



BUSINESS
SCHOOL

Departamento de Marketing, Operações e Gestão Geral

Previsão do fluxo de Turismo para uma Gestão Estratégica em Portugal: Explorando Metodologias de *Machine Learning*

Maria Luísa Pinto Basto Pereira de Lacerda

Mestrado em Gestão Empresarial

Orientador:

Professor auxiliar convidado, Filipe Roberto de Jesus Ramos
ISCTE Business School, DMQGE

Co-Orientador:

Professor associado (com Agregação), José Joaquim Dias Curto
ISCTE Business School, DMQGE

Setembro, 2022



BUSINESS
SCHOOL

Departamento de Marketing, Operações e Gestão Geral

Previsão do fluxo de Turismo para uma Gestão Estratégica em Portugal: Explorando Metodologias de *Machine Learning*

Maria Luísa Pinto Basto Pereira de Lacerda

Mestrado em Gestão Empresarial

Orientador:

Professor auxiliar convidado, Filipe Roberto de Jesus Ramos
ISCTE Business School, DMQGE

Co-Orientador:

Professor associado (com Agregação), José Joaquim Dias Curto
ISCTE Business School, DMQGE

Setembro, 2022

Agradecimentos

A melhor forma de começar um trabalho desta dimensão e exigência só pode ser a agradecer. Não acredito que algum trabalho de investigação seja feito por uma só pessoa. É preciso ter orientação, apoio e quem nos distraia para poder voltar a focar.

Começo por agradecer ao professor Filipe, que foi fundamental para que a construção desta tese fosse um percurso tão curioso e interessante. Esteve sempre disponível e motivado para tornar este trabalho cada vez mais completo. Agradeço também ao professor Dias Curto, por me ter aberto os olhos e a curiosidade para o ramo da análise de dados e por ter sido tão prestável e assertivo nos comentários e melhorias.

Agradeço muito aos meus queridos pais, que me abriram as portas de casa para me suportarem nas fases mais críticas, sempre com uma leveza e alegria que descansa e dá muita confiança. Quero ainda agradecer ao avô Né por fazer perguntas e querer saber sempre mais sobre o que estava a descobrir e a estudar. É um grande exemplo de como nos podemos manter vivos e jovens de espírito enquanto nos mantivermos interessados e atualizados. Agradeço também aos meus manos e às minhas queridas sobrinhas, com quem me sinto sempre tão eu, porque posso espairecer para poder voltar ao foco.

Por fim, agradeço ao meu Miguel, minha companhia de todos os dias, que está sempre presente e que foi, é e será, um apoio incondicional.

Resumo

Face à importância que o Turismo tem no tecido económico e social de Portugal, dado que se trata de um setor bastante volátil e em constante mudança, impõe-se a definição de uma estratégia de atuação futura para perceber como nos podemos reinventar e, assim, poder fazer face a situações de dependência externa.

Com aumento do Turismo em Portugal, bem como das empresas afetas ao mesmo, surge a necessidade de análise e previsão do fluxo de turistas para que a gestão dos negócios seja dotada de uma estratégia competitiva. Perante as alterações na ‘recente’ dinâmica dos dados afetos ao Turismo, decorrentes da pandemia, discutimos o recurso a metodologias clássicas de previsão (modelos de alisamento exponencial) e a metodologias de *Machine Learning* (com recurso a redes neuronais), quando aplicadas a este setor.

O estudo incide na modelação e previsão de séries temporais afetas ao número de dormidas mensais, em estabelecimentos de alojamento turístico em Portugal, entre janeiro de 2002 e março de 2022. Pretendemos assim procurar modelos adequados de previsão (passíveis de serem utilizados pela indústria afeta ao Turismo), sendo uma ferramenta de medida de apoio à tomada de decisão.

Ao estudar as séries temporais referentes ao Turismo, as quais apresentam um padrão de tendência e sazonalidade, constatamos um insucesso das metodologias clássicas na previsão das mesmas. Por outro lado, as metodologias de *Machine Learning* mostraram-se promissoras e uma boa ferramenta para um planeamento estratégico dos negócios associados ao Turismo.

Palavras-chave: Gestão de Empresas, Gestão Estratégica, Turismo, Séries Temporais, *Machine Learning*, Previsão

Abstract

Given the importance that Tourism has in Portugal with a major economic and social impact, being such a volatile and constantly changing sector, it is imperative to define a strategy for future action. We need to understand how we can reinvent ourselves and be able to deal with situations of external dependence.

With the increase of Tourism in Portugal, as well as the companies involved in it, there is a need to analyze and forecast the flow of tourists so that management of the businesses related to it, is endowed with a competitive strategy. The changes in the 'recent' dynamics of data related to Tourism, resulting from the pandemic, were used to discuss the use of classical forecasting methodologies (exponential smoothing models) and Machine Learning methodologies (using neural networks), when applied to this sector.

The study focuses on the modeling and forecasting of time series related to the number of monthly overnight stays, in tourist accommodation establishments in Portugal, between January 2002 and March 2022. Thus, we intend to look for adequate forecast models (which can be used by the industry related to Tourism), being a measurement tool to support decision making.

Studying time series related to Tourism, as they present a trend and seasonal pattern, we noticed the failure of classical methodologies to forecast them. On the other hand, Machine Learning methodologies proved to be promising and a good tool for strategic planning of businesses related to Tourism.

Key words: Business Management, Strategic Management, Tourism, Time Series, Machine Learning, Forecasting

Índice Geral

Agradecimentos	i
Resumo	iii
Abstract.....	v
Índice Geral	vii
Índice de Figuras.....	ix
Índice de Tabelas.....	xi
Introdução.....	1
1. Revisão da Literatura.....	3
1.1. Turismo	3
1.1.1. Turismo: Visão Geral	4
1.1.2. Turismo em Portugal.....	6
1.1.3. Turismo: relação com o setor Empresarial	10
1.1.4. Gestão Estratégica	11
1.2. Previsão	12
1.2.1. Previsão do fluxo de Turismo: contexto de Pandemia	12
1.2.2. Métodos de Previsão	14
2. Enquadramento Teórico	17
2.1. Séries Temporais	17
2.1.1. Noções Gerais	17
2.1.2. Componentes de uma série temporal	18
2.1.3. Quebras de estrutura	19
2.1.4. Previsão em séries temporais	20
2.2. Modelos de Alisamento Exponencial (ETS).....	21
2.2.1. Noções Gerais	21
2.2.2. <i>Single Exponential Smoothing</i>	22
2.2.3. <i>Double Exponential Smoothing</i>	22
2.2.4. <i>Triple Exponential Smoothing</i>	23
2.2.5. Critérios de seleção de Modelos.....	24
2.3. <i>Deep Learning</i>	25

2.3.1.	Da Inteligência Artificial ao <i>Deep Learning</i>	25
2.3.2.	<i>Deep Learning</i> na modelação e previsão de séries temporais	26
2.3.3.	Arquitecturas <i>Perceptron Multilayer</i> e <i>Long Short-Term Memory</i>	28
2.3.4.	<i>Hybrid Model Box & Jenkins</i> – DNN (BJ – DNN)	29
3.	Metodologia	31
3.1.	Apresentação dos Dados	31
3.2.	Implementação Computacional	31
3.2.1	Análise Exploratória de Séries Temporais	32
3.2.2	Modelos de Previsão: Modelos de Alisamento Exponencial (ETS) e Modelos de Redes Neurais	33
3.2.3	<i>Forecasting</i> e Avaliação do Erro (Métricas).....	34
4.	Resultados.....	37
4.1.	Apresentação de Resultados.....	37
4.1.1	Séries Temporais.....	37
4.1.2	Previsão: Alisamento Exponencial (ETS).....	41
4.1.3	Previsão: <i>Deep Learning</i>	45
4.2.	Discussão dos Resultados.....	46
5.	Conclusão.....	49
6.	Referências Bibliográficas	53
7.	Anexos.....	61
7.1	Anexo A	61
7.2	Anexo B.....	65
7.3	Anexo C	67
7.4	Anexo D	69

Índice de Figuras

Figura 1.1 – “The Travelling Companions”	3
Figura 1.2 – Interligação entre Turismo, Voos e Economia.	4
Figura 1.3 – Índice de Competividade do Turismo	5
Figura 1.4 – Número de chegadas (Internacional à esquerda, Europa à direita)	5
Figura 1.5 – Número de chegadas a Portugal de 2007 a 2019.....	6
Figura 1.6 – Alojamentos turísticos.....	7
Figura 1.7 – Dormidas nos estabelecimentos de Turismo de habitação e rural (milhares)	7
Figura 1.8 – Dormidas de residentes por região – Dezembro 2021 vs Dezembro 2019 (%)	8
Figura 1.9 – Nº de dormidas de residentes por região	8
Figura 1.10 – Exemplos de Campanhas “Vá para fora cá dentro”	9
Figura 1.11 – Prémios atribuídos ao Turismo de Portugal	10
Figura 1.12 – Distribuição das empresas afetas ao Turismo em Portugal em 2019	11
Figura 2.1 – Da Inteligência Artificial ao <i>Deep Learning</i>	26
Figura 2.2 – Modelo BJ-DNN (Fonte: Ramos <i>et al.</i> (2022)).....	29
Figura 3.1 – Metodologia de implementação computacional dos modelos de ETS.....	34
Figura 4.1 – Série Dormidas Total ($\times 10^7$) – Representação gráfica (por anos).....	37
Figura 4.2 – Série Dormidas Residentes – Representação gráfica (por anos).....	37
Figura 4.3 – Série Dormidas Total: Representação gráfica dos <i>box-plots</i> anuais	38
Figura 4.4 – Série Dormidas Residentes: Representação gráfica dos <i>box-plots</i> anuais.....	38
Figura 4.5 – Série Dormidas Total: Representação gráfica dos <i>box-plots</i> mensais.....	39
Figura 4.6 – Série Dormidas Residentes: Representação gráfica dos <i>box-plots</i> mensais.....	39
Figura 4.7 – Mudanças/quebras estruturais (série Total à esquerda, série Residentes à direita) ..	41
Figura 4.8 – Critérios de Informação para modelos ETS da série Dormidas Total.....	42
Figura 4.9 – Critérios de Informação para modelos ETS da série Dormidas Residentes	42
Figura 4.10 – Modelos ETS (<i>fitting</i> e <i>forecasting</i>) da série Dormidas Total	43
Figura 4.11 – Modelos ETS (<i>fitting</i> e <i>forecasting</i>) da série Dormidas Residentes.....	43
Figura 4.12 – Resíduos de modelos ETS da série Dormidas Total	44
Figura 4.13 – Resíduos de modelos ETS da série Dormidas Residentes.....	45
Figura 4.14 – Modelos BJ-DNN: LSTM da série Dormidas Total.....	45

