., & El-Hawary, M. (n.d.). SARIMA Model Forecasting of Short-Term Electrical Load Data Augmented by Fast Fourier Transform Seasonality Detection.
- Narasimha Murthy, K. V., Saravana, R., & Vijaya Kumar, K. (2018). Modeling and forecasting rainfall patterns of southwest monsoons in North–East India as a SARIMA process. *Meteorology and Atmospheric Physics*, 130(1), 99–106.
<https://doi.org/10.1007/s00703-017-0504-2>
- Nogueira, S. M. C., & Moreira, M. A. (2015). Avaliação das previsões de temperatura do modelo ETA para o Estado do Paraná.
- Peinado Gonzalo, A., Pliego Marugán, A., & García Márquez, F. P. (2020a). Survey of maintenance management for photovoltaic power systems. In *Renewable and Sustainable Energy Reviews* (Vol. 134). Elsevier Ltd.
<https://doi.org/10.1016/j.rser.2020.110347>
- Peinado Gonzalo, A., Pliego Marugán, A., & García Márquez, F. P. (2020b). Survey of maintenance management for photovoltaic power systems. In *Renewable and*

- Sustainable Energy Reviews (Vol. 134). Elsevier Ltd.
<https://doi.org/10.1016/j.rser.2020.110347>
- pmdarima. (2017). Função auto_arima ndiffs de R. <https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.ndiffs.html#pmdarima.arima.ndiffs>
- Python Software Foundation. (2022). Python.org. Python. <https://www.python.org/>
- Rabbani, M. B. A., Musarat, M. A., Alaloul, W. S., Rabbani, M. S., Maqsoom, A., Ayub, S., Bukhari, H., & Altaf, M. (2021). A Comparison Between Seasonal Autoregressive Integrated Moving Average (SARIMA) and Exponential Smoothing (ES) Based on Time Series Model for Forecasting Road Accidents. *Arabian Journal for Science and Engineering*, 46(11), 11113–11138. <https://doi.org/10.1007/s13369-021-05650-3>
- Ramalho, D., Oliveira, G., & Miranda, P. (2022). TIPOS DE ENERGIAS RENOVÁVEIS E NÃO RENOVÁVEIS.
- Ramsami, P., & Oree, V. (2015). A hybrid method for forecasting the energy output of photovoltaic systems. *Energy Conversion and Management*, 95, 406–413. <https://doi.org/10.1016/j.enconman.2015.02.052>
- Rossi, R., Murari, A., Gaudio, P., & Gelfusa, M. (2020). Upgrading model selection criteria with goodness of fit tests for practical applications. *Entropy*, 22(4). <https://doi.org/10.3390/E22040447>
- Runge, J., & Zmeureanu, R. (2019). Forecasting energy use in buildings using artificial neural networks: A review. In *Energies* (Vol. 12, Issue 17). MDPI AG. <https://doi.org/10.3390/en12173254>
- Secreto, H. T. M. (2020). Criação de Modelos de Aprendizado de Máquina para Previsão de Eficiência de Automóveis.
- Sharkawy, A.-N. (2020). Principle of Neural Network and Its Main Types: Review. In *Journal of Advances in Applied & Computational Mathematics* (Vol. 7).
- Thomaz, A. :, & Favarini, O. (2018). Análise de Séries Temporais: Comparação entre modelos preditivos em estudo de caso.
- Trojan, F., & Marçal, R. F. M. (2017). Proposal of Maintenance-types Classification to Clarify Maintenance Concepts in Production and Operations Management. *Journal of Business and Economics*, 8(7), 560–572. [https://doi.org/10.15341/jbe\(2155-7950\)/07.08.2017/005](https://doi.org/10.15341/jbe(2155-7950)/07.08.2017/005)
- Vagropoulos, S., Chouliaras, G., Kardakos, E., Simoglou, C., & Bakirtzis, A. (2016). Comparison of SARIMAX, SARIMA, Modified SARIMA and ANN-based Models for Short-Term PV Generation Forecasting. *IEEE*.
- Ventura, A. A. V. (2012). Aplicação do método Action Research ao processo de apoio à decisão com o uso do sistema VIP Analysis.
- Vieira da Silva, M., & Carolina Nicolosi da Rocha Gracioso, A. (2021). MODELO DE PREVISÃO DE DEMANDA PARA ATMS UTILIZANDO REDE NEURAL ARTIFICIAL DO TIPO

MULTILAYER PERCEPTRON. COLLOQUIUM EXACTARUM, 12(4), 54–62.

<https://doi.org/10.5747/ce.2020.v12.n4.e339>

Wang, J., Zhang, W., Shi, Y., & Duan, S. (n.d.). SUBMITTED TO IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING 1 Industrial Big Data Analytics: Challenges, Methodologies, and Applications.

Yang, Z., & Ge, Z. (2022). On Paradigm of Industrial Big Data Analytics: From Evolution to Revolution. IEEE Transactions on Industrial Informatics, 18(12), 8373–8388.

<https://doi.org/10.1109/TII.2022.3190394>

ANEXO A – FUNÇÕES ANÁLISE NULOS

Função `contar_nulos`, apresentada na Figura 108 e Figura 109 tem como objetivo analisar e contabilizar os valores nulos de cada variável, em cada ano, e posteriormente a soma dos dois anos. E para além disso, emitir alerta de falha de comunicação se a condição for verificada. A função admite que:

- Input: `df` (conjunto de dados) ; `doc`(documento word para guardar o gráfico, é opcional) ; `consecutivo_minimo` (valor mínimo de valores nulos consecutivos)

```
def contar_nulos(df, doc, consecutivo_minimo):
    # Criar um DataFrame auxiliar com anos separados
    df_por_ano = df.copy()
    df_por_ano['Ano'] = df['Timestamp'].dt.year # Substitua 'Data' pelo nome da sua coluna de data
    df_por_ano = df_por_ano.drop(columns=['Timestamp']) # Remova a coluna 'Data' se desejar

    # Contar valores nulos em cada coluna por ano
    null_counts_por_ano = df_por_ano.groupby('Ano').apply(lambda group: group.isnull().sum())

    # Adicionar um título
    doc.add_heading('Valores Nulos por Ano', level=1)

    # Criar uma tabela
    table = doc.add_table(rows=1, cols=4)
    table.allow_autofit = False

    # Definir o cabeçalho da tabela
    header_cells = table.rows[0].cells
    header_cells[0].text = 'Ano'
    header_cells[1].text = 'Variável'
    header_cells[2].text = 'Quantidade de Nulos'

    # Adicionar as variáveis e a quantidade de valores nulos na tabela por ano
    for ano, null_counts_ano in null_counts_por_ano.iterrows():
        for variable, null_count in null_counts_ano.items(): # Substitua iteritems() por items()
            row_cells = table.add_row().cells
            row_cells[0].text = str(ano)
            row_cells[1].text = variable
            row_cells[2].text = str(null_count)
```

