



Instituto Politécnico
de Castelo Branco
Escola Superior
de Tecnologia

Análise e Previsão de Consumos de Energia Utilizando Algoritmos de *Machine Learning* para Eficiência Sustentável

Nuno André Couto Melo

Nº: 20211553

Orientador

Professor Doutor Pedro Miguel Baptista Torres

Relatório de Projeto Final de Curso, apresentado à Escola Superior de Tecnologia do Instituto Politécnico de Castelo Branco para cumprimento dos requisitos necessários à obtenção do grau de Licenciado em Engenharia das Energias Renováveis, realizada sob a orientação científica do Professor Doutor Pedro Miguel Baptista Torres, do Instituto Politécnico de Castelo Branco.

Julho de 2024

Composição do júri

Presidente do júri

Professora Doutora, Paula Cristina Alves Pereira

Professor Adjunto, Escola Superior de Tecnologia do Instituto Politécnico de Castelo Branco

Arguente

Professor Doutor, António Cerejo da Silva

Professor Adjunto, Escola Superior de Tecnologia do Instituto Politécnico de Castelo Branco

Orientador

Professor Doutor, Pedro Miguel Baptista Torres

Professor Adjunto, Escola Superior de Tecnologia do Instituto Politécnico de Castelo Branco

Dedicatória

Aos meus pais, que são a minha maior motivação pelos enormes sacrifícios que fazem por mim. Aspiro orgulhá-los com esta conquista, de ser a primeira geração na academia.

Para os meus amigos, que fizeram um estudante deslocado nunca esquecer a sua casa, os Açores.

À minha madrinha e à sua mãe e irmã, o meu obrigado pelo amor que tenho desde criança.

Aos meus colegas, pelas memórias criadas nestes 3 anos, que nunca me irei esquecer.

Aos meus professores, o meu agradecimento por contribuírem para o meu crescimento, como pessoa.

Ao meu orientador, Professor Doutor Pedro Miguel Baptista Torres, pela sua paciência perante eu neste projeto.

E a todos, o meu eterno obrigado.

Resumo

O aumento do consumo de energia elétrica dos setores de transportes, doméstico e indústria nos últimos anos, mostra que o recurso a energias renováveis (RE) é crucial e imperativo, de modo a atingir as metas climáticas. Para esse efeito, no dimensionamento e gestão de uma rede elétrica verde e autossustentável, é necessário diminuir as perdas na distribuição e transporte, sob um lema de maior eficiência, visando a preservação da biodiversidade e do ecossistema, impedindo a acumulação de gases com efeito de estufa (GEE). O projeto centra-se na aplicação de algoritmos de *Machine Learning* (ML) para a previsão de consumos de energia em curto prazo (*short-term load forecasting* (STLF)), em Portugal entre janeiro e abril de 2024. Ao mesmo tempo, explorando a complexa relação entre a previsão da carga e os parâmetros meteorológicos, por exemplo, a temperatura. As conclusões do projeto permitem deduzir que algoritmos como o *Long Short-Term Memory* (LSTM) e o *Multilayer Perceptron* (MLP), em contraste com outros mais tradicionais, o *Seasonal Autoregressive Integrated Moving Average* (SARIMA) e *Seasonal Autoregressive Integrated Moving Average Exogenous* (SARIMAX), tiveram um desempenho 50% superior nas métricas de desempenho, *Mean Absolute Percentage Error* (MAPE), *Mean Absolute Percentage* (MEAN) e *Root Mean Square Error* (RMSE). Os resultados mostram que a utilização de algoritmos de ML, são uma mais-valia no equilíbrio do sistema elétrico e permitem por antecipação fornecer inputs de consumos, importante para os gestores da rede, tomarem as medidas necessárias que garantam a qualidade de serviço.

Palavras-chave

Consumos de Energia Renovável, Previsão, Inteligência Artificial, *Machine Learning*, Eficiência Energética.

Abstract

The rise in energy consumption driven by the transport, residential, and industrial sectors has made the electrification of power grids by renewable energies (RE) vital and unavoidable, as we aspire to reach climate goals. To design and manage a green and self-sustainable electrical network, it is necessary to reduce losses in distribution and transportation, under a motto of higher efficiency, to preserve biodiversity and the ecosystem, deterring the accumulation of greenhouse gases (GEE). The research is centered around the application of *Machine Learning* (ML) algorithms for *short-term load forecasting* (STLF) relative to Portugal between January and April 2024. At the same time, exploring the complex relation between predicting load with weather parameters, for example, temperature. The current study found that algorithms such as *Long Short-Term Memory* (LSTM) and *Multilayer Perceptron* (MLP), as opposed to more traditional ones, *Seasonal Autoregressive Integrated Moving Average* (SARIMA) and *Seasonal Autoregressive Integrated Moving Average Exogenous* (SARIMAX), have performed 50% better in the metrics of *Mean Absolute Percentage Error* (MAPE), *Mean Absolute Percentage* (MEAN), and *Root Mean Square Error* (RMSE). These findings provide evidence that the use of ML algorithms are an added value in the balancing of the electric system and allows for consumption inputs, which is paramount for grid operators to take the necessary measures to guarantee quality of service.

Keywords

Renewal Energy Consumption, Forecasting, Artificial Intelligence, Machine Learning, Energy Efficiency.

Índice geral

1.	Introdução.....	1
1.1	Enquadramento do Trabalho.....	1
1.2	Contextualização.....	3
1.3	Objetivos do Trabalho.....	5
1.4	Estrutura do Relatório.....	5
2.	Machine Learning.....	7
2.1	Sistemas de Machine Learning.....	9
2.2	Métodos de Aprendizagem.....	9
2.2.1	Supervised learning.....	10
2.2.2	Unsupervised learning.....	10
2.2.3	Reinforced learning.....	11
2.3	Algoritmos utilizados para previsão de consumos de energia.....	11
2.4	Aplicação no campo das Energias Renováveis.....	12
2.4.1	<i>Mercado de eletricidade</i>	13
2.4.2	<i>Grids do futuro</i>	17
2.5	Python.....	21
3.	Metodologia.....	23
3.1	Recolha de dados.....	23
3.2	Pré-processamento dos dados.....	23
3.3	Métodos de análise.....	24
3.4	Caso em estudo.....	25
3.5	Licenças.....	25
3.6	Métricas.....	26
4.	Análise de Consumo de Energia.....	27
4.1	Análise exploratória dos dados.....	30
5.	Previsão de Consumos de Energia com Machine Learning.....	35
5.1	Metodologia dos Modelos.....	35
6.	Resultados e Discussão.....	49
6.1	Comparação dos modelos.....	49
6.2	Melhorias.....	57
7.	Conclusões.....	59

7.1 Contribuições do estudo para o campo das energias renováveis e eficiência energética.....	59
7.2 Sugestões para trabalhos futuros.....	59
Referências Bibliográficas	61

Índice de figuras

Figura 1 – Anomalia da temperatura do ar, entre o intervalo de 1940 a 2024. Fonte:(Copernicus, 2024).	1
Figura 2 - Consumo energético final conforme o cenário 1,5 °C. Fonte:(IRENA, 2023).	4
Figura 3 - Etapas de processamento da rede elétrica. Fonte:(Farhangi, 2010).	4
Figura 4 – Processo genérico de Machine Learning. Fonte:(Raschka e Mirjalili, 2017).	7
Figura 5 - Comparação entre uma decisão linear e não linear e um respetivo equilíbrio, relativamente ao modelo. Fonte:(Raschka e Mirjalili, 2017).	8
Figura 6 – A integração de ML, desde as suas etapas iniciais, que vai desde a obtenção de dados (data federation) até à criação do modelo, com a componente de software, com o propósito de garantir que o sistema esteja em níveis de desempenho adequados. Fonte:(Benton, 2020).	9
Figura 7 – Métodos de aprendizagem de Machine Learning. Fonte:(Raschka e Mirjalili, 2017).	10
Figura 8 - Processo de supervised learning. Fonte:(Raschka e Mirjalili, 2017).	10
Figura 9 - Processo de “reinforced learning”. Fonte:(Raschka e Mirjalili, 2017).	11
Figura 10 – Exemplificação de como é determinado preço de mercado do dia seguinte, resultante do cruzamento entre a produção e a procura. Fonte:(Pinhão, Fonseca e Covas, 2022).	13
Figura 11 - Energia gerada no mercado MIBEL e o seu respetivo custo, no dia 3 de abril de 2024. Fonte:(<i>Day-ahead hourly price / OMIE</i> , [s.d.])	14
Figura 12 - Produção de eletricidade, por tipo, para Portugal Continental no dia 4 de abril de 2024. Fonte:(REN, 2024).	15
Figura 13 – Acréscimo verificado nas TAR de maio para junho de 2024. Fonte:(ERSE, 2024)	16
Figura 14 –Representação da repartição dos custos, aos quais são associados às TAR. Fonte:(João Lopes Veloso Ribeiro da Silva <i>et al.</i> , 2011).	16
Figura 15 –Demonstração de um sistema UPAC, ligado tanto a uma infraestrutura doméstica como à rede pública, permitido pelo contador digital. Fonte:(E-REDES, [s.d.])	19
Figura 16 – Dados disponíveis na operação de Smart Grids. Fonte:(Kaplan, Tehrani e Jamshidi, 2021)	19
Figura 17 – A Demand Response pode-se caracterizar pela necessidade de uma tarifa dinâmica, Price-based Programs. Ou, por outro lado, pelo controlo de aparelhos, de forma direta, imposta pelo operador. Como também com a partição incentivada do cliente na manutenção da operação da rede. Fonte:(Alotaibi <i>et al.</i> , 2020)	20
Figura 18 – Etapas que descrevem o passo inicial de recolha de dados bem como o seu processamento consecutivo.	24
Figura 19 – Distribuição da ocorrência do vetor vento, pela escala visual e a sua respetiva direção com indicação da velocidade.	28

Figura 20 – Representação da função sinusoidal e função cosseno, em termos de hora e dia. A função sinusoidal retrata a ocorrência de picos de carga.....	28
Figura 21 - Estudo do nível de impacto que as variáveis têm sobre a previsão da variável carga, por um modelo de Random Forest.....	29
Figura 22 – Estudo da correlação entre variáveis, pelo método de Pearson, com a sua máxima semelhança representado a vermelho e o inverso a azul. Este sistema permite averiguar quais variáveis podem entrar em conflito pela sua semelhança....	30
Figura 23 - Representação gráfica da variável carga, na totalidade. Em dois formatos por hora e dia.	31
Figura 24 - Representação, simultânea, das variáveis carga e temperatura.	31
Figura 25 - Carga mediana por dia da semana.	32
Figura 26 - Carga mediana por hora.....	32
Figura 27 - Distribuição da frequência para as variáveis de temperatura, humidade, pressão, nebulosidade e carga.	33
Figura 28 - Identificação da mediana, quartis, mínimo e máximo e outliers para as variáveis de temperatura, humidade, pressão, nebulosidade e carga. Pela sua representação em box plots.....	34
Figura 29 - Representação da variável Carga após a sua normalização para a escala de 0 a 1.....	36
Figura 30 – Diagrama que enquadra a metodologia aplicada na construção do modelo baseado no algoritmo SARIMA.....	38
Figura 31 – A figura tem como interesse a exemplificação do processo de previsão e como este é efetuado. Onde os Inputs, através de $t=N$, representam os dados e a sua respetiva entrada. Que serão processados pelo modelo de forma a fazer as previsões e comparar aos Labels a que correspondem aos valores reais.	39
Figura 32 - Diagrama que enquadra a metodologia aplicada na construção do modelo baseado no algoritmo SARIMAX.....	40
Figura 33 – A figura tem como interesse a exemplificação do processo de previsão e como este é efetuado para o Modelo 2. Onde os Inputs, através de $t=N$, representam os dados e a sua respetiva entrada. Que serão processados pelo modelo, em conjunto com as variáveis exógenas que auxiliarão o processo das previsões. Com os respetivos Labels, que armazenam os valores reais.	40
Figura 34 – Estado de uma célula de LSTM e o modo de processamento da sequência de dados. Com a representação das portas, o estado da célula e camada oculta. Fonte:(Alsabban, Salem e Malik, 2021).....	43
Figura 35 - Demonstração do processo de previsão do modelo. Com a utilização de uma sequência de 24 valores, dos dados de teste, para realizar a previsão do valor consequente. No caso out of sample, diferencia-se ao utilizar os últimos valores no registo de teste, onde após prever $t=24$, esse valor é adicionado à sequência. O que, em ambos os casos, remove o valor inicial.....	43
Figura 36 - Diagrama que enquadra a metodologia aplicada na construção do modelo baseado no algoritmo LSTM.....	45

Figura 37 - Perdas associadas à aprendizagem do modelo 3. A escala em uso será de 0 a 1, pela questão dos dados se encontrarem normalizados, com o número de <i>epochs</i> na escala inferior.....	45
Figura 38 - Representação da estrutura que compõem o MLP. Fonte:(Bento, 2021).	46
Figura 39 - Demonstração do processo de previsão do modelo. Com a utilização de uma sequência de 24 valores, dos dados de teste, para realizar a previsão do valor consequente. No caso out of sample, diferencia-se ao utilizar os últimos valores no registo de teste, onde após prever t=24, esse valor é adicionado à sequência.	47
Figura 40 - Diagrama que enquadra a metodologia aplicada na construção do modelo baseado no algoritmo MLP.....	48
Figura 41 - Perdas associadas à aprendizagem do modelo 4. A escala em uso será de 0 a 1, pela questão dos dados se encontrarem normalizados, com o número de <i>epochs</i> na escala inferior.	48
Figura 42 - Previsão efetuada sobre os dados de teste do Modelo 1.....	52
Figura 43 - Previsão efetuada out of sample (5 dias após os dados de teste), num formato horário, para o Modelo 1.....	52
Figura 44 - Previsão efetuada sobre os dados de teste do Modelo 2.....	53
Figura 45 - Previsão efetuada out of sample (5 dias após os dados de teste), num formato horário, para o Modelo 2.....	54
Figura 46 - Previsão efetuada sobre os dados de teste do Modelo 3.....	54
Figura 47 - Previsão efetuada out of sample (5 dias após os dados de teste), num formato horário, para o Modelo 3.....	55
Figura 48 - Previsão efetuada sobre os dados de teste do Modelo 4.....	56
Figura 49 - Previsão efetuada out of sample (5 dias após os dados de teste), num formato horário, para o Modelo 4.....	56

Índice de mapas

Mapa 1 - Cartas do Índice de Aridez de Portugal Continental de 1960-1990 e de 2000 a 2010. Fonte:(Coelho e Valente, 2015).	2
Mapa 2 - Intensidade de emissão de dióxido de carbono na produção de eletricidade, em 2023. Fonte:(<i>Electricity Maps - Europe- 2023 - Carbon Intensity Data, 2024</i>).	13

Lista de tabelas

Tabela 1 - Primeira entrada dos parâmetros meteorológicos para as respetivas estações.....	27
Tabela 2 – A Tabela 2 demonstra as cinco primeiro entradas dos do dataframe final, ou seja, os dados utilizados para prever a carga.....	28
Tabela 3 - Parâmetros do algoritmo ARIMA.....	35
Tabela 4 – Resultados obtidos para o teste ADF, com o intuito de verificar a hipótese que a série seria estacionária.....	37
Tabela 5 – A seguinte tabela, enquadra os MAE e RMSE respetivos a cada modelo. Onde valores superiores, significa um pior desempenho.....	50
Tabela 6 - A seguinte tabela, enquadra o MAPE respetivo a cada modelo. Onde valores superiores, significa um pior desempenho.	50
Tabela 7 - A seguinte tabela, enquadra o R2 respetivo aos Modelos 3 e 4. Onde valores superiores, significa um melhor desempenho.	51
Tabela 8 - Registo dos cinco primeiros valores de previsão dos dados de teste para o Modelo 1, com os respetivos intervalos de confiança inferior e superior.....	52
Tabela 9 - Registo dos cinco primeiros valores de previsão out of sample para o Modelo 1, com os respetivos intervalos de confiança inferior e superior.....	53
Tabela 10 - Registo dos cinco primeiros valores de previsão dos dados de teste para o Modelo 2, com os respetivos intervalos de confiança inferior e superior.....	53
Tabela 11 - Registo dos cinco primeiros valores de previsão out of sample para o Modelo 2, com os respetivos intervalos de confiança inferior e superior.....	54
Tabela 12 - Registo dos cinco primeiros valores de previsão dos dados de teste para o Modelo 3.....	55
Tabela 13 - Registo dos cinco primeiros valores de previsão out of sample para o Modelo 3.....	55
Tabela 14 - Registo dos cinco primeiros valores de previsão dos dados de teste para o Modelo 4.....	56
Tabela 15 – Registo dos cinco primeiros valores de previsão out of sample para o Modelo 4.....	57

Lista de equações

Equação 1	26
Equação 2	26
Equação 3	26
Equação 4	26
Equação 5	29
Equação 6	29
Equação 7	36
Equação 8	36
Equação 9	36
Equação 10	36
Equação 11	37
Equação 12	37
Equação 13	39
Equação 14	41
Equação 15	41
Equação 16	41
Equação 17	41
Equação 18	42
Equação 19	42
Equação 20	42
Equação 21	44
Equação 22	46

Lista de abreviaturas, siglas e acrónimos

Unidades

ppm – Partes por milhão
ppb– Partes por bilião
GW- GigaWatt
Gt - Gigatonelada
km - Quilómetros
CET – Central European Time

Termos relativos a governança

NDC - Nationally Determined Contributions
PIB – Produto Interno Bruto
UE – União Europeia

Elementos químicos

GEE – Gases de Efeito Estufa
CO₂ - Dióxido de Carbono
CH₄ - Metano

Machine Learning

MAE - Mean Absolute Error
MAPE - Mean Absolute Percentage Error
 R^2 - Coefficient of Determination
RMSE - Root Mean Squared Error
ML – Machine Learning
ARIMA – Autoregressive Integrated Moving Average
SARIMA – Seasonal Autoregressive Integrated Moving Average
ADF – Augmented Dickey–Fuller
AIC – Akaike Information Criterion
SARIMAX - Seasonal Autoregressive Integrated Moving Average Exogenous
LSTM – Long Short-Term Memory

RNN - Recurrent Neural Network

SVR – Support Vector Regression

ETS – Exponential Smoothing

MLP - Multilayer Perceptron

Energias Renováveis

UPAC – Unidade de Produção para Autoconsumo

CER – Comunidade de Energia Renovável

AMI - Advanced Metering Infrastructure

RE – Energias Renováveis

EES – Electrical Energy Storage

DER – Distributed Energy Resources

Setor elétrico

FIT - Feed-In Tariffs

ENTSO-E - European Network of Transmission System Operators for Electricity

FACTS - Flexible AC Transmission System

TSO - Transmission System Operator

PPA - Power Purchase Agreements

MIBEL - Mercado Ibérico de Eletricidade

ERSE – Entidade Reguladora dos Serviços Energéticos

CIEG – Custos de Interesse Económico Geral

TAR – Tarifas de Acesso às Redes

CAE - Contrato de Aquisição de Energia

PRG - Produção de Remuneração Garantida

Cfd - Contracts for Differences

OPF – Optimal Power Flow

STLF - Short Term Load Forecasting

LCOE - Levelized Cost of Energy

Licenças

ODbL - Open Data Commons Open Database License

CC-BY 4.0 - Creative Commons Attribution 4.0 International License

Ferramentas de desenvolvimento

API – Application Programming Interface

JSON - JavaScript Object Notation

IDE - Integrated Development Environment

CSV - Comma-Separated Values

HTTP - Hypertext Transfer Protocol

1. Introdução

1.1 Enquadramento do Trabalho

O efeito acumulativo de emissões de GEE (Gases de Efeito de Estufa), como o CO₂ (Dióxido de Carbono) e CH₄ (Metano), vem alterando consequentemente o paradigma do clima a nível global. A presença destes dois gases na atmosfera tem aumentado anualmente, os últimos valores registavam 424,55 ppm e 1934,16 ppb, para CO₂ e CH₄, respetivamente (Lan, 2024, 2024).

No ano de 2023, constatou-se um agravamento do aumento das temperaturas, tendo uma anomalia de 1,3 a 1,5 °C. Essa tendência tem vindo a agravar-se e no ano de 2024, como destacado na **Figura 1**, é possível evidenciar que estamos perante temperaturas sem precedentes, onde se registou o mês de janeiro mais quente, tendo uma anomalia de 1,66 °C (Copernicus, 2024).

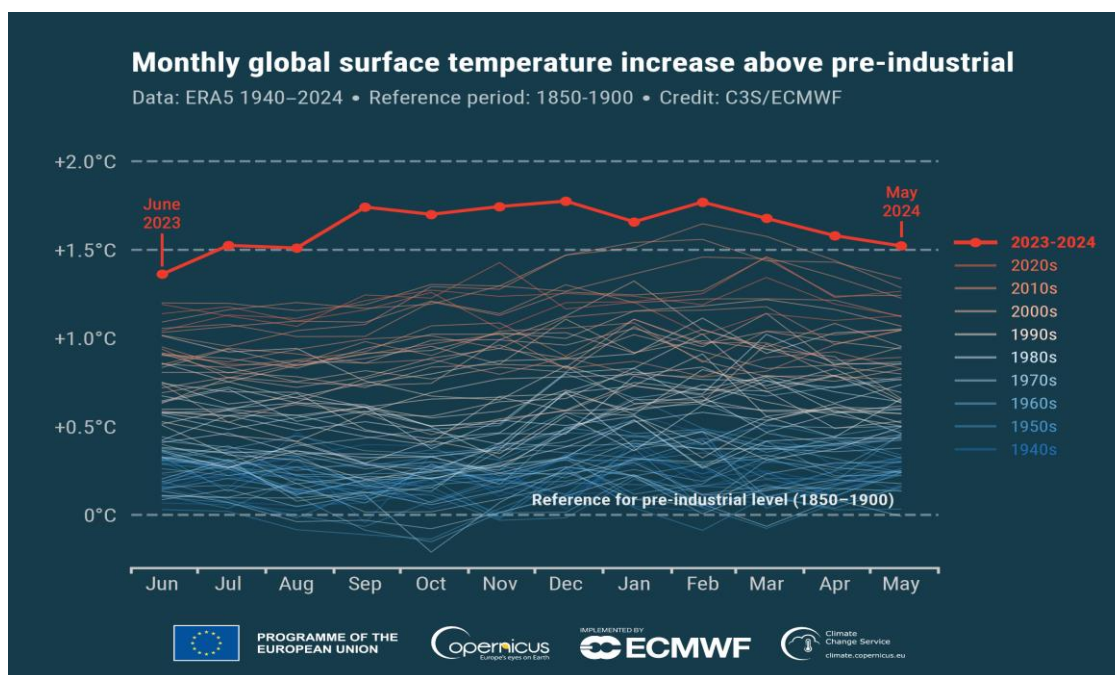
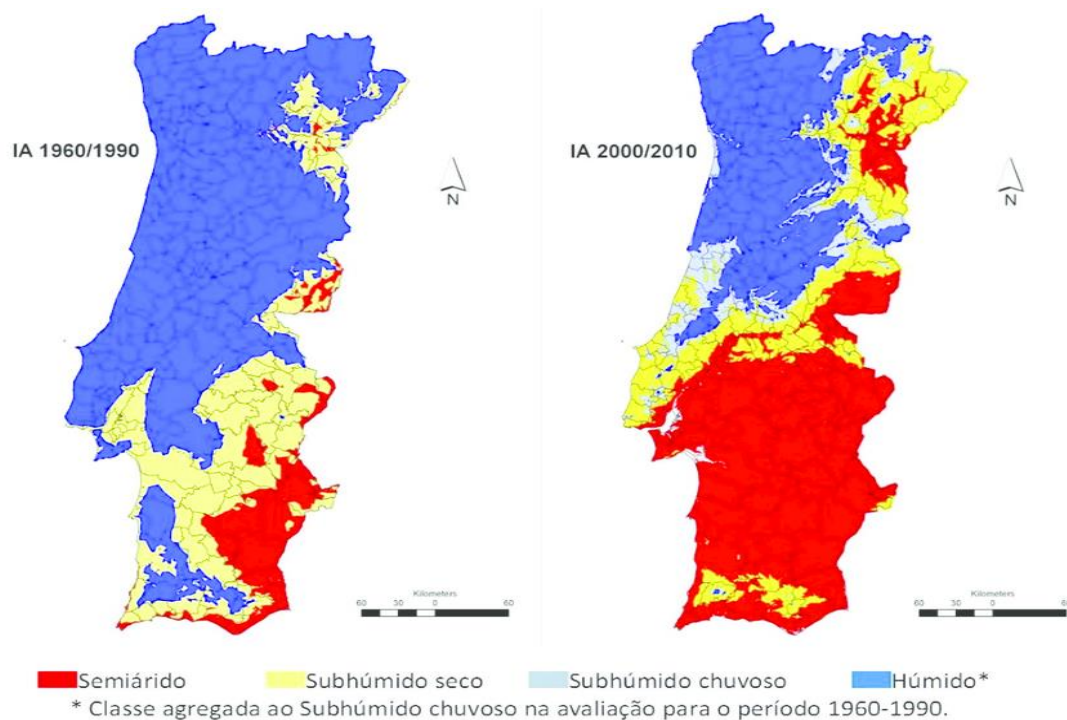


Figura 1 - Anomalia da temperatura do ar, entre o intervalo de 1940 a 2024. Fonte: (Copernicus, 2024).

A crescente desertificação dos terrenos, provocado por ação humana e subida de temperaturas, que em conjunto com o aumento da população, onde aumentou 2 biliões desde 2001, para 7,91 biliões em 2021 (Ritchie *et al.*, 2023), são fatores que põem em causa o acesso a alimento e diminuem a área de solo disponível para cultivo. Nacionalmente a produção de cereais, em 2023, sofreu uma quebra, apresentando um défice comercial de 1.100 biliões (GPP, 2024), o que rotula Portugal como insuficiente e dependente do mercado estrangeiro, levantando sérias questões sobre o futuro da agricultura em Portugal e como poderá tornar-se maioritariamente consumidor.