Figura 4.15 – Modelos BJ-DNN: LSTM da série Dormidas Residentes.....	46
Figura 7.1 – Série Total: Decomposição (aditiva e multiplicativa)	65
Figura 7.2 – Número Total: Histograma (com curva de densidade).....	66
Figura 7.3 – Série Total – Função cumulativa de distribuição	66
Figura 7.4 – Série Residentes: Decomposição (aditiva e multiplicativa)	67
Figura 7.5 – Número Residentes: Histograma (com curva de densidade).....	68
Figura 7.6 – Série Residentes: Função cumulativa de distribuição	68
Figura 7.7 – Modelo ETS da Série Total	69
Figura 7.8 – Modelo BJ-DNN da Série Total	69
Figura 7.9 – Modelo ETS da Série Residentes	70
Figura 7.10 – Modelo BJ-DNN da Série Residentes	70

Índice de Tabelas

Tabela 3.1 – Classificação dos modelos ETS	33
Tabela 4.1 – Estatísticas descritivas.....	39
Tabela 4.2 – Testes de normalidade, estacionariedade e independência	40
Tabela 4.3 – Valores da “força de tendência” e da “força de sazonalidade”	41
Tabela 4.4 – Previsões segundo os modelos ETS (MAPE).....	44
Tabela 4.5 – Previsões segundo os modelos BJ-DNN (MAPE).....	46
Tabela 4.6 – Comparações entre valores reais e previsões	47

Introdução

O setor do Turismo é um motor para se criar, manter e desenvolver empresas, sendo uma atividade económica fundamental para a geração de emprego e de riqueza em Portugal. Por se tratar de uma atividade em constante crescimento, as empresas devem ter uma Gestão Estratégica que sustente o negócio. Como se sabe, as previsões que poderiam ser feitas relativamente ao Turismo sofreram uma quebra impensável com a pandemia que começou em 2020. Surge assim o problema relativamente à forma de se fazer previsões, dado que estas permitem que os gestores tomem decisões informadas e que garantam, dentro do possível, a sustentabilidade do negócio em Portugal.

Este estudo tem como objetivo discutir como metodologias de *Machine Learning*, nomeadamente de *Deep Learning*, comparativamente a metodologias clássicas, em particular de Alisamento Exponencial, podem potenciar a forma de prever o fluxo de Turismo. Com efeito, pretende-se alavancar o conhecimento científico sobre o fluxo de Turismo, para que as decisões tomadas na gestão dos negócios deste setor estejam assentes em dados concretos, garantindo o sucesso das organizações e das pessoas envolvidas. Para tal, partindo do tópico central, definiram-se os objetivos específicos da investigação a seguir identificados.

1. Apresentar uma perspetiva geral do Turismo em Portugal;
2. Analisar o paradigma associado ao Turismo interno, antes e durante a pandemia;
3. Discutir os benefícios da previsão para uma Gestão Estratégica em negócios do setor do Turismo;
4. Analisar e caracterizar séries temporais relacionadas com as dormidas em alojamentos turísticos nos últimos 20 anos;
5. Encontrar modelos de previsão de séries temporais (com recurso a metodologias clássicas e de *Machine Learning*) capazes de captar a dinâmica das séries temporais afetas ao Turismo;
6. Apontar diretrizes numa perspetiva de Gestão Estratégica, decorrente dos resultados observados.

Consequentes dos objetivos, foram definidas também as questões de investigação:

1. Quais as alterações (caso existam) na dinâmica do fluxo de Turismo decorrente da pandemia?
2. Que potencialidades/limitações encontramos nas metodologias clássicas e de *Machine Learning* na previsão do fluxo de Turismo?
3. Qual a importância da utilização da informação obtida na modelação e previsão para as empresas afetas ao setor do Turismo?
4. Que diretrizes podem ser apontadas, numa perspectiva de Gestão Estratégica, para desenvolver o Turismo interno?

O presente estudo torna-se surpreendente na medida em que a aplicação de metodologias clássicas de alisamento exponencial mostram resultados ineficientes na previsão do fluxo de turistas, dada a quebra abrupta dos dados no ano de 2020. O que reforça a mais-valia que traz apostar em metodologias modernas de *Machine learning*.

A comparação entre a previsão de turistas no geral e de turistas residentes, mostra-se interessante por espelhar de forma evidente o impacto dos portugueses no Turismo do seu próprio país, especialmente no período da pandemia, fazendo-nos refletir sobre a melhor estratégia para os negócios associados ao setor.

Na estrutura da dissertação encontra-se a Revisão da Literatura, onde é explorada a definição de Turismo e o impacto deste em Portugal. É ainda aprofundado o que é a Gestão Estratégica bem como a importância de previsão para a mesma. Seguidamente é feito um levantamento relativamente a metodologias de previsão. Pela bibliografia consultada, selecionam-se duas metodologias, uma clássica e uma de *Machine Learning*, com o intuito de aplicar as duas.

No capítulo seguinte é feito o Enquadramento Teórico relativo ao estudo de séries temporais, metodologias de Alisamento Exponencial e metodologias de *Machine Learning* (com ênfase no *Deep Learning*). Seguem-se a este os capítulos afetos à Metodologia, onde é descrita a implementação computacional, e à apresentação/discussão dos Resultados, onde os mesmos são apresentados e discutidos, de forma a comparar as metodologias implementadas neste trabalho.

Por fim é apresentada a Conclusão onde se pretende dar resposta às questões de investigação e são levantadas as limitações e contributos deste estudo. O trabalho termina com a Bibliografia e Anexos.

1. Revisão da Literatura

Neste capítulo desenvolvemos o conceito de Turismo, desde uma visão geral até ao impacto que tem em Portugal. Seguidamente relacionamos o Turismo com o setor empresarial e explicamos em que consiste a Gestão Estratégica. Temos também uma secção onde introduzimos o que é a Previsão, e os respetivos métodos, e explicamos como esta pode ser aplicada ao fluxo de Turismo.

1.1. Turismo

A movimentação humana é uma característica inata demonstrada pela humanidade ao longo da História (Var & Gunn, 2002). Segundo os autores, as viagens podiam acontecer por diversas razões: visitar amigos, curiosidade, melhorar a qualidade de vida, peregrinações, comércio e desejo de viajar.



Figura 1.1 – “The Travelling Companions”

Fonte: Egg (1962)

À medida que a qualidade das viagens foi evoluindo, acabou por se definir o conceito de Turismo como o movimento temporário de pessoas para destinos fora dos habituais locais de trabalho e de residência, bem como as atividades realizadas durante a estadia nesses destinos e as instalações criadas para satisfazer essas necessidades (Mathieson & Wall, 1982). Esta definição foi completada e atualmente considera-se Turismo quando “a viagem dura de 24 horas até 12 meses consecutivos para lazer, negócios e outros fins” (BCcampus, 2008).

O Turismo acabou por se tornar num dos mais importantes setores económicos a nível mundial. Por isso, a Organização das Nações Unidas estabeleceu, em 1980, no dia em que foi celebrado o 10º aniversário da adoção dos Estatutos da Organização Mundial do Turismo das Nações Unidas, que 27 de setembro seria o Dia Mundial do Turismo (ONU, 2021).

1.1.1. Turismo: Visão Geral

O Turismo é um dos principais setores em termos económicos para os países desenvolvidos ou em vias de desenvolvimento (Ramos & Costa, 2017). Segundo os mesmos autores, está em constante interligação com todos os outros setores da vida (económico, social, cultural, ambiental e político). O Turismo de lazer representa mais de três quartos das receitas mundiais do setor, mas em 2019 o peso do Turismo de negócios é mais do dobro de 2010 (Vieira, 2021). Posto isto, verifica-se como o Turismo faz parte de um ciclo que se interliga tanto com o tráfego aéreo bem como o crescimento económico, como se pode ver na Figura 1.2.



Figura 1.2 – Interligação entre Turismo, Voos e Economia.

Fonte: Adaptado de Cristina (2017)

Ainda, segundo Cristina (2017), o crescimento económico contribui para o desenvolvimento do Turismo, proporcionando infraestruturas melhoradas e padrões de vida mais elevados. Verificamos seguidamente o posicionamento de Portugal comparativamente com outros países.