Figura 108 : Função `contar_nulos` em *Python*

```

# Ajustar a largura das colunas
table.columns[0].width = Inches(1.0)
table.columns[1].width = Inches(1.5)
table.columns[2].width = Inches(1.5)
table.columns[3].width = Inches(1.5)
plt.close()

# Contar valores nulos em cada coluna no df no total dos dois anos
null_counts = df.isnull().astype(int)

# Contar vezes de nulos consecutivos maiores que consecutivo_minimo
nulos_consecutivos = null_counts.rolling(window=consecutivo_minimo).sum()
excedimentos = (nulos_consecutivos >= consecutivo_minimo).astype(int)

# Verificar se há algum excedimento em qualquer variável
excedimentos_total = excedimentos.sum()

# Adicionar um título
doc.add_heading('Valores Nulos ', level=1)

# Criar uma tabela
table = doc.add_table(rows=1, cols=4)
table.allow_autofit = False

# Definir o cabeçalho da tabela
header_cells = table.rows[0].cells
header_cells[0].text = 'Variável'
header_cells[1].text = 'Quantidade de Nulos'
header_cells[2].text = 'Quantidade de Excedimentos'

# Adicionar as variáveis e a quantidade de valores nulos na tabela
for variable in df.columns:
    row_cells = table.add_row().cells
    row_cells[0].text = variable
    row_cells[1].text = str(null_counts[variable].sum())
    row_cells[2].text = str(excedimentos_total[variable])

# Emitir o alerta se houver algum excedimento em qualquer variável
if any(excedimentos_total > 0):
    root = tk.Tk()
    root.withdraw()
    messagebox.showwarning("Alerta",
        f"FALHA DE COMUNICAÇÃO POR EXISTIREM EXCEDIMENTOS DO LIMITE DE NULOS CONSECUTIVOS!")
    root.destroy()
else:
    root = tk.Tk()
    root.withdraw()
    messagebox.showinfo("Notificação", "NÃO EXISTEM FALHAS DE COMUNICAÇÃO")
    root.destroy()

# Ajustar a largura das colunas
table.columns[0].width = Inches(1.5)
table.columns[1].width = Inches(1.5)
table.columns[2].width = Inches(1.5)
plt.close()

```

Figura 109 : Continuação da função contar_nulos em Python

Função `get_daytime`, apresentada na Figura 110 tem como objetivo estabelecer o período do dia, e admite que:

- Input: `timestamp` (variável temporal)

```
def get_daytime(timestamp):  
    hour = timestamp.hour  
    if 6 <= hour < 20:  
        return 'Dia'  
    else:  
        return 'Noite'
```

Figura 110 : Função `get_daytime` em *Python*

Função `substituir_nulos`, apresentada na Figura 111 tem como objetivo substituir os valores nulos de uma determinada variável pela média mensal do ano em análise e de acordo com o período do dia correspondente, e admite que:

- Input: `df` (conjunto de dados) ; variável (variável a estudar) ; ano (ano de análise)

```
def substituir_nulos(df, variavel, ano):
    df['Daytime'] = df['Timestamp'].apply(get_daytime)

    # Calcular a média agrupada por ano e mês (considerando apenas valores não nulos)
    df_ano = df[df['Timestamp'].dt.year == ano]
    df_ano_sem_nulos = df_ano.dropna(subset=[variavel])

    df_media_mes_diurna = \
        df_ano_sem_nulos[df_ano_sem_nulos['Daytime'] == 'Dia'].groupby(df_ano_sem_nulos['Timestamp'].dt.month)[
            variavel].mean()
    df_media_mes_noturna = \
        df_ano_sem_nulos[df_ano_sem_nulos['Daytime'] == 'Noite'].groupby(df_ano_sem_nulos['Timestamp'].dt.month)[
            variavel].mean()

    df_media_mes_diurna_arredondados = round(df_media_mes_diurna, 2)
    df_media_mes_noturna_arredondados = round(df_media_mes_noturna, 2)

    # Substituir valores nulos da variável pela média correspondente (diurna ou noturna)
    for mes, media_diurna in df_media_mes_diurna_arredondados.items():
        media_noturna = df_media_mes_noturna_arredondados[mes]

        mask_diurna = (df['Timestamp'].dt.year == ano) & (df['Timestamp'].dt.month == mes) & (df['Daytime'] == 'Dia')
        mask_noturna = (df['Timestamp'].dt.year == ano) & (df['Timestamp'].dt.month == mes) & (df['Daytime'] == 'Noite')

        df.loc[mask_diurna & df[variavel].isnull(), variavel] = media_diurna
        df.loc[mask_noturna & df[variavel].isnull(), variavel] = media_noturna

    return df
```

Figura 111 : Função `substituir_nulos` em *Python*

ANEXO B – FUNÇÕES ANÁLISE EXPLORATÓRIA

Função `criar_grafico_linhas`, apresentada na Figura 112 e Figura 113 tem como objetivo construir o gráfico de linhas de uma ou duas variáveis, tendo em conta os valores médios mensais dessa variável no ano especificado como dado de entrada. Esta função admite ainda o seguinte:

- Input: `df` (conjunto de dados) ; `variavel1`(variável 1 a ser representada) ; `variavel2` (variável 2 a ser representada no mesmo gráfico, é opcional) ; `xlabel` (legenda do eixo x do gráfico, é opcional) ; `ylabel` (legenda do eixo y do gráfico, é opcional) ; `title`(título do gráfico, é opcional) e `doc`(documento word para guardar o gráfico, é opcional)

```
def criar_grafico_linhas(df, variavel1, variavel2=None, xlabel=None, ylabel=None, title=None, doc=None, ano=None, cor1='blue', cor2='red'):
    if ano is not None:
        # Filtrar os dados para o ano especifico
        df_ano = df[df['Timestamp'].dt.year == ano]
        df_media1 = df_ano.groupby(df_ano['Timestamp'].dt.month)[variavel1].mean()

    if variavel2 is not None:
        # Calcular a média mensal da variável 2
        df_media2 = df_ano.groupby(df_ano['Timestamp'].dt.month)[variavel2].mean()

    # Criar uma nova figura para cada gráfico
    fig, ax1 = plt.subplots()

    # Obter os meses como símbolos
    meses = df_media1.index

    # Definir os nomes dos meses
    nomes_meses = [calendar.month_name[mes] for mes in meses]

    # Definir as preferências das barras
    positions = range(len(meses))

    # Criar a primeira variável
    ax1.plot(positions, df_media1, label='Variável 1', color=cor1)
    ax1.set_xlabel(xlabel)
    ax1.set_ylabel(ylabel)
    ax1.set_title(title)
    ax1.set_xticks(positions)
    ax1.set_xticklabels(nomes_meses)

    # Criar a segunda variável com escala secundária
    ax2 = ax1.twinx()
    ax2.plot(positions, df_media2, label='Variável 2', color=cor2)
    ax2.set_ylabel('Variável 2')

    # Adicionar as legendas
    lines, labels = ax1.get_legend_handles_labels()
    lines2, labels2 = ax2.get_legend_handles_labels()
    ax2.legend(lines + lines2, labels + labels2, loc='upper right')

    plt.xticks(rotation=45)
    plt.tight_layout()

    # Salvar o gráfico como imagem BytesIO
    img_buffer = BytesIO()
    plt.savefig(img_buffer, format='png')
    img_buffer.seek(0)

    # Adicionar a imagem do gráfico ao documento Word
    if doc is not None:
        doc.add_picture(img_buffer, width=Inches(6), height=Inches(4))
    else:
        print("Documento Word não fornecido.")
    plt.close()
```

Figura 112 : Função `criar_grafico_linhas` em *Python*

```
else:
    # Criar uma nova figura para cada gráfico
    plt.figure()

    # Obter os meses como símbolos
    meses = df_medial.index

    # Definir os nomes dos meses
    nomes_meses = [calendar.month_name[mes] for mes in meses]

    # Definir as preferências das barras
    positions = range(len(meses))

    # Criar a primeira variável
    plt.plot(positions, df_medial, label='Variável 1', color=cor1)

    # Definir os rótulos do gráfico, se fornecidos
    if xlabel is not None:
        plt.xlabel(xlabel)
    if ylabel is not None:
        plt.ylabel(ylabel)
    if title is not None:
        plt.title(title)

    # Girar os rótulos dos meses para melhor visualização
    plt.xticks(rotation=45)

    # Definir os rótulos do eixo x como os meses
    plt.gca().set_xticks(positions)
    plt.gca().set_xticklabels(nomes_meses)

    # Salvar o gráfico como imagem BytesIO
    img_buffer = BytesIO()
    plt.savefig(img_buffer, format='png')
    img_buffer.seek(0)

    # Adicionar a imagem do gráfico ao documento Word
    if doc is not None:
        doc.add_picture(img_buffer, width=Inches(6), height=Inches(4))
    else:
        print("Documento Word não fornecido.")
    plt.close()
```