Os avanços tecnológicos e práticas mais sustentáveis têm permitido aumentar o rendimento das plantações, sem ter de aumentar necessariamente o solo para cultivos.

Ao incidir nestes dois pontos, a melhor utilização e práticas de regeneração do solo, para diminuir a área desertificada e agricultura intensiva, podendo corresponder ao aumento da procura por alimento. Sendo que é previsto um aumento do consumo calórico de 2880 kcal para 3250 kcal, de 2015 para 2050 (Flies *et al.*, 2018), o que se traduz na necessidade de produzir mais 50% de alimento, até 2050, no entanto, o rendimento da produção irá baixar até 30% (Quiggin *et al.*, 2021).



Mapa 1 - Cartas do Índice de Aridez de Portugal Continental de 1960-1990 e de 2000 a 2010.
Fonte: (Coelho e Valente, 2015).

A procura de um mundo sustentável depende não só dos órgãos governamentais, mas também das pessoas, uma mensagem, clara e positiva, que indique que uma ação no presente, será o espelho de um futuro melhor, o que será fulcral para uma maior adesão a políticas ambientais, que visam contrariar as alterações climáticas (Vlasceanu *et al.*, 2024).

A vertente ambiental tem tido uma maior ponderação no crescimento económico, para esse efeito tem-se considerado a pegada ecológica, que quantifica o consumo humano pelos recursos naturais da Terra, tendo em conta a sua capacidade regenerativa (Guarnido-Rueda e Amate-Fortes, 2021). A alocação de recursos em medidas que visam a proteção do ambiente, em conjunto com taxas sobre o transporte, são dois fatores que levam a uma menor pegada ecológica, criando uma harmonização entre esse e o crescimento económico (Carmona *et al.*, 2023).

O PIB alocado para a Proteção do Ambiente, na EU entre 2013 e 2022, foi, em média, 0,8%, exceto em 2020, onde foi 0,9% (General Government Expenditure by Function, 2024). Ao considerar um período temporal de 10 anos, 2013 a 2023, verificou-se uma disjunção entre o crescimento da economia e da emissão de CO₂, globalmente, onde o

PIB cresceu 2,9% e a emissão 0,7%(International Energy Agency, 2023). Salienta-se o segundo valor por ser relativamente baixo, ao ter sido instalado uma potência combinada de eólica e fotovoltaica, no valor de 540 GW, um aumento de 75%, relativamente a 2022(International Energy Agency, 2023), o que se traduz numa redução da emissão de CO₂, de 2022 para 2023, em cerca de 1,7 Gt (International Energy Agency, 2023).

Assumir como diretiva uma economia verde e circular, será a única forma de ser possível obter um desenvolvimento sustentável e equitativo. A transição não pode ser dependente de uma extração intensiva, sendo esperada crescer 60% até 2060(UN Environment Program, 2024) e questionável a nível ético, que agrava as desigualdades em países em desenvolvimento.

Com um culminar numa sociedade justa, concedendo condições dignas, independentemente das diferentes situações sociais. Tendo como pilar uma vertente socioambiental, com um reforço da cooperação global e das NDC. Pela crise climática não se tratar de um problema de apenas um “eu”, mas sim de um coletivo que tem de unir e globalizar os seus esforços.

1.2 Contextualização

O setor energético terá de ser reformulado, com o propósito de o munir com a capacidade de integrar fontes renováveis, permitindo acompanhar o aumento do consumo energético necessário para atingir emissões neutras em 2050, o que resulta num crescimento de 20%(Energy Agency, 2023) superior ao da última década. Atingir as metas propostas de eletrificação por renováveis do consumo energético final, implica ter uma quota de 90%(IRENA, 2023), como consta na **Figura 2**. No entanto, alerta-se para o facto de haver barreiras que retardarão o processo de eletrificação dos vários setores, os quais podem ser desde um deficiente plano energético, condições geográficas, situação económica e entre outros(Sovacool, Rio, Del e Zhang, 2023).

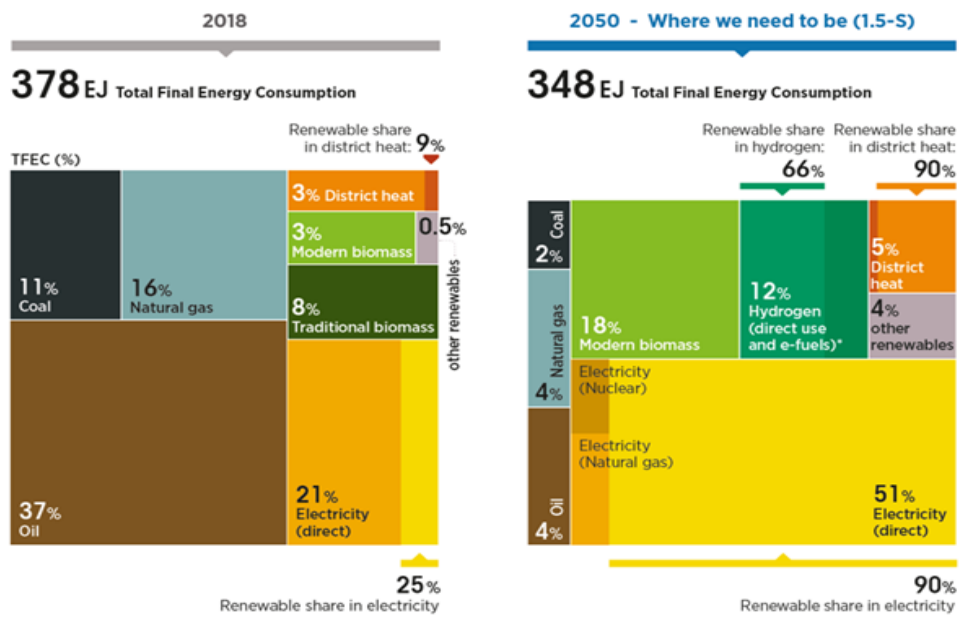


Figura 2 - Consumo energético final conforme o cenário 1,5 °C. Fonte:(IRENA, 2023).

O que poderia ser considerado anteriormente um entrave para novos projetos, sejam fotovoltaicos ou eólicos, o LCOE, atingiu valores inferiores aos de combustíveis fósseis, em 2022, menos 29% e 52%, respetivamente(Renewable Energy Agency, 2023).

Na **Figura 3**, podemos constatar as várias etapas da rede, desde a produção, transporte e distribuição de eletricidade até ao consumidor. A atual rede elétrica apresenta lacunas em termos de eficiência e gestão, ao perder 8% da energia produzida nas linhas de transporte e 20% da sua capacidade tem como intuito corresponder a picos de consumo(Kataray *et al.*, 2023).

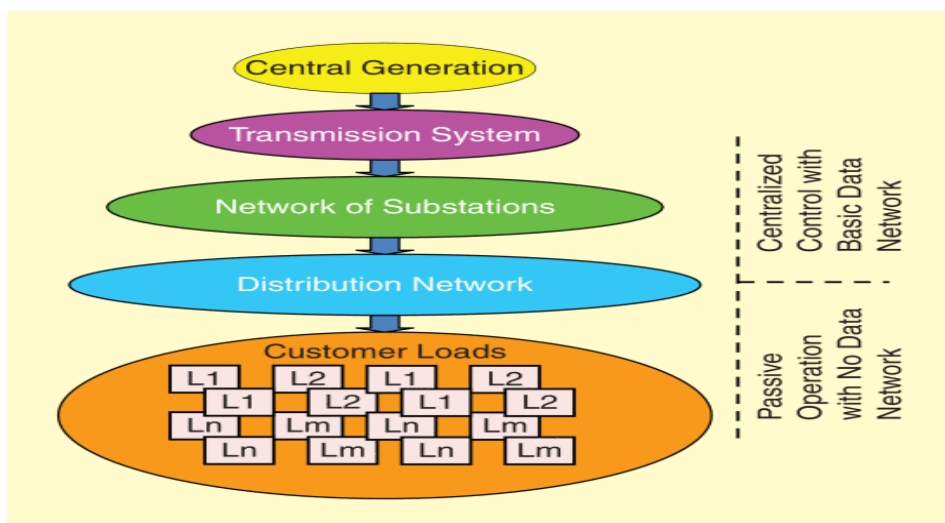


Figura 3 - Etapas de processamento da rede elétrica. Fonte:(Farhangi, 2010).

A combinação das tecnologias de informação e comunicação com a rede elétrica, permite o desenvolvimento da mesma, sem a necessidade inicial de a expandir. Ao

dotar a rede com estes equipamentos, podemos melhorá-la em termos de gestão e eficiência, fulcral para a integração de fontes intermitentes, devido à sua variabilidade e imprevisibilidade (Kataray *et al.*, 2023), pelo acesso a uma multitude de informação em tempo real.

Contudo, alerta-se para o facto da grande necessidade de investir na expansão/substituição da rede, especificamente em 80 milhões de km (Energy Agency, 2023). Sabendo que este é um processo de longa duração, devido aos vários componentes empregados para o correto funcionamento, desde transformadores a sistemas de controlo, o que leva a constrangimentos na incorporação de novas centrais.

A aplicação de novas tecnologias como a Inteligência Artificial, em particular na automação de processos e previsão de eventos através de técnicas de *Machine Learning (ML)*, com aplicação nas *Smart grids*, vem trazer a otimização dos sistemas de produção e consumo. Prever o consumo energético, requer a análise de uma grande quantidade de dados de consumos passados, bem como ter em conta a variabilidade dos dados em função de fatores externos, como o clima e/ou eventos socioeconómicos.

A análise do comportamento dos consumos e previsão via algoritmos de ML, proporciona uma visão mais abrangente da rede, facilitando a sua gestão e otimização. É sabido também que, com o crescimento do consumo energético, as perdas associadas ao funcionamento da rede têm tendência a aumentar, especialmente nos picos de consumo, sendo que através destas técnicas de *forecasting* (previsão), consegue-se identificar esses picos, levando a que por antecipação possam ser tomadas algumas ações na rede, a fim de minimizar desperdícios ou falhas.

1.3 Objetivos do Trabalho

O intuito do trabalho, foca-se na aplicação de algoritmos de *Machine Learning* na análise de consumos energéticos e previsão (*forecasting*) de consumos a curto prazo (*short-term*), que permitam perceber o comportamento energético do sistema em análise. Esta informação é primordial para as entidades gestoras da rede, providenciando informação útil sobre quais vão ser as necessidades de consumo num período próximo e que ações devem ser tomadas para o garantir. São analisados diferentes algoritmos a fim de estabelecer uma comparação em termos de precisão e exatidão dos resultados obtidos, bem como o seu custo computacional.

1.4 Estrutura do Relatório

O trabalho está organizado em 7 capítulos onde no Capítulo 1 é introduzido a ideia da necessidade da transição energética, sustentada num princípio de sustentabilidade, como meio de combater as alterações climáticas e o seu entrelaço com os pilares da

economia e sociais. Ainda no mesmo capítulo, aborda-se a necessidade do projeto na transição verde e os seus respetivos objetivos.

No Capítulo 2, apresenta-se o fundamento teórico que servirá como espinha dorsal, relevante para o contexto abordado. São introduzidas as ideias iniciais, conceptuais de ML e a sua respetiva aplicação na área das Energias Renováveis (RE), em concreto, de como o ML desempenhará um papel importante no setor elétrico.

O Capítulo 3, reúne a metodologia do trabalho com a descrição deste, a obtenção dos dados e o seu tratamento, para posterior utilização nos algoritmos de ML, com enumeração das ferramentas de Python utilizadas.

No conseqüente capítulo, Capítulo 4, é feita uma análise aos dados de Carga (Load) e meteorológicos, com o intuito de identificar padrões ou tendências nos mesmos, com o recurso de ferramentas visuais.

O Capítulo 5, apresenta os algoritmos utilizados para realizar as previsões de Carga e a sua respetiva metodologia de processamento.

A seguir, no Capítulo 6 são expostos os resultados dos modelos, e com auxílio das métricas mencionadas no último capítulo, a caracterização dos seus desempenhos em termos de precisão, assim como, considerações sobre percalços enfrentados.

Por último, o Capítulo 7 aborda os resultados anteriores, tomando elações sobre esses mesmos e a sua perspectiva perante o campo das RE, bem como o trabalho pode ser expandido e desenvolvido.

2. Machine Learning

A capacidade de detetar padrões em amostras empíricas, que não são possíveis ser detetadas a olho humano, é a essência do que é o *ML*. Tem a capacidade de extrapolar informação, a grande escala, e de continuamente atualizar os resultados com integração de novos dados, novos conhecimentos(Wittek, 2014; Woolf, 2009).

No entanto, antes de poder efetuar qualquer ação com algoritmos de *ML*, é necessário trabalhar os dados, nomeadamente fazer alguma análise estatística e limpar o “*dataset*” através da eliminação de “*outliers*”, para não enviesar os resultados. No processo de aprendizagem o “*dataset*” é dividido em dados de treino, teste e validação, antes de serem injetados como entradas do algoritmo de *ML*.

A escolha dos algoritmos é em função do tipo de aplicação que se pretende, podendo eles ser mais vocacionados para estimação ou para classificação. No caso concreto deste trabalho, estamos perante séries temporais, ou seja, valores de consumos energéticos em função do tempo, logo são selecionados algoritmos específicos para conseguir trabalhar com séries temporais e com capacidade para fazer inferências.

O *ML* é uma subárea da Inteligência Artificial, que pode ser caracterizada pelas seguintes etapas(Edgar e Manz, 2017):

- Processamento de dados;
- Análise de comportamento de dados através da relação causal dos dados ou correlação;
- Tendo em conta essa relação, selecionar um algoritmo que melhor se enquadra para extrair determinada característica ou padrão;
- Obter um modelo, que reúna as relações dos dados e com capacidade de os prever. Este modelo é selecionado conforme o seu desempenho, por parâmetros;
- Capacidade contínua de detetar anomalias, para diminuir o erro no modelo.

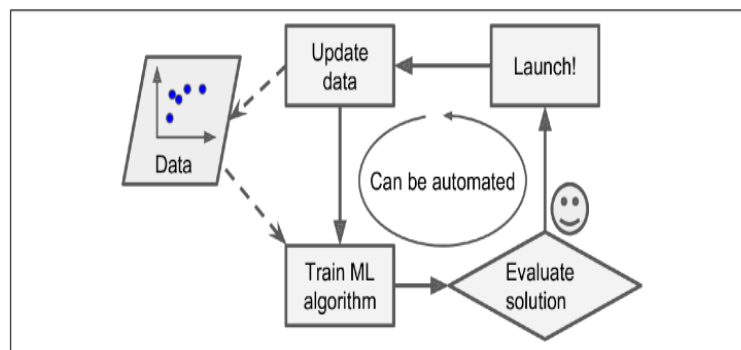


Figura 4 - Processo genérico de Machine Learning. Fonte: (Raschka e Mirjalili, 2017).

O processamento dos dados é certamente algo crucial para o funcionamento do modelo, de forma a poder selecionar e extrair as características relevantes. Podendo melhor representar os dados, ao transformar os “*inputs*” em um menor número de

dimensões e diminuindo a sua complexidade(Alpaydin, 2014). Para esse efeito, é necessário ter em conta o dilema de “*bias/variance*”.

No ML, vai se verificar a existência de noções predeterminadas presentes nos dados. No “*dataset*” de treino, será necessário escolher um que generalize os dados dando um melhor enquadramento a esses, ou seja, diminuindo o “*sampling bias*”(Raschka e Mirjalili, 2017). No entanto, existam várias formas de “*bias*”, mesmo do próprio utilizador ao detetar padrões na amostra, “*inductive bias*”. Assim estará a definir os resultados para uma certa hipótese, e, ao mesmo tempo, a generalizar a amostragem, com o intuito de a simplificar(Mitchell, 1997).

Por outra instância, a “*variance*” representa a distribuição da amostra e a sua variabilidade. Onde será necessária uma quantidade mínima de dados, para melhor representar o padrão, tendo de se enquadrar, a variabilidade presente na amostra, para evitar que os resultados, tenham um erro baixo num “*dataset*” de treino, e ainda assim, não represente os vários dados em questão. Ao aplicar o teorema do Limite Central, implica que uma amostra tenha um número de dados significativo, poderemos considerar que a distribuição presente se aproximará de uma distribuição normal(Mitchell, 1997).

Mediante “*cross-validation*”, como dito anteriormente, é necessário separar a amostra em dados de teste e de treino, inferir hipóteses no treino, verificar qual é a hipótese com menor erro, comparando com o teste. (Alpaydin, 2014).

O equilíbrio entre o *bias* (viés) e a “*variance*”, vai contribuir para a construção de um modelo mais preciso, ao mesmo tempo, que se obtém uma amostra processada, com diminuição de características que estejam em excesso. Para esse efeito, uma técnica pertinente será a regularização, com utilidade para filtrar os dados, especialmente quando estes possuem várias características semelhantes, penalizando essa relação(Raschka e Mirjalili, 2017).

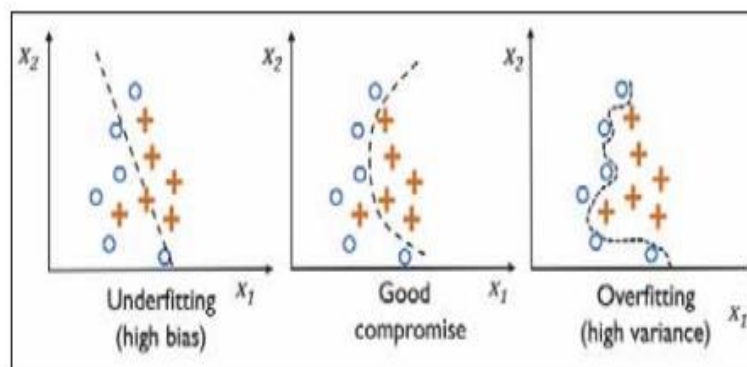


Figura 5 - Comparação entre uma decisão linear e não linear e um respetivo equilíbrio, relativamente ao modelo. Fonte:(Raschka e Mirjalili, 2017).

2.1 Sistemas de Machine Learning

A complexidade de um sistema de ML está normalmente associada ao modelo criado, como também nas etapas que suportaram a sua criação. O artigo (Benton, 2020), introduz o conceito aplicações inteligentes, esta abordagem difere ao considerar que as equipas responsáveis por estes, devem ter diferentes aptitudes, em termos de engenharias que possam ser vocacionadas para ML.

Esta visão multifacetada, permite abordar um sistema de ML, não só como uma fração isolada, mas como um sistema complexo ao incorporar as etapas associadas a ML, como mencionadas anteriormente, como também o sistema operativo. Ao requerer, a interação entre equipas. O mesmo sistema que, outrora, podia ser precário, eleva a sua robustez, tanto pela forma como a sua contínua monitorização é mantida, como também, pela sua integração no fluxo de processamento do *software*, o que acarreta um valor adicional, pela sua alta aplicabilidade, e como (Benton, 2020) refere, ML assegura a funcionalidade crítica do sistema.

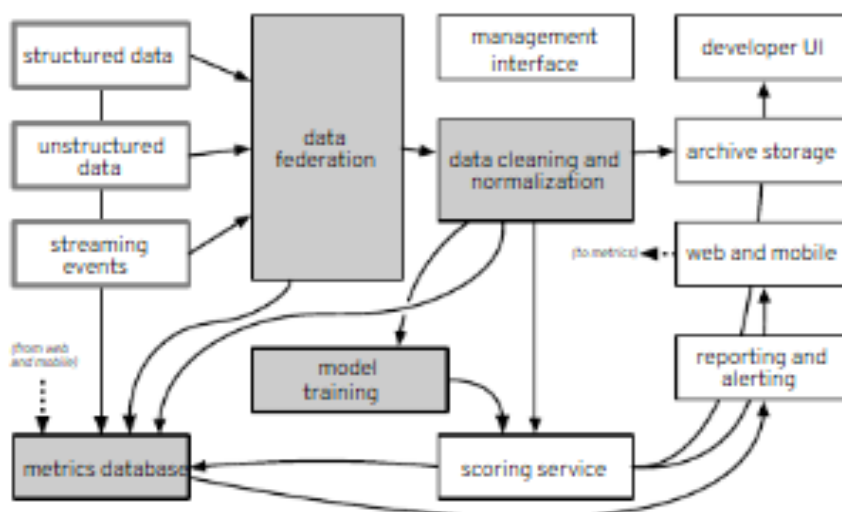


Figura 6 - A integração de ML, desde as suas etapas iniciais, que vai desde a obtenção de dados (data federation) até à criação do modelo, com a componente de software, com o propósito de garantir que o sistema esteja em níveis de desempenho adequados. Fonte: (Benton, 2020).

2.2 Métodos de Aprendizagem

Os métodos de aprendizagem utilizados em ML visam a utilização de um *dataset* de treino para detetar padrões existentes. Contudo, diferem no facto de possuírem, ou não, variáveis originadas pelo seu vetor de formação (Golden, 2001).

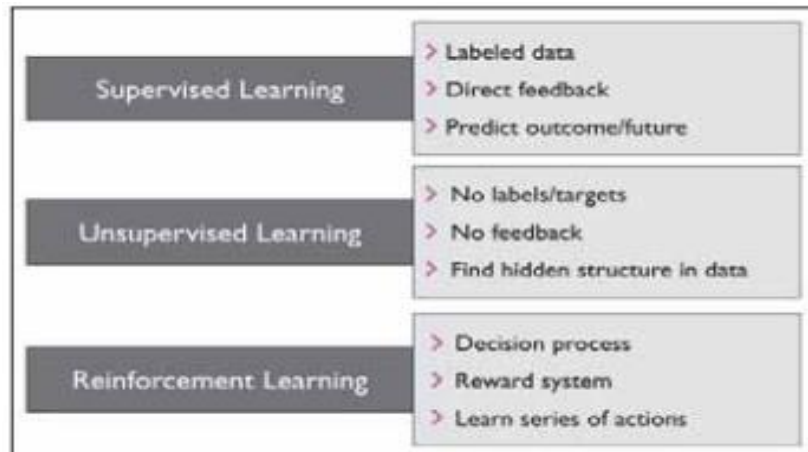


Figura 7 - Métodos de aprendizagem de Machine Learning. Fonte:(Raschka e Mirjalili, 2017).

2.2.1 Supervised learning

O método de aprendizagem *supervised learning*, implica o conhecimento de causa – efeito, contempla um conjunto de dados de entrada que leva a um determinado resultado. O intuito é formar um modelo capaz de prever novos dados, baseado em histórico, com associação aos respetivos “*labels*”. A formação de um modelo permitirá analisar e prever novas variáveis, e obter o respetivo *label* (Raschka e Mirjalili, 2017), como demonstra a **Figura 8**.

As variáveis previstas poderão ser dividida em dois grupos, em *regression task* ou *classification task*, conforme a função produzida. As variáveis de *regression task* são contínuas e as de *classification*, categóricas(Wang *et al.*, 2021).

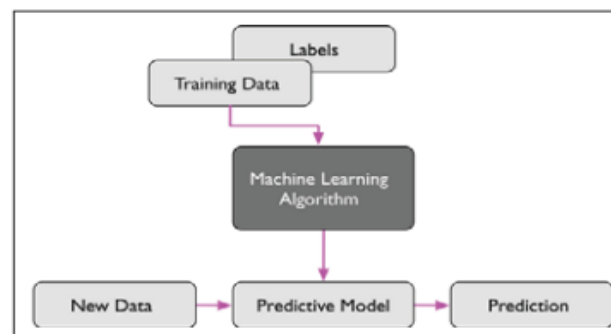


Figura 8 - Processo de supervised learning. Fonte:(Raschka e Mirjalili, 2017).

2.2.2 Unsupervised learning

No caso de *unsupervised learning*, apenas possuímos os dados de entrada, que ao contrário de *supervised learning*, não terá os correspondentes “*labels*” (consequências).

O processo passa por procurar características semelhantes nos dados, agrupando-os em função das suas características. (Alpaydin, 2014; Raschka e Mirjalili, 2017).

A informação presente nos dados pode ser detetada pelas seguintes técnicas:

- *Clustering;*
- *Principal component analysis (PCA);*
- *Factor Analysis;*
- *Density estimation;*
- *Dimensionality reduction.*

2.2.3 Reinforced learning

Por último, temos o *reinforced learning* onde o intuito será a criação de um agente que possa classificar, não a nível global, mas sim, a ação final e a sequência de escolhas que culminaram nesse fim. Este processo cogita revelar os melhores resultados, incentivado por um sistema de recompensas, ao otimizar continuamente o seu processo(Alpaydin, 2014; Bishop, 2006).

O processo visa a obtenção das etapas que levem a um certo objetivo, no estado em que este seja feito eficientemente. Ao ser caracterizado de uma forma positiva ou negativa, através do sistema de recompensas numérico presente. As etapas que resultem na ação pretendida, seriam classificadas como positivas, mas só na conclusão do processo(Mitchell, 1997).

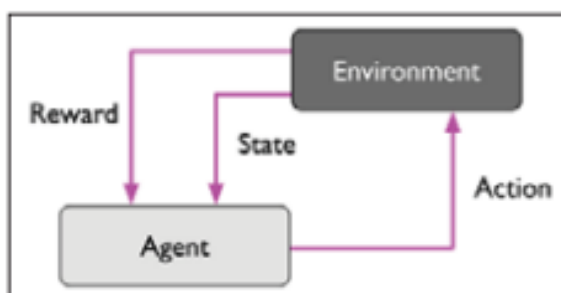


Figura 9 - Processo de “reinforced learning”. Fonte:(Raschka e Mirjalili, 2017).