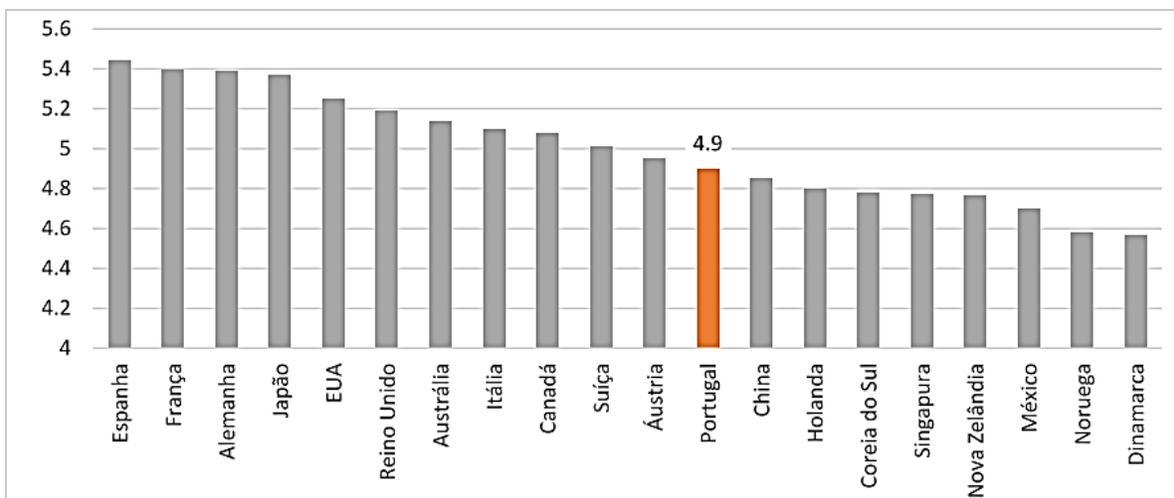


Figura 1.3 – Índice de Competitividade do Turismo

Fonte: BPI Research (2022)

A Figura 1.4 mostra como os países mais desenvolvidos têm números superiores relativamente às chegadas, o que se reflete também no Turismo.

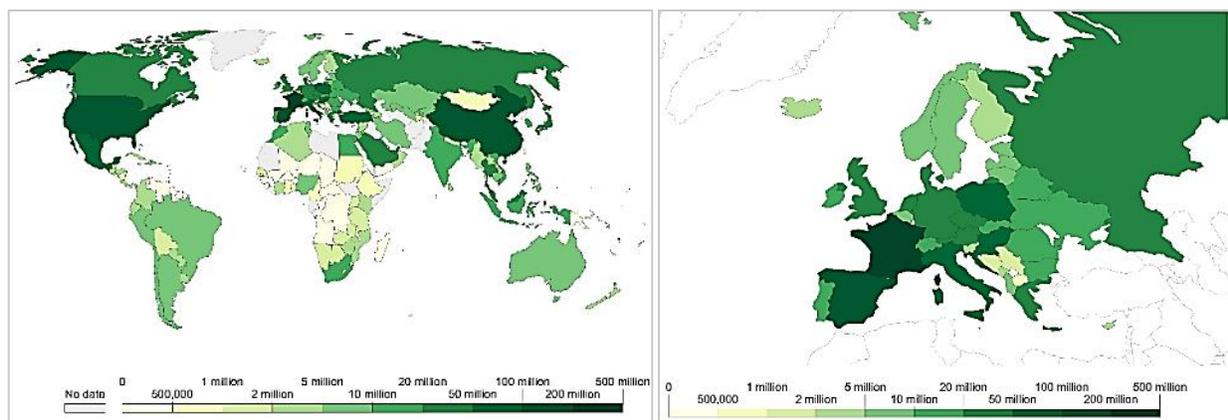


Figura 1.4 – Número de chegadas (Internacional à esquerda, Europa à direita)

Fonte: Roser (2019)

Como se pode verificar na Figura 1.4, o Turismo atinge valores mais elevados no Hemisfério Norte, nomeadamente nos EUA, Europa e China. Especificamente na Europa, a maioria dos países teve acima de 5 milhões de chegadas em 2019, podendo chegar, no caso de França, aos 212 milhões de chegadas.

De acordo com a mesma fonte, é possível verificar que em Portugal, de 2015 para 2019 houve mais 5,45 milhões de visitantes, o que corresponde a um aumento de 46% no espaço de quatro anos (Roser, 2019).

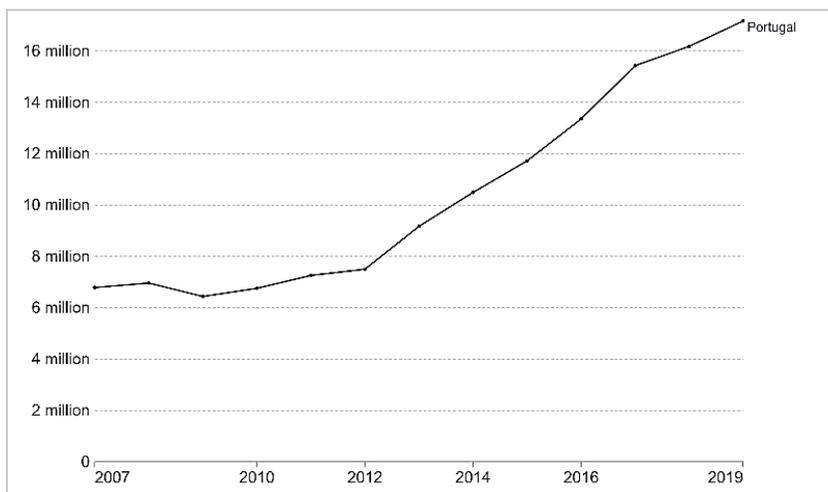


Figura 1.5 – Número de chegadas a Portugal de 2007 a 2019

Fonte: Roser (2019)

1.1.2. Turismo em Portugal

O setor do Turismo é uma atividade económica fundamental para a geração de riqueza e emprego em Portugal, de tal forma que se encontra sob a tutela do Ministério da Economia e da Transição Digital, com uma Secretaria de Estado do Turismo, responsável pela definição de políticas na área do Turismo. O Turismo de Portugal é a Autoridade Turística Nacional, organismo técnico central integrado no Ministério da Economia e da Transição Digital, com jurisdição sobre todo o território nacional (Turismo de Portugal, 2022). Em Portugal, o Turismo tem sido, nas últimas décadas, um dos motores da economia nacional. As previsões são de crescimento num país com culturas e tradições variadas onde a principal aposta deve ser no autêntico e diferenciador que nos distingue como país, onde a multiculturalidade é um fator determinante e convidativo (Ramos & Costa, 2017).

O Turismo tem crescido ano após ano devido a características únicas deste pequeno país – Algarve, vale do Douro, cidades históricas como Lisboa e Porto, variedade gastronómica e de vinho, um povo simpático e acolhedor, preços baratos para os padrões europeus e segurança (Magano & Cunha, 2020).

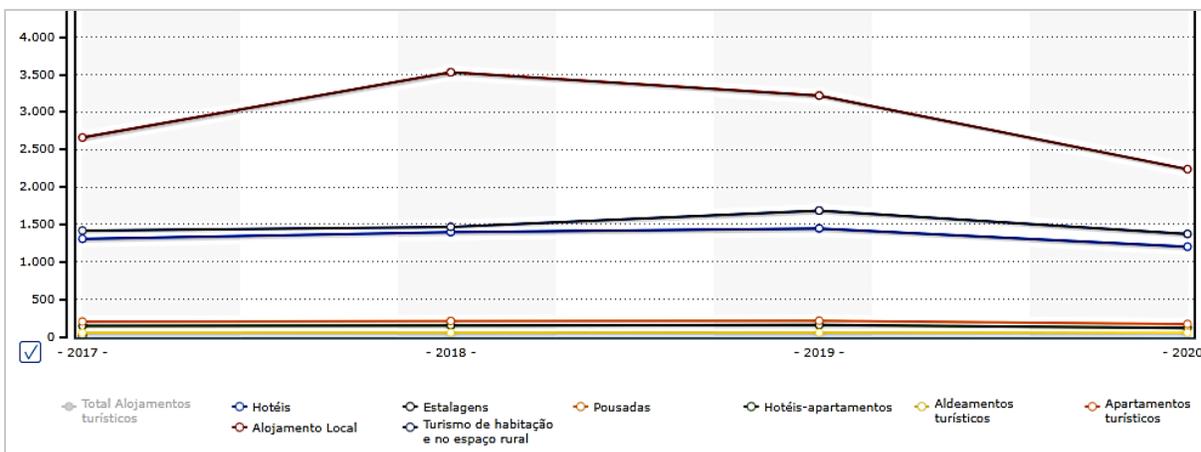


Figura 1.6 – Alojamentos turísticos

Fonte: PORDATA (2022)

De acordo com a Figura 1.6, os alojamentos turísticos com mais dormidas, entre 2017 e 2020, foram o Alojamento Local, Turismo de Habitação e Espaço Rural e Hotéis. Considerando as dormidas em Turismo de Habitação e Espaços Rurais, verifica-se no gráfico em seguida que, desde 2014, as dormidas tanto em estabelecimentos de Turismo de Habitação como de Turismo Rural têm crescido de ano para ano. Nota-se uma quebra em 2012 e 2020, devido, respetivamente, a questões económicas e pandémicas.