Figura 113 : Continuação da função `criar_grafico_linhas` em *Python*

Função `criar_matriz_correlacao`, apresentada na Figura 114 tem como objetivo desenvolver e apresentar a matriz correlação de um determinado conjunto de dados. Admite o seguinte:

- Input: `df` (conjunto de dados) ; `colunas_selecionadas` (conjunto de variáveis a ser inserida nesta matriz) e `doc`(documento word para guardar a matriz, é opcional).

```
def criar_matriz_correlacao(df, colunas_selecionadas, doc=None):
    #novo dataframe apenas com as colunas necessárias para a matriz correlação
    dados_selecionados = df[colunas_selecionadas]

    # Calcular a matriz de correlação
    matriz_correlacao = dados_selecionados.corr().round(2)
    plt.figure(figsize=(20, 20))

    # Adicionar os nomes das variáveis na primeira coluna: coluna vazia, em branco na primeira coluna da matriz de correlação
    matriz_correlacao.insert(0, ' ', colunas_selecionadas)

    # Converter a matriz de correlação numa tabela com a biblioteca pandas
    tabela = pd.DataFrame(matriz_correlacao)

    if doc is not None:

        # Estilo e tamanho da tabela
        estilo_tabela = doc.styles['Table Grid']
        estilo_tabela.font.size = Pt(9)

        # Adicionar a tabela em branco ao documento Word e as suas dimensões
        doc.add_table(tabela.shape[0] + 1, tabela.shape[1], style=estilo_tabela)

        # Preencher os cabeçalhos das colunas com a função enumerate assim temos o indice i e o nome da coluna da tabela (matriz correlação)
        #com o ".text" atribuímos o nome à coluna vazia que está no word na posição ( 1, i)
        for i, coluna in enumerate(tabela.columns):
            doc.tables[0].cell(0, i).text = coluna

        # Preencher os valores da tabela
        for i in range(tabela.shape[0]):
            for j in range(tabela.shape[1]):
                doc.tables[0].cell(i + 1, j).text = str(tabela.iloc[i, j])

        for row in doc.tables[0].rows:
            for cell in row.cells:
                cell.paragraphs[0].runs[0].font.size = Pt(5) # Definir tamanho da letra

    if doc is not None:
        # Salvar o gráfico de tabela como uma imagem
        plt.savefig('matriz_correlacao.png')

        # Adicionar a imagem da matriz de correlação ao documento Word
        doc.add_picture('matriz_correlacao.png', width=Inches(6), height=Inches(4))
        plt.close()
```

Figura 114 : Função `criar_matriz_correlacao` em *Python*

Função `criar_grafico_barras` apresentadas na Figura 115 e Figura 116 tem como principal objetivo criar o gráfico de barras da média mensal da variável especificada para cada mês. Esta função admite que:

- Input: `df` (conjunto de dados) ; variável (variável a ser representada) ; ano (referir o ano de análise) ; cor (especificar a cor das barras de acordo com a variável) ; `xlabel`, `ylabel`, `title`(legendas do eixo x, eixo y e título do gráfico e são opcionais) ; `doc`(documento word para guardar a matriz, é opcional).

```
def criar_grafico_barras(df, variavel, ano, cor='blue', xlabel=None, ylabel=None, title=None, doc=None):
    # Filtrar os dados para o ano especifico
    df_ano = df[df['Timestamp'].dt.year == ano]

    # Calcular a média mensal da variável
    df_media = df_ano.groupby(df_ano['Timestamp'].dt.month)[variavel].mean()
    df_media_arredondados = round(df_media, 2)

    # Calcular a quantidade de dados nulos por mês
    df_nulos = df_ano.groupby(df_ano['Timestamp'].dt.month)[variavel].apply(lambda x: x.isnull().sum())

    # figura para gráfico de barras
    plt.figure()

    # Obter os meses como símbolos
    meses = df_media.index

    # Definir os nomes dos meses
    nomes_meses = [calendar.month_name[mes] for mes in meses]

    # Definir as preferências das barras
    positions = range(len(meses))

    # Plotar as barras da média mensal
    plt.bar(positions, df_media, color=cor)

    # Definir os rótulos do gráfico, fornecidos
    if xlabel is not None:
        plt.xlabel(xlabel)
```

Figura 115 : Função `criar_grafico_barras` em *Python*

```
if ylabel is not None:
    plt.ylabel(ylabel)
if title is not None:
    plt.title(title)

# Girar os rótulos dos meses para melhor visualização
plt.xticks(rotation=45)

# Definir os rótulos do eixo x como os meses
plt.gca().set_xticks(positions)
plt.gca().set_xticklabels(nomes_meses)

# Salvar o gráfico como imagem
img_buffer = BytesIO()
plt.savefig(img_buffer, format='png')
img_buffer.seek(0)
# Adicionar a imagem do gráfico ao documento Word

if doc is not None:
    doc.add_picture(img_buffer, width=Inches(6), height=Inches(4))
    # Criar uma tabela com os valores médios para cada mês
    table_data = pd.DataFrame({'Mês': nomes_meses, 'Valor Médio': df_media_arredondados.values, 'Nulos': df_nulos})
    table = doc.add_table(rows=1, cols=3)
    table.cell(0, 0).text = 'Mês'
    table.cell(0, 1).text = 'Valor Médio'
    table.cell(0, 2).text = 'Nulos'

    for i, row in table_data.iterrows():
        # Adiciona uma nova linha à tabela
        row_cells = table.add_row().cells
        # Define o texto das células da nova linha
        row_cells[0].text = str(row['Mês'])
        row_cells[1].text = str(row['Valor Médio'])
        row_cells[2].text = str(row['Nulos'])

else:
    print("Documento Word não fornecido.")

plt.close()
```

Figura 116 : Continuação da função `criar_grafico_barras` em `Python`

Função `criar_grafico_barras_eficiencia_word` apresenta o gráfico de barras da média mensal da variável especificada para cada mês, neste caso, será a eficiência dos dois inversores, Figura 117. Esta função admite que:

- Input: `df` (conjunto de dados) ; `variável` (variável a ser representada) ; `ano` (referir o ano de análise) ; `cor` (especificar a cor das barras de acordo com a variável) ; `xlabel`, `ylabel`, `title`(legendas do eixo x, eixo y e título do gráfico e são opcionais) ; `doc`(documento word para guardar a matriz, é opcional).

```

def criar_grafico_barras_eficiencia_word(df, variavel, ano, cor='blue', xlabel=None, ylabel=None, title=None, doc=None):
    # Filtrar os dados para o ano especifico
    df_ano = df[df['Timestamp'].dt.year == ano]

    # Filtrar os valores não nulos e diferentes de zero da variável de eficiência
    df_nao_nulos = df_ano[df_ano[variavel].notnull() & (df_ano[variavel] != 0)]

    # Calcular a média mensal da variável (apenas valores não nulos e diferentes de zero)
    df_media = df_nao_nulos.groupby(df_nao_nulos['Timestamp'].dt.month)[variavel].mean()
    df_media_arredondados = round(df_media, 2)

    # figura para gráfico de barras
    plt.figure()

    # Obter os meses como símbolos
    meses = df_media.index

    # Definir os nomes dos meses
    nomes_meses = [calendar.month_name[mes] for mes in meses]

    # Definir as preferências das barras
    positions = range(len(meses))

    # barras da média mensal
    plt.bar(positions, df_media, color=cor)

    # Definir títulos e rótulos
    if title:
        plt.title(title)
    if xlabel:
        plt.xlabel(xlabel)
    if ylabel:
        plt.ylabel(ylabel)

    # Girar os rótulos dos meses para melhor visualização
    plt.xticks(rotation=45)
    plt.gca().set_xticks(positions)
    plt.gca().set_xticklabels(nomes_meses)

    # Criar um objeto BytesIO para salvar o gráfico
    img_buffer = BytesIO()
    plt.savefig(img_buffer, format='png')
    img_buffer.seek(0)

    plt.close()

    if doc is not None:
        doc.add_picture(img_buffer, width=Inches(6), height=Inches(4))
        # Criar uma tabela com os valores médios para cada mês
        table_data = pd.DataFrame({'Mês': nomes_meses, 'Valor Médio': df_media_arredondados.values})
        table = doc.add_table(rows=1, cols=3)
        table.cell(0, 0).text = 'Mês'
        table.cell(0, 1).text = 'Valor Médio'

        for i, row in table_data.iterrows():
            # Adiciona uma nova linha à tabela
            row_cells = table.add_row().cells
            # Define o texto das células da nova linha
            row_cells[0].text = str(row['Mês'])
            row_cells[1].text = str(row['Valor Médio'])

    else:
        print("Documento Word não fornecido.")

    plt.close()