2.3 Algoritmos utilizados para previsão de consumos de energia

A previsão do consumo energético por meio de ML e com base em(Han, Shu e Wang, 2023; Salam e Hibaoui, 2018), pode ser realizado com os seguintes algoritmos:

- *Linear Regression;*
- *K Nearest Neighbours;*
- *Random Forest;*
- *Decision Tree;*
- *Extreme Gradient Boost Regressor;*

- *Artificial Neural Network*;
- *Support Vector Machine*;

2.4 Aplicação no campo das Energias Renováveis

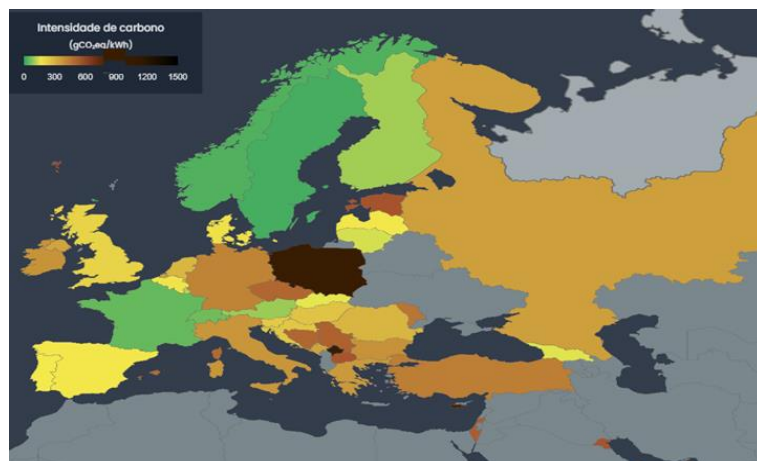
A utilização de ML no campo das RE, resume-se ao facto de ser uma ferramenta com inúmeras utilidades e certamente que pode ter um enorme contributo, numa área que se encontra em desenvolvimento, dando robustez para acelerar a sua implementação. O seu potencial de utilização estende-se a outros domínios da sociedade, desde, por exemplo, saúde e marketing(Allal *et al.*, 2024).

Com uma possível utilização de ML, na prevenção e deteção de anomalias/avarias, que tanto possam ser causados no sistema elétrico, pelo ambiente ou a nível estrutural(Kaplan, Tehrani e Jamshidi, 2021; Ren *et al.*, 2024), procura-se uma melhor previsão do consumo e da produção de eletricidade(Alazemi, Darwish e Radi, 2024; Ledmaoui *et al.*, 2023). Por outro lado, a possibilidade de averiguar o potencial de um sistema híbrido renovável, a nível ambiental e económico, ao considerar o LCOE e as emissões de CO₂ tendo em conta 4 combinações diferentes(Roy *et al.*, 2024).

Ao estudar variáveis, que possuam algum grau de incerteza, desde a irradiação solar e a velocidade do vento, avaliar edifícios com o propósito de determinar a sua eficiência energética, sendo uma solução aos problemas que causam um constrangimento para o funcionamento da rede e que afetam a relação de consumidor-cliente(Tharani *et al.*, 2020).

A análise e previsão no setor das RE, é essencial ao diminuir a variabilidade presente nos dados, ao ser empregue inúmeras tecnologias, neste campo. Este problema técnico retarda a implementação de uma maior capacidade instalada de renováveis, e a eventual transformação de um perfil energético que dependa em hidrocarbonetos, em um com menor intensidade de produção de CO₂ e com um eventual maior *share* de RE.

O caso de Portugal é certamente um exemplo, no que toca à transformação do seu perfil de produção de eletricidade, ao ter uma emissão de carbono de 153 gCO₂eq por kWh, em 2023, o que representa uma redução de 68%, comparativamente a 2017 (*Electricity Maps - Portugal - 2023 - Carbon Intensity Data*, 2024).



Mapa 2 - Intensidade de emissão de dióxido de carbono na produção de eletricidade, em 2023.
 Fonte: (Electricity Maps - Europe- 2023 - Carbon Intensity Data, 2024).

2.4.1 Mercado de eletricidade

A previsão de carga é essencial para o funcionamento do mercado de eletricidade, em principal o STLF (*Short Term Load Forecasting*) para o mercado diário. O STLF caracteriza-se por efetuar a previsão de carga, num período temporal reduzido, desde minutos até a uma semana.

No mercado diário, até às 12:00 horas, do fuso horário CET (*Central European Time*), é efetuada a comercialização de eletricidade, para o dia seguinte. A finalidade é estabelecer os preços de eletricidade, ao encontrar uma convergência entre a procura e a produção, **Figura 10**, tendo como referência os valores do mercado de compra e venda, para esse efeito. Contudo, os métodos de previsão empregados, possuem um erro associado ao não captar as relações não lineares nos dados (Loizidis *et al.*, 2023).

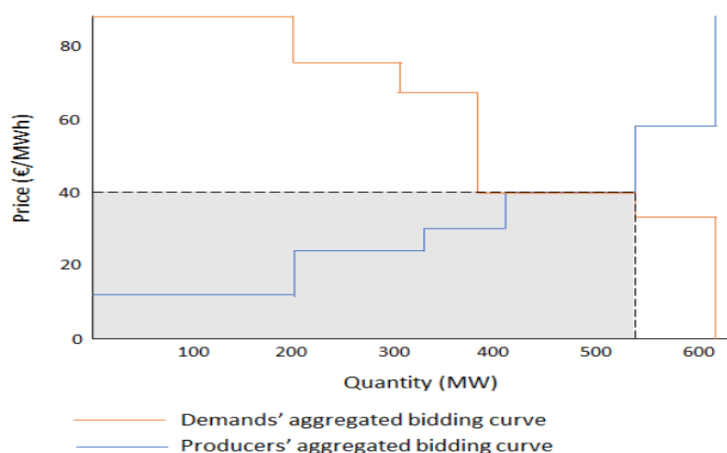


Figura 10 - Exemplificação de como é determinado preço de mercado do dia seguinte, resultante do cruzamento entre a produção e a procura. Fonte: (Pinhão, Fonseca e Covas, 2022).

Para esse efeito, tem-se verificado que o preço *spot* no MIBEL (Mercado Ibérico de Eletricidade) como constam no **Figura 11**, atingiram, no dia 3 de abril, entre as 13 e 18

horas, valor zero, o que resultou de uma maior oferta do que procura, em conjunto do modo de operação do mercado, algo ainda incomum no MIBEL, mas já recorrente no mercado europeu.

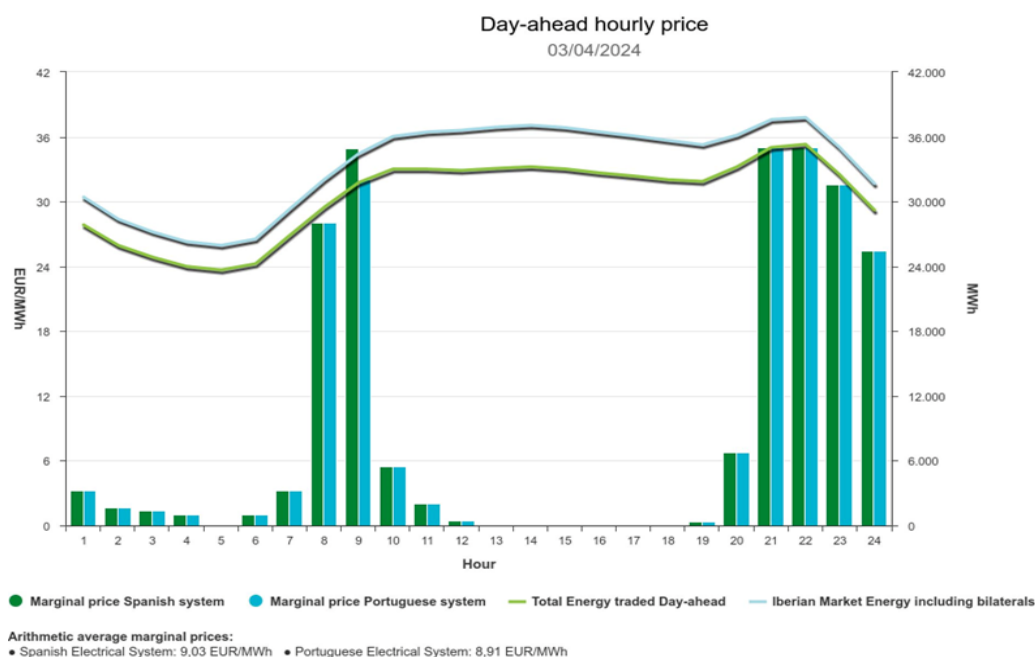


Figura 11 - Energia gerada no mercado MIBEL e o seu respetivo custo, no dia 3 de abril de 2024.
 Fonte: (Day-ahead hourly price | OMIE, [s.d.]).

Na **Figura 12** é apresentada a produção de eletricidade para o dia 3 de abril em Portugal Continental. A conjugação de:

- Consumo ser inferior à produção;
- As FIT (*Feed In Tariffs*), que garantem uma remuneração acima do preço ao produtor de RE, embora abolidas, as já estabelecidas contratualmente ainda vigoram, como para novos leilões de energia por parte do Governo de Portugal (Vitorino, 2022);
- Um sistema inflexível e sem capacidade suficiente de EES para responder ao fluxo de carga.

Têm sido dos principais contribuidores para os preços baixos que se verificam no MIBEL, em conjunto a elevada produção renovável.

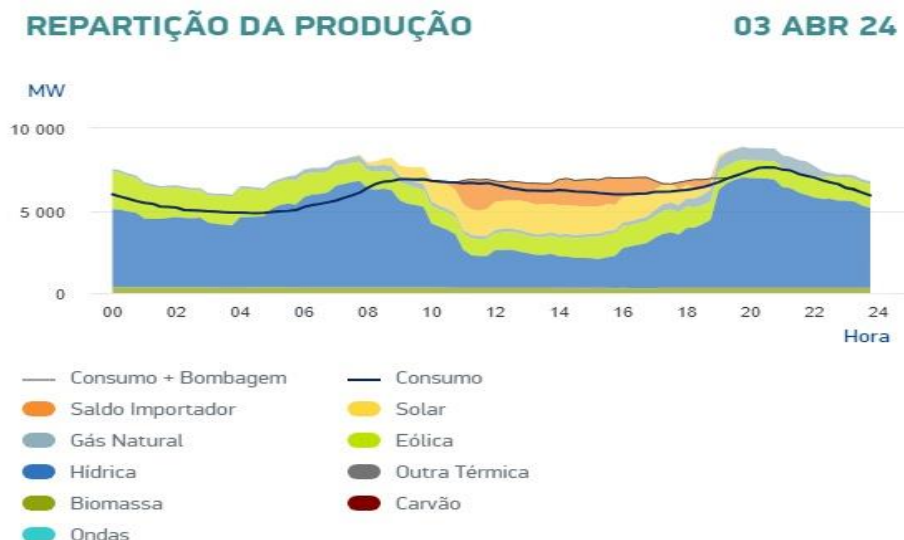


Figura 12 - Produção de eletricidade, por tipo, para Portugal Continental no dia 4 de abril de 2024. Fonte:(REN, 2024).

Se assumirmos que irá implicar uma relação causal de um menor preço da eletricidade, estaremos a entrar num paradoxo, pelo que o desvio previsto entre consumo e produção, reflete-se em taxas superiores para o consumidor(Vargas, 2024).

Os valores aplicados remetem para desincentivar a produção, no entanto, existindo mínimos impostos aos produtores e de que os de fonte renovável têm prioridade na ordem de venda, causa a desarmonização do mercado ao não refletir os custos marginais de operação(Atănăsoae, Pentiuc e Hopulele, 2020).

A necessária reestruturação do mercado, desde regulação existente e de leilão, para assegurar o incremento da capacidade instalada de renováveis, em cerca do triplo atual, assegurando que as medidas fiscais em apoio para o setor renovável não provoquem os preços negativos, garantido o retorno financeiro para os investidores(Prokhorov e Dreisbach, 2022).

Com a recente aprovação de medidas que visam a reformulação do mercado elétrico, com a diligência de PPA (*Power Purchase Agreements*) e CfD (*Contracts for Differences*), tendo como visão uma União Europeia, resiliente a crises e com uma menor volatilidade no setor, bem como, a erradicação da pobreza energética e proteção do consumidor(*Electricity market reform: Council signs off on updated rules*, 2024).

Conforme apresentado na **Figura 13**, a ERSE (Entidade Reguladora dos Serviços Energéticos), teve de preservar o CIEG (Custos de Interesse Económico Geral), com o aumento das TAR (Tarifa de Acesso às Redes). Pelo baixo preço do mercado grossista, em 2024, ser de 44,4 euros por MWh, comparativamente aos 88,3 euros por MWh estipulados pela ERSE. O que se denomina um défice tarifário, ampliada por ter sido calculada semestralmente, o que levou a ser calculada a cada 3 meses. Neste caso, a enorme divergência, exigiu a ação incisiva da ERSE, de forma a balancear o mercado, com a fixação excepcional das tarifas(ERSE, 2024).

euros por MWh	Varição mensal Jun 2024 / Mai 2024
MAT	+5,2
AT	+7,9
MT	+15,0
BTE	+28,3
BTN	+35,6

Figura 13 - Acréscimo verificado nas TAR de maio para junho de 2024. Fonte:(ERSE, 2024)

A existência de CAE (Contratos de Aquisição de Energia) e (PRG) produção com remuneração garantida, resulta em sobrecustos e o diferencial entre o preço grossista e o de retalho tenha de ser compensado, nas TAR e CIEG, **Figura 14**, independentemente da sua ou não divergência com a flutuação do preço de eletricidade, havendo essa relação inversa. Sendo que 60% da produção renovável, em 2023, seja de PRG(ERSE, 2024).

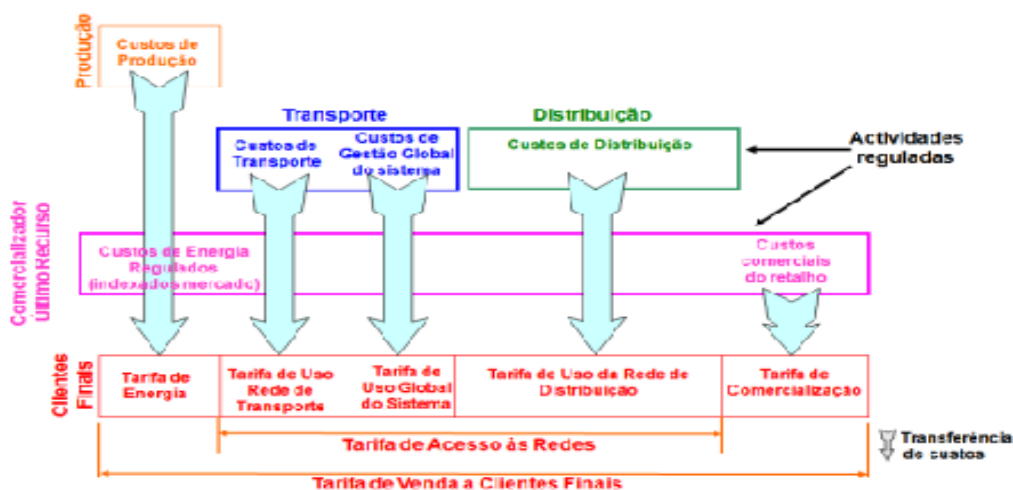


Figura 14 -Representação da repartição dos custos, aos quais são associados às TAR. Fonte:(João Lopes Veloso Ribeiro da Silva *et al.*, 2011).

Uma das formas imediatas de tornar o consumo mais adaptável, será através de EES (*Electrical energy storage*). O estudo (Keck *et al.*, 2019) analisa a viabilidade económica para a implementação de EES na Austrália num cenário de 100% de RE. Ao que concluiu, a redução de 70% de energia desperdiçada para 15 a 23%, em simultâneo, com a redução da capacidade instalada em 22%.

Os EES acoplados com métodos de *forecast* precisos, enfatiza uma melhor gestão de energia, ao identificar períodos vantajosos economicamente, tendo o benefício de evitar o seu carregamento em picos de carga, tendo um comportamento mais proativo no mercado, que em efeito reduzirá os custos de operação, pelo correto

dimensionamento dos EES, onde há uma relação direta do dimensionamento, com a precisão dos métodos de *forecast*, como relatado em (Emde *et al.*, 2023).

Ocorrer um desvio nos valores previstos e os ocorridos, *over-forecasting* e *under-forecasting*, e com base em (Ortega-Vazquez e Kirschen, 2006) afeta em particular a:

- *Start-up* do sistema de energia (Unidades de geração);
- A alocação da capacidade de transmissão;
- Energia não entregue.

Um menor erro de previsão de apenas no intervalo de 1 a 5% pode significar uma poupança nos custos de operação em tornos dos 0,36%(Ortega-Vazquez e Kirschen, 2006). Ao que se pode tomar as seguintes deduções:

- Ter uma previsão superior ao consumo, incute maiores custos económicos, atribuídos pela ativação de unidades de geração em excesso, embora tenha um menor risco de falha, ao providenciar um maior SR (*Spinning Reserve*) e aumentar consecutivamente a fiabilidade, o menor custo de energia não entregue não compensa esta situação.
- Ao que uma previsão inferior, aumenta o risco de falha e interrupção do sistema, pois as unidades de geração serão inferiores à carga, acentuado pelo baixo SR. Resultando a que a energia não entregue seja um custo mais proeminente nesta situação.

A aplicação de técnicas de ML, consolidará o setor energético, onde neste trabalho, já foram enumeradas as suas superioridades relativamente aos modelos em uso, podendo aperfeiçoar os modelos de previsão, viabilizando uma melhor gestão dos recursos energéticos. Tem ainda a capacidade de fundir, o mais próximo possível, os valores de produção e consumo, antecipando-se a qualquer desequilíbrio que se verifique(Jain e Mahajan, 2022), onde de 2021 para 2022, houve mais 21% de ocorrências na rede de distribuição que deturparam os serviços de fornecimento(DGEG, 2023).

2.4.2 *Grids* do futuro

Uma massificação de fontes de RE na rede, suscita questões sobre a sua operação. Ao serem não despacháveis, ou seja, estes geradores não conseguem ser interrompidos de acordo com o nível de carga, no exato momento. Com tendência a ocorrer *curtailment*, que é o impedimento da injeção da produção de RE, por exemplo, deficiências no design do mercado, um excesso de produção, a falta de EES ou de linhas para exportar, tendo a repercussão de aumentar os custos da integração de RE entre 10 a 20%, como apresentado em(Newbery, 2023).

Como (Liu *et al.*, 2023) referência, as RE podem suscitar a um impacto em várias camadas na rede, tais como:

- Subida da tensão;
- A introdução de bidirecionalidade, ao introduzir carga na rede não só a partir da central, mas sim do consumidor;
- Distorções harmónicas, pela baixa inércia inerente a RE, como observado em (Saha e Saleem, 2022), onde há uma correlação direta entre o aumento de RE e o decréscimo de inércia;

E entre outras, que originam pelo seu fluxo volátil e imprevisível, o que afeta diretamente a rede e a sua estabilidade, onde de apenas uma falha pode originar uma consecutiva aparição de falhas, denominado como falha em cascata.

A mitigação da ocorrência de falhas na rede, poderá ser pretensa na criação de modelos que captem as interdependências entre os sistemas de comunicação e elétrico, essencial para prevenir falhas de cascata, como (Filipe e Ferreira, 2018) identifica. Também apresenta medidas para mitigar as falhas da rede, que vão desde o aumento da capacidade das linhas de transmissão, a redistribuição de carga de áreas não críticas e a criação de secções isoladas.

A atual proteção na transmissão, faz com que não seja possível implementá-la sem fazer um *retool*, onde a atual proteção, não equaciona a bidirecionalidade, nem a distribuição da capacidade instalada, com um maior acesso distribuído. Para esse efeito, e com auxílio de FACTS (*Flexible AC Transmission Systems*), analisa-se a falha da rede por função de energia e não de corrente ou fase (Li *et al.*, 2012). Seja em qualquer modo em que se encontre, conectada à rede ou isolada.

A implementação de *microgrids*, contribui para a estabilidade da rede, mas também, para a sua expansão para áreas com acessos dificultados, com possibilidade de funcionar como uma *Virtual Power Plant* (João Lopes Veloso Ribeiro da Silva *et al.*, 2011). Sem a necessidade de construir novas linhas de transmissão, que aumentariam o nível de interdependências. Essencialmente, com uma melhor distribuição dos recursos energéticos (DER).

Como verificado em UPAC (Unidade de Produção para Autoconsumo) ou CER (Comunidade de Energia Renovável), onde a energia gerada é consumida no próprio local da instalação. Onde aliadas com EES, como (Battula, Vuddanti e Salkuti, 2021) lista, desde via térmica ou química, entre outros, diminui a variabilidade causada pelas fontes intermitentes, ao possibilitar um consumo posterior. A produção descentralizada, permite reduzir o estresse na rede pública nas horas de pico.

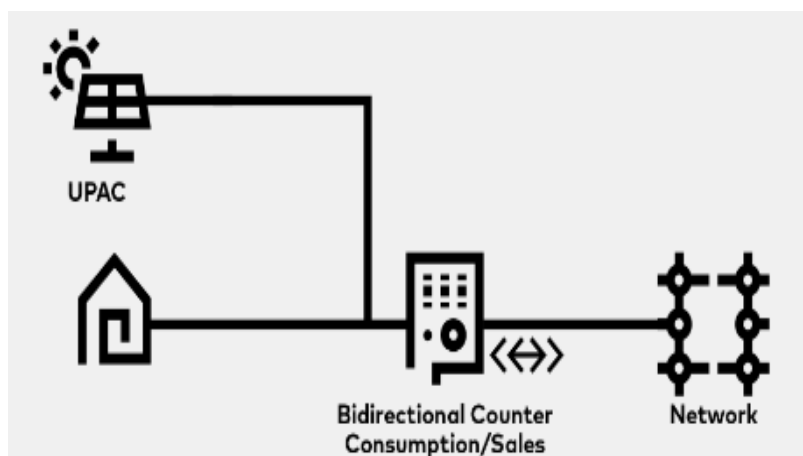


Figura 15 - Demonstração de um sistema UPAC, ligado tanto a uma infraestrutura doméstica como à rede pública, permitido pelo contador digital. Fonte: (E-REDES, [s.d.]).

O modo de funcionamento das *Smart grids* introduz o consumidor na equação de operação da rede. Com as capacidades analíticas, proporcionadas por contadores digitais, especialmente a tecnologia AMI (*Advanced Metering Infrastructure*).

A integração de sistemas de informação na rede, como também, sistemas de comunicação, caracteriza-se por uma constante corrente de informações e energia a fluir, **Figura 16**, que possibilita a tomada de decisões, tanto para o cliente e/ou operador, mais propicia a uma melhor gestão da comodidade que é a energia. O que será relevante para mudar comportamentos do consumidor, com o acréscimo de informações (Vitiello *et al.*, 2022).

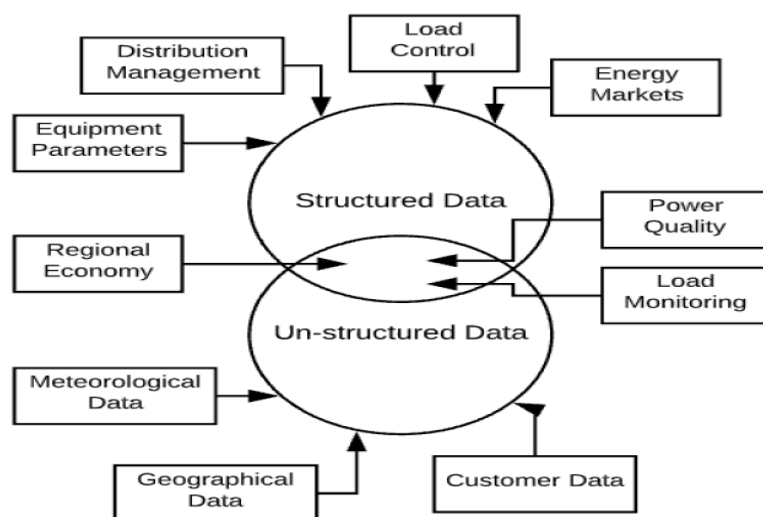


Figura 16 - Dados disponíveis na operação de Smart Grids. Fonte: (Kaplan, Tehrani e Jamshidi, 2021)

O estudo de (João Lopes Veloso Ribeiro da Silva *et al.*, 2011) apresenta um plano de negócio para o gradual processo de implementação de contadores digitais. Ao acrescentar valor pela maior poupança para o cliente e diminuição da imprevisibilidade do seu consumo. Atualmente, em Portugal, os contadores digitais

estão disponíveis para 94% dos clientes, estando quase implementados na sua totalidade(E-REDES, 2024).

Neste contexto, o ML desempenha um papel importante, ao assegurar que a relação entre consumidor e produtor esteja equiparada, assegurando um OPF (*Optimal Power Flow*)(Keyhani, 2012). E que os TSO (*Transmission System Operator*), sejam aptos a administrar a rede, consoante os dados entregues aos mesmos, identificando os preços marginais de operação. Ao diminuir as perdas nas redes de transporte e distribuição, que para Portugal Continental, no ano de 2024, se situam em 10,5%(ERSE, 2024).

Uma possibilidade seria a implementação de um mecanismo com base em *Demand Response*, onde Portugal poderá ter um sistema flexível, apto para 100% da sua capacidade ser de RE. Com vista, uma redução de custos por uma melhor gestão de energia, integrando simultaneamente as RE, com o almejo das metas de carbono. Com uma possível redução de 7% a 26% dos picos de consumo, na Europa(Anjo *et al.*, 2018).