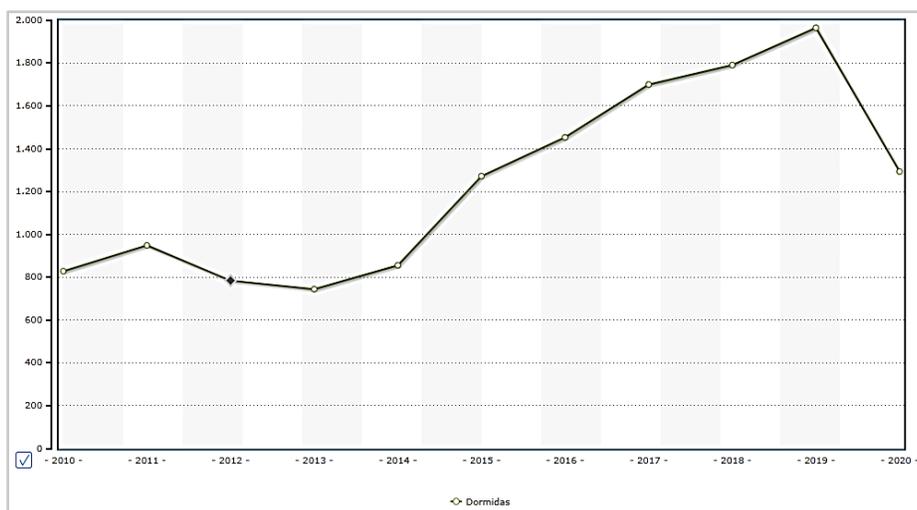


Figura 1.7 – Dormidas nos estabelecimentos de Turismo de habitação e rural (milhares)

Fonte: PORDATA (2022)

De 2009 até 2015 houve um aumento de 5,8% na capacidade de alojamento em Hotéis de 4 e 5 estrelas (Turismo de Portugal, 2017). Ao analisar as dormidas nas diferentes regiões de Portugal continental e ilhas, verificamos seguidamente que a procura se manteve, exceto no Algarve, que teve cerca de 70% menos dormidas em 2021 do que 2019 e a Madeira que aumentou quase 130%.

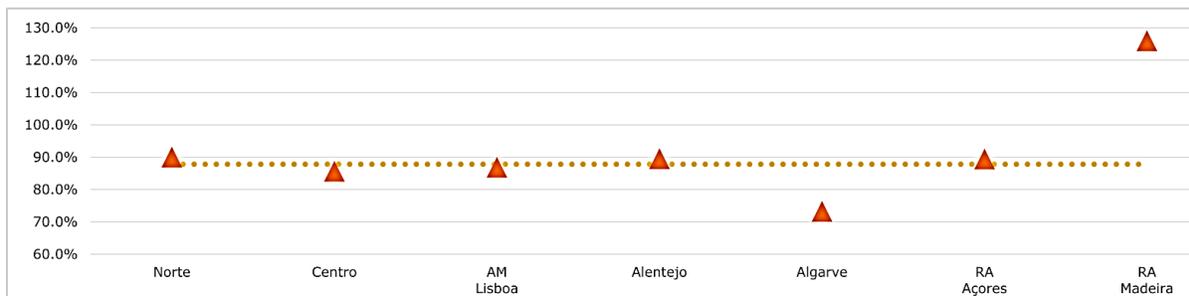


Figura 1.8 – Dormidas de residentes por região – Dezembro 2021 vs Dezembro 2019 (%)

Fonte: BPI Research (2022)

O Turismo, nesta fase da história, apresenta um crescimento muito acentuado, nomeadamente em algumas zonas do país como o Algarve, a zona de Lisboa e a Madeira (Ramos & Costa, 2017). Os invernos amenos, temperaturas quentes de verão e uma oferta diversificada de destinos costeiros e de montanha fizeram de Portugal um destino de férias ideal na Europa — tanto que em 2021 o país liderou o ranking dos melhores destinos europeus para visitar de carro, entre outras coisas, devido seu clima, segurança, acessibilidade e pontos de referência (López, 2021).

Na Figura 1.9 é demonstrada a média de estadia bem como a percentagem de hóspedes residentes e não residentes.

	Estada média (dias)	Taxa líquida de ocupação-cama	% de Hóspedes		% Dormidas	
			Residentes	Não Residentes	Residentes	Não Residentes
2019	2,58	47,3%	39,5%	60,5%	30,0%	70,0%
2020	2,47	24,1%	62,5%	37,5%	52,7%	47,3%
2021	2,58	27,6%	59,0%	41,0%	50,0%	50,0%

Figura 1.9 – Nº de dormidas de residentes por região

Fonte: BPI Research (2022)

Dado que o Turismo a nível mundial tem tido um crescimento tanto no setor do lazer, mas também de negócios, o nosso país tem acompanhado essa tendência. Portugal entrou no top 10 do ranking mundial dos destinos mais procurados para a realização de congressos, conferências e

eventos internacionais (Larguesa, 2017). Assistiu-se a uma alteração do *mix* na tipologia de hóspedes face ao contexto pré-pandémico, com incremento do peso dos hóspedes residentes, que se atenuou ligeiramente em 2021 (BPI Research, 2022).

Assim, assistimos a um novo paradigma, onde não só os estrangeiros querem desfrutar destes destinos para férias, como os próprios *locals* têm aproveitado para explorar e conhecer mais a fundo o país onde vivem. Têm surgido, assim, diferentes campanhas com o *slogan* “Vá para fora cá dentro”, em revistas, redes sociais e canais de televisão.



Figura 1.10 – Exemplos de Campanhas “Vá para fora cá dentro”
Fontes: RTP (2021), TV Guia (2018), Turismo de Portugal (2013)

Ao ter mais portugueses a explorar e conhecer novas zonas do país, desde o litoral ao interior, os próprios portugueses acabam por desempenhar o papel de turistas e de desfrutar programas culturais, gastronómicos e lúdicos, o que se revela benéfico para a economia do país.

Devido ao seu crescimento, o Turismo de Portugal tem recebido vários prémios, distinções e incentivos para se desenvolver e tornar o país mais próspero. Em 2021 recebeu 3 prémios, em 2020 e 2019 recebeu 7 e 9 prémios respetivamente, em 2018 e 2017 foram recebidos 4 prémios em cada ano, e de 2008 a 2016 recebeu 8 prémios (Turismo de Portugal, 2021a). Os diferentes prémios estão descritos na seguinte figura, cujos detalhes poderão ser consultados no Anexo A. Considera-se relevante destacar os *World Travel Awards*. Considerados os “óscars” do Turismo, estes prémios são atribuídos anualmente, desde 1993, pelos profissionais do setor a uma escala mundial, e na edição de 2021 a Madeira continua a ser o Melhor Destino Insular do Mundo, e o Algarve o Melhor destino de Praia do Mundo (BPI Research, 2022).

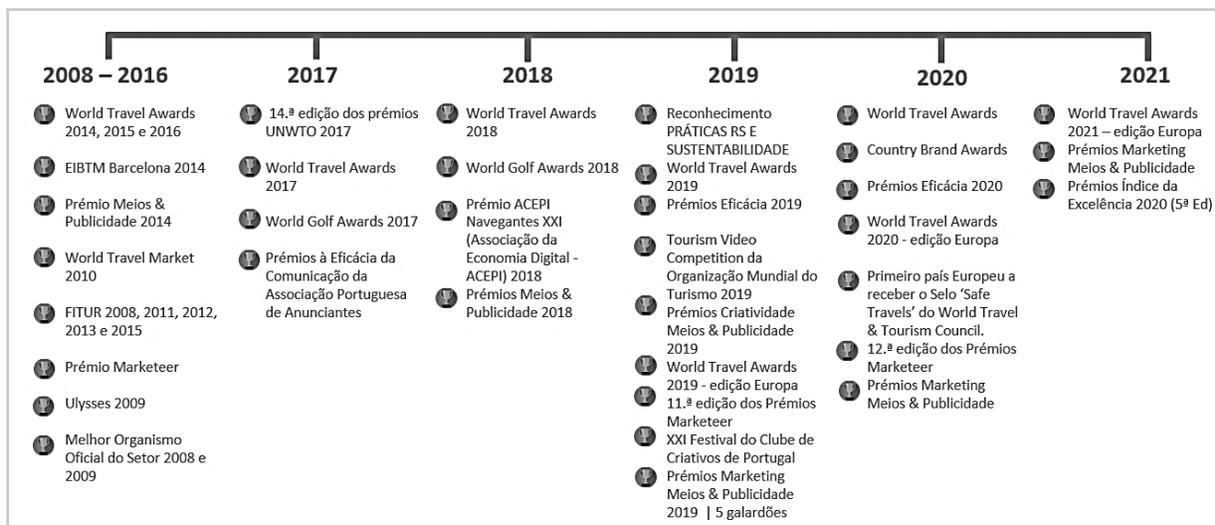


Figura 1.11 – Prémios atribuídos ao Turismo de Portugal

Adaptado de: Turismo de Portugal (2021a)

Estes reconhecimentos e conseqüente crescimento exige uma Gestão Estratégica dos negócios associados ao Turismo.