```

Figura 117 : Função criar_grafico_barras_eficiencia_word em Python

Função `media_mensal`, Figura 118, é capaz de calcular a média mensal de uma variável em específico. Admite que:

- Input: `df` (conjunto de dados) ; `variavel` (variável a ser estudada) ; `title`(título da tabela, é opcional) ; `doc`(documento word para guardar a matriz, é opcional).

```
def media_mensal (df, variavel , title= None ,doc=None):
    df_media = df.groupby(df['Timestamp'].dt.month)[variavel].mean()
    df_media_arredondados = round(df_media,2)

    # Obter os meses como rótulos
    meses = df_media.index

    # Definir os nomes dos meses
    nomes_meses = [calendar.month_name[mes] for mes in meses]

    if doc is not None:

        # Criar uma tabela com os valores médios para cada mês
        table_data = pd.DataFrame({'Mês': nomes_meses, 'Valor Médio': df_media_arredondados.values})
        table = doc.add_table(rows=len(table_data) + 2, cols=2)

        # Definir o título da tabela
        if title is not None:
            table_title = table.cell(0, 0).merge(table.cell(0, 1))
            table_title.text = title

        table.cell(1, 0).text = 'Mês'
        table.cell(1, 1).text = 'Valor Médio'

        for i, row in table_data.iterrows():
            table.cell(i + 2, 0).text = str(row['Mês'])
            table.cell(i + 2, 1).text = str(row['Valor Médio'])

    else:
        print("Documento Word não fornecido.")
```

Figura 118 : Função `media_mensal` em *Python*

ANEXO C – FUNÇÕES PRÉ-PROCESSAMENTO

De seguida, são apresentadas as linhas de código do Ficheiro `pre-processamento`, Figura 119 e Figura 120, que tem como principal função normalizar e preparar o conjunto de dados para previsão.

```
#import analise_nulos
from sklearn.preprocessing import MinMaxScaler
import joblib
import pandas as pd
import analise_nulos

# carregar dataframe
df = analise_nulos.df
#df = pd.read_excel('C:/Users/helen/Desktop/dados_total/dados_totais_limpos.xlsx')

# Excluir a coluna "daytime"
df = df.drop("Daytime", axis=1)

variaveis = df.columns[1:-2]

df_media = pd.DataFrame(columns=['Ano', 'Mes'] + variaveis.tolist() + ['Eficiencia_INV_A_Media', 'Eficiencia_INV_B_Media'])

anos = df['Timestamp'].dt.year.unique()
meses = df['Timestamp'].dt.month.unique()

for ano in anos:
    for mes in meses:
        df_filtrado = df[(df['Timestamp'].dt.year == ano) & (df['Timestamp'].dt.month == mes)]
        valores_medios = df_filtrado[variaveis].mean().values

        # Calcular a eficiência média para o mês atual, excluindo os zeros
        eficiencia_inv_a_mes = df_filtrado['Eficiencia_INV_A']
        eficiencia_inv_b_mes = df_filtrado['Eficiencia_INV_B']

        eficiencia_inv_a_mes_sem_zeros = eficiencia_inv_a_mes[eficiencia_inv_a_mes != 0]
        eficiencia_inv_b_mes_sem_zeros = eficiencia_inv_b_mes[eficiencia_inv_b_mes != 0]

        media_eficiencia_mes_inv_a = eficiencia_inv_a_mes_sem_zeros.mean() if not eficiencia_inv_a_mes_sem_zeros.empty else 0
        media_eficiencia_mes_inv_b = eficiencia_inv_b_mes_sem_zeros.mean() if not eficiencia_inv_b_mes_sem_zeros.empty else 0

        nova_linha = pd.DataFrame(
            [[ano, mes] + list(valores_medios) + [media_eficiencia_mes_inv_a, media_eficiencia_mes_inv_b]],
            columns=df_media.columns)
        df_media = df_media._append(nova_linha, ignore_index=True)

print(df_media)

'''
for ano in anos:
    for mes in meses:
        df_filtrado = df[(df['Timestamp'].dt.year == ano) & (df['Timestamp'].dt.month == mes)]
        valores_medios = df_filtrado[variaveis].mean().values
        nova_linha = pd.DataFrame([[ano, mes] + list(valores_medios)], columns=df_media.columns)
        df_media = df_media._append(nova_linha, ignore_index=True)

print(df_media)
'''
df_media.to_excel('C:/Users/helen/Desktop/dados_total/dataset_previsao.xlsx', index=False)

'''
GERAR DADOS QUE NÃO FOI UTILIZADO
Chamar a função gerar_dados e gerar dados até 2010 tendo por base o ano de 2022
df_estimado = gerar_dados(df_media)
print(df_estimado)
df_estimado.to_excel('C:/Users/helen/Desktop/dados_total/teste.xlsx', index=False)

Juntar o df_media (2021-2022) e o df_estimado(2010-2020) = df_media (2010-2022)
df_media = pd.concat([df_estimado, df_media], ignore_index=True) #
ordenar o df
df_media= df_media.sort_values(['Ano', 'Mes']).reset_index(drop=True)
df_media.to_excel('C:/Users/helen/Desktop/dados_total/dataset_2010_2022.xlsx')
'''
```

Figura 119 : Linhas de código do Ficheiro `pre-processamento`

```

#NORMALIZAÇÃO DOS DADOS
scaler = MinMaxScaler()

)df_media.loc[:, 'Ano'] = scaler.fit_transform(df_media[['Ano']])
)df_media.loc[:, 'Mes'] = scaler.fit_transform(df_media[['Mes']])
df_media.loc[:, 'E_TOTAL_INVERSOR_A [kWh]'] = scaler.fit_transform(df_media[['E_TOTAL_INVERSOR_A [kWh]']])
df_media.loc[:, 'P_AC_INV_A [W]'] = scaler.fit_transform(df_media[['P_AC_INV_A [W]']])
df_media.loc[:, 'P_DC_INV_A [W]'] = scaler.fit_transform(df_media[['P_DC_INV_A [W]']])
df_media.loc[:, 'T_INV_A [°C]'] = scaler.fit_transform(df_media[['T_INV_A [°C]']])
df_media.loc[:, 'E_TOTAL_INVERSOR_B [kWh]'] = scaler.fit_transform(df_media[['E_TOTAL_INVERSOR_B [kWh]']])
df_media.loc[:, 'P_AC_INV_B [W]'] = scaler.fit_transform(df_media[['P_AC_INV_B [W]']])
df_media.loc[:, 'P_DC_INV_B [W]'] = scaler.fit_transform(df_media[['P_DC_INV_B [W]']])
df_media.loc[:, 'T_INV_B [°C]'] = scaler.fit_transform(df_media[['T_INV_B [°C]']])
df_media.loc[:, 'P_AC_SAIDA_REDE [W]'] = scaler.fit_transform(df_media[['P_AC_SAIDA_REDE [W]']])
df_media.loc[:, 'IRR_MEDIA [Wh/m²]'] = scaler.fit_transform(df_media[['IRR_MEDIA [Wh/m²]']])
df_media.loc[:, 'IRE [Wh/m²]'] = scaler.fit_transform(df_media[['IRE [Wh/m²]']])
df_media.loc[:, 'IRO [Wh/m²]'] = scaler.fit_transform(df_media[['IRO [Wh/m²]']])
df_media.loc[:, 'T_MEDIA [°C]'] = scaler.fit_transform(df_media[['T_MEDIA [°C]']])
df_media.loc[:, 'T_ESTE [°C]'] = scaler.fit_transform(df_media[['T_ESTE [°C]']])
df_media.loc[:, 'T_OESTE [°C]'] = scaler.fit_transform(df_media[['T_OESTE [°C]']])
)df_media.loc[:, 'Eficiencia_INV_A_Media'] = scaler.fit_transform(df_media[['Eficiencia_INV_A_Media']])
)df_media.loc[:, 'Eficiencia_INV_B_Media'] = scaler.fit_transform(df_media[['Eficiencia_INV_B_Media']])

) # Salvar o scaler para mais à frente se necessário
scaler_filename = 'scaler.save'
joblib.dump(scaler, scaler_filename)

# Salvar os dados normalizados
df_media.to_excel('C:/Users/helen/Desktop/dados_total/dataset_previsao_normalizado.xlsx', index=False)

# OUTPUT: df_media