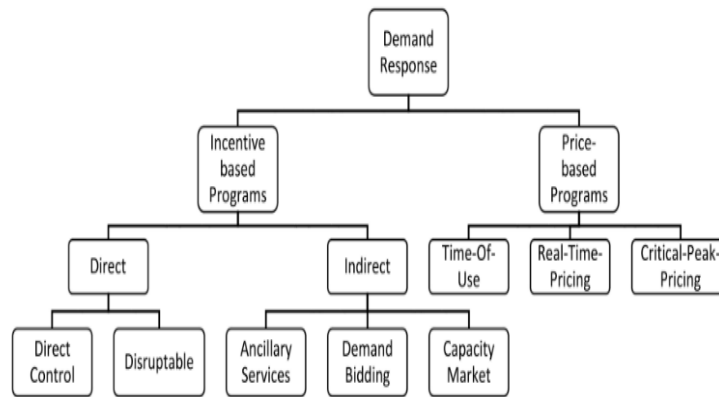


Figura 17 - A Demand Response pode-se caracterizar pela necessidade de uma tarifa dinâmica, Price-based Programs. Ou, por outro lado, pelo controlo de aparelhos, de forma direta, imposta pelo operador. Como também com a partição incentivada do cliente na manutenção da operação da rede. Fonte:(Alotaibi *et al.*, 2020).

A envolvimento do cliente, é feito de forma autónoma, possibilitado pelas ferramentas de comunicação inteligente, em termos pré-estabelecidos com o operador. O que coincide numa melhor gestão do consumo energético e flexibilização do mesmo. Podendo ser estruturado de uma forma económica ou hierárquica, tendo como premissa um eventual melhor uso da capacidade da rede e dos seus recursos energéticos de forma a corresponder à carga, em condições ideais. Com o acrescido benefício de diminuir a ocorrência de eventos que causem constrangimentos na rede(Battula, Vuddanti e Salkuti, 2021).

Para esse efeito e como se tem mencionando, relativamente aos próximos passos do setor elétrico e a sua consolidação, o cliente será requisitado a desempenhar um papel fundamental, em *simbiose* com o operador. Tendo o operador a capacidade de oferecer incentivos para o consumidor final alterar o seu consumo, para a situação mais oportuna. Tendo a finalidade de reduzir o congestionamento cíclico que ocorre nas horas de pico(Vaccaro *et al.*, [s.d.]).

O que poderá trazer em primeiro plano, com certo idealismo, não tão distante numa vida humana, que são as *Smart Cities*, onde, em (Swain, Salkuti e Swain, 2021), é apresentado um sistema descentralizado, com base em *prosumers* (Clientes que produzem e consomem energia). A combinação de vários sistemas tecnológicos, por exemplo:

- Redes de sensores sem fio que permitem o controlo dos dados;
- Tecnologia *Blockchain* que assegura a segurança do sistema e dos seus respetivos e eventuais ataques;
- Como a utilização de algoritmos que visam a otimização da fluidez do sistema energético e ir de encontro para a averiguação das estratégias ótimas, no menor tempo possível.

Este estudo encapsula, de uma certa forma, a revolução tecnológica que se verifica no setor energético. Numa premissa de que a distribuição seria feita de uma forma mais direta e delineada, de pessoa para pessoa (Abdella e Shuaib, 2018), num sistema elétrico *self sustainable*.

2.5 Python

O Python é uma das linguagens de programação mais utilizadas atualmente. No ranking de linguagens de programação da IEE, em 2023, o Python reforçou a sua liderança (Cass, 2023). A simplicidade desta linguagem, permite ao utilizador uma utilização mais intuitiva, que no campo científico é essencial, ao assegurar praticabilidade e um rápido entrosamento com a utilização do Python (Halvorsen, 2019).

Para o efeito de desenvolver projetos na área de investigação, o Python revela-se como ferramenta bastante preponderante, ao possuir vastas bibliotecas e pacotes, “*open-source*”, possuindo logo uma grande variabilidade de ferramentas, que se traduz em várias utilizações desde o tratamento de dados e a sua respetiva visualização, scripts, desenvolvimento Web e entre outros (Müller e Guido, 2017).

Como discutido anteriormente, acerca da necessidade de processamento dos dados, Python é bastante proficiente ao possuir vários instrumentos para esse efeito, dando a possibilidade de um rápido tratamento dos dados, a sua análise e a sua respetiva visualização.

Em termos de bibliotecas existentes em Python utilizadas no campo de ML temos:

- Scikit-learn;
- Keras;
- SciPy;
- TensorFlow.

No caso de pacotes disponíveis têm-se as seguintes possibilidades:

- Pandas (para análise de dados);
- Matplotlib (representação de dados);
- SciPy (computação científica);
- Numpy (computação científica).

O que leva a esta ser linguagem de código selecionada para este projeto, pela capacidade analítica que requer o tema.

3. Metodologia

3.1 Recolha de dados

A obtenção dos dados via API (*Application programming interfaces*) *call*, mais propriamente para as plataformas da ENTSO-E *Transparency* (*European Network of Transmission System Operators for Electricity*) e da *OpenWeather*, remete-se ao facto que estas duas fontes serem extramente fidedignas, o que se traduz em: dados mais fiáveis, precisos e recentes, sendo crucial para os algoritmos, que requerem dados robustos para obter uma maior interpretabilidade dos resultados.

A relevância dessas qualidades, é crucial para o presente estudo, permitindo que as relações formadas dos resultados sejam confiáveis, como também, possam elevar a sua relevância com uso de dados recentes, com o propósito de ir em concordância com os objetivos do projeto, sendo coerente para esses pontos.

Através do uso dos API presentes no repositório de (Friederich, 2017), e com um ajuste na escolha da plataforma meteorológica, pela não disponibilidade da que se situava no código, *Dark Sky*, ficheiro JSON (JavaScript Object Notation), armazenava os *Tokens Ids*, como os *Endpoints* que foram requisitados em cada plataforma. Também possibilita a seleção das estações meteorológicas e dos feriados.

3.2 Pré-processamento dos dados

Após a obtenção dos dados, e o seu posterior armazenamento em ficheiros individuais do tipo CSV (*Comma-Separated Values*), foi necessário agregar os parâmetros meteorológicos, não por estação, mas sim por hora, ao calcular a sua média. Foram removidos os valores identificados como *Nan* e 0,0.

As introduções de variáveis temporais de hora, do dia e dia da semana, visam a identificação de padrões de carga consoante esses dois critérios. Via *Feature engineering*, adicionou-se um vetor que caracterizava a variável vento em sentido e velocidade, como as funções sinusoidais e cosseno da variável em estudo, culminando, no enriquecimento dos dados que serão utilizados nos algoritmos.

De forma a tornar a utilização dos dados mais acessível entre algoritmo, foi agrupado pelas suas características desde:

- *Weather* para os dados extraídos da *OpenWeather*;
- *Calendar* para os parâmetros adicionais temporais;
- *Actual* para os dados obtidos da ENTSO-E;
- *All* que agregava os 3 grupos.

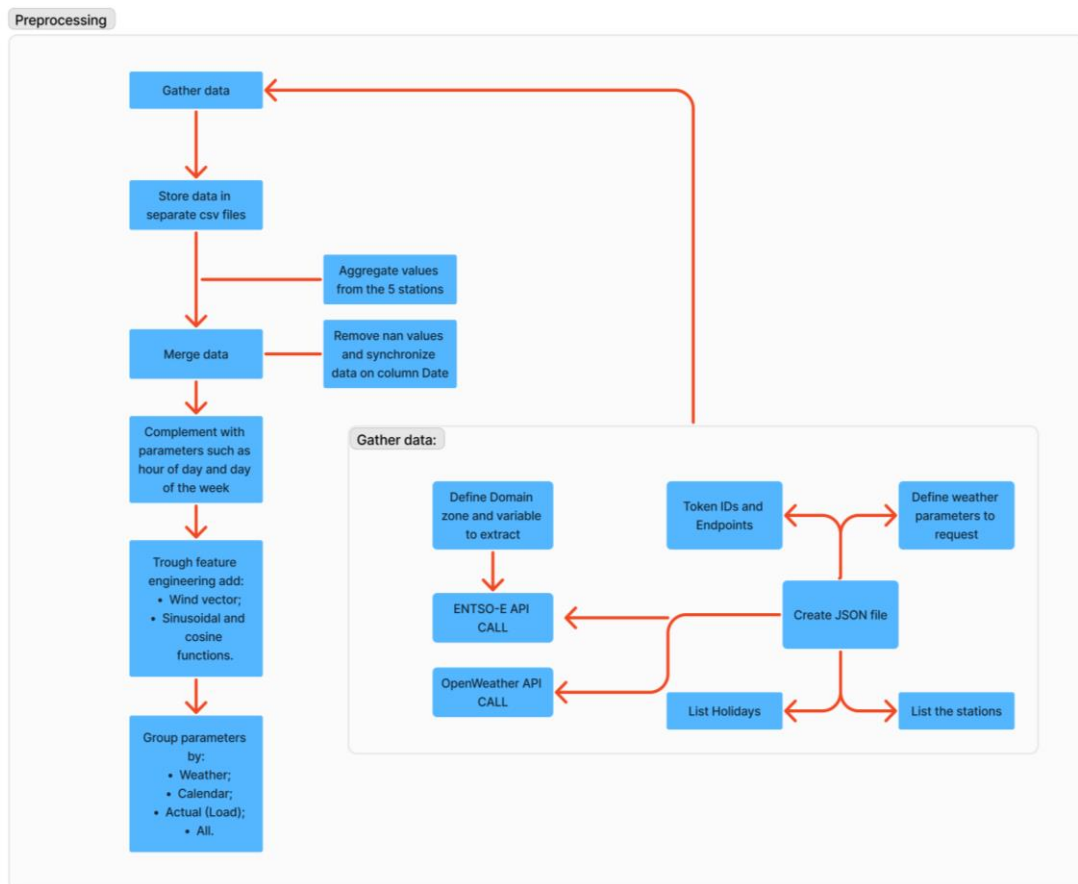


Figura 18 - Etapas que descrevem o passo inicial de recolha de dados bem como o seu processamento consecutivo.

3.3 Métodos de análise

A análise desenvolvida ao longo deste projeto foi realizada no IDE (*Integrated Development Environment*) *Spyder*, que possibilitou a utilização de Python como linguagem de programação.

No decorrer dessa análise, foram utilizados vários pacotes e bibliotecas com diferentes propósitos. A extração de dados requereu o uso conjunto de *requests* e *BeautifulSoup*, que através da sua junção permite a solicitação via HTTP (*Hypertext Transfer Protocol*) de informação e o seu posterior *parsing*.

Em termos de tratamentos dos dados temos:

- *Pandas*, que permite a manipulação dos dados, especialmente do género série temporal.
- *NumPy*, possui várias ferramentas matemáticas, sendo aplicado, por exemplo, na etapa de normalização dos dados, no cálculo de estatísticas para identificar outliers e entre outros.

- *Matplotlib*, permitiu ao longo do projeto a representação gráfica de forma acessível dos dados, como demonstrado nos Capítulos 4 e 5.
- *Seaborn*, teve a mesma funcionalidade que o *Matplotlib*, sendo aplicado na visualização do mapa de correlação e da ordenação hierárquica por importância das variáveis, **Figura 21 e 22**.
- *scikit-learn*, esta biblioteca possui várias ferramentas proíficas, como as requeridas neste projeto desde: o *MinMaxScaler* para normalizar os dados, a importação do modelo de *RandomForest* para determinar a importâncias das variáveis até às métricas do subcapítulo 3.10.
- *os*, *csv*, com o efeito de guardar os ficheiros gerados num diretório específico.
- *Json*, para a utilização de ficheiros em formato JSON.
- *DateFormatter* e *DayLocator*, têm a particularidade de realizar o ajuste dos eixos dos gráficos.

A construção do modelo a partir de LSTM, requereu o uso do *Tensorflow*, o qual é uma biblioteca direcionada para ML, especialmente para redes neuronais. O seu uso conjunto com *Keras*, associado a *deep learning*, têm a premissa de criar e modelar o modelo, para posterior previsão de carga.

3.4 Caso em estudo

Os algoritmos de ML referidos serão utilizados para modelar a carga para Portugal, mais propriamente, será testado a sua capacidade de interpretar os dados e de os prever com precisão, ao mesmo tempo, introduzindo parâmetros meteorológicos podendo melhor caracterizar a variável em estudo, no sentido de verificar comportamentos subjacentes entre a sua relação.

Ainda, será explorado a previsão de valores fora da amostra temporal, num período de 5 dias.

3.5 Licenças

Os usos transformativos, de forma legal e ética, dos dados meteorológicos do *OpenWeather* estão certificados pela *ODbL (Open Data Commons Open Database License)*. Para a plataforma ENTSO-E os dados de carga, estão assegurados pela *CC-BY 4.0 (Creative Commons Attribution 4.0 International License)*. Ao seguir os termos impostos, é dado o devido crédito aos agregadores dos dados e assegurando a integridade do estudo.

3.6 Métricas

Para averiguar a credibilidade dos resultados obtidos que incidiram sobre o teste, foram utilizadas as seguintes métricas demonstradas matematicamente em (Abad *et al.*, 2020; Chicco, Warrens e Jurman, 2021):

- *Mean Absolute Error* (MAE) indica o erro absoluto.

Equação 1

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

- *Mean Absolute Percentage Error* (MAPE) indica a percentagem do erro em questão.

Equação 2

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{x_i - x_f}{x_i} \right|}{n} \quad (2)$$

- *Coefficient of Determination* (R^2) indica a variância entre a variável em estudo e os valores previstos.

Equação 3

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - x_f)^2}{\sum_{i=1}^n (x_i - x_f)^2} \quad (3)$$

- *Root mean square error* (RMSE) através da raiz quadrada do valor diferencial entre o valor factual e previsto, ao quadrado., indica a precisão do modelo.

Equação 4

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - x_f)^2}{n}} \quad (4)$$

Onde x_i será o valor que servirá como a variável de controlo, que será os valores factuais de carga e x_f o valor previsto pelo modelo, com n o número de previsões.

4. Análise de Consumo de Energia

Os *datasets* que serão explorados neste projeto, serão relativos a Portugal. Em concreto, a carga (MW) total e os seguintes parâmetros meteorológicos:

- Temperatura e o seu respetivo mínimo e máximo (K);
- Pressão(hPa);
- Humidade (%);
- Velocidade do vento(m/s) e a sua direção(°);
- Nebulosidade (%);
- Volume de precipitação(mm).

Tabela 1 - Primeira entrada dos parâmetros meteorológicos para as respetivas estações.

Date	Station	Temperature	Min Temperature	Max Temperature	Pressure	Humidity	Wind Speed	Wind Degrees	Cloudiness	Rain Volume
01/01/2024 00:00	Braga	282.11	280.63	283.72	1021	89	1	121	100	0
	Coimbra	281.81	281.81	281.81	1021	87	0.74	96	100	0
	Faro	288	286.03	288.13	1020	77	3.09	40	20	0
	Lisbon	284.88	282.16	286.08	1020	66	1.79	225	40	0
	Porto	283.52	282.15	283.77	1021	89	3.09	130	75	0

Ambos *datasets* encontram-se na mesma escala de tempo, em concreto horária, relativo ao período de 1 de janeiro a 16 de abril, 2024.

Para ter uma melhor representatividade das condições meteorológicas, os dados foram obtidos em função de 5 localidades, Braga, Coimbra, Faro, Lisboa e Porto. Consecutivamente, estes foram estabelecidos, tendo em conta um valor médio.

O processamento, em termos de *NaN* e de valores “*none*”, 0,0, procedeu-se com a remoção dessas linhas. Com a respetiva utilização da função *dropna()* e pela filtragem de linhas onde se localizava 0,0. O que fez com que se removesse 229 entradas *NaN*, relacionados com valores meteorológicos, e 469 linhas na coluna de Nebulosidade, pela presença de valores iguais a 0,0.

Consecutivamente, como o volume de precipitação, não reunia nenhum valor, removeu-se este de posterior consideração. Como também, converteram-se os parâmetros de mínimo e máximo de temperatura em um, transformando-os num parâmetro que indicava a sua respetiva média por hora.

Para estabelecer parâmetros, a partir de engenharia de recursos, que capturem melhor a relação com a carga, converteu-se as duas características do vento num vetor, que indicava o seu sentido e velocidade. E analisando a **Figura 19**, o vento, predominantemente, tem direção norte, com uma velocidade de 2 metros por segundo.

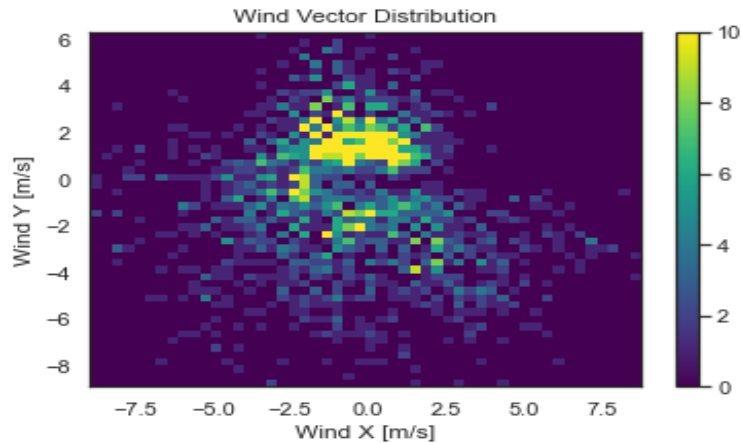


Figura 19 - Distribuição da ocorrência do vetor vento, pela escala visual e a sua respetiva direção com indicação da velocidade.

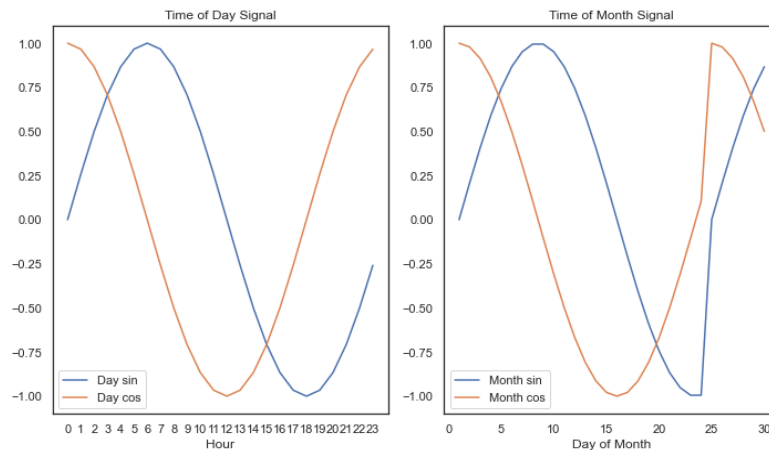


Figura 20 - Representação da função sinusoidal e função cosseno, em termos de hora e dia. A função sinusoidal retrata a ocorrência de picos de carga.

Na totalidade, o *dataframe* final (junção entre os dois datasets, **Tabela 2**), depois do seu processamento, registou 1.800 observações e 15 parâmetros. Estes mesmos parâmetros, que foram selecionados com os métodos aplicados descritos a seguir.

Tabela 2 - A Tabela 2 demonstra as cinco primeiro entradas dos do dataframe final, ou seja, os dados utilizados para prever a carga.

Date	Actual_Load	Temperature	Mean Temperature	Pressure	Humidity	Wx	Wy	Cloudiness	Weekday	Hour	Holiday	Day sin	Day cos	Month sin	Month cos
01/01/2024 00:00	5135	283.52	282.96	1021	87	-0.92	1.53	75	0	0	1	0	1	0	1
01/01/2024 01:00	4962	282.66	282.61	1021	89	-0.08	0.67	40	0	1	0	0.26	0.97	0.21	0.98
01/01/2024 02:00	4684	282.08	281.80	1020	91	-0.99	1.18	20	0	2	0	0.5	0.87	0.41	0.91
01/01/2024 03:00	4414	282.09	281.50	1020	92	0.29	1.28	59	0	3	0	0.71	0.71	0.59	0.81
01/01/2024 04:00	4248	281.82	282.00	1021	92	0.18	1.13	30	0	4	0	0.87	0.5	0.74	0.67

Com o propósito de capturar a variação da carga de acordo com eventos de tempo, introduziu-se as variáveis “Hour”, “Weekday” e “Holiday”.

De forma a verificar os parâmetros mais pertinentes para o estudo da variável carga, realizou-se, mediante um modelo de *Random Forest*, a quantificação dos

restantes parâmetros, em termos de importância. O modelo pode ser matematicamente definido como(Parashar, 2020):

Equação 5

$$RF_{fi} = \frac{\sum_{j \in \text{all trees}} \text{normfi}_{ij}}{T} \quad (5)$$

Para esse efeito, normfi_{ij} representa a importância normalizada da característica i na árvore j , com T a indicar o número total de árvores.

Ao apurar os parâmetros mais pertinentes para o estudo da previsão da variável carga, verificou-se que “Holiday” é insignificante, por isso não foi tida em consideração no estudo realizado. A razão por esta exclusão, resume-se ao simples facto de ocorrer apenas 4 feriados nacionais, no período em questão.

No inverso, a variável “Month cos”, é o parâmetro com maior relevância, com 0,28, seguido de “Hour” e “Weekday” com 0,25, e 0,16, respetivamente. Assim, o parâmetro “Month cos”, afirma-se como o parâmetro mais relevante para o estudo, em termos de potencial para previsão de carga.

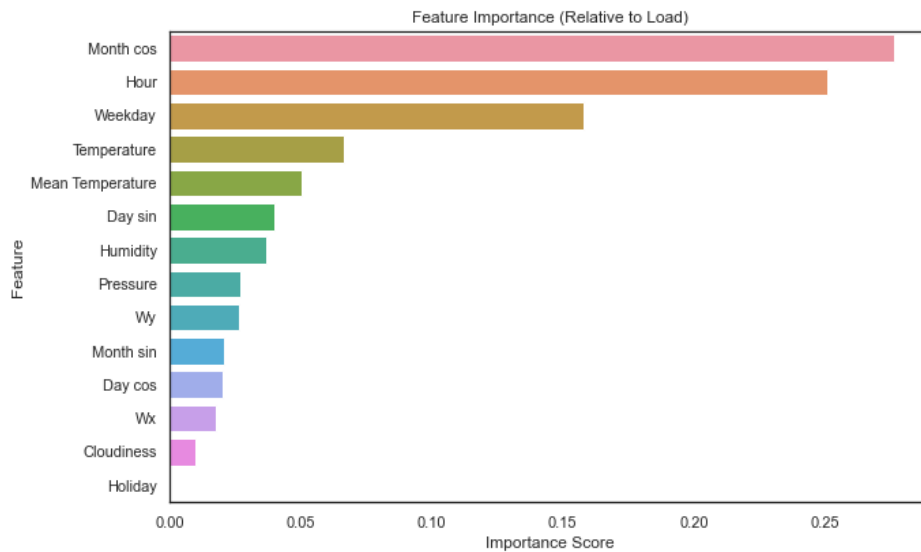


Figura 21 - Estudo do nível de impacto que as variáveis têm sobre a previsão da variável carga, por um modelo de Random Forest.

A propósito de evitar multicolinearidade, projetou-se um mapa de calor com base no método de Pearson, **Figura 22**, que traduz a interdependência entre variáveis, ou seja, a sua semelhança. De forma a reduzir a redundância dos parâmetros e de descomplicar o treino do modelo. Podendo ser representado pela seguinte equação(University, 2023):

Equação 6

$$r = \frac{\sum_{i=1}^n (x_i - x)(y_i - y)}{\sqrt{\sum_{i=1}^n (x_i - x)^2} \sqrt{\sum_{i=1}^n (y_i - y)^2}} \quad (6)$$

Nesse sentido, apurou-se que as variáveis “Temperature”, “Max Temperature” e “Min Temperature”, apresentavam um alto valor de correlação. Como mencionado anteriormente, “Max Temperature” e “Min Temperature” foram transformadas num novo parâmetro “Mean Temperature”.

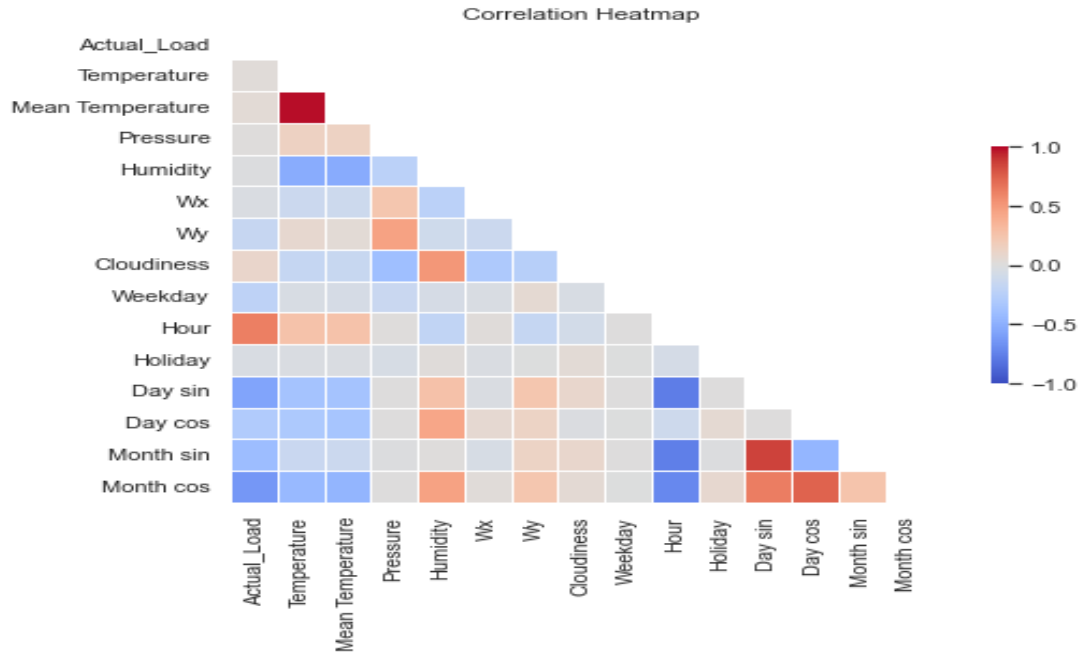


Figura 22 - Estudo da correlação entre variáveis, pelo método de Pearson, com a sua máxima semelhança representado a vermelho e o inverso a azul. Este sistema permite averiguar quais variáveis podem entrar em conflito pela sua semelhança.