1.1.3. Turismo: relação com o setor Empresarial

A atividade turística emprega aproximadamente meio milhão de pessoas, e tem um forte impacto nas exportações nacionais (Ramos & Costa, 2017). Portugal passou a ter um rácio superior a 1 (um turista por cada habitante) a partir do final de 2014, e este indicador não tem parado de subir (Botelho, 2019). Segundo López (2021), Lisboa foi por larga margem o destino de viagem mais popular de Portugal para turistas internacionais em 2019 e 2020 e como resultado, teve o maior número de aberturas de hotéis e albergues no último ano. A mesma autora publicou ainda um estudo relativo à dimensão das empresas afetas ao Turismo, em Portugal, mostrando que a esmagadora maioria são microempresas (90,2%).

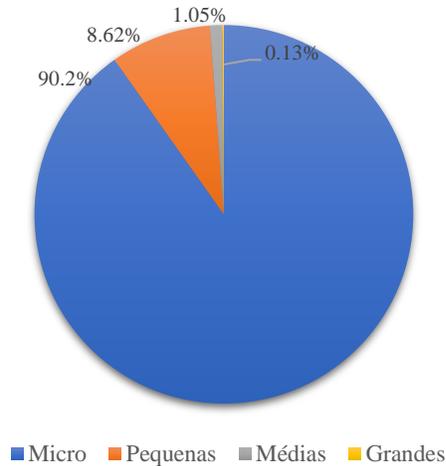


Figura 1.12 – Distribuição das empresas afetas ao Turismo em Portugal em 2019
Adaptado de: López (2021)

Em termos absolutos, o número de empresas de Turismo a operar em Portugal em 2019 ascendeu a mais de 70 mil, o que representa um incremento de cerca de 8% face ao ano anterior (López, 2021). Mais, segundo a mesma autora, ao longo da década de 2010 – 2020, o número de empreendimentos turísticos no país europeu cresceu de forma constante em relação aos quase 48 mil registados em 2010. Assim, o Turismo tornou-se um fator impulsionador da economia local/regional do mesmo, o qual gerou riqueza, emprego e dinamismo a uma região. Contudo deu-se uma situação pandémica desde 2020 até agora. O setor do alojamento turístico registou que os níveis atingidos em setembro de 2021 foram inferiores aos observados em setembro de 2019, tendo diminuído o número de hóspedes e de dormidas, 28,9% e 26,6%, respetivamente (INE, 2021).

A evolução do Turismo permite ou tem a capacidade de permitir que as populações conheçam a sua história, as suas origens e que se possam reinventar e transcender em termos de desenvolvimento e ideias inovadoras, por isso o Turismo funciona como um “motor” das economias, agregando princípios base de genuinidade, tradição, cultura e de novas estratégias de gestão e organização do setor (Ramos & Costa, 2017).

1.1.4. Gestão Estratégica

A Gestão Estratégica consiste em continuamente planear, monitorizar, analisar e avaliar todas as necessidades que uma organização precisa para atingir as suas metas e objetivos. Tem como

objetivo ajudar o empresário a enfrentar a complexidade do presente e a incerteza do futuro, proporcionando-lhe meios com os quais possa medir as suas decisões, na resolução de problemas estratégicos relativos à sobrevivência e prosperidade da empresa (Faulkner & Campbell, 2003). Compreender as fontes de vantagem competitiva sustentada para as empresas, tornou-se uma importante área de investigação no campo da Gestão Estratégica (Porter, 1985). A Gestão Estratégica é o motor da atividade turística, sendo o que promove a eficiência organizacional em qualquer entidade envolvida no Turismo (Kirovska, 2011).

Com o aumento da concorrência e da globalização, a dinâmica e o desenvolvimento do Turismo são cada vez mais intensificados, a Gestão Estratégica deve assegurar operações comerciais eficientes e eficazes, tais como a gestão de recursos, sistemas de informação de gestão, manutenção e desenvolvimento das relações com o Turismo, expansão e desenvolvimento do negócio turístico, gestão constante a projetos de Turismo novos e inovadores (Kirovska, 2011). Em determinadas regiões, o Turismo tornou-se um pilar estratégico da indústria, dadas as suas contribuições cada vez mais significativas para o produto interno bruto local (Zhang *et al.*, 2021). Assim, deve estudar-se de que forma a previsão do fluxo de Turismo pode levar a uma perspetiva estratégica do desenvolvimento do negócio e, conseqüentemente, do local e do país.

1.2. Previsão

1.2.1. Previsão do fluxo de Turismo: contexto de Pandemia

A previsão da procura de Turismo é fundamental para a estratégia de planeamento, uma vez que permite a expansão dos serviços turísticos, tanto no investimento em recursos humanos como em estruturas físicas, para poder atender às necessidades das pessoas (Casanova, 2012). Os serviços turísticos têm uma natureza perecível, o que torna necessária uma previsão precisa da procura do Turismo (Gunter & Önder, 2015). Outro aspeto significativo de dados da procura turística é a natureza da sazonalidade que deve ser identificada antes da seleção do modelo de previsão ideal (Goh & Law, 2002). De um modo geral, a procura turística tem apresentado um crescimento sustentado, no entanto, trata-se de um mercado que tem sofrido diversas oscilações devido à volatilidade de fatores determinantes e intervenções externas (Song *et al.*, 2019).

A indústria do Turismo experimentou efeitos negativos significativos durante os chamados eventos de crise do "cisne negro", como a crise financeira em 1997 e 2008, a epidemia de síndrome

respiratória aguda grave (SARS) em 2003 e vários terremotos e episódios de agitação social (Zhang *et al.*, 2021). E em 2020 foi declarada a pandemia do COVID-19. O Turismo está entre as indústrias mais afetadas negativamente por esta pandemia, e tanto as empresas de Turismo como as organizações contam com a previsão de recuperação ao preparar os seus planos de recuperação de crise (Zhang *et al.*, 2021). A Organização Mundial de Turismo das Nações Unidas (OMTNU) relatou que, até 20 de abril de 2020, todos os principais destinos turísticos haviam implementado restrições de viagem em resposta à pandemia COVID-19 (OMTNU, 2020). A previsão precisa da escala total do impacto [da pandemia COVID-19] na indústria do Turismo e a recuperação do mercado é crítica para o planeamento estratégico de destinos turísticos e empresas relacionadas com o Turismo (Song & Li, 2021). De acordo com López (2021), antes da pandemia do COVID-19, o setor do Turismo representava quase um quinto da economia portuguesa, mas a crise global fez com que a contribuição do Turismo para o PIB caísse mais de metade em 2020 ¹. Esta quebra, alinhada com a literatura citada, poderá consistir num entrave e gerar dificuldades ao fazer previsões. Assim, muitas organizações contam com a previsão para preparar os seus planos de recuperação de crise (Zhang *et al.*, 2021).

Nas últimas cinco décadas, houve uma grande quantidade de literatura sobre modelação e previsão da procura turística através de várias metodologias (Song *et al.*, 2019). Segundo Atkinson (2016), as metodologias estatísticas não podem capturar os impactos de eventos inesperados repentinos, como doenças, desastres ou outras crises, nas previsões. Conclui-se que não existe apenas um método que possa desenvolver as melhores previsões para todos os problemas, tal que os métodos de previsão combinados são melhores em comparação com os métodos tradicionais (Ghalekhondabi *et al.*, 2019).

Devido à utilização de tecnologias sofisticadas, as empresas são cada vez mais capazes de melhorar a sua eficiência e o desempenho dos seus negócios (Antonova, 2014). A mesma autora defende que as tecnologias de informação desempenham um papel essencial nas crescentes reduções de custos de transação, permitindo o aparecimento de novos modelos de negócio, e criando fontes adicionais de criação de valor para os clientes, bem como incentivam as empresas a tornarem-se mais diretas, sem fronteiras, empreendedoras, orientadas a processos e projetos e a

¹ Apesar deste revés, Portugal manteve-se entre os três países mais dependentes do Turismo da União Europeia nesse ano. O Turismo de Portugal lança o Programa Empresas Turismo 360, uma medida inscrita no Plano “Reativar o Turismo. Construir o Futuro” que pretende colocar as empresas no centro do processo de transformação sustentável do Setor (Turismo de Portugal, 2021b).

desenvolver modelos de negócio globais, complexos e inovadores. Assistimos, deste modo, a uma recente revolução, possibilitada em parte pela utilização de sistemas inteligentes, e que dá origem a uma nova onda de tecnologias (Atkinson, 2016).

1.2.2. Métodos de Previsão

A previsão pode ser feita através de diferentes vias, sendo uma das possibilidades o recurso a séries temporais que de algum modo traduzem “fluxos” afetos ao Turismo em função do tempo. São inúmeras as famílias de modelos propostas na literatura para a análise de uma série temporal (Gonçalves & Lopes, 2008). Outra das possibilidades utilizadas neste trabalho é o *Deep Learning*, que é uma área do *Machine Learning*. Classificar os métodos de previsão como “*Machine Learning*” ou “Estatístico” tornou-se comum na literatura e também na Comunidade de Previsão (Januschowski *et al.*, 2020). Diversos estudos têm sido desenvolvidos na área da previsão, e são variados os métodos utilizados:

- Num estudo de Camara *et al.* (2021), com o objetivo de analisar os principais fatores económicos determinantes da procura turística internacional para o estado do Rio de Janeiro, foram utilizados modelos econométricos de Regressão Linear Múltipla, com dados anuais de 2000 a 2017. Neste estudo, os resultados contrariam proposições teóricas e evidências empíricas amplamente estabelecidas na literatura, o que pode estar associado à fragilidade das estatísticas de Turismo no Brasil e/ou à natureza simplificada do modelo estatístico adotado (Camara *et al.*, 2021).
- Por outro lado, Silva (2021) apresenta o processo de criação e uso de Modelos de Regressão Linear Simétricos para amostras de tamanho pequeno. Segundo o autor, existem métodos de regressão linear para modelar funções que possuem erros aleatórios com distribuição simétrica.
- Existe ainda um estudo onde é proposta a identificação de modelos de séries temporais baseada no algoritmo de Informação Mútua Parcial (PMI), um critério que leva em conta tanto as relações lineares, como as relações não-lineares existentes entre as variáveis consideradas na análise (Luna *et al.*, 2006).
- Encontrou-se um estudo que compara previsões da taxa de inflação mensal brasileira a partir de diferentes modelos lineares e não lineares de séries temporais e da curva de Phillips (Arruda *et al.*, 2011).