```

Figura 120 : Continuação das linhas de código do Ficheiro pre-processamento

ANEXO D – FUNÇÕES REDES NEURONAIIS

De seguida, são apresentadas as linhas de código de implementação da rede neuronal MLP para a variável de temperatura do inversor A em graus celsius, Figura 121 e Figura 122.

```

#BIBLIOTECAS IMPORTANTES
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error, mean_squared_error
from keras import Sequential
from keras.layers import Dense
from tensorflow import keras
import tensorflow
import pandas as pd
from docx import Document
from funcoes_analise_exploratoria import criar_grafico_barras_previsao
from funcoes_analise_exploratoria import criar_grafico_linhas
from keras.callbacks import EarlyStopping
import matplotlib.pyplot as plt
import numpy as np
import pre_processamento

print('-----DATASET-----')

# carregar dataframe
df = pre_processamento.df_media
#df = pd.read_excel('C:/Users/helen/Desktop/dados_total/dataset_previsao_normalizado.xlsx')

print('-----VARIÁVEL A PREVER-----')
variavel_x = 'T_INV_A [°C]'

print('-----SELEÇÃO COLUNAS- CORRELAÇÃO-----')
#definir limite correlação
limite_correlacao = 0.5
# Calcula a matriz de correlação
matriz_correlacao = df.corr()

# Seleciona as variáveis com correlação em relação à variável X
colunas_selecionadas = matriz_correlacao.index[abs(matriz_correlacao[variavel_x]) >= limite_correlacao]

# Filtra o dataset original com as colunas selecionadas
df = df[colunas_selecionadas]

# Verifica as colunas selecionadas
print(df.columns)

dfmL = df.copy()

print('-----PREPARAR DADOS DE ENTRADA E SAÍDA-----')
#dados de saida
y = dfmL[variavel_x]
#dados de entrada
X = dfmL.drop([variavel_x], axis=1) # TIRAR TODAS AS VARIÁVEIS QUE QUERO PREVER NOS OUTROS MODELOS PARA O LSTM
print(y)

print('-----SEMENTE DE INICIAÇÃO-----')
tensorflow.random.set_seed(2)

print('-----CONJ. TREINO E TESTE-----')
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, shuffle=False) #, random_state=40)
print(X_train)
print(X_test)
print(y_train)
print(y_test)

```

Figura 121 : Linhas de código de implementação da rede neuronal MLP para a variável de temperatura do inversor A em graus celsius

```

print('-----DEF. MODELO-----')
model = Sequential()
# camadas ocultas
model.add(Dense(40, activation='relu', kernel_initializer='random_normal', input_dim=len(X.columns)))
model.add(Dense(38, activation='relu', kernel_initializer='random_normal'))
model.add(Dense(36, activation='relu', kernel_initializer='random_normal'))
#model.add(Dense(34, activation='relu', kernel_initializer='random_normal'))

model.add(Dense(1, kernel_initializer='random_normal'))

#early_stopping = EarlyStopping(monitor='mean_absolute_error', patience=80, verbose=1)
#early_stopping = EarlyStopping(monitor='mean_absolute_error', min_delta=0.09, patience=100, verbose=1)

print('-----COMPILAR O MODELO-----')
model.compile(optimizer='adam', loss='mean_squared_error', metrics=['mean_absolute_error'])

print('-----TREINAR O MODELO-----')
model.fit(X_train, y_train, batch_size=128, epochs=80, verbose=False)
#batch_size = 32 e 64
# epochs = 50

print('-----FAZER PREVISÕES-----')

# Faz previsões para o conjunto de teste
previsoes = model.predict(X_test)
print(previsoes)

print('-----AVALIAÇÃO DO MODELO-----')
mae = mean_absolute_error(y_test, previsoes)
mse = mean_squared_error(y_test, previsoes)
print("Erro Absoluto Médio (MAE):", mae)
print("Erro Quadrático Médio (MSE):", mse)

# valores reais do conjunto de teste e as previsões
valores_totais = df[variavel_x].values
valores_reais = y_test.values
valores_treino = y_train.values
valores_previstos = previsoes.flatten()

plt.figure(figsize=(12, 6))
# gráfico os valores de treinamento em verde
plt.plot(y_train.index, valores_treino, color='green', label='Conjunto de treino', marker='o', linestyle='-')
# gráfico os valores de teste em azul
plt.plot(y_test.index, valores_reais, color='blue', label='Conjunto de teste', marker='o', linestyle='-')
# gráfico as previsões em laranja
plt.plot(y_test.index, valores_previstos, color='orange', label='Previsões', marker='o', linestyle='-')
# Adicionar uma linha de ligação entre o último ponto do treinamento e o primeiro ponto do teste
plt.plot([y_train.index[-1], y_test.index[0]], [valores_treino[-1], valores_reais[0]], color='green', linestyle='-')
plt.ylim(-2, 2)
plt.title('Previsões Futuras do Modelo SARIMA')
plt.ylabel(f'Valores de {variavel_x}')
plt.xlabel('Índice')
plt.legend()
plt.grid(True)
future_forecast_plot_path = "Previsões futuras do Modelo RN_T_INV_Apng"
plt.savefig(future_forecast_plot_path)
plt.show()

```

Figura 122 : Continuação da implementação da rede neuronal MLP para a variável de temperatura do inversor A em graus celsius

ANEXO E – FUNÇÕES SARIMA

Função `pre_processamento_SARIMA`, Figura 123, é responsável por criar o dataset para o modelo, e admite:

- Input: `df_normalizado` (conjunto de dados normalizado) e `variavel_x` (variável a estudar)
- Output: `df_processado` (conjunto de dados a ser usado para o modelo)

```
def pre_processamento_SARIMA(df_normalizado, variavel_x):  
    df_preprocessado = df_normalizado.copy()  
    df_preprocessado['Data'] = df_preprocessado['Ano'].astype(str) + '-' + df_preprocessado['Mes'].apply(lambda x: str(x).zfill(2))  
    df_preprocessado = df_preprocessado[['Data', variavel_x]]  
    df_preprocessado['Data'] = pd.to_datetime(df_preprocessado['Data'])  
    # Definir o índice como a coluna "Data"  
    df_preprocessado.set_index('Data', inplace=True)  
  
    return df_preprocessado
```