4.1 Análise exploratória dos dados

Ao estudar a variável primária, carga, num período de hora não se consegue inferir nenhuma dedução pertinente, no entanto, ao alterar para uma escala diária é possível retirar 3 conclusões, tais que são: a carga desde 1 de janeiro tem diminuído, comparativamente aos seguintes meses, o que, inicialmente, poderá ser atribuído às subidas de temperatura e que os períodos de menor consumo vão incidir no período do fim de semana.

Estas afirmações tornam-se factos quando visualizamos as variáveis de carga e temperatura no mesmo plano, e quando consultamos a média de carga por dia de semana, como é possível observar na **Figura 23**.

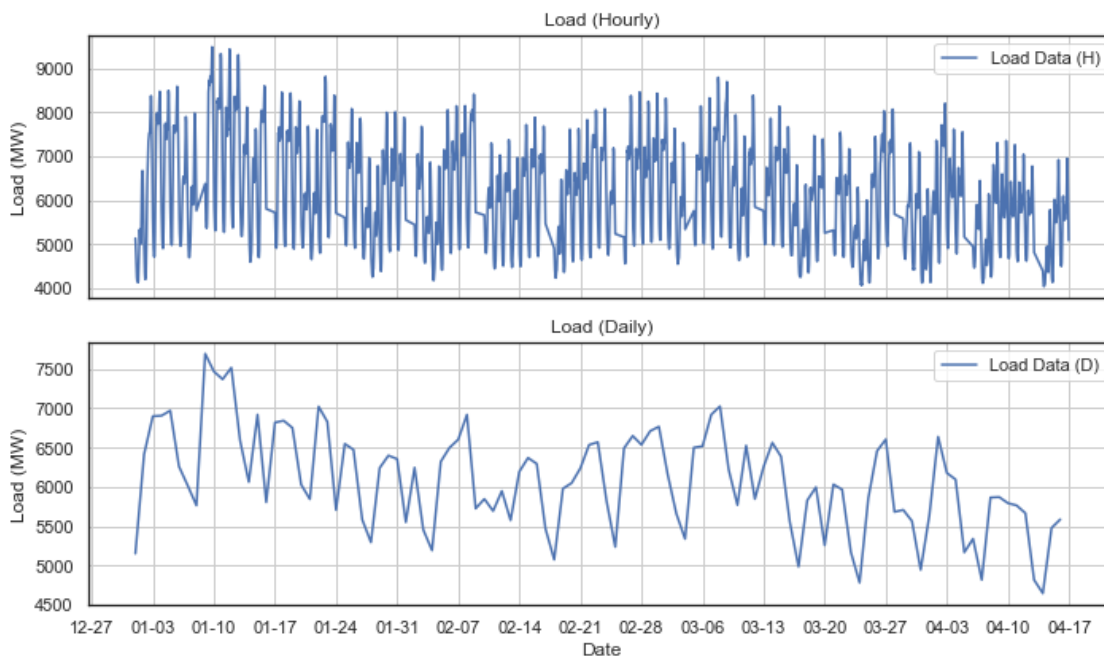


Figura 23 - Representação gráfica da variável carga, na totalidade. Em dois formatos por hora e dia.

A partir da subsequente **Figura 24**, os picos de maior consumo vão coincidir nos períodos de menor temperatura, e o oposto observa-se na cauda do gráfico, onde a partir de 15 °C, nota-se essa inversão, na relação entre temperatura e carga. O que se apura, mais evidentemente, nos períodos entre 13 a 27 de março e, mais recentemente, de 10 a 16 de abril.

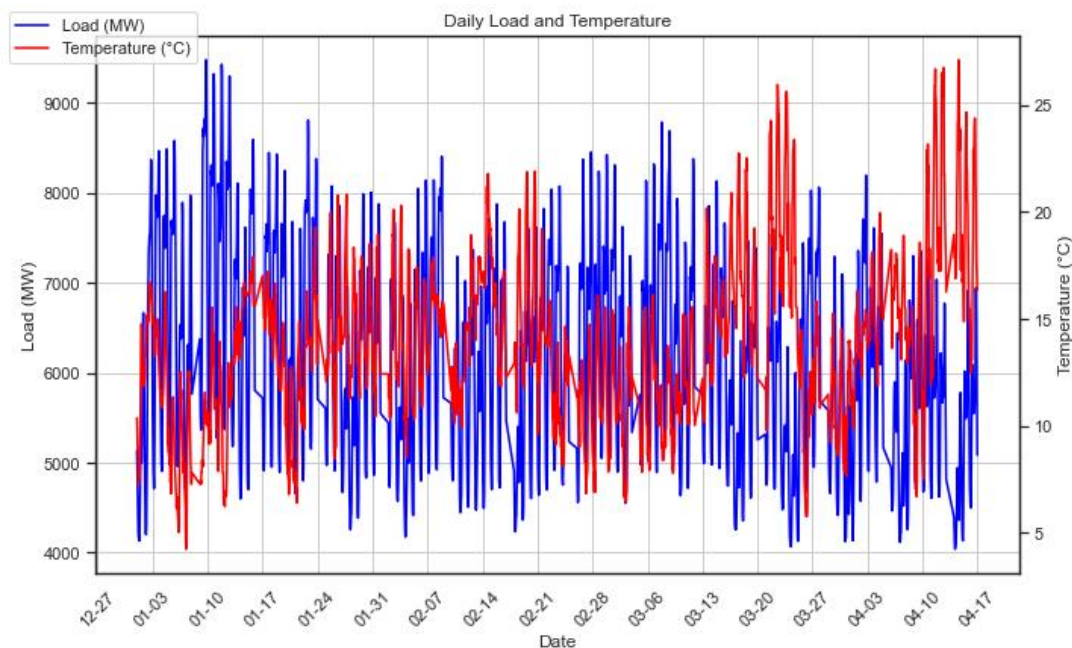


Figura 24 - Representação, simultânea, das variáveis carga e temperatura.

É possível apontar domingo, como o dia que ocorrerá o menor consumo e o fim de semana como o período onde ocorre consecutivamente baixos consumos. Havendo a

tendência de divergir por menos 1.000 MW, relativamente aos valores de segunda a sexta.

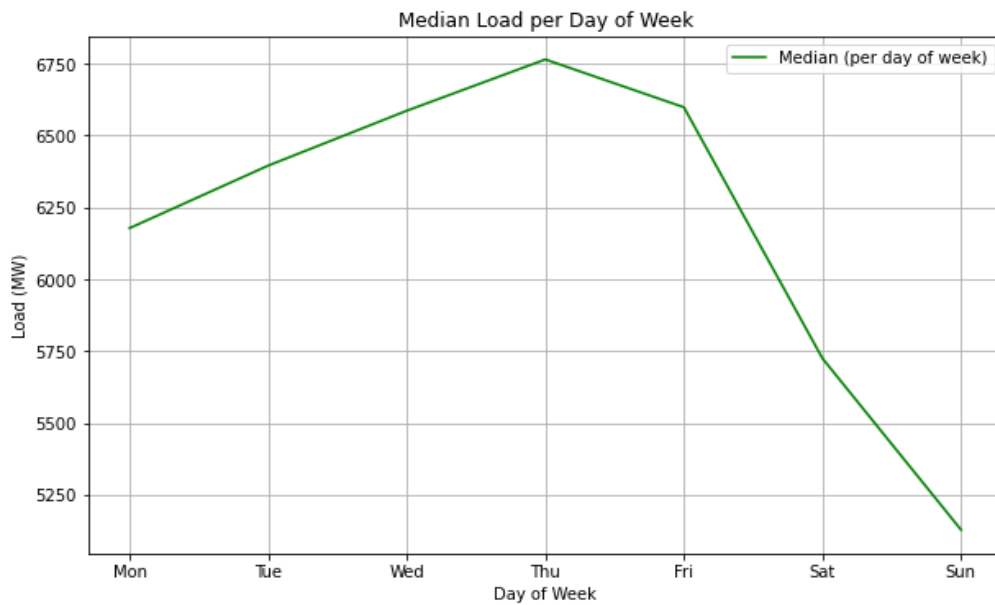


Figura 25 - Carga mediana por dia da semana.

Por meio de uma análise mais precisa, numa menor escala, de 24 horas, como é feita da carga na **Figura 26**, destaca-se a possibilidade de definir períodos horários em termos de consumo.

No período noturno, verifica-se que é onde ocorre o menor consumo, como esperado. A partir, relativamente, das 6 horas, descreve-se um período de crescimento, que estabiliza nas 2 horas seguintes. Esse mesmo consumo mantém-se estável durante o período das 8 às 17 horas, aproximadamente 6.500 MW. Posteriormente, às 19 horas, é onde se regista o valor mais elevado de consumo, um valor aproximado de 7.500 MW. Seguindo-se de um decréscimo, até ao intervalo onde ocorrerá o menor consumo.

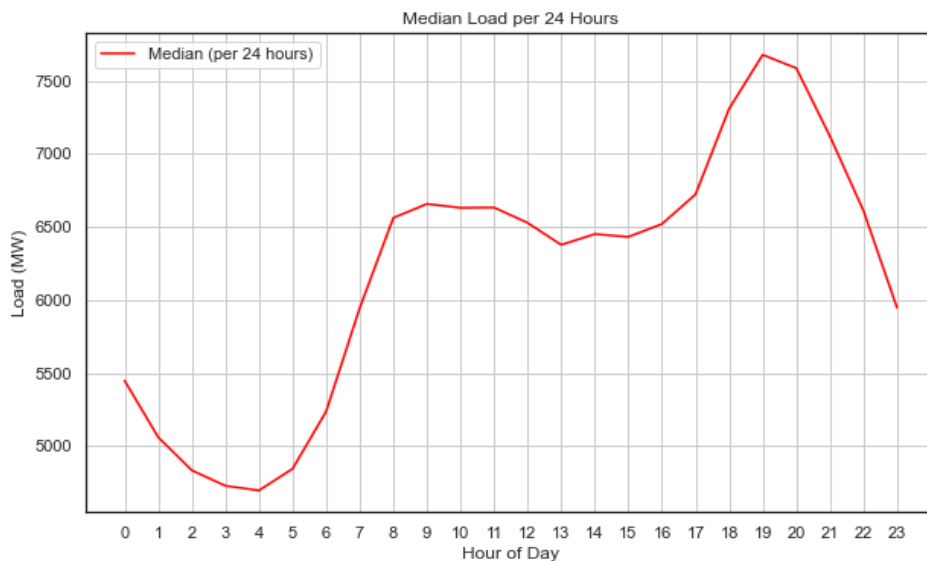


Figura 26 - Carga mediana por hora.

O uso de histogramas permite, de uma forma nítida, captar as tendências nos dados, especificamente para os presentes na **Figura 27**.

Os valores de temperatura, vão de acordo, com os esperados em Portugal Continental, para os 4 primeiros meses, concentrando-se no intervalo de 10 °C a 15 °C. Tendo uma baixa ocorrência para valores superiores a 15 °C.

Para o seguinte caso, a humidade, concentra-se na parte superior da sua escala, ocorrendo em mais frequência acima de 70% e antes desse valor, a frequência diminui acentuadamente.

Na pressão, os valores concentram-se, após 1.015 hPa e certamente, podemos afirmar a existência de *outliers* na zona dos 1.000 hPa.

O caso de nebulosidade, é certamente distinto dos outros ao concentrar em apenas 5 valores: 0%, 25%, 45%, 75% e em menor quantidade nos 100%.

A carga, também tem a sua distinção, ao ser a primeira variável onde não é possível inferir nenhuma tendência, onde nenhuma se sobrepõe a outra destacadamente, havendo uma certa simetria. Podemos afirmar que os valores se encontram no intervalo de 5.000 MW a 7.000 MW. Por estarmos num período transitório de inverno-primavera verifica-se a existência de alguns valores superiores, em torno dos 8.000 MW. Sendo os que se situam em 9.000 MW, têm uma baixa frequência, sendo possível ser considerados *outliers*.

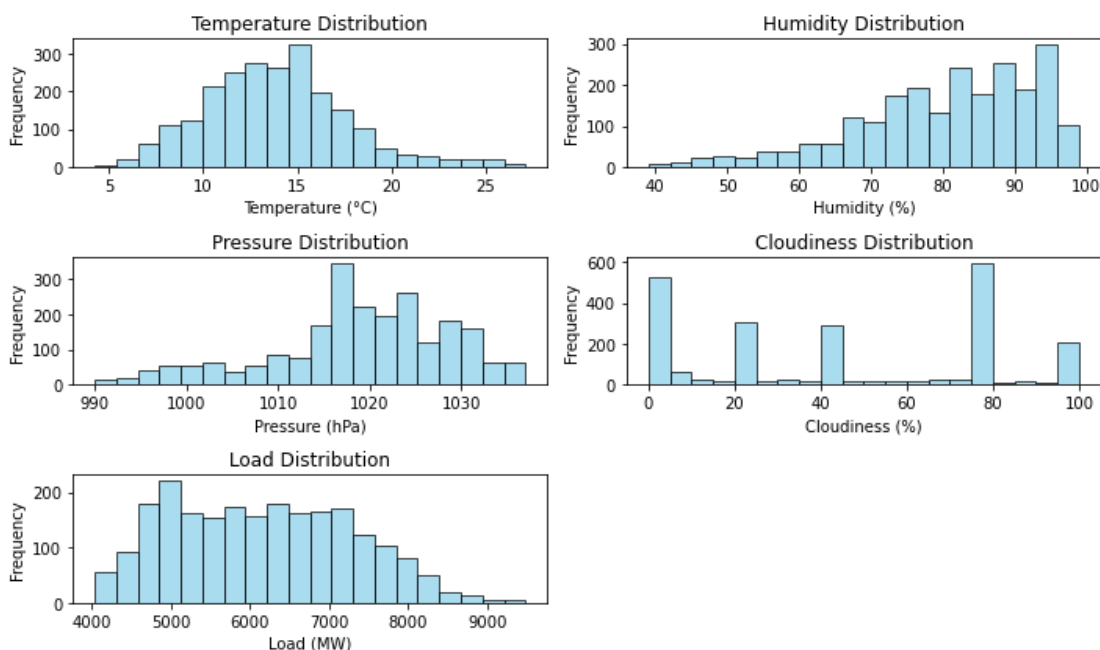


Figura 27 - Distribuição da frequência para as variáveis de temperatura, humidade, pressão, nebulosidade e carga.

A análise feita anteriormente, pode ser reforçada por meio de *box plots*. A centralidade, verifica-se claramente em qualquer dos dois tipos de gráficos temos:

- A temperatura a rondar os 10 °C a 15 °C;

- A humidade no intervalo de 70% a 90%;
- A concentração de pressão nos 1020 hPa;
- A enorme dispersão dos valores registados para nebulosidade, com os quartis em polos opostos;
- Os quartis, para o caso de carga, incidem perfeitamente no intervalo mencionado de 5.000 MW a 7.000 MW, com a sua mediana de 6.124 MW.

Em termos de *outliers*, que se encontram identificados a vermelho, e como dito anteriormente, a existência destes valores anormais para temperatura na faixa dos 25 °C. A humidade, pela sua concentração existente, de uma certa forma torna-se clarividente a existência de *outliers*, especificamente em torno dos 40%. O mesmo caso acontece para a pressão, pela sua distribuição, que evidencia uma baixa frequência até 1010 hPa. Por fim, carga, a análise feita em torno do histograma discorda em torno dos outliers, com a da box plot. Mas, certamente, seria mais ajustado afirmar que os valores que ocorrem em volta de 8.500 MW, são tendenciosamente valores fora da mediana, que devem ser considerados anormais.

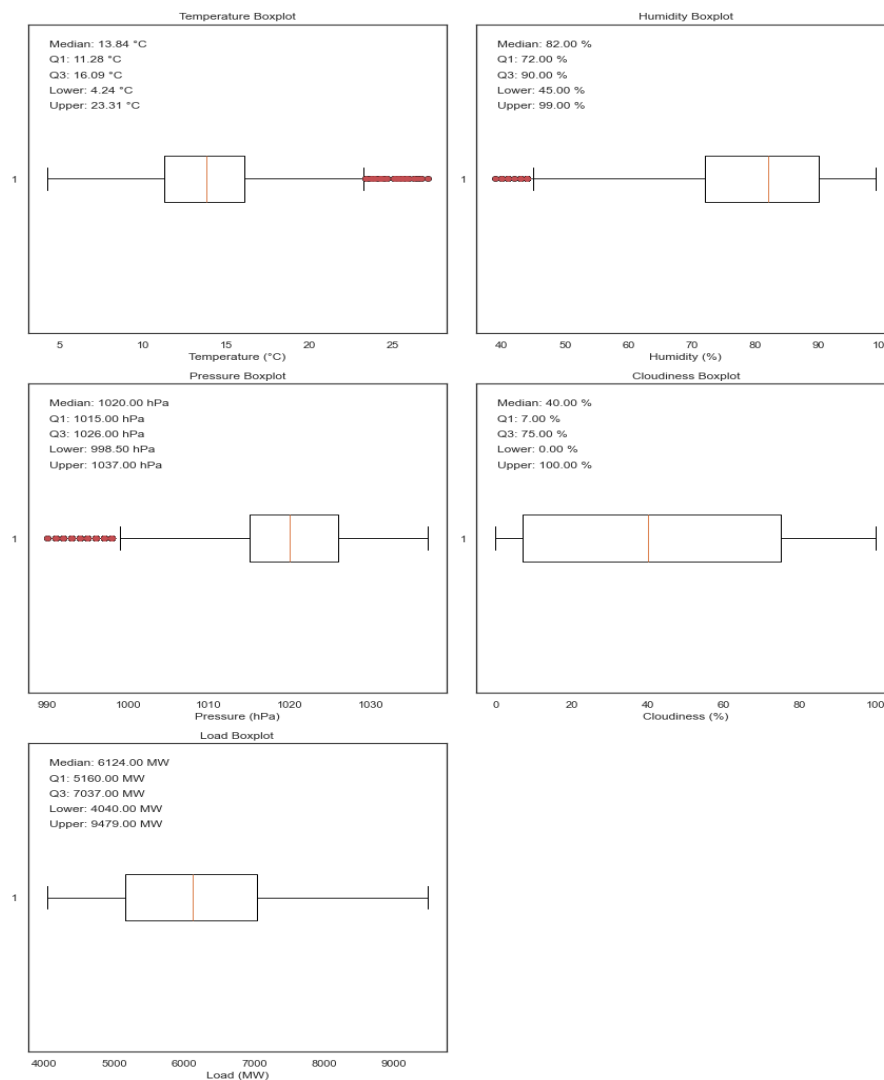


Figura 28 - Identificação da mediana, quartis, mínimo e máximo e outliers para as variáveis de temperatura, humidade, pressão, nebulosidade e carga. Pela sua representação em box plots.

5. Previsão de Consumos de Energia com Machine Learning

5.1 Metodologia dos Modelos

Modelo 1

O algoritmo do Modelo 1 é a variante sazonal do ARIMA (*AutoRegressive Integrated Moving Averages*) mais apropriado para analisar a periodicidade da variável carga, chamado SARIMA. Mas antes de descrever, em concreto, o conceito do SARIMA (*Seasonal AutoRegressive Integrated Moving Averages*), enquadremos o algoritmo ARIMA, um método linear. O próprio, ramifica-se em 3 componentes:

- Autorregressivo (AR): Esta componente representa a ligação entre uma observação e um certo número de observações anteriores. O parâmetro associado é p .
- Filtro de Integração (I): O componente I lida com a diferenciação da série temporal para torná-la estacionária. O parâmetro associado é d .
- Médias Móveis (MA): A componente de MA demonstra entre uma observação e os resíduos de erros de previsões anteriores. O parâmetro associado é q .

O ARIMA possui 3 parâmetros não sazonais, p , d e q . A **Tabela 3** estabelece a relação entre cada parâmetro, e o seu respetivo componente.

Tabela 3 - Parâmetros do algoritmo ARIMA.

Parâmetros	Conceito	Componente do ARIMA
p	Número de termos autorregressivos	AR
d	Número de diferenciações	I
q	Ordem de médias móveis	MA

O algoritmo SARIMA, possibilita a representação da periodicidade pela variável m . Sendo incluído mais 3 parâmetros, sazonais, P , D e Q .

Os algoritmos ARIMA e SARIMA podem ser representados, respetivamente, pelas seguintes equações (*From AR to SARIMAX: Mathematical Definitions of Time Series Models*, [s.d.]):

ARIMA

Equação 7

$$\begin{aligned}\Delta^d y_t &= \theta(L)^p \Delta^d y_t + \Phi(L)^q \Delta^d \epsilon_t + \Delta^d \epsilon_t \quad (7) \\ y_t &= \theta(L)^p y_t + \Phi(L)^q \epsilon_t + \epsilon_t \quad (7)\end{aligned}$$

SARIMA

Equação 8

$$\begin{aligned}\theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D y_t &= \Phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \epsilon_t \quad (8) \\ \theta(L)^p \theta(L^s)^P y_t &= \Phi(L)^q \phi(L^s)^Q \epsilon_t \quad (8)\end{aligned}$$

Com y_t a variável responsável por indicar os dados em uso, $(L)^p$ e $(L)^q$ os termos não sazonais e $(L^s)^P$ e $(L^s)^Q$ os termos sazonais, a representar a parte de AR e MA, respetivamente, com ϵ_t a indicar o erro. Assim, são representados os parâmetros não sazonais e sazonais. Os termos Δ^d e Δ_s^D , indicam as diferenciações da parte não sazonal e sazonal, respetivamente, mas podem ser excluídas. Pela série não ter sido diferenciada.

Antes de proceder com a previsão da carga, normalizou-se os dados numa escala de 0 a 1, por meio de *Min-Max Scaling*, representado matematicamente por:

Equação 9

$$X' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (9)$$

Tendo como x o valor da variável, e x_{max} , x_{min} o seu, respetivo máximo e mínimo. Assim enquadrando os dados na escala de 0 a 1.

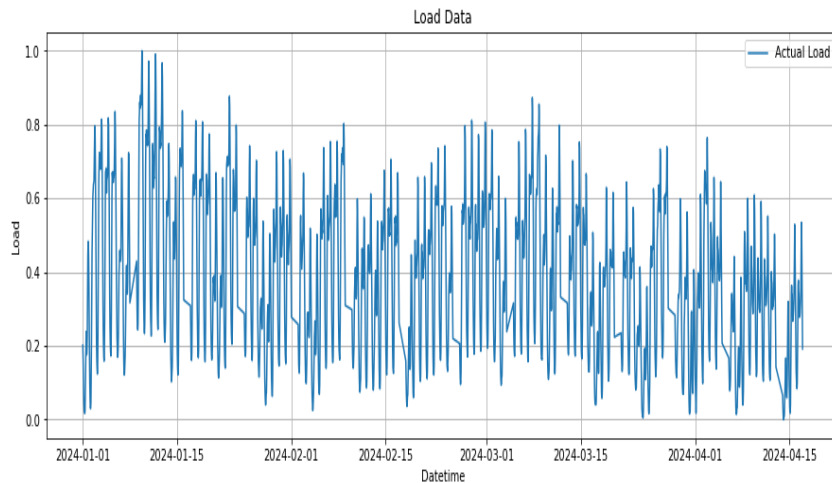


Figura 29 - Representação da variável Carga após a sua normalização para a escala de 0 a 1.

A normalização foi invertida, posteriormente, para verificar os valores, nas unidades iniciais. Com o uso da equação:

Equação 10

$$x = X' * (x_{max} - x_{min}) + x_{min} \quad (10)$$

É indispensável verificar se a série é estacionária, a mesma, foi verificada com auxílio do teste de *Dickey-Fuller Aumentado* (ADF). Tendo como representação a seguinte equação(ADF Test: *Augmented Dickey Fuller Equation*, 2023):

Equação 11

$$Y_t = \alpha Y_{t-1} + \sum_{k=1}^k \theta_k \Delta Y_{t-k} + \mu_t \quad (11)$$

O termo α é uma constante de Y que indica o atraso do aprendizado. A parte $\sum_{k=1}^k \theta_k \Delta Y_{t-k}$, os atrasos são representados por k e μ_t o termo de erro.

Ao que se apurou que a série era estacionária, pois, p não superou os limites críticos de 1%, 5% e 10%. Como também o resultado do teste ADF foi inferior a esses limites. O que confirma e reforçar a hipótese, de que a série é estacionária e não necessitará de diferenciação, sendo d/D considerado 0.

Tabela 4 - Resultados obtidos para o teste ADF, com o intuito de verificar a hipótese que a série seria estacionária.

Teste ADF	
Estatística ADF	-5,58
p	$1,38 * 10^{-6}$
1%	-3,43
5%	-2,86
10%	-2,56

Os dados foram repartidos em 80% para treino e os restantes 20% para teste.

De uma forma autónoma, para apurar os valores de cada parâmetro para a construção do modelo, foi apropriado o uso da função `auto_arima`, de `pmдарima`. O que complementado, com o Critério de Informação de Akaike (AIC) permitiu obter os parâmetros ideais, utilizados para a construção do modelo, com base na sua respetiva equação(Bevans, 2020):

Equação 12

$$AIC = 2k - 2 \ln(L) \quad (12)$$

O k representa o número de parâmetros no modelo e $\ln(L)$ a função *likelihood*(Misra e Han, 2020), que representa a probabilidade de o modelo produzir os dados iniciais, quantificando a capacidade de o modelo prever esses valores.

O menor valor registado de AIC, $-7.972,060$, corresponde ao modelo com melhor desempenho, o que resultou na seguinte combinação de parâmetros:

$$(p, d, q)(P, D, Q)[m]$$

↓

$$(3, 0, 0)(2, 0, 1)[24]$$

O algoritmo SARIMA, tem uma metodologia onde para prever $t=N+1$, recorre a todos os dados (N) que o precedem, como demonstrado na **Figura 31**.