- Num estudo de Simões *et al.* (2019) foi utilizado o método de Alisamento Exponencial. Os autores referem como as principais características estão relacionadas com a facilidade de uso e a aplicabilidade a um grande número de séries temporais, o que pode devolver resultados muito satisfatórios dependendo de condições como tendência, sazonalidade, ciclo e componente aleatório, além de permitir uma boa compreensão dos resultados.

- Foi ainda encontrado um estudo com o objetivo de gerar previsões de curto, médio e longo prazo e comparar a precisão dos modelos em cada horizonte de previsão. Este estudo compara diferentes modelos, entre os quais o modelo de série temporal *Holt-Winters* (HW) e a rede neural (Porto *et al.*, 2020). O modelo HW tem muita utilidade para os decisores e planeadores estratégicos que visam um modelo económico relativamente preciso com a finalidade de amplo horizonte de previsão e significativo número de séries temporais (Brandon *et al.*, 1987). Quanto aos resultados obtidos no estudo de Porto *et al.* (2020), mostrou-se que a rede neural fornece previsões mais precisas da procura do Turismo do que *Holt-Winters*. Outro ponto levantado pelos autores, refere-se ao facto de se terem limitado a apenas uma série temporal, sugerindo assim a utilização de várias séries.

Tendo em conta a diversidade de estudos desenvolvidos na área da previsão, foram seleccionadas metodologias baseadas no estudo de Ramos (2021). Assim, segue-se o Enquadramento Teórico relativamente às metodologias utilizadas no estudo das Séries Temporais: Modelos de Alisamento Exponencial (*Exponential Smoothing*) e Modelos de *Deep Learning*.

2. Enquadramento Teórico

Ao longo deste Capítulo serão referidos tópicos afetos ao estudo de séries temporais. Toda a secção tem por base Ramos (2021), cujo trabalho serviu de consulta e orientação para os conceitos, esclarecimentos e definições apresentados seguidamente.

2.1. Séries Temporais

2.1.1. Noções Gerais

O registo de fenómenos a variar com o tempo tem o nome de série temporal (Kitagawa, 2010), sendo um conceito que se caracteriza pelo conjunto de observações feitas em períodos consecutivos e durante um determinado intervalo de tempo – discreto ou contínuo. Este capítulo procura introduzir alguns conceitos fundamentais em séries temporais e que, de algum modo, serão fundamentais para a abordagem aqui desenvolvida ².

Uma série temporal tem um grande interesse prático em muitas disciplinas, pois permite descobrir, com alguma margem de erro, os valores futuros de uma série a partir de seus valores passados (Tealab, 2018). As séries temporais podem ser constituídas por registo de um único fenómeno ao longo do tempo (séries temporais univariadas) ou obtidas pelo registo simultâneo de dois ou mais fenómenos (séries temporais multivariadas). No nosso caso será seguida uma metodologia univariada (ou extrapolativa), com base na modelação/previsão de séries temporais.

De uma forma geral, qualquer modelo contém “quantidades” que são observáveis e outras que não são diretamente observáveis. A regressão, enquanto técnica estatística de modelação, é um modelo probabilístico onde a variável endógena (dependente) é a soma duas componentes (uma determinística e uma estocástica ou aleatória, u_t), como se pode verificar na Equação (1).

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + u_t \quad (1)$$

onde y_t representa a variável endógena (dependente) e o membro da direita é soma da expressão linear (em que os x_{it} , com $i = 1, \dots, k$, representam as k variáveis exógenas e os β_j , com $j = 0, \dots, k$, os coeficientes de regressão a estimar) com a componente aleatória, u_t (convencionalmente designada por erro do modelo).

² Existem modelos fundamentais que não serão aprofundados neste trabalho por não se considerarem relevantes para a análise feita.

Porém, sendo o nosso foco de interesse a modelação/previsão em séries temporais, será seguida uma metodologia univariada ou extrapolativa, pelo que se despreza a possibilidade de existirem relações entre a variável de interesse, y_t , e outras variáveis (Ramos, 2021) sendo o comportamento de uma série temporal baseia-se nos valores passados, ou seja, no seu histórico, e ainda nos valores correntes e passados do erro do modelo ³. Quando se considera uma série temporal y no instante t (num intervalo de tempo T), $\{y_t\}$ com $t \in T$, esta pode ser descrita como função (linear) dos valores passados e alguma perturbação (u_t), conforme descrito na Equação (2):

$$y_t = f(y_{t-1}, y_{t-2}, \dots) + u_t \quad (2)$$

2.1.2. Componentes de uma série temporal

Uma série temporal desempenha um comportamento em função das suas componentes – tendência, ciclicidade, sazonalidade e aleatoriedade ⁴. A decomposição da série temporal deve ser considerada de início (no momento da sua caracterização), porque além de melhorar a compreensão da série, pode ser necessária no processo de modelação/previsão. Segundo Makridakis *et al.* (1998), deve-se analisar e extrair dinâmicas (tendência, ciclicidade e sazonalidade) para melhor compreender o comportamento da série e aumentar a precisão das previsões.

Em termos matemáticos, a decomposição da série temporal $\{y_t\}$ pode ser escrita de duas perspectivas: decomposição aditiva e decomposição multiplicativa. Dado que neste estudo se trata de uma variação no padrão sazonal que evidencia uma proporcionalidade, é mais apropriado utilizar a decomposição multiplicativa Equação (3).

$$y_t = CT_t \times S_t \times R_t \quad (3)$$

³ No entanto, um modelo só é válido se a componente aleatória, u_t , corresponder a um processo estocástico puro, ruído branco (*White Noise – WN*), ou seja, se toda a informação relevante estiver contida na parte determinística (Ramos, 2021).

⁴ **Tendência** reflete uma evolução, a longo prazo, no sentido da monotonia da série (crescente/decrescente), podendo apresentar uma dinâmica linear ou não linear, sendo facilmente identificada através de uma análise gráfica; **Ciclicidade** é traduzida por um padrão de flutuações de médio prazo, podendo ou não ser periódicas, as quais podem afetar diretamente a tendência global da série; **Sazonalidade** reflete-se em variações cíclicas regulares que ocorrem em períodos constantes e menores comparativamente à ciclicidade; **Aleatoriedade** está associada a flutuações imprevisíveis, não modeladas pelas componentes anteriores (Ramos, 2021).

onde CT_t , representa a componente do ciclo de tendência, S_t representa a componente sazonal, e R_t representa a componente residual (aleatória).

A decomposição de séries temporais pode, ainda, ser usada para medir a “força da tendência” e a “força da sazonalidade” presentes na série temporal (Wang *et al.*, 2006). A “força de tendência” é definida pela Equação (4), que é medida pelo quociente entre a variância de R_t , $Var(R_t)$, e a variância da soma de CT_t com R_t , $Var(CT_t + R_t)$. Uma série com valor de \mathcal{F}_T próximo de 1 evidencia uma forte tendência, enquanto numa série sem tendência, o valor será 0.

$$\mathcal{F}_T = \max \left(0, 1 - \frac{Var(R_t)}{Var(CT_t + R_t)} \right) \quad (4)$$

A “força de sazonalidade” é definida pela Equação (5), onde uma série com valor de \mathcal{F}_S próximo de 0 exibe pouca ou nenhuma sazonalidade, enquanto numa série com forte sazonalidade, o valor será próximo de 1. Comparativamente à \mathcal{F}_T , neste caso em vez de usar a componente relativa ao ciclo de tendência, CT_t , utiliza-se a componente relativa à sazonalidade, S_t .

$$\mathcal{F}_S = \max \left(0, 1 - \frac{Var(R_t)}{Var(S_t + R_t)} \right) \quad (5)$$

Ainda, um ponto de interesse na previsão de séries temporais prende-se com o facto de a previsão poder ser feita *in-sample* – produz resultados para o intervalo temporal da amostra da série temporal em estudo; e/ou *out-of-sample* – é utilizada para a obtenção de valores futuros para uma data “fora” da janela temporal da série (Marques, 2017). Segundo o mesmo autor, um dos problemas, comumente referido na literatura, encontrado no estudo de séries temporais é a presença de quebras de estrutura. Segundo a literatura científica existe uma forte relevância (com possíveis consequências) em estudos empíricos onde é detetada a presença de quebras de estrutura em séries temporais (Ramos, 2021).

2.1.3. Quebras de estrutura

Tal como o nome indica, a “existência de quebras de estrutura” costuma estar associado a situações que destoam do padrão. Considera-se quebra de estrutura o momento em que algum dos

parâmetros presentes no modelo sofre uma alteração (anormal) – data de quebra ⁵. Torna-se assim evidente que quanto maior o intervalo temporal dos dados a serem analisados, maior a probabilidade de existirem quebras de estrutura.