Figura 123: Função `pre_processamento_SARIMA` em Python

Função `calcular_media_desvio_padrao`, Figura 124, é responsável por desenvolver os gráficos da média móvel e do desvio padrão móvel da variável a estudar, e admite:

- Input: `df_processado` (conjunto de dados) e `variavel_x` (variável a estudar)
- Output: `df_preprocessado` (conjunto de dados a ser usado para o modelo) e `plot_path` (gráfico da média móvel e do desvio padrão móvel ao longo do tempo)

```
def calcular_media_desvio_padrao(df_preprocessado, variavel_x):  
  
    df_preprocessado["rolling_avg"] = df_preprocessado[variavel_x].rolling(window=12).mean()  
    df_preprocessado["rolling_std"] = df_preprocessado[variavel_x].rolling(window=12).std()  
  
    plt.figure(figsize=(15,7))  
    plt.plot(df_preprocessado[variavel_x], color='#379BDB', label='Original')  
    plt.plot(df_preprocessado["rolling_avg"], color='#D22A0D', label='Rolling Mean')  
    plt.plot(df_preprocessado["rolling_std"], color='#142039', label='Rolling Std')  
    plt.legend(loc='best')  
    plt.title('Rolling Mean & Standard Deviation')  
  
    # Salvar o gráfico como imagem  
    plot_path = "media_desvio_plot.png"  
    plt.savefig(plot_path)  
    plt.close() # Fechar o gráfico para evitar exibições indesejadas  
  
    return df_preprocessado, plot_path
```

Figura 124: Função `pre_processamento_SARIMA` em Python

Função `teste_dickey_fuller`, Figura 125, é capaz de aplicar o teste ADF através da biblioteca `statsmodels.tsa.stattools`, e admite:

- Input: `df_processado` (conjunto de dados) e `variavel_x` (variável a estudar) e `output_path` (ainda admite o nome da figura com os resultados do teste para guardar)
- Output: `df_output` (resultados do teste ADF)

```
def teste_dickey_fuller(df_preprocessado, variavel_x, output_path):
    print('Resultados Dickey Fuller Test:')
    dftest = adfuller(df_preprocessado[variavel_x], autolag='AIC')

    dfoutput = pd.Series(dftest[0:4], index=['Test Statistic', 'p-value', '#Lags Used', 'Number of Observations Used'])
    for key, value in dftest[4].items():
        dfoutput['Critical Value (%)' % key] = value
    print(dfoutput)

    # Criar uma tabela
    fig, ax = plt.subplots(figsize=(10, 6))
    ax.axis('tight')
    ax.axis('off')
    the_table = ax.table(cellText=np.array([dfoutput]), colLabels=dfoutput.index, cellLoc='center', loc='center')
    the_table.auto_set_font_size(False)
    the_table.set_fontsize(10)
    the_table.scale(2, 1.2)

    # GUARDAR A TABELA COMO UMA IMAGEM
    plt.savefig(output_path, bbox_inches='tight', pad_inches=0.5)
    plt.close()

    return dfoutput
```

Figura 125: Função `teste_dickey_fuller` em Python

Função `plot_acf_pacf_to_word`,

Figura 126, desenvolvida através da biblioteca `statsmodels.graphics.tsaplots`, esta função é capaz de representar os gráficos de ACF e PACF de uma determinada variável, e admite:

- Input: `series` (dataframe em análise e a variável, exemplo: `df_processado[variável_x]`) e `lags` (a quantidade de defasagens para incorporar nos gráficos).

```
def plot_acf_pacf_to_word(series, lags=None):  
    # Crie um gráfico ACF  
    fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(12, 8))  
    plot_acf(series, lags=lags, ax=ax1)  
    plot_pacf(series, lags=lags, ax=ax2)  
  
    # Salve o gráfico como imagem  
    plot_path = "acf_pacf.png"  
    plt.savefig(plot_path)  
    plt.close()
```

Figura 126: Função `plot_acf_pacf_to_word` em Python

Função `Previsao_Sarima_treino_teste`, Figura 127 e Figura 128, utilizada para construir previsões para o conjunto de teste e admite:

- Input: `df_processado` (conjunto de dados) e `variavel_x` (variável a estudar), `test_start_date` (data de começo do conjunto de teste), `Seasonal` (`Seasonal = True` se admitir sazonalidade, caso contrário, `False`), `d` (parâmetro de diferenciação não sazonal), `D` (parâmetro de diferenciação sazonal), `m` (representa a sazonalidade, sazonalidade anual, `m=12`), `doc` (nome do documento word para guardar as figuras).
- Output: `SARIMA_model` (informações sobre o modelo SARIMA ajustado), `SARIMA_model.order` (parâmetros não sazonais), `SARIMA_model.seasonal_order` (parâmetros sazonais do modelo), `SARIMA_model.aic()` (valor de AIC), `mae` e `rmse` (resultados das métricas de avaliação), `diag_plot_path` (nome para guardar a figura do gráfico de diagnostico), `forecast_plot_path` (nome para guardar a figura do gráfico de previsão para o conjunto de teste)

```

def Previsao_Sarima_treino_teste(df_preprocessado,variavel_x, test_start_date, Seasonal, d, D, m,doc=None):

    print('-----DIVISÃO EM TREINO E TESTE-----')
    # Dividir os dados em treino e teste
    print(type(test_start_date))
    train_data = df_preprocessado[df_preprocessado.index <= pd.to_datetime(test_start_date)]
    test_data = df_preprocessado[df_preprocessado.index >= test_start_date]

    print('-----MODELO SARIMA-----')
    # Ajustar o modelo SARIMA com os parâmetros selecionados usando os dados de treino
    SARIMA_model = pm.auto_arima(train_data[variavel_x],
                                start_p= 1, start_q=1,
                                #test='adf',
                                max_p= 2, max_q=2,
                                m=m, # Frequência do ciclo (sazonalidade anual)
                                start_P= 0, start_Q=0,
                                max_P=1, max_Q=1,
                                seasonal= Seasonal, # Sazonalidade definida como True
                                d=d, # Diferenciação não sazonal de primeira ordem
                                D=D, # Diferenciação sazonal de primeira ordem
                                trace=False,
                                error_action='ignore',
                                suppress_warnings=True,
                                stepwise=True,
                                #
                                information_criterion='aic')

    # Visualizar os parâmetros ótimos encontrados
    print("Parâmetros Ótimos do Modelo SARIMA:")
    print(SARIMA_model.order) # (p, d, q) do modelo
    print(SARIMA_model.seasonal_order) # (P, D, Q, s) do modelo (caso sazonal)

    # Valor do AIC
    print("Valor do AIC:")
    print(SARIMA_model.aic())

    print('-----DIAGNÓSTICO DO MODELO-----')

    SARIMA_model.plot_diagnostics(lags=min(10, len(test_data)-2), figsize=(15, 12))

    diag_plot_path = "diagnostics_plot.png"
    plt.savefig(diag_plot_path)
    plt.close()

    print('-----PREVISÕES PARA O CONJUNTO TESTE-----')
    # Fazer previsões no conjunto de teste usando o modelo SARIMA ajustado
    n_periods = len(test_data)
    print(n_periods)

    forecast_dates = pd.date_range(start=test_start_date, periods=n_periods+1, freq='MS')[1:]
    print(forecast_dates)

    forecasts, conf_int = SARIMA_model.predict(n_periods=n_periods, return_conf_int=True, alpha=0.05)
    # Inclua a previsão para julho no início da lista
    print(forecasts)