Inicialmente, através dos dados de treino, previram-se os respetivos valores dos dados de teste. Isto, até apurar todos os valores existentes nos dados de teste e os seus respetivos Intervalos de confiança.

Para a previsão *out of sample*, a previsão como incide sobre valores não apurados, a sua previsão é feita através da utilização simultânea dos conjuntos de treino e teste.

Para ambos os casos, é importante realçar que se um valor for previsto, este será imediatamente adicionado aos inputs, para prever o conseqüente valor.

A posterior avaliação do modelo recorreu às métricas de MAE, MAPE, RMSE e R^2 , indicando a precisão sobre as previsões efetuadas sobre a variável em estudo. Todas as etapas estão representadas na **Figura 30**.

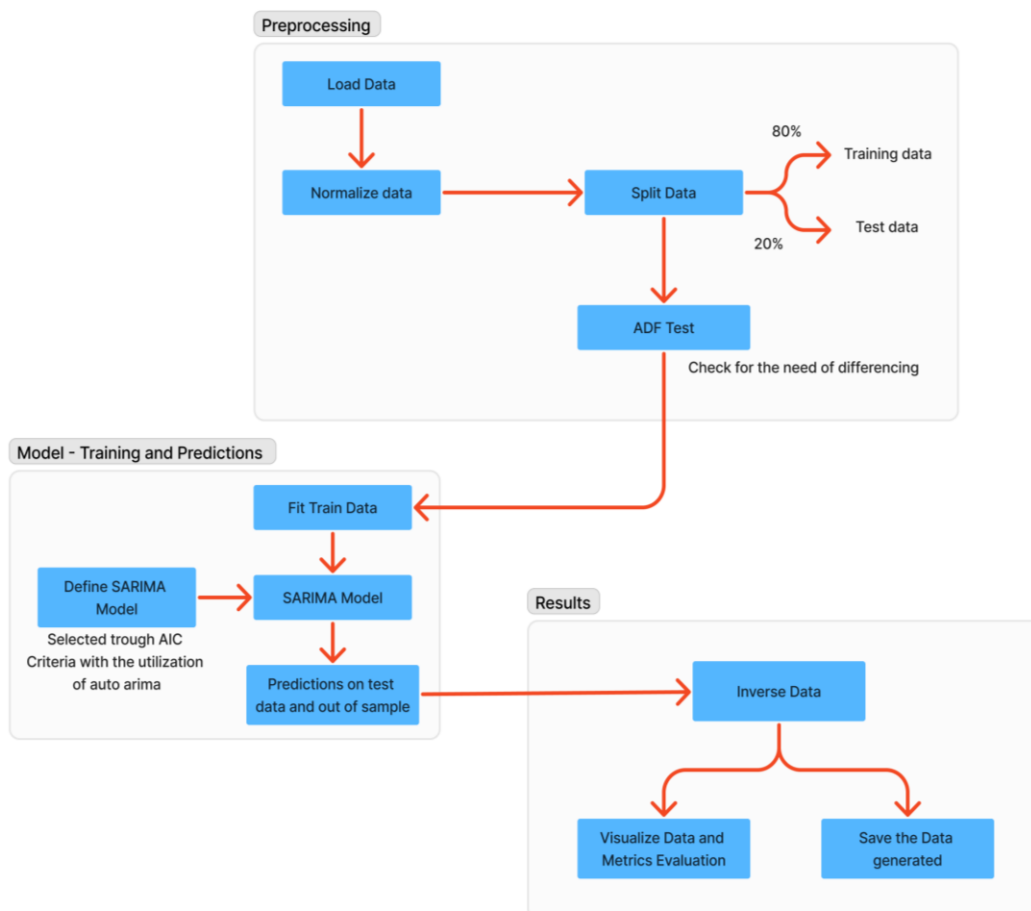


Figura 30 - Diagrama que enquadra a metodologia aplicada na construção do modelo baseado no algoritmo SARIMA.

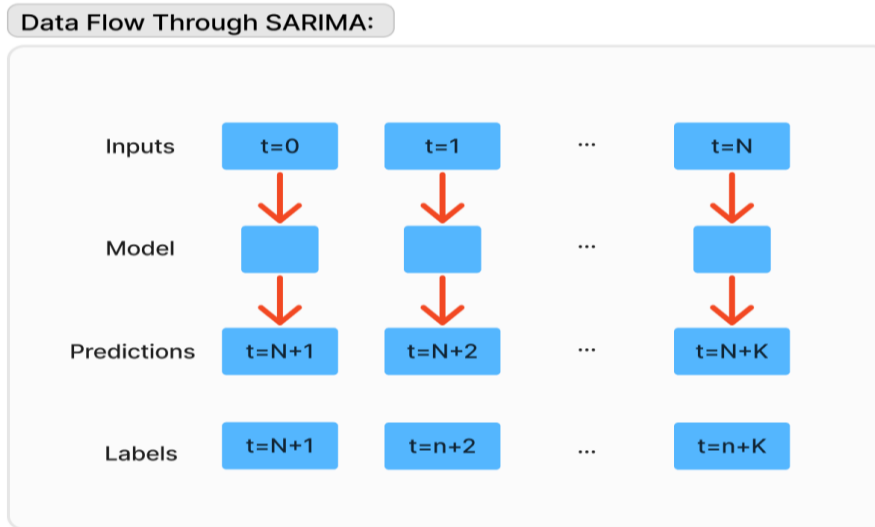


Figura 31 - A figura tem como interesse a exemplificação do processo de previsão e como este é efetuado. Onde os Inputs, através de t=N, representam os dados e a sua respetiva entrada. Que serão processados pelo modelo de forma a fazer as previsões e comparar aos Labels a que correspondem aos valores reais.

Modelo 2

O segundo modelo, surge no seguimento do Modelo 1, trata-se do SARIMAX (*Seasonal Autoregressive Integrated Moving Average Exogenous*), que se diferencia ao introduzir variáveis exógenas, variáveis que servirão como um auxílio na previsão da carga, tais como representadas na **Figura 33**.

O algoritmo representa-se, na sua generalidade, por (*From AR to SARIMAX: Mathematical Definitions of Time Series Models*, [s.d.]):

Equação 13

$$\theta(L)^p(L^s)^P\Delta^d\Delta_s^D y_t = \Phi(L)^q\phi(L^s)^Q\Delta^d\Delta_s^D\epsilon_t + \sum_{i=1}^n \beta_i x_t^i \quad (13)$$

$$\theta(L)^p(L^s)^P y_t = \Phi(L)^q\phi(L^s)^Q\epsilon_t + \sum_{i=1}^n \beta_i x_t^i \quad (13)$$

Ao contrário das equações dos algoritmos ARIMA e SARIMA, as variáveis exógenas são somadas pelo termo $\sum_{i=1}^n \beta_i x_t^i$.

A metodologia da construção do modelo anterior, foi semelhante à aplicada neste como revela a **Figura 32**. Onde os parâmetros foram definidos como:

$$(p, d, q)(P, D, Q)[m]$$

↓

$$(3, 0, 0)(2, 0, 1)[24]$$

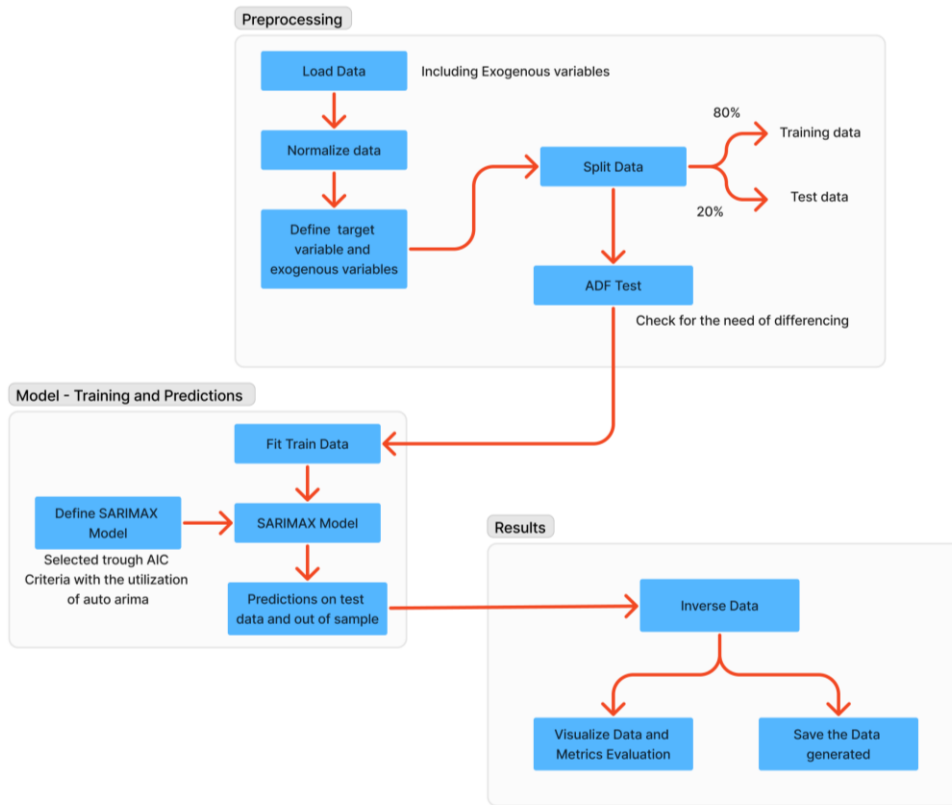


Figura 32 - Diagrama que enquadra a metodologia aplicada na construção do modelo baseado no algoritmo SARIMAX.

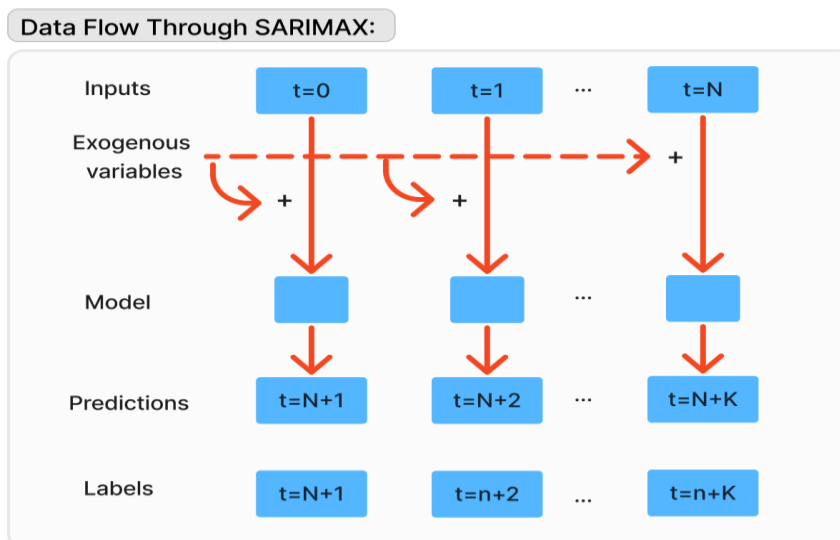


Figura 33 - A figura tem como interesse a exemplificação do processo de previsão e como este é efetuado para o Modelo 2. Onde os Inputs, através de t=N, representam os dados e a sua respetiva

entrada. Que serão processados pelo modelo, em conjunto com as variáveis exógenas que auxiliarão o processo das previsões. Com os respectivos Labels, que armazenam os valores reais.

Modelo 3

O algoritmo empregado neste caso, é um modelo de *Deep Learning* (redes neuronais profundas), denominado de *Long Short-Term Memory* (LSTM), do grupo de RNN (*Recurrent Neural Network*).

Um RNN, possui uma rede neuronal que possibilita a comunicação entre neurónios e unidades de processamento. As ligações, existentes na rede neuronal, estão associadas a um peso, de forma intuitiva, adaptadas à aprendizagem, sendo possível o seu ajuste, consoante a troca de sinal entre as camadas de saída e entrada e o respetivo gradiente de erro (Muhuri *et al.*, 2020).

A RRN pode ser representados pelas seguintes equações (Muhuri *et al.*, 2020):

Equação 14

$$h_t = \sigma (W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (14)$$

$$y_t = W_{hy}h_t + b_y \quad (14)$$

O sinal de entrada, a sequência, representada por x , h as “*hidden layers*”, com os seus respetivos pesos dados por W_{xh} e W_{hh} e b como o termo de *viés*. É acompanhado por uma função de ativação não linear, σ . A saída de uma rede RNN, y_t , acompanha-se do peso para o passo oculto e o termo de *viés*.

O que infere ao modelo e dota-o da capacidade de armazenar, num maior número, os dados anteriores, e, ao mesmo tempo, equacionando os padrões e relações existentes, num maior período de retenção. Como dito em (Rajasekhar, Hemanth e SK, 2023), os modelos tradicionais são omissos na sua capacidade de aprendizagem em termos autossuperação, e é onde LSTM preenche essa lacuna ao diminuir o erro do gradiente.

A célula de LSTM pode ser repartida em 3 portas tais que: *Forget gate*, *Input gate* e *Output gate*. Representadas, na mesma ordem de enumeração, matematicamente por (Alsabban, Salem e Malik, 2021):

Equação 15

$$f_t = \sigma (W_f * |h_{t-1} \oplus x_t| + b_f) \quad (15)$$

Equação 16

$$i_t = \sigma (W_i * |h_{t-1} \oplus x_t| + b_i) \quad (16)$$

Equação 17

$$o_t = \sigma (W_o * |h_{t-1} \oplus x_t| + b_o) \quad (17)$$

Onde σ é a função sigmoide, onde W corresponde à matriz de pesos da respetiva porta, $|h_{t-1} \oplus x_t|$ representa a concatenação do passo oculto e da entrada.

O *Input gate*, servirá como operador de controlo, ao determinar que informações adicionais são pertinentes ou não, representado por uma função tangente hiperbólica:

Equação 18

$$b_t = \tanh(W_C * |h_{t-1} \oplus x_t| + b_C) \quad (18)$$

O *Forget gate* tem a responsabilidade verificar quais informações devem passar para a seguinte etapa da célula expresso por:

Equação 19

$$C_t = f_t * C_{t-1} + i_t * b_t \quad (19)$$

Que se pode separar em duas partes: as informações retidas $f_t * C_{t-1}$ e as adicionadas $i_t * b_t$.

A camada oculta, que armazena as informações do passo anterior. A célula diverge desse pelo facto, de agregar as informações das várias etapas. A atualização da camada oculta dá-se pela equação:

Equação 20

$$h_t = o_t * \tanh(C_t) \quad (20)$$

Tendo o_t como o vetor de ativação da porta e $\tanh(C_t)$ a função tangente hiperbólica que incide sobre o estado da célula.

A **Figura 34** encapsula o processo de aprendizagem, a nível da célula, com o controlo da informação perante 3 portas, e pelos 2 estados que na figura estão como *Cell state* e *Hidden state*, ou, estado da célula e camada oculta. Tendo o estado da célula a informação, de uma forma temporal mais prolongada, enquanto, que na camada oculta esse tempo seja de uma menor quantidade.

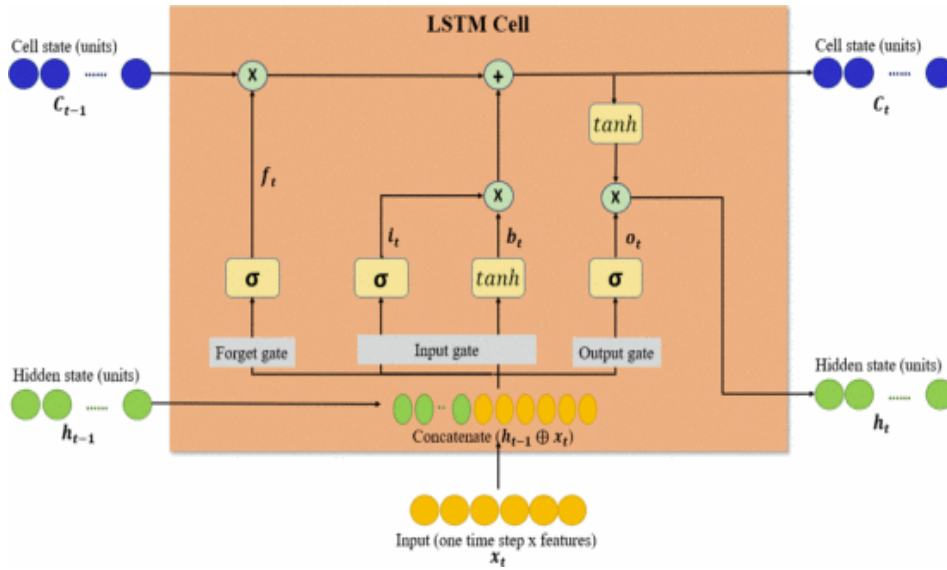


Figura 34 - Estado de uma célula de LSTM e o modo de processamento da sequência de dados. Com a representação das portas, o estado da célula e camada oculta. Fonte: (Alsabban, Salem e Malik, 2021).

Em termos de metodologia de operação do modelo, para previsão, consistia na agregação de 24 valores históricos para prever o seguinte valor, sendo essa a sequência adotada, como identificado na **Figura 35**.

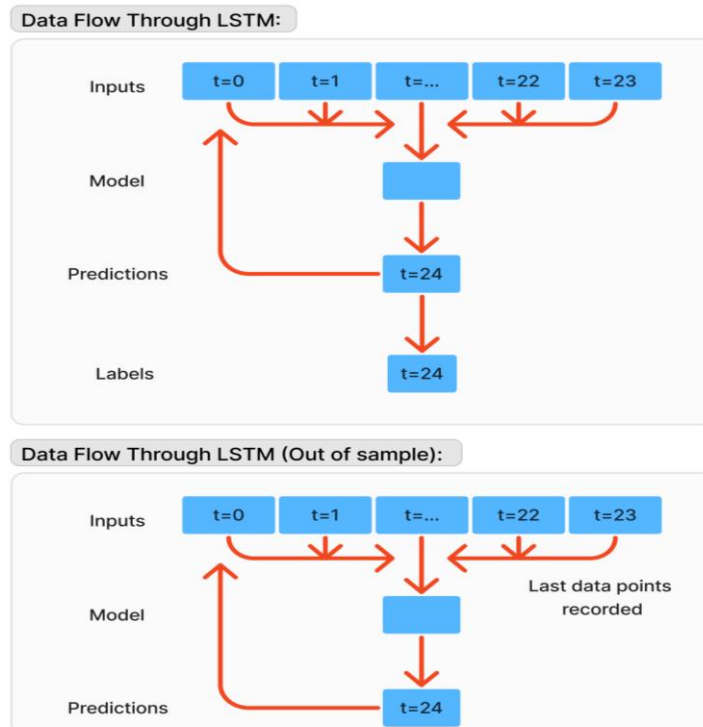


Figura 35 - Demonstração do processo de previsão do modelo. Com a utilização de uma sequência de 24 valores, dos dados de teste, para realizar a previsão do valor consequente. No caso out of sample, diferencia-se ao utilizar os últimos valores no registo de teste, onde após prever $t=24$, esse valor é adicionado à sequência. O que, em ambos os casos, remove o valor inicial.

Os dados foram repartidos em treino e validação, com a distinção do *Tensor-Flow* para esse efeito, que posteriormente seriam utilizados para avaliar o desempenho de treino dos modelos e posterior seleção com melhor desempenho. As repartições dos dados de validação foram igualitárias para os dados de teste. Tendo como função a realização de previsões sobre esses.

O parâmetro *epoch* como definido em (Alsabban, Salem e Malik, 2021), dita o número de iterações do algoritmo nos dados de treino. Um *epoch* é composto por x *batches*, que indica o número de steps. A divisão dos dados de treino deve ser feita consoante o *batch size*, 256 neste caso, que repartem os dados segundo a seguinte equação:

Equação 21

$$\text{N}^{\circ} \text{ of Batches} = \frac{\text{Training Sample}}{\text{Batch Size}} \quad (21)$$

A construção do modelo, teve como base *grid search*, com o intuito de selecionar os parâmetros mais adequados, especificamente, o número de células por camada, a taxa de aprendizagem, que determina os pesos presentes na rede, e a percentagem de *Dropout*, que previne *overfitting*.

O modelo LSTM, para este caso, possui 6 camadas na totalidade, 3 de LSTM, prosseguindo-se com a aplicação de duas camadas de *dropout*, posterior à primeira e segunda camada de LSTM, culminando numa camada dense final. Foi compilado usando ADAM e MAE como métrica de desempenho para as perdas.

Culminando, no modelo a ter configuração identificada na **Figura 36**, em específico na secção Define Network.

De seguida, repetiu-se o processo de treino para o melhor modelo, ao dar, novamente, os dados de treino para fazer o ajuste do modelo. Para esse efeito, introduziu-se *Earlystopping* para o seu desempenho não estagnar ou piorar, num intervalo de 10 epochs. O número definido de *epochs* foi de 100.

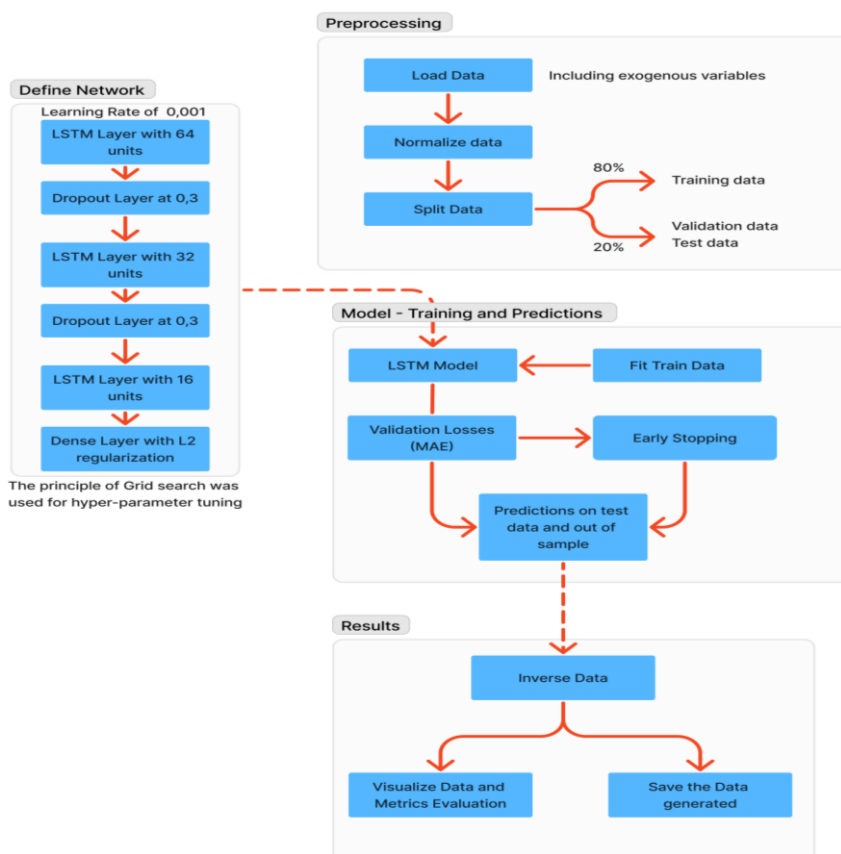


Figura 36 - Diagrama que enquadra a metodologia aplicada na construção do modelo baseado no algoritmo LSTM.

Em termos de ajustar os dados de treino ao modelo, verificou-se que houve uma evolução positiva na sua aprendizagem. No entanto, em termos das perdas de validação, o modelo estagnou não havendo uma melhoria recorrente na previsão de dados sem *labels*, o que poderá revelar algum *overfitting*, ao comparar com a outra variável em estudo.

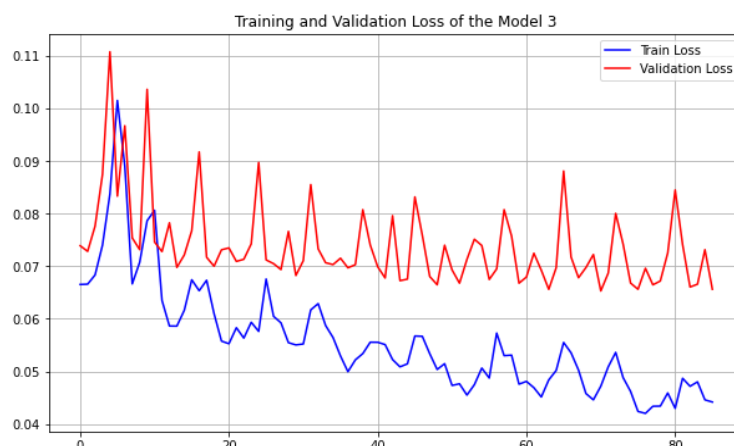


Figura 37 - Perdas associadas à aprendizagem do modelo 3. A escala em uso será de 0 a 1, pela questão dos dados se encontrarem normalizados, com o número de *epochs* na escala inferior.

Modelo 4

Para o último modelo do estudo, recorreu-se ao algoritmo MLP (*Multilayer Perceptron*). Um algoritmo de redes neuronais, como descrito em (Cordeiro-Costas *et al.*, 2023), é composto por 3 camadas interligadas, **Figura 38**, que são:

- *Input Layer* que aglomera, neste caso, os valores de carga;
- *Hidden Layers*, compõem a estrutura interna da rede neural;
- *Output Layer*, os valores previstos;

O algoritmo pode ser matematicamente representado por (Cordeiro-Costas *et al.*, 2023):

Equação 22

$$z^{[i]} = w^{[i-1,i]} * g * (z^{[i-1]}) + b^{[i-1,i]} \quad (22)$$

Para $w^{[i-1,i]}$ temos o peso do neurónio precedente e para $b^{[i-1,i]}$ o peso do viés, com $g * (z^{[i-1]})$ a representar a função de ativação do neurónio precedente, terminado na obtenção da função para o neurónio, $z^{[i]}$.