Na literatura é possível encontrar diversos testes de quebras de estrutura, contudo iremos centrar-nos no teste CUSUM, desenvolvido por Brown, Durbin e Evans (1975). Neste teste é feita uma análise de variações (ao longo do tempo) na componente residual, de modo a encontrar quebras de estrutura. De modo sucinto, este teste, tomando como hipótese nula a estabilidade do(s) parâmetro(s), pode ser visto como um teste à estabilidade na variância da componente residual, u_t , (Ramos, 2021)

$$Var(u_t) = \begin{cases} \sigma_1^2, & t \leq T_1 \\ \sigma_2^2, & t > T_1 \end{cases} \quad (6)$$

Deste modo, para cada momento $1 \leq t \leq n$, consideramos o seguinte teste de hipóteses

TESTE DE QUEBRAS DE ESTRUTURA – CUSUM	
$H_0: \forall i \neq j, \sigma_i^2 = \sigma_j^2, i, j = 1, \dots, n$ (sem evidências de quebra de estrutura)	$H_1: \exists i \neq j, \sigma_i^2 \neq \sigma_j^2, i, j = 1, \dots, n$ (evidências de existência de quebra de estrutura)

2.1.4. Previsão em séries temporais

Segundo Hyndman & Athanasopoulos (2018) no estudo de séries temporais, boas previsões capturam os padrões que existem nos dados históricos, sendo importante distinguir entre uma flutuação aleatória, que deve ser ignorada, e um padrão genuíno, que deve ser modelado e extrapolado. Em séries temporais, os métodos de previsão mais simples usam apenas informações sobre a variável a ser prevista procurando, em particular, extrapolar tendências e padrões sazonais (Ramos, 2021). Deste modo, para a modelação de séries temporais, as observações passadas têm

⁵ As alterações estruturais podem-se apresentar de diferentes formas: (i) afetarem todos os parâmetros do modelo ou apenas parte deles; (ii) ocorrerem de forma abrupta ou gradual; (iii) ocorrerem numa data conhecida ou desconhecida; (iv) ocorrem uma única vez, ou como um fenómeno múltiplo num determinado período.

de conter informação fundamental da série e que tenham padrões recorrentes ao longo do tempo. Acontece que, tal como já referido acima, parte da eficácia dos modelos está em conseguir lidar e distinguir a parte da informação que é padrão da parte que é ruído (Makridakis *et al.*, 1998).

Se a série apresenta uma dinâmica simples, com clara evidência de tendência e/ou sazonalidade, as metodologias clássicas (muito em particular as metodologias de Alisamento Exponencial) mostram-se suficientes por apresentarem previsões relativamente ajustadas (Ramos, 2021). Com efeito, dentro dos modelos lineares (clássicos), foi nossa opção a família de modelos de Alisamento Exponencial (*Exponential Smoothing*), abordados seguidamente ⁶.

Como explica Ramos (2021), nesta linha de raciocínio, os modelos de previsão em séries temporais poderão ser encarados como modelos quantitativos que, tendo por base as observações passadas da série (e no seu inter-relacionamento), procuram fornecer previsões para os momentos futuros. Ou seja, independentemente do modelo de previsão considerado e formalizando o conceito de previsão, considerando a série temporal $\{y_t\}$ com $t \in T$, o objetivo é prever o valor para um período futuro $t + h$ (com $h \geq 1$), \hat{y}_{t+h} , supondo que são conhecidas todas as observações até ao momento t , ou seja, $y_1, y_2, \dots, y_{t-1}, y_t$.

2.2. Modelos de Alisamento Exponencial (ETS)

2.2.1. Noções Gerais

A metodologia de *Exponential Smoothing* (ETS) surgiu na década de 1950, enquanto os autores estavam a trabalhar na criação de modelos de previsão para sistemas de controlo de stock (Marques, 2017). Refere-se à modelação explícita do erro, tendência e sazonalidade, por apresentar previsões confiáveis de rápida implementação e compreensão simples, sendo uma vantagem para o mundo empresarial (Ord, 2004; Hyndman & Athanasopoulos, 2018).

Hyndman & Athanasopoulos (2018) referem ainda que a maior vantagem desta metodologia está na identificação de modelo melhor se ajusta aos dados, através da análise gráfica de cada componente (tendência e sazonalidade), depois de decompor a série. É, por isso, um método que fornece uma boa previsão e que, por ter requisitos de armazenamento de dados mínimos, torna-se o método ideal para a aplicação em tempo real (Marques, 2017). Trata-se assim de modelos que

⁶ Para mais detalhes consultar Ramos (2021, pg 40 – 46).

se baseiam em médias ponderadas de observações anteriores. Importa ainda referir que cada modelo deverá ter um ou mais hiperparâmetros, que corresponde aos parâmetros ajustáveis que irão controlar o modelo.

2.2.2. *Single Exponential Smoothing*

O modelo mais simples de *exponential smoothing* é o *single exponential smoothing* (SETS), para uma dada série temporal $\{y_t\}_{t \in T}$, é descrito pela equação seguinte

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha) \hat{y}_{t|t-1}, \forall t \in T \quad (7)$$

onde $\hat{y}_{t+1|t}$ corresponde à previsão para o momento $t+1$ (até ao instante $t \in T$), e $0 \leq \alpha \leq 1$, corresponde ao parâmetro de alisamento. O coeficiente α chama-se coeficiente de alisamento para o nível. Quando α apresenta valores próximos de 1, o modelo privilegia as observações mais recentes, tornando-se sensível, caso existam, a oscilações recentes dos dados. Por outro lado, quando o α for mais próximo de 0, serão privilegiados os valores mais antigos, o que indica um tratamento uniforme e com previsões com menores oscilações.

O SETS apresenta como único hiperparâmetro o α , o que torna um modelo indicado realizar previsões para uma série que não apresente tendência e/ou sazonalidade.

2.2.3. *Double Exponential Smoothing*

O *double exponential smoothing* (DETS) é visto como uma extensão do SETS para previsões que apresentem tendência linear (Holt, 1957). É um método que recorre a duas constantes de alisamento, α e β , com valores compreendidos entre 0 e 1. Em termos de *component form*⁷, além da equação de previsão (8) envolve duas equações de suavização, uma para o nível (9) e outra para a tendência (10)

$$\hat{y}_{t+h|t} = \ell_t + h \tau_t, \forall t \in T \quad (8)$$

onde $\hat{y}_{t+1|t}$ corresponde à previsão para o momento $t+1$, e h corresponde ao horizonte de previsão;

⁷ O DETS pode exibir tendências modeladas de duas formas – modelo aditivo e modelo multiplicativo, dependendo de a tendência ser linear ou exponencial. Nesta caso será apenas referido o modelo aditivo.

$$\ell_t = \alpha y_t + (1-\alpha) (\ell_{t-1} + \tau_{t-1}), \forall t \in T \quad (9)$$

onde ℓ_t , corresponde ao nível de tendência da série no instante t , e α corresponde ao parâmetro de alisamento;

$$\tau_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta) \tau_{t-1}, \forall t \in T \quad (10)$$

onde τ_t , corresponde à estimativa de tendência da série no instante t , e β corresponde ao parâmetro de alisamento.

Em suma, o modelo DETS (com amortecimento ⁸) é apropriado para séries que possuem uma trajetória com tendência, considerando-se como hiperparâmetros: α – Coeficiente de alisamento para o nível; β – Coeficiente de alisamento para a tendência; δ – Coeficiente de amortecimento, com tipo de amortecimento aditivo ou multiplicativo.

2.2.4. Triple Exponential Smoothing

O *triple exponential smoothing* (TETS) é utilizado quando as séries temporais, além de poderem apresentar padrões com tendência linear, exibem sazonalidade (Holt, 1957; Winters, 1960). Assim, no processo de previsão, além de se estimar o nível, tendência e sazonalidade da série, o modelo inclui também um novo parâmetro que controla a influência da componente sazonal (s_t).

Existem duas abordagens distintas, consoante a modelação da sazonalidade. Contudo neste estudo será explorado apenas o método multiplicativo ⁹

$$\hat{y}_{t+h|t} = (\ell_t + h \tau_t) \times s_{t+h-m(k+1)}, \forall t \in T \quad (11)$$

$$\ell_t = \alpha \left(\frac{y_t}{s_{t-m}} \right) + (1 - \alpha)(\ell_{t-1} + \tau_{t-1}), \forall t \in T \quad (12)$$

$$\tau_t = \beta (\ell_t - \ell_{t-1}) + (1 - \beta) \tau_{t-1}, \forall t \in T \quad (13)$$

$$s_t = \gamma \left(\frac{y_t}{\ell_t} \right) + (1 - \gamma) s_{t-m}, \forall t \in T \quad (14)$$

⁸ A juntar aos parâmetros de alisamento α e β , o modelo pode integrar um parâmetro de amortecimento, $0 \leq \delta \leq 1$. Para mais detalhes consultar Ramos (2021, pg 42 – 44).

⁹ Para mais detalhes consultar Ramos (2021).

onde (11) corresponde à equação da previsão e as seguintes correspondem às equações de suavização: (12) para o nível, (13) para a tendência e (14) para a componente sazonal. Os parâmetros de suavização (com valores entre 0 e 1, inclusive) correspondentes, α , β e γ , e ainda m para indicar a frequência da sazonalidade, que corresponde ao número de partes do ano (por exemplo, $m = 4$ para dados trimestrais e $m = 12$ para dados mensais).

2.2.5. Critérios de seleção de Modelos

Para uma seleção equilibrada dos modelos de previsão, têm sido propostos vários critérios na literatura. De acordo com Ramos (2021), procurando evitar uma escolha *ad hoc* do ‘melhor modelo’ para efetuar cenários de previsão, optou-se por considerar uma forma mais robusta analisando-se, para cada série, todos os modelos que apresentavam convergência, em paralelo com a informação obtida a partir dos critérios *Akaike Information Criteria* (AIC) e *Bayesian Information Criteria* (BIC).

A estatística subjacente ao AIC, desenvolvido por Akaike (1974), permite fazer uma seleção do(s) modelo(s) que melhor se ajusta(m) aos dados com base na qualidade de ajuste dos modelos estimados. O valor da estatística referente ao AIC incorpora duas componentes, uma que valoriza a precisão do ajuste e outra que penaliza, por meio de uma função crescente, os modelos com maior número de parâmetros.

$$\mathcal{S}_{AIC} = -2\log(L) + 2k \quad (15)$$

Por sua vez o BIC, desenvolvido por Schwarz (1978), está relacionado com o AIC, mas incorpora um formalismo bayesiano. Se ambos os critérios permitem a comparação entre modelos, penalizando os modelos com maior número de parâmetros, no modelo BIC essa penalização é mais rigorosa.

Em suma, aquando da etapa fundamental de seleção do(s) melhor(es) modelo(s), importa olhar para a informação que resulta da aplicação de cada um dos três critérios (AIC, BIC e HQIC). Quando comparados os três critérios, a estatística \mathcal{S}_{BIC} tende a ser a mais parcimoniosa. Daí que, segundo a literatura, o “... *BIC has been widely used for model identification in time series (...)* [and] *be applied quite widely to any set of maximum likelihood-based models...*” (Clement, 2014, p. 216).

2.3. *Deep Learning*

2.3.1. *Da Inteligência Artificial ao Deep Learning*

Apesar da facilidade de implementação e rapidez na execução computacional das metodologias clássicas, na literatura científica têm-lhes sido apontadas algumas limitações, nomeadamente a dificuldade em lidar com alterações fora dos padrões e em captar as informações realmente relevantes sobre os dados, com vista a efetuar previsões (Ramos, 2021). Assim, além de estar presente neste estudo a aplicação de metodologias clássicas, será explorado também o *Deep Learning*. Este torna-se especialmente relevante no presente estudo dado que em séries com comportamentos mais irregulares (que não apresentam padrões definidos de tendência e/ou sazonalidade, por exemplo) as metodologias clássicas não exibem boa qualidade preditiva.

Dado os avanços tecnológicos e desenvolvimento dos computadores, é relevante discutir como a Inteligência Artificial pode ser uma alternativa à previsão feita através de modelos clássicos. A Inteligência artificial (IA) consiste na forma que uma máquina tem de imitar o comportamento humano inteligente. Um sistema de IA, integrando componentes de representação, raciocínio e aprendizagem, deve ser capaz de cumulativamente: (i) armazenar conhecimento; (ii) aplicar o conhecimento armazenado para resolver problemas e (iii) adquirir novo conhecimento através da experiência (Ramos, 2021).

Machine learning (ML) é um ramo da IA que dá aos computadores a capacidade de “aprender” – geralmente a partir de dados – sem serem programados explicitamente (Gupta, 2018). Os modelos de ML tornaram-se, na última década, concorrentes sérios aos modelos estatísticos clássicos na área de previsão (Ahmed *et al.*, 2010). Por sua vez, o *Deep Learning* (DL) é um subcampo do ML que usa algoritmos chamados Redes Neurais Artificiais, que são inspirados na estrutura e função do cérebro e são capazes de autoaprendizagem, e as Redes Neurais Artificiais são treinadas para “aprender” modelos e padrões, em vez de serem explicitamente informadas sobre como resolver um problema (Gupta, 2018). O mesmo autor explica como, apesar de terem sido descobertas nos anos 1950, as redes neuronais permaneceram em desuso até aos dias de hoje, devido à capacidade limitada dos computadores da altura.

Numerosos estudos foram publicados em modelos de ML com desempenho relativamente melhor do que as técnicas clássicas de previsão de séries temporais (Sezer *et al.*, 2020). Nos

últimos anos, DL emergiu fortemente como a classe de predictor de melhor desempenho no campo ML em várias áreas de implementação (Sezer *et al.*, 2020).

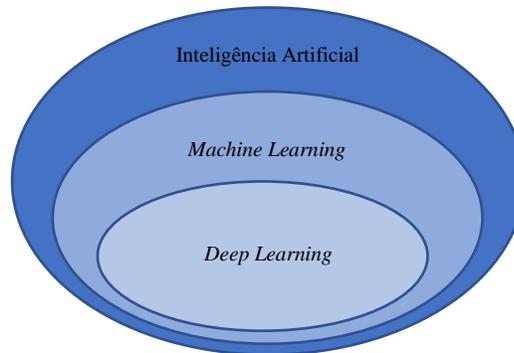


Figura 2.1 – Da Inteligência Artificial ao *Deep Learning*

2.3.2. *Deep Learning* na modelação e previsão de séries temporais

Segundo os autores Tkáč & Verner (2016) e Zhang *et al.* (2017), nas últimas décadas tem-se assistido a diferentes paradigmas de abordagem, em particular a aplicação de metodologias baseadas em Inteligência Artificial, articulando técnicas de *Data Science* e *Machine Learning*, nomeadamente as baseadas em *Deep Learning*, ou *Deep Neural Networks* (DNN). São identificados dois fatores determinantes para o sucesso destas metodologias: (i) as redes neuronais são capazes de extrair recursos abstratos de dados brutos, pois combinam várias fontes de informação, processam dados heterogéneos e conseguem captar alterações dinâmicas e (ii) caso se trate um grande volume de dados, é possível desenhar arquiteturas de redes neuronais mais complexas capazes de aprender e treinar com amostras de dimensão considerável (Zhang *et al.*, 2017).

Deste modo, a investigação científica, aliada ao grande progresso computacional observado nos últimos anos, tem assumido um papel fundamental na procura de soluções eficazes, não só com a melhoria das técnicas já existentes como na procura de metodologias alternativas, onde as DNN têm sido apontadas com uma opção promissora na modelação e previsão de séries temporais (Costa *et al.*, 2019). Nesse seguimento, várias arquiteturas de redes recorrentes, baseadas em mecanismos capazes de aprender dependências de longo prazo – a chamada memória da rede – têm sido propostas na literatura, evidenciando a sua ampla aplicabilidade (Koutník *et al.*, 2014). Contudo, através do estudo de Jozefowicz *et al.* (2015) alinhado com Greff *et al.* (2015), percebe-

se que apesar das diferenças existentes nos vários trabalhos, a base arquitetural é a mesma. Desta forma, será explicada de forma sucinta o funcionamento destas arquiteturas de redes neuronais, baseadas em mecanismos de aprendizagem profunda (Ramos, 2021).

Da multiplicidade de algoritmos existentes, o recurso a redes neuronais é tópico de particular interesse, em particular no que respeita a processos de *Deep Learning*. Segundo Ramos (2021) tirando partido do gradual e notável progresso tecnológico/computacional, sobretudo pelo notável aumento da ‘capacidade’ das máquinas, várias arquiteturas de redes neuronais baseadas em mecanismos de aprendizagem profunda – DNN – têm sido propostas na literatura nos últimos anos, das quais se destacam as popularizadas redes *Long Short-Term Memory* (LSTM).

A unidade fundamental das redes neuronais artificiais é o neurónio artificial (McCulloch & Pitts, 1943; Rosenblatt, 1958; Minsky & Papert, 1969). Os neurónios encontram-se organizados em camadas, sendo cada camada constituída por um conjunto de neurónios semelhantes. Partindo dos conceitos de “neurónio” e “camada”, bem como das ligações que se estabelecem entre eles, podemos considerar a construção da rede neuronal, vista como um conjunto de algoritmos computacionais, com uma arquitetura semelhante à do cérebro humano, desenhado para reconhecer padrões numéricos extraídos de dados reais (sejam eles imagens, sons, texto ou séries temporais) (Ramos, 2021). Parte do sucesso das redes neuronais em conseguir ‘aprender’ os dados respeita ao processo de *Machine Learning*, envolvendo o treino e aprendizagem da rede.

Centrando particular atenção na aprendizagem da rede, embora esta seja parcialmente ditada por se considerarem mais camadas ocultas na rede (Haykin, 2009; Data Science Academy, 2019), importa referir que o número de camadas ocultas na rede deve ser suficientemente grande para captar a dinâmica do modelo, sem ser demasiado grande, para não prejudicar a capacidade de generalização, ou seja, para que a rede não armazene toda a informação da aprendizagem, mas consiga generalizá-la. É desta forma que se evita *overfitting*, fenómeno que corre quando os pesos levam o sistema a aprender demasiados detalhes sobre os dados em vez de compreender a sua ‘estrutura global’ (Hawkins, 2004; Brownlee, 2017). Existem vários métodos para evitar este problema (*overfitting*), sendo um dos mais utilizados a avaliação da capacidade de generalização da rede durante o próprio treino (Hastie *et al.*, 2009). Esta técnica, usualmente designada de *cross-validation* (Arlot & Celisse, 2