```

Figura 127: Função Previsao_Sarima_treino_teste em Python

```

print('-----COMPARAR PREVISÕES Vs DADOS REAIS-----')
# gráfico com as previsões e os valores reais (conjunto de teste)

plt.plot(train_data.index, train_data[variavel_x], label='Conj. treino', color='blue')
plt.plot(test_data.index, test_data[variavel_x], label='Conj. teste', color='green')
plt.plot(forecast_dates, forecasts, color='red', label='Previsões')
plt.fill_between(forecast_dates, conf_int[:, 0], conf_int[:, 1], color='pink', alpha=0.3)
plt.legend()
plt.ylim(-2, 2)
plt.title('Previsões do Modelo SARIMA')
plt.ylabel(f'Valores de {variavel_x}')
plt.xlabel('Data')

forecast_plot_path = "Previsões do Modelo SARIMA.png"
plt.savefig(forecast_plot_path)
plt.close()

print('-----Qualidade das previsões-----')
# Encontrar as datas em comum entre test_data e forecasts
common_dates = test_data.index.intersection(forecasts.index)
# Filtrar os dados de test_data e forecasts apenas para as datas em comum
test_data_common = test_data.loc[common_dates]
forecasts_common = forecasts.loc[common_dates]

# Calcular o MAE apenas para as datas em comum
mae = mean_absolute_error(test_data_common[variavel_x], forecasts_common.values)
# Calcular o RMSE apenas para as datas em comum
mse = mean_squared_error(test_data_common[variavel_x], forecasts_common.values)
# Cálculo do Erro Médio Quadrático da Raiz (RMSE)
rmse = np.sqrt(mse)

print("Erro Médio Absoluto (MAE):", mae)
print("Raiz do Erro Médio Quadrático (RMSE):", rmse)

if doc is not None:
    # Adicionar os valores de MAE e MSE ao documento Word
    doc.add_paragraph("Erro Médio Absoluto (MAE): " + str(mae))
    doc.add_paragraph("Raiz Do Erro Médio Quadrático (RMSE): " + str(rmse))

return {
    'modelo': SARIMA_model, # Retorna o modelo ajustado
    'order': SARIMA_model.order,
    'seasonal_order': SARIMA_model.seasonal_order,
    'aic': SARIMA_model.aic(),
    'mae': mae,
    'rmse': rmse,
    'diag_plot_path': diag_plot_path,
    'forecast_plot_path': forecast_plot_path
}

```

Figura 128: Continuação da função Previsao_Sarima_treino_teste em Python

Função `fazer_previsoes_futuras`, Figura 129, é responsável por desenvolver as previsões para datas futuras e um gráfico das mesmas, e admite:

- Input: `df_processado` (conjunto de dados), `modelo_sarima` (informações sobre o modelo SARIMA ajustado anteriormente), `variavel_x` (variável a estudar), `num_periods` (quantidade de períodos para previsão), `output_excel_path` (caminho do ficheiro Excel para guardar as previsões)
- Output: `future_dates` (datas de previsão futuras), `forecasts` (valores de previsão), `conf_int` (intervalo de confiança, valores máximos e mínimos de previsão), `future_forecast_plot_path` (nome da figura das previsões futuras)

```

def fazer_previsoes_futuras(df_preprocessado, modelo_sarima, variavel_x, start_date, num_periods, output_excel_path):
    print('-----Desenvolvimento de previsões futuras')

    # Fazer previsões futuras usando o modelo SARIMA reajustado
    forecast, conf_int = modelo_sarima.predict(n_periods=num_periods, return_conf_int=True, alpha=0.05)

    print("Tamanho do forecast:", len(forecast))
    print("Tamanho do conf_int:", len(conf_int))

    last_date = df_preprocessado.index[-1]
    future_dates = pd.date_range(start=last_date + pd.DateOffset(months=1), periods=num_periods, freq='MS')
    #future_dates = pd.date_range(start=last_date, periods=num_periods+1, freq='MS')
    print(future_dates)
    print(forecast)
    print(conf_int)

    # Salvar as previsões num DataFrame
    previsoes_df = pd.DataFrame({
        'Data': future_dates,
        'Previsao': forecast,
        'Intervalo_Conf_Inf': conf_int[:, 0],
        'Intervalo_Conf_Sup': conf_int[:, 1]
    })

    # Guardar as previsões num arquivo Excel
    previsoes_df.to_excel(output_excel_path, index=False)

    plt.plot(df_preprocessado.index, df_preprocessado[variavel_x], label='original')
    plt.plot(future_dates, forecast, color='red', label='forecast')
    plt.fill_between(future_dates, conf_int[:, 0], conf_int[:, 1], color='pink', alpha=0.3)
    plt.ylim(-2, 2)
    plt.title('Previsões futuras do Modelo SARIMA')
    plt.ylabel(f'Valores de {variavel_x}')
    plt.xlabel('Data')

    future_forecast_plot_path = "Previsões futuras do Modelo SARIMA.png"
    plt.savefig(future_forecast_plot_path)
    plt.close()

    return {
        'future_dates': future_dates,
        'forecasts': forecast,
        'conf_int': conf_int,
        'future_forecast_plot_path': future_forecast_plot_path
    }
}

```

Figura 129: Função fazer_previsoes_futuras em Python

ANEXO F – RESULTADOS ANÁLISE EXPLORATÓRIA

De seguida, são apresentadas figuras relativas à análise exploratória de várias variáveis.

Análise dos valores de temperatura do módulo a este

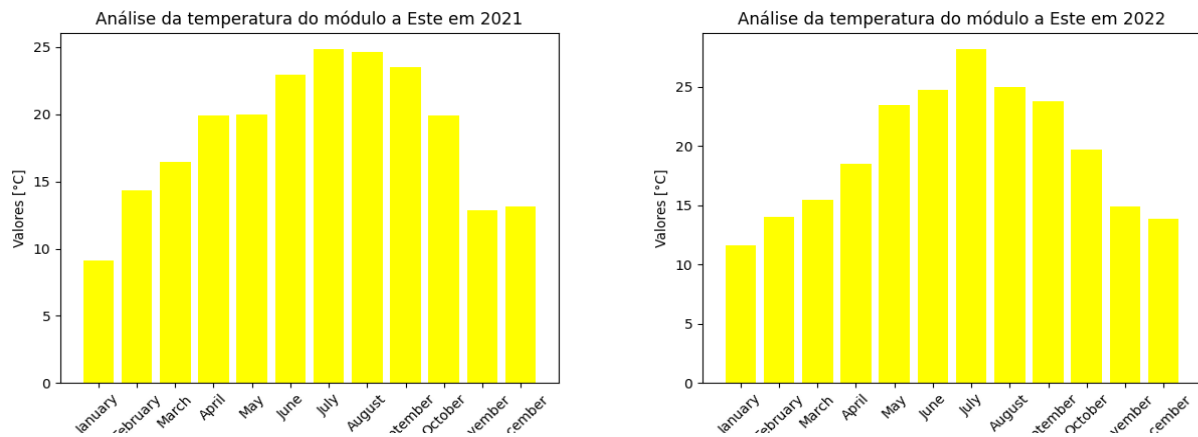


Figura 130 : Análise da temperatura do módulo a Este em 2021 (lado esquerdo) e 2022 (lado direito)

Análise dos valores de temperatura do módulo a oeste

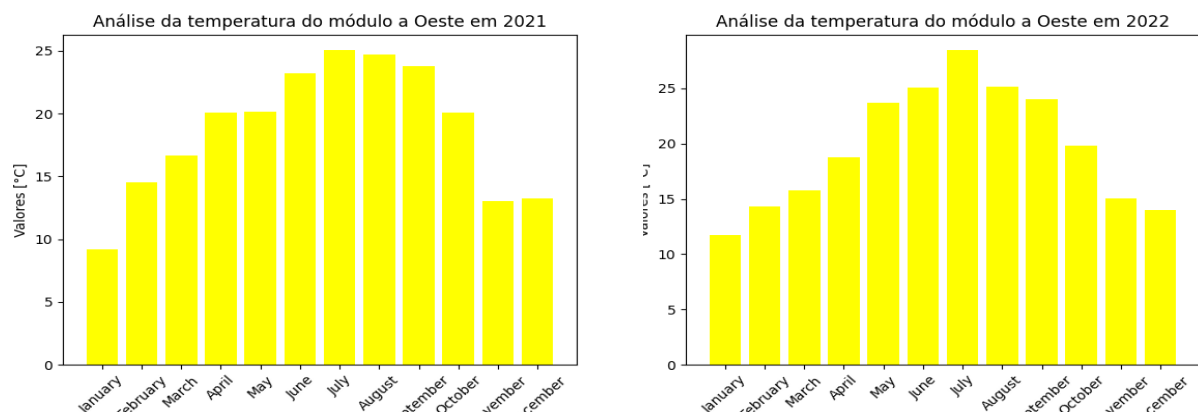


Figura 131: Análise da temperatura do módulo a Oeste em 2021 (lado esquerdo) e 2022 (lado direito)

Análise dos valores do sensor de radiação a este

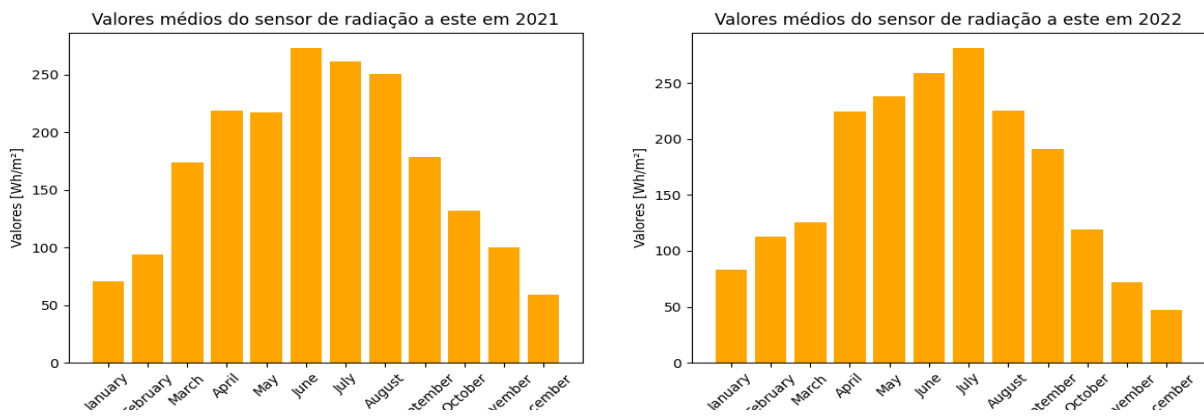


Figura 132: Análise dos valores médios mensais da radiação a Este em 2021 (lado esquerdo) e 2022 (lado direito)

Análise dos valores do sensor de radiação a oeste

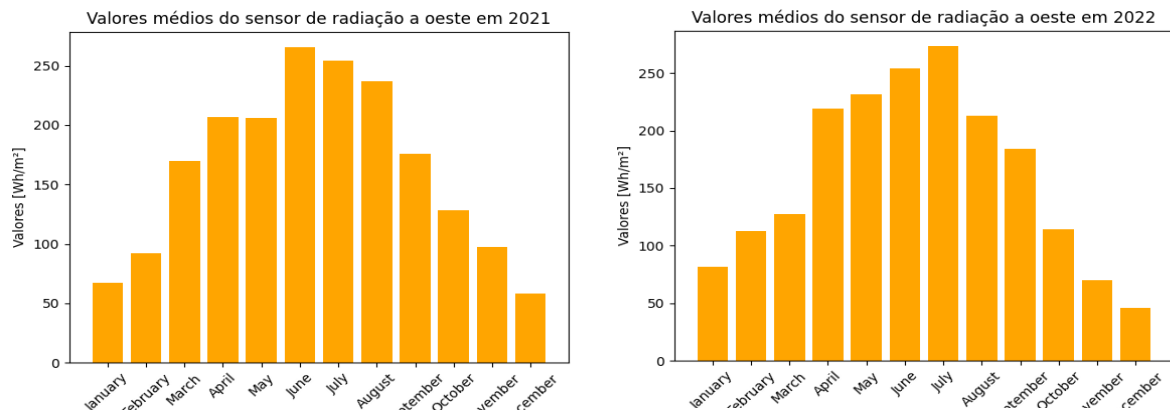


Figura 133: Análise dos valores médios mensais da radiação a Oeste em 2021 (lado esquerdo) e 2022 (lado direito)

Análise dos valores de eficiência do inversor A

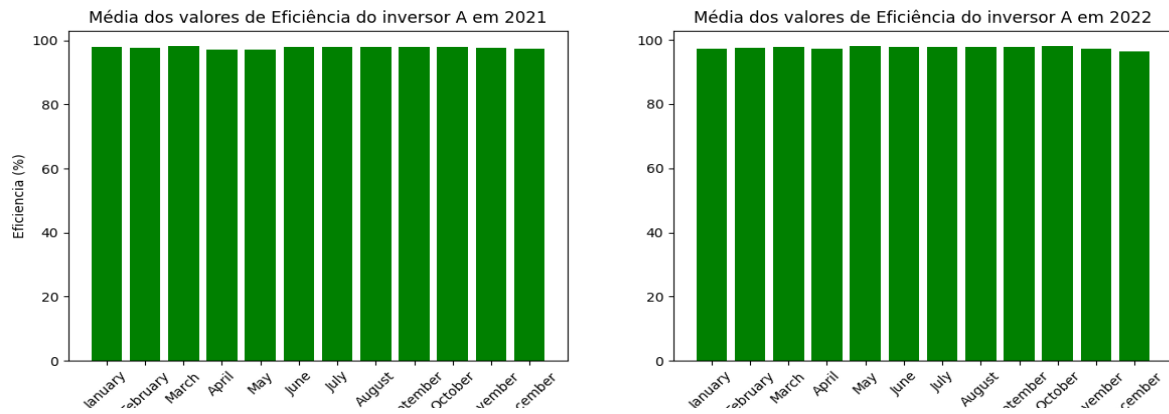


Figura 134: Análise dos valores médios mensais da eficiência do inversor A em 2021 (lado esquerdo) e 2022 (lado direito)

Análise dos valores de eficiência do inversor B

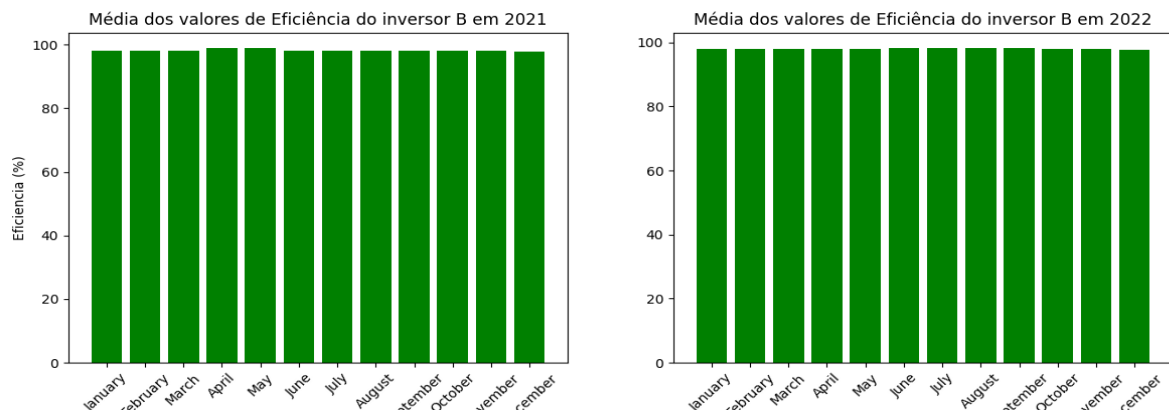


Figura 135: Análise dos valores médios mensais da eficiência do inversor B em 2021 (lado esquerdo) e 2022 (lado direito)