O modo de como o algoritmo processa informação, assenta na vertente *Feedforward*, que em outros termos significa o modo como é feito esse processamento é unidirecional, pela ordem de listagem mencionada nas 3 camadas. O procedimento de como este algoritmo diminui as perdas, neste fluxo de processamento, incide no método *Backward Propagation*.

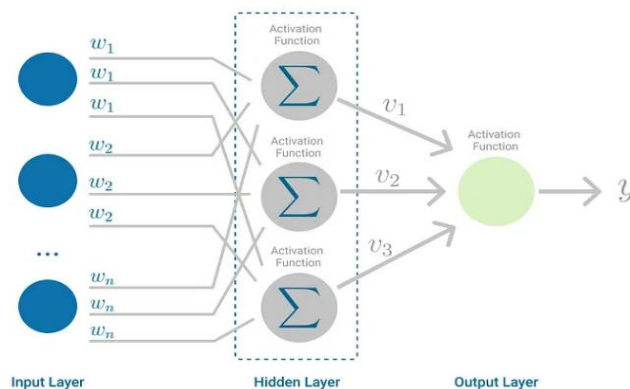


Figura 38 - Representação da estrutura que compõem o MLP. Fonte: (Bento, 2021).

Portanto, o Modelo 4 difere num ponto na previsão dos dados de carga, relativamente ao Modelo 3, pois não recorre a valores previstos para efetuar as suas previsões, pelo seu fluxo unidirecional, como presente na **Figura 39**.

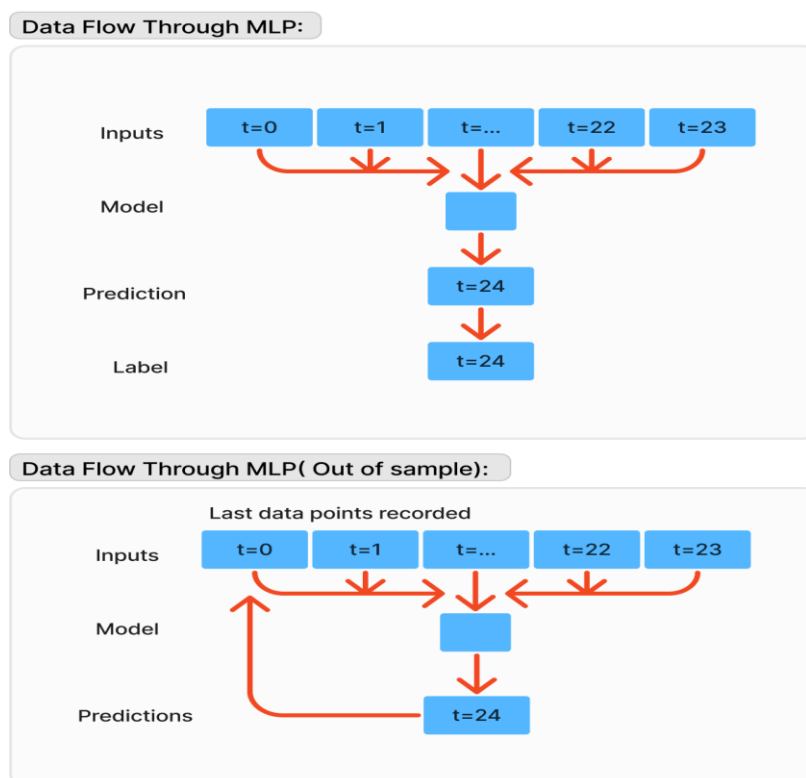


Figura 39 - Demonstração do processo de previsão do modelo. Com a utilização de uma sequência de 24 valores, dos dados de teste, para realizar a previsão do valor consequente. No caso out of sample, diferencia-se ao utilizar os últimos valores no registo de teste, onde após prever $t=24$, esse valor é adicionado à sequência.

A metodologia aplicada na construção do Modelo 3 foi apropriada para este caso, bem como o número de epochs de 100 e o intervalo de 10, como descrito na **Figura 40**. Com o mesmo princípio de *grid search* aplicado, o modelo teve os seguintes parâmetros: de 3 camadas Dense, interpostas por 2 camadas de Dropout e uma camada final de Dense, totalizando 6 camadas como o Modelo 3. Tendo sido, também, compilado por ADAM e MAE como métrica de desempenho para as perdas.

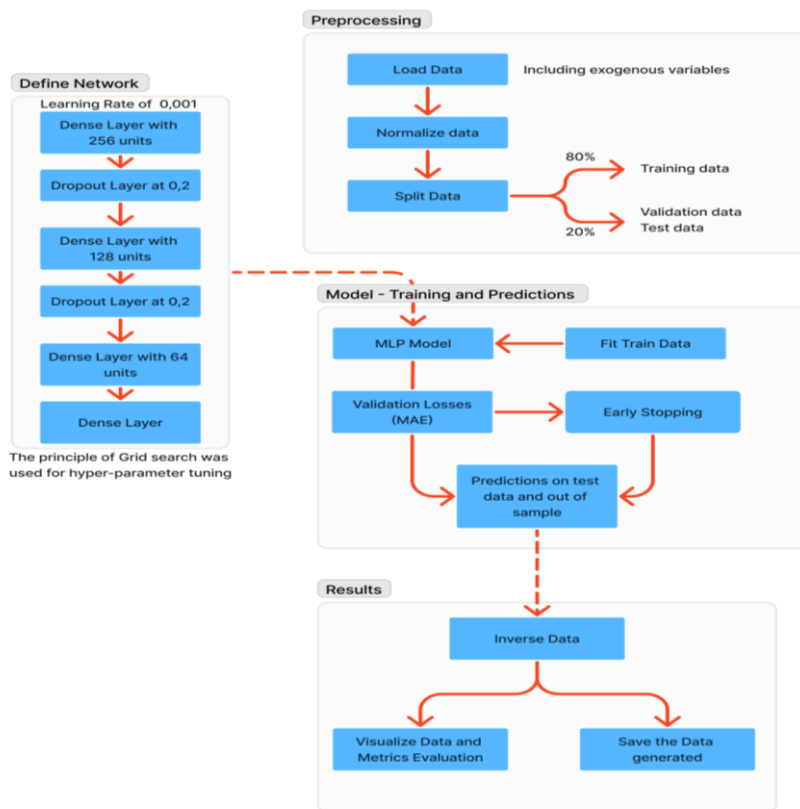


Figura 40 - Diagrama que enquadra a metodologia aplicada na construção do modelo baseado no algoritmo MLP.

Ao comparar a **Figura 41** com a **Figura 37**, vemos, que apesar de um ponto de partida com maiores perdas, o modelo teve um bom desempenho, ao conseguir generalizar os dados. Como desejado, a perda de validação foi inferior à de treino, evitando *overfitting*.

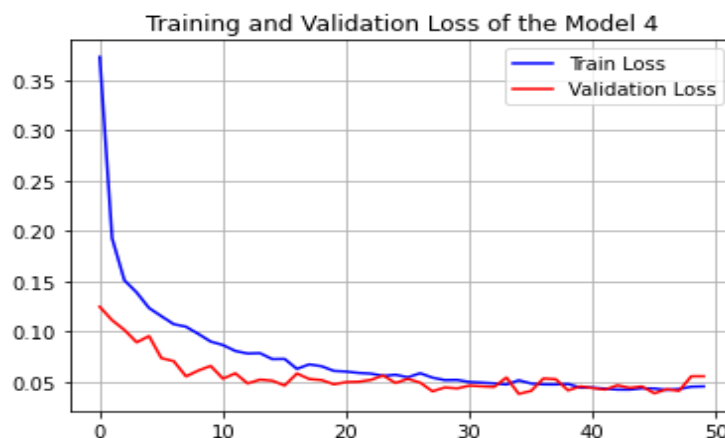


Figura 41 - Perdas associadas à aprendizagem do modelo 4. A escala em uso será de 0 a 1, pela questão dos dados se encontrarem normalizados, com o número de epochs na escala inferior.

6. Resultados e Discussão

6.1 Comparação dos modelos

Em termos de resultados obtidos, ambos os Modelos 1 e 2 tiveram o mesmo desempenho, resultando em:

- MAE – 677,28 MW;
- MAPE – 12,76%;
- RMSE – 828,85.

Ao estabelecer um termo de comparação com os Modelos 3 e 4, esses resultados são claramente inferiores aos enunciados nas **Tabelas 5 e 6**, tais quais:

- MAE:

Para o Modelo 3 o valor situou-se em 333,84 MW e 244,38 para o Modelo 4.

- RMSE:

Para o Modelo 3 o valor situou-se em 432,29 MW e 318,21 para o Modelo 4.

- MAPE:

Para o Modelo 3 o valor situou-se em 5,77% e 4,39% para o Modelo 4.

Os resultados obtidos permitem inferir a superioridade dos modelos de redes neurais, Modelo 3 e Modelo 4, que aos Modelo 1 e Modelo 2. Pela razão, que são variantes de ARIMA, um algoritmo tradicionalmente aplicado a problemas lineares.

Por isso a sua aptidão a este problema, será inferior, dado à volatilidade e não linearidade da carga. No entanto, como as variantes de ARIMA, são apenas comparados a 2 algoritmos, essa generalidade retratada pelos resultados pode não se refletir em outros casos

Tabela 5 - A seguinte tabela, enquadra os MAE e RMSE respetivos a cada modelo. Onde valores superiores, significa um pior desempenho.

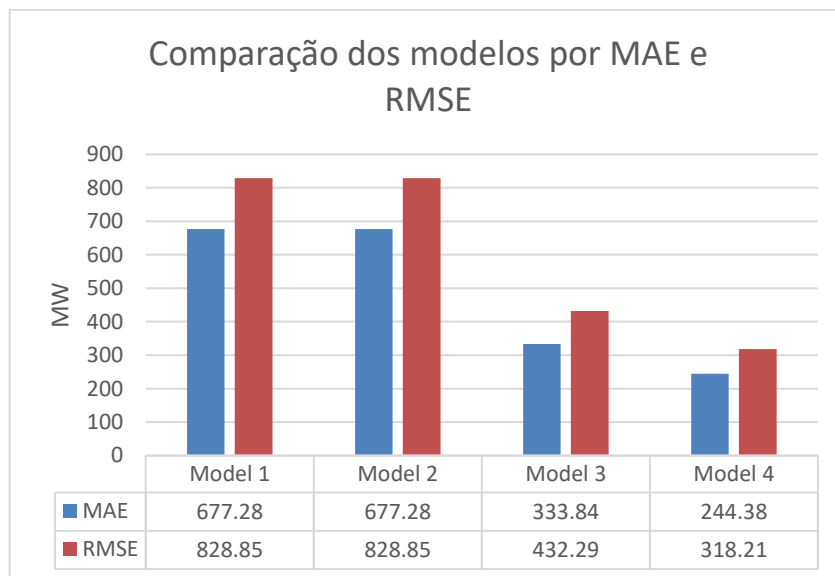
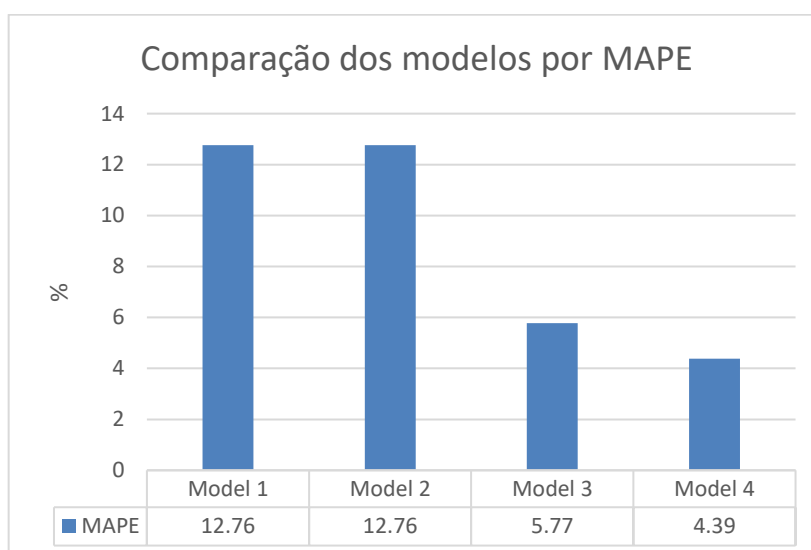


Tabela 6 - A seguinte tabela, enquadra o MAPE respetivo a cada modelo. Onde valores superiores, significa um pior desempenho.



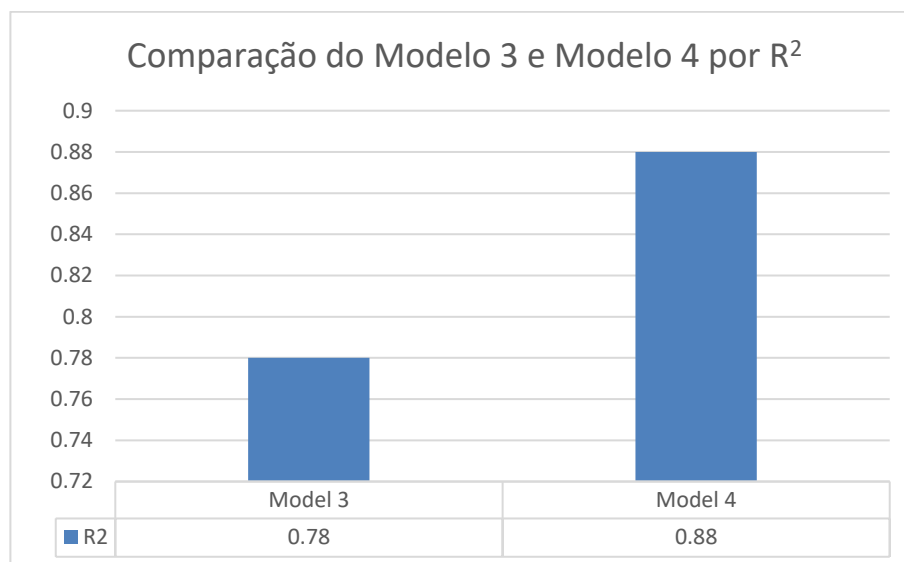
Entres os Modelo 1 e Modelo 2, embora se tenha introduzido variáveis exógenas no referido por último, verificou-se uma convergência entre os resultados. Pode-se assumir que o desempenho de SARIMAX ao introduzir variáveis não lineares, pode refletir em que o modelo não capte esses padrões.

O que se acentua pelo facto de que no teste de variáveis com maior importância para previsão da variável em estudo, **Figura 21**, não houve um resultado edificante, com nenhuma variável a superar a metade de escala de 0,5. No entanto ambos os modelos, conseguiram captar as variações presentes nos dados de treino, no seu intervalo de confiança de 95%, **Figuras 42 e 44**.

Os modelos de redes neuronais, tiveram desempenhos semelhantes, com uma ligeira superioridade para o Modelo 3, MLP, ao Modelo 4, LSTM. Sendo o referido em

primeiro, o modelo superior deste estudo. Ambos tiveram a aptidão de prever a variabilidade da variável carga, de acordo com os resultados de 0,78 e 0,88, para o Modelo 3 e Modelo 4, respetivamente

Tabela 7 - A seguinte tabela, enquadra o R^2 respetivo aos Modelos 3 e 4. Onde valores superiores, significa um melhor desempenho.



Modelo 1

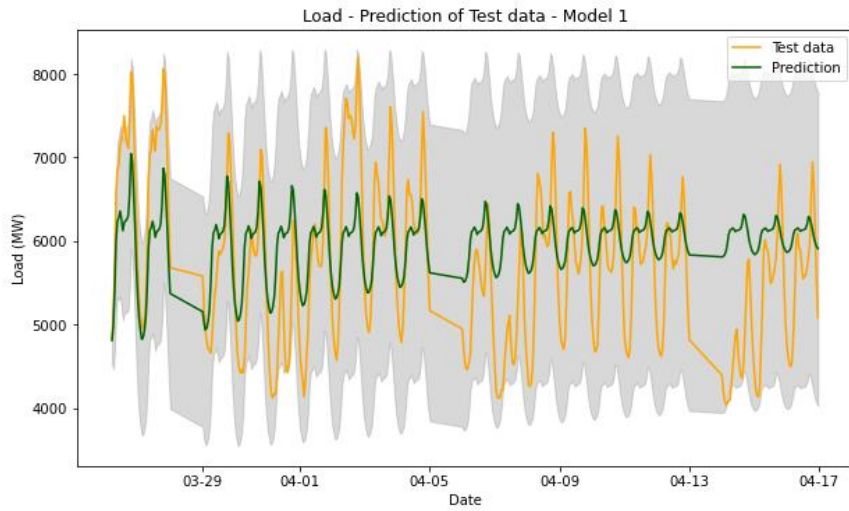


Figura 42 - Previsão efetuada sobre os dados de teste do Modelo 1.

Tabela 8 - Registo dos cinco primeiros valores de previsão dos dados de teste para o Modelo 1, com os respetivos intervalos de confiança inferior e superior.

Data	Previsões (MW)	Carga (MW)	IC inferior (MW)	IC superior (MW)
26/03/2024 05:00	4810.90	4817.00	4527.41	5094.40
26/03/2024 06:00	5000.80	5092.00	4467.22	5534.39
26/03/2024 07:00	5423.99	5642.00	4706.55	6141.42
26/03/2024 08:00	5999.87	6558.00	5160.20	6839.55
26/03/2024 09:00	6241.87	6847.00	5321.36	7162.38

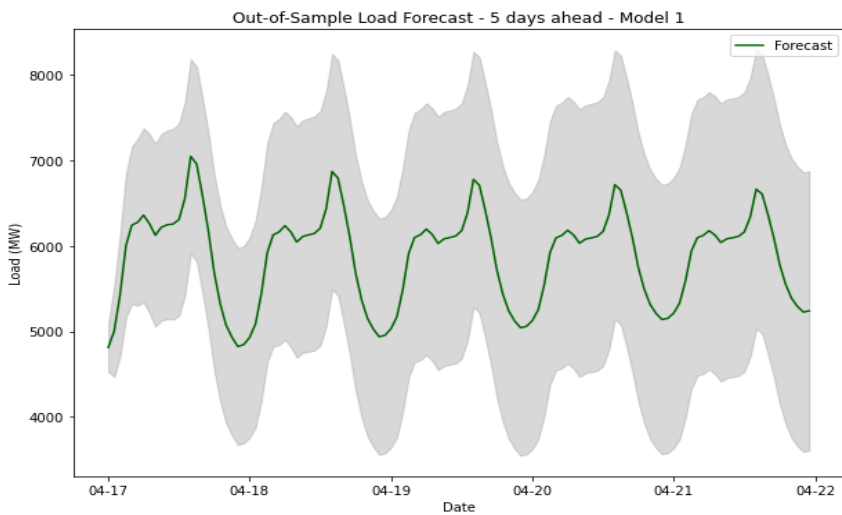


Figura 43 - Previsão efetuada out of sample (5 dias após os dados de teste), num formato horário, para o Modelo 1.

Tabela 9 - Registo dos cinco primeiros valores de previsão out of sample para o Modelo 1, com os respetivos intervalos de confiança inferior e superior.

Data	Previsões (MW)	IC inferior (MW)	IC superior (MW)
17/04/2024 00:00	4810.91	4527.41	5094.40
17/04/2024 01:00	5000.81	4467.22	5534.40
17/04/2024 02:00	5423.99	4706.56	6141.43
17/04/2024 03:00	5999.88	5160.20	6839.55
17/04/2024 04:00	6241.87	5321.37	7162.38

Modelo 2

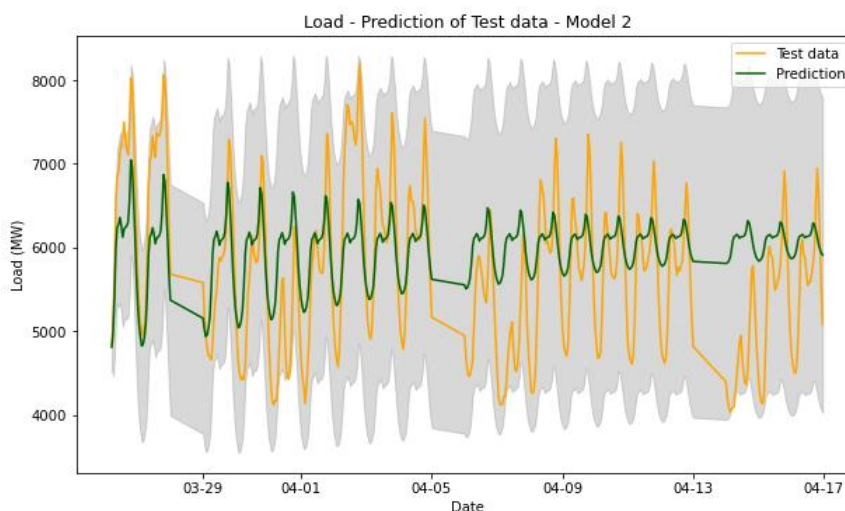


Figura 44 - Previsão efetuada sobre os dados de teste do Modelo 2.

Tabela 10 - Registo dos cinco primeiros valores de previsão dos dados de teste para o Modelo 2, com os respetivos intervalos de confiança inferior e superior.

Data	Previsões (MW)	Carga (MW)	IC inferior (MW)	IC superior (MW)
26/03/2024 05:00	4810.91	4817.00	4527.41	5094.40
26/03/2024 06:00	5000.81	5092.00	4467.22	5534.40
26/03/2024 07:00	5423.99	5642.00	4706.56	6141.43
26/03/2024 08:00	5999.88	6558.00	5160.20	6839.55
26/03/2024 09:00	6241.87	6847.00	5321.37	7162.38

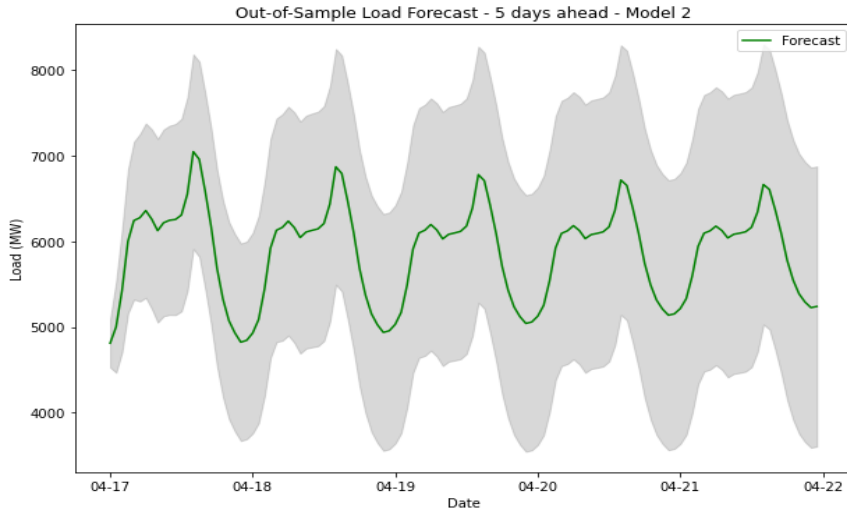


Figura 45 - Previsão efetuada out of sample (5 dias após os dados de teste), num formato horário, para o Modelo 2.

Tabela 11 - Registo dos cinco primeiros valores de previsão out of sample para o Modelo 2, com os respetivos intervalos de confiança inferior e superior.

Data	Previsões (MW)	IC inferior (MW)	IC superior (MW)
17/04/2024 00:00	4810.91	4527.41	5094.40
17/04/2024 01:00	5000.81	4467.22	5534.40
17/04/2024 02:00	5423.99	4706.56	6141.43
17/04/2024 03:00	5999.88	5160.20	6839.55
17/04/2024 04:00	6241.87	5321.37	7162.38

Modelo 3



Figura 46 - Previsão efetuada sobre os dados de teste do Modelo 3.

Tabela 12 - Registo dos cinco primeiros valores de previsão dos dados de teste para o Modelo 3.

Data	Previsões (MW)	Carga (MW)
27/03/2024 05:00	4917.35	5036.00
27/03/2024 06:00	5135.00	5277.00
27/03/2024 07:00	5670.25	5839.00
27/03/2024 08:00	6319.74	6708.00
27/03/2024 09:00	6773.49	7023.00

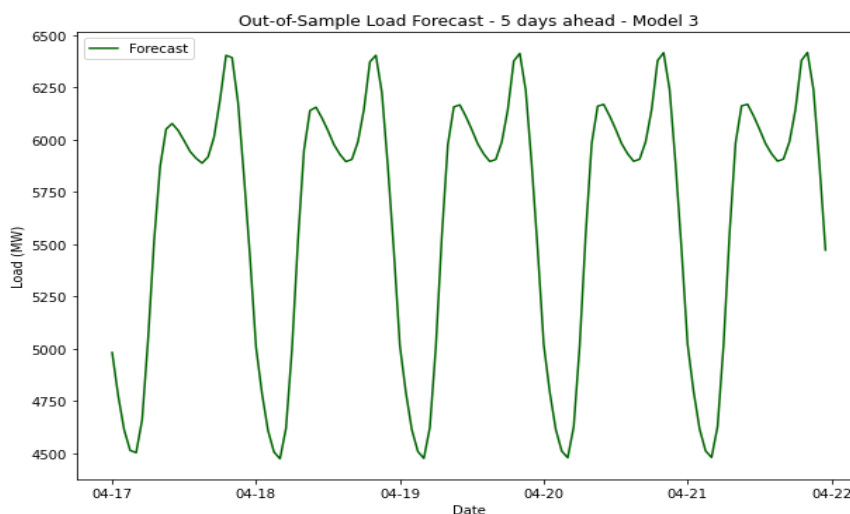


Figura 47 - Previsão efetuada out of sample (5 dias após os dados de teste), num formato horário, para o Modelo 3.

Tabela 13 - Registo dos cinco primeiros valores de previsão out of sample para o Modelo 3.

Data	Previsões (MW)
17/04/2024 00:00	5449.55
17/04/2024 01:00	7536.51
17/04/2024 02:00	6980.46
17/04/2024 03:00	6891.54
17/04/2024 04:00	4782.82

Modelo 4

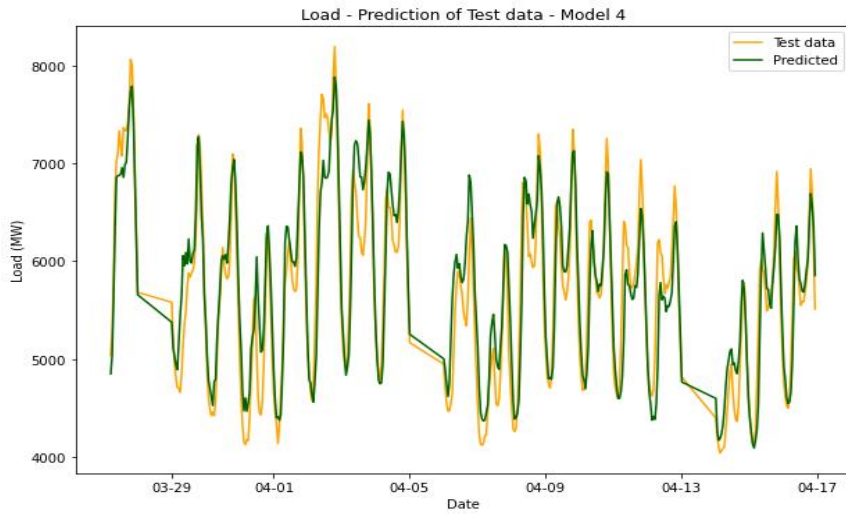


Figura 48 - Previsão efetuada sobre os dados de teste do Modelo 4.

Tabela 14 - Registo dos cinco primeiros valores de previsão dos dados de teste para o Modelo 4.

Data	Previsões (MW)	Carga (MW)
27/03/2024 05:00	5036.00	5000.45
27/03/2024 06:00	5277.00	5311.51
27/03/2024 07:00	5839.00	6020.40
27/03/2024 08:00	6708.00	6987.71
27/03/2024 09:00	7023.00	7127.96

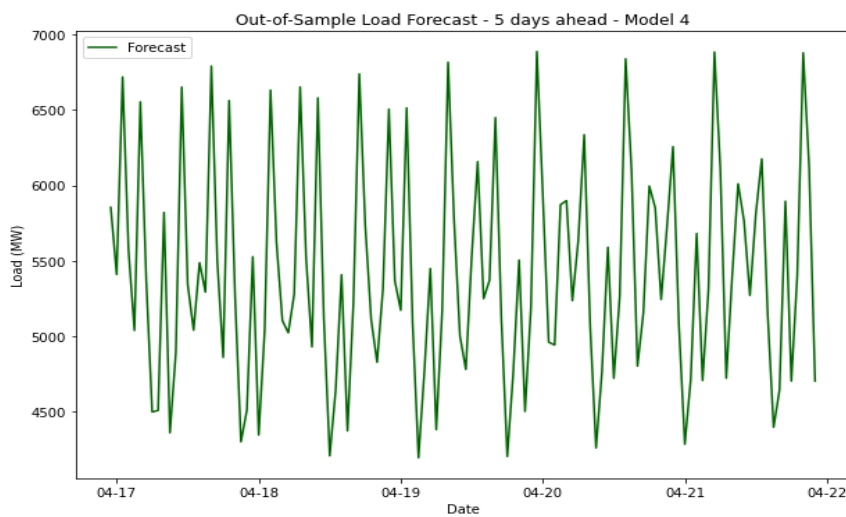


Figura 49 - Previsão efetuada out of sample (5 dias após os dados de teste), num formato horário, para o Modelo 4.

Tabela 15 - Registo dos cinco primeiros valores de previsão out of sample para o Modelo 4.

Data	Previsões (MW)
17/04/2024 00:00	5449.55
17/04/2024 01:00	7536.51
17/04/2024 02:00	6980.46
17/04/2024 03:00	6891.54
17/04/2024 04:00	4782.83

6.2 Melhorias

A utilização de uma série temporal mais extensiva será benéfica para a aprendizagem dos algoritmos de redes neuronais, ao requerem em maior número parâmetros, pela sua elevada complexidade. De modo a evitar o ocorrido *overfitting*, **Figura 37**, no Modelo 3.

Os resultados expostos em (Farsi *et al.*, 2021), onde visou a utilização de vários algoritmos desde: ARIMA, LSTM, ETS (Exponential Smoothing), SVR (Support Vector Regression) e entre outros. Para previsão de carga, com recurso a dados referentes do setor elétrico da Alemanha e Malásia. Confirmam o problema exposto no último parágrafo. Pelo caso de a Malásia ter uma escala horária, enquanto na Alemanha era diário, tendo um número semelhante de entradas, 2168, ao presente estudo.

Culminando, no melhor desempenho dos modelos para o caso asiático e desempenhos semelhantes, na métrica R^2 , entre os Modelos 3 e 4 para o caso alemão, rondando os 0,80. Contudo, com menor desempenho em MAPE, exceto no caso PLCNET com 2,08%, com uma diferenciação, pelo menos, de mais de 264%, sendo similar ao valor de 12,76% para os Modelos 1 e 2.

Algo pertinente a mencionar será, que não foi possível obter o nível de precipitação via o *OpenWeather*. Uma variável que poderia ser pertinente para o presente estudo. Também a obtenção de outras variáveis, poderá ser relevante, como por exemplo, radiação solar. Não só a nível meteorológico, mas também de outras como, a carga da semana passada para o mesmo dia, como explorado em (Ibrahim *et al.*, 2022). A distinção do perfil de carga, ou seja, em termos de doméstico, industrial e entre outros.

7. Conclusões

Este trabalho, visava a previsão de consumos energéticos, em *short-term Load forecasting*, para Portugal, recorrendo a algoritmos de *Machine Learning*. Tais algoritmos que são: ARIMA, SARIMAX, LSTM e MLP.

Os resultados para previsão dos valores da carga demonstraram, que o modelo construído com base em MLP foi o superior, seguido do LSTM. Assim demonstrando a capacidade superior dos algoritmos de *deep learning*, a algoritmos mais tradicionais e com utilização mais própria para problemas lineares, que não será o caso deste pela volatilidade da própria carga.

Como também foi para o caso *out of sample*, onde houve uma melhor interpretação da variabilidade da carga, pelos modelos construídos com MLP e LSTM.

Para esse efeito, foi apurado a relevância que têm variáveis exógenas na flutuação da carga, tanto meteorológicas, como padrões de consumo. Destas destacam-se a temperatura e o valor referente à carga por hora e dia.

7.1 Contribuições do estudo para o campo das energias renováveis e eficiência energética

Estas deduções reforçam a ideia, da eventualidade da integração de *Machine Learning* no setor elétrico e certamente no dia a dia. Ao aperfeiçoar os modelos de previsão de carga, tendo um causa efeito direto na fiabilidade e sustentabilidade da rede e no mercado de eletricidade, onde outrora baseado nos combustíveis fósseis, está a ser permutado, em um sustentado em energias renováveis e as implicações inerentes a essa transformação, podem ser combatidas com *Machine Learning*.

É necessário realçar, a necessidade de proteção dos dados tanto para o proprietário, como seu comercializador, em termos legais. A inteligência artificial está na vanguarda da presente revolução tecnológica, mas só poderá ser implementada quando for assegurada a privacidade e confidencialidade dos dados. Tanto para viés dos próprios modelos, com possíveis implicações sociais, como para ameaças de segurança.

7.2 Sugestões para trabalhos futuros

O projeto, com base na previsão de consumos energéticos em Portugal, pode ser aprofundado e mais abrangente, algumas das possibilidades passam por:

- A utilização de um maior número de algoritmos, bem como a exploração de métodos *ensemble*, que consiste na junção de algoritmos;
- Analisar comportamentos de consumos energéticos, de acordo com a renumeração salarial e ao mesmo tempo delimitando por cluster esse

critério, por exemplo, em municípios. Com o objetivo de identificar zonas desfavorecidas, em termos de conforto térmico, como combate à pobreza energética;

- Estudar a integração de EES na rede nacional, em torno de uma eventual expansão da capacidade instalada de energias renováveis;
- O estudo da criação de um modelo que consiga prever falhas na rede;

Referências Bibliográficas

ABAD, L. A. *et al.* - A Short-Term Load Forecasting Algorithm Using Support Vector Regression & Artificial Neural Network Method (SVR-ANN). Em **2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC)**

ABDELLA, Juhar; SHUAIB, Khaled - Peer to peer distributed energy trading in smart grids: A survey. **Energies**. . ISSN 19961073. 11:6 (2018). doi: 10.3390/en11061560.

ADF Test: Augmented Dickey Fuller Equation - (2023).

ALAZEMI, Talal; DARWISH, Mohamed; RADI, Mohammed - Renewable energy sources integration via machine learning modelling: A systematic literature review. **Heliyon**. . ISSN 24058440. 10:4 (2024) e26088. doi: 10.1016/j.heliyon.2024.e26088.

ALLAL, Zaid *et al.* - Machine learning solutions for renewable energy systems: Applications, challenges, limitations, and future directions. **Journal of Environmental Management**. . ISSN 03014797. 354:2024) 120392. doi: 10.1016/j.jenvman.2024.120392.

ALOTAIBI, Ibrahim *et al.* - A comprehensive review of recent advances in smart grids: A sustainable future with renewable energy resources. **Energies**. . ISSN 19961073. 13:23 (2020). doi: 10.3390/en13236269.

ALPAYDIN, Ethem. - Introduction to Machine Learning. (2014).

ALSABBAN, Maha S.; SALEM, Nema; MALIK, Hebatullah M. - Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) Power Forecasting. Em **2021 13th IEEE PES Asia Pacific Power & Energy Engineering Conference (APPEEC)**

ANJO, João *et al.* - Modeling the long-term impact of demand response in energy planning: The Portuguese electric system case study. **Energy**. . ISSN 0360-5442. 165:2018) 456–468. doi: <https://doi.org/10.1016/j.energy.2018.09.091>.

ATĂNĂSOAE, Pavel; PENTIUC, Radu Dumitru; HOPULELE, Eugen - Considerations Regarding the Negative Prices on the Electricity Market. **Proceedings**. . ISSN 2504-3900. 63:1 (2020). doi: 10.3390/proceedings2020063026.

BATTULA, Amrutha Raju; VUDDANTI, Sandeep; SALKUTI, Surender Reddy - Review of Energy Management System Approaches in Microgrids. **Energies**. . ISSN 1996-1073. 14:17 (2021). doi: 10.3390/en14175459.

BENTO, Carolina - Multilayer Perceptron Explained with a Real-Life Example and Python Code: Sentiment Analysis. (2021).

BENTON, William - Machine Learning Systems and Intelligent Applications. **IEEE Software**. PP:2020). doi: 10.1109/MS.2020.2985224.

BEVANS, Rebecca - Akaike Information Criterion | When & How to Use It. (2020).

BISHOP, Christopher M. - **Pattern Recognition and Machine Learning (Information Science and Statistics)**. Berlin, Heidelberg : Springer-Verlag, 2006. ISBN 0387310738.

CARMONA, Pedro *et al.* - Climate change and government policy: Fresh insights from complexity theory. **Journal of Environmental Management**. . ISSN 03014797. 338:2023) 117831. doi: 10.1016/j.jenvman.2023.117831.

CASS, Stephen - The Top Programming Languages 2023 - IEEE Spectrum. 2023).

CHICCO, Davide; WARRENS, Matthijs; JURMAN, Giuseppe - The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. **PeerJ Computer Science**. 7:2021) e623. doi: 10.7717/peerj-cs.623.

COELHO, Celeste; VALENTE, Sandra - Envolvimento dos agentes locais na identificação e na implementação de soluções para o combate à desertificação e degradação do solo. Em . ISBN 978-972-745-193-7. p. 37-46.

COPERNICUS - Surface air temperature for May 2024. 2024).

CORDEIRO-COSTAS, Moisés *et al.* - Load Forecasting with Machine Learning and Deep Learning Methods. **Applied Sciences**. 13:2023) 7933. doi: 10.3390/app13137933.

Day-ahead hourly price | OMIE - [s.d.]).

DGEG - **Relatório de Monitorização da Segurança de Abastecimento do Sistema Elétrico Nacional 2024-2040 (RMSA-E 2023)**

EDGAR, Thomas W.; MANZ, David O. - Machine Learning. Em **Research Methods for Cyber Security**. [S.l.] : Elsevier, 2017. p. 153-173.

Electricity Maps - Europe- 2023 - Carbon Intensity Data - [Em linha], atual. 2024. Disponível em WWW:<URL:https://app.electricitymaps.com/map?s=&solar=false&lang=pt-br>.

Electricity Maps - Portugal - 2023 - Carbon Intensity Data - [Em linha], atual. 2024. Disponível em WWW:<URL:https://app.electricitymaps.com/map?s=&solar=false&lang=pt-br>.

Electricity market reform: Council signs off on updated rules - 2024).

EMDE, Alexander *et al.* - Effects of Load Forecast Deviation on the Specification of Energy Storage Systems. **Designs**. . ISSN 2411-9660. 7:5 (2023). doi: 10.3390/designs7050107.

ENERGY AGENCY, International - **Electricity Grids and Secure Energy Transitions Enhancing the foundations of resilient, sustainable and affordable power systems** [Em linha] Disponível em WWW:<URL:www.iea.org>.

E-REDES - E-REDES instala 6 milhões de contadores inteligentes . 2024).

E-REDES - Produção de energia para Autoconsumo . [s.d.].

ERSE - **Proposta de fixação excecional de tarifas de eletricidade a vigorarem a partir de 1 de junho de 2024**

ERSE - **Caracterização da Procura de Energia Elétrica em 2024** [Em linha] Disponível em WWW:<URL:www.erse.pt>.

ERSE - Fixação excecional de tarifas de eletricidade a partir de 1 de junho de 2024. 2024).

FARHANGI, H. - The path of the smart grid. **IEEE Power and Energy Magazine**. . ISSN 1540-7977. 8:1 (2010) 18–28. doi: 10.1109/MPE.2009.934876.

FARSI, Behnam *et al.* - On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach. **IEEE Access**. . ISSN 21693536. 9:2021) 31191–31212. doi: 10.1109/ACCESS.2021.3060290.

FILIPE, João; FERREIRA, Beja - **Smart Grids Análise de Falhas em Cascata**

FLIES, Emily J. *et al.* - Forecasting future global food demand: A systematic review and meta-analysis of model complexity. **Environment International**. . ISSN 01604120. 120:2018) 93–103. doi: 10.1016/j.envint.2018.07.019.

FRIEDERICH, David - Electricity load forecasting with LSTM. 2017).

From AR to SARIMAX: Mathematical Definitions of Time Series Models - [s.d.].

General Government Expenditure by Function - [Em linha], atual. 2024. Disponível em WWW:<URL:https://ec.europa.eu/eurostat/databrowser/view/gov_10a_exp_custom_10307189/default/table?lang=en>.

GOLDEN, R. M. - Statistical Pattern Recognition. Em **International Encyclopedia of the Social & Behavioral Sciences**. [S.l.] : Elsevier, 2001. p. 15040–15044.

GPP - **Newsletter - Cereais**

GUARNIDO-RUEDA, Almudena; AMATE-FORTES, Ignacio - Social indicators of sustainable resource management. Em **Sustainable Resource Management**. [S.l.] : Elsevier, 2021. p. 273–288.

HALVORSEN, Hans-Petter - **Python for Science and Engineering** [Em linha] Disponível em WWW:<URL:https://www.halvorsen.blog>.

HAN, Jiaming; SHU, Kunxin; WANG, Zhenyu - Predicting energy use in construction using Extreme Gradient Boosting. **PeerJ Computer Science**. . ISSN 2376-5992. 9:2023) e1500. doi: 10.7717/peerj-cs.1500.

IBRAHIM, Bibi *et al.* - Machine Learning for Short-Term Load Forecasting in Smart Grids. **Energies**. . ISSN 19961073. 15:21 (2022). doi: 10.3390/en15218079.

INTERNATIONAL ENERGY AGENCY - **CO2 Emissions in 2023** [Em linha] Disponível em WWW:<URL:www.iea.org>.

IRENA - How to Decarbonise end-use Sectors with Renewable power? 2023).

JAIN, Ritu; MAHAJAN, Vasundhara - Load forecasting and risk assessment for energy market with renewable based distributed generation. **Renewable Energy Focus**. . ISSN 1755-0084. 42:2022) 190–205. doi: <https://doi.org/10.1016/j.ref.2022.06.007>.

JOÃO LOPES VELOSO RIBEIRO DA SILVA, Miguel *et al.* - **Smart Grids em Portugal**

KAPLAN, Halid; TEHRANI, Kambiz; JAMSHIDI, Mo - Fault Diagnosis of Smart Grids Based on Deep Learning Approach. Em **2021 World Automation Congress (WAC)**. [S.l.] : IEEE, 1 Ago. 2021. ISBN 978-1-68524-111-7

KATARAY, Tarun *et al.* - Integration of smart grid with renewable energy sources: Opportunities and challenges – A comprehensive review. **Sustainable Energy Technologies and Assessments**. . ISSN 22131388. 58:2023) 103363. doi: [10.1016/j.seta.2023.103363](https://doi.org/10.1016/j.seta.2023.103363).

KECK, Felix *et al.* - The impact of battery energy storage for renewable energy power grids in Australia. **Energy**. . ISSN 03605442. 173:2019) 647–657. doi: [10.1016/j.energy.2019.02.053](https://doi.org/10.1016/j.energy.2019.02.053).

KEYHANI, Ali - Smart Power Grids. Em KEYHANI ALI AND MARWALI, MUHAMMAD (Ed.) - **Smart Power Grids 2011** [Em linha]. Berlin, Heidelberg : Springer Berlin Heidelberg, 2012 Disponível em WWW:<[URL:https://doi.org/10.1007/978-3-642-21578-0_1](https://doi.org/10.1007/978-3-642-21578-0_1)>. ISBN 978-3-642-21578-0. p. 1–25.

LAN, X. , K. W. Thoning, And E. J. Dlugokencky - **Trends in globally-averaged CH₄, N₂O, and SF₆ determined from NOAA Global Monitoring Laboratory measurements**, atual. nov. 2024.

LAN, X. , Tans, P. And K. W. Thoning - **Trends in globally-averaged CO₂ determined from NOAA Global Monitoring Laboratory measurements**, atual. nov. 2024.

LEDMAOUI, Younes *et al.* - Forecasting solar energy production: A comparative study of machine learning algorithms. **Energy Reports**. . ISSN 23524847. 10:2023) 1004–1012. doi: [10.1016/j.egy.2023.07.042](https://doi.org/10.1016/j.egy.2023.07.042).

LI, Zhongzhou *et al.* - Directional protection based on fault component energy function in micro-grid. Em **IEEE PES Innovative Smart Grid Technologies**

LIU, Wenxia *et al.* - Vulnerability Assessment of Power Grids with High Penetration of Renewable Energy. Em **2023 5th Asia Energy and Electrical Engineering Symposium, AEEES 2023**. [S.l.] : Institute of Electrical and Electronics Engineers Inc., 2023. ISBN 9781665490542

LOIZIDIS, Stylianos *et al.* - Electricity Day-Ahead Market Conditions and Their Effect on the Different Supervised Algorithms for Market Price Forecasting. **Energies**. . ISSN 1996-1073. 16:12 (2023). doi: [10.3390/en16124617](https://doi.org/10.3390/en16124617).

MISRA, Siddharth; HAN, Yifu - Chapter 13 - Characterization of subsurface hydrocarbon/water saturation using Markov-chain Monte Carlo stochastic inversion of broadband electromagnetic logs. Em MISRA, SIDDHARTH; LI, HAO; HE, JIABO (Eds.) - **Machine Learning for Subsurface Characterization** [Em linha]. [S.l.] : Gulf Professional Publishing, 2020 Disponível em WWW:<URL:https://www.sciencedirect.com/science/article/pii/B9780128177365000132>. ISBN 978-0-12-817736-5. p. 369–402.

MITCHELL, Tom - **Machine learning** [Em linha] Disponível em WWW:<URL:https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-1689%28199909%299%3A3%3C191%3A%3AAID-STVR184%3E3.0.CO%3B2-E>.

MUHURI, Pramita Sree *et al.* - Using a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to Classify Network Attacks. **Information**. . ISSN 2078-2489. 11:5 (2020). doi: 10.3390/info11050243.

MÜLLER, Andreas C.; GUIDO, Sarah - **Introduction to Machine Learning with Python**

NEWBERY, David M. - High renewable electricity penetration: Marginal curtailment and market failure under “subsidy-free” entry. **Energy Economics**. . ISSN 01409883. 126:2023). doi: 10.1016/j.eneco.2023.107011.

ORTEGA-VAZQUEZ, M. A.; KIRSCHEN, D. S. - Economic impact assessment of load forecast errors considering the cost of interruptions. Em **2006 IEEE Power Engineering Society General Meeting, PES**. [S.l.] : IEEE Computer Society, 2006. ISBN 1424404932

PARASHAR, Priyanka - Random Forest Classification and it's Mathematical Implementation. (2020).

PINHÃO, Miguel; FONSECA, Miguel; COVAS, Ricardo - Electricity Spot Price Forecast by Modelling Supply and Demand Curve. **Mathematics**. . ISSN 22277390. 10:12 (2022). doi: 10.3390/math10122012.

PROKHOROV, Oleksandr; DREISBACH, Dina - The impact of renewables on the incidents of negative prices in the energy spot markets. **Energy Policy**. . ISSN 03014215. 167:2022). doi: 10.1016/j.enpol.2022.113073.

QUIGGIN, Daniel *et al.* - **Climate change risk assessment 2021**

RAJASEKHAR, J.; HEMANTH, T.; SK, Anjuman - SMS Spam Classification and Through Recurrent Neural Network (LSTM) model. Em **2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)**

RASCHKA, Sebastian.; MIRJALILI, Vahid. - **Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow**. [S.l.] : Packt Publishing, 2017. ISBN 9781787125933.

REN - Eletricidade - Balanço Diário. 2024).

REN, Bo *et al.* - Machine learning applications in health monitoring of renewable energy systems. **Renewable and Sustainable Energy Reviews**. . ISSN 13640321. 189:2024) 114039. doi: 10.1016/j.rser.2023.114039.

RENEWABLE ENERGY AGENCY, International - **Renewable Power Generation Costs in 2022** [Em linha] Disponível em WWW:<URL:www.irena.org>. ISBN 978-92-9260-544-5.

RITCHIE, Hannah *et al.* - Population Growth. **Our World in Data**. 2023).

ROY, Dibyendu *et al.* - Techno-economic and environmental analyses of hybrid renewable energy systems for a remote location employing machine learning models. **Applied Energy**. . ISSN 03062619. 361:2024) 122884. doi: 10.1016/j.apenergy.2024.122884.

SAHA, Sajeeb; SALEEM, M. Ismail - Inertia Sensitivity Analysis of Power Grids with High Penetration of Renewable Energy Sources. Em **2022 IEEE Industry Applications Society Annual Meeting (IAS)**

SALAM, Abdulwahed; HIBAOUI, Abdelaaziz El - Comparison of Machine Learning Algorithms for the Power Consumption Prediction : - Case Study of Tetouan city -. Em **2018 6th International Renewable and Sustainable Energy Conference (IRSEC)**

SOVACOOOL, Benjamin K.; RIO, Dylan Furszyfer DEL; ZHANG, Weimin - The political economy of net-zero transitions: Policy drivers, barriers, and justice benefits to decarbonization in eight carbon-neutral countries. **Journal of Environmental Management**. . ISSN 03014797. 347:2023) 119154. doi: 10.1016/j.jenvman.2023.119154.

SWAIN, Ayusee; SALKUTI, Surender Reddy; SWAIN, Kaliprasanna - An Optimized and Decentralized Energy Provision System for Smart Cities. **Energies**. . ISSN 1996-1073. 14:5 (2021). doi: 10.3390/en14051451.

THARANI, Kusum *et al.* - Machine learning models for renewable energy forecasting. **Journal of Statistics and Management Systems**. . ISSN 0972-0510. 23:1 (2020) 171–180. doi: 10.1080/09720510.2020.1721636.

UN ENVIRONMENT PROGRAM - Global Resource Outlook 2024. 2024).

UNIVERSITY, Newcastle - Strength of Correlation. 2023).

VACCARO, Alfredo *et al.* - **Research Trends and Challenges in Smart Grids**

VARGAS, Brandon - Negative Electricity Prices Explained. 2024).

VITIELLO, Silvia *et al.* - Smart Metering Roll-Out in Europe: Where Do We Stand? Cost Benefit Analyses in the Clean Energy Package and Research Trends in the Green Deal. **Energies**. . ISSN 19961073. 15:7 (2022). doi: 10.3390/en15072340.

VITORINO, João - The New Portuguese Electrical System Rules. 2022).

VLASCEANU, Madalina *et al.* - Addressing climate change with behavioral science: A global intervention tournament in 63 countries. **Science Advances**. 10:6 (2024) ead5778. doi: 10.1126/sciadv.adj5778.

WANG, Lizhi *et al.* - A Deep-forest based approach for detecting fraudulent online transaction. Em HURSON, ALI R.; WU, SHENG (Eds.) - **AI and Cloud Computing** Advances in Computers. [Em linha]. [S.l.] : Elsevier, 2021 Disponível em WWW:<URL:https://www.sciencedirect.com/science/article/pii/S0065245820300851>.v. 120. p. 1-38.

WITTEK, Peter - **Quantum Machine Learning**. [S.l.] : Elsevier, 2014. ISBN 9780128009536.

WOOLF, Beverly - **Building Intelligent Interactive Tutors**. [S.l.] : Elsevier, 2009. ISBN 9780123735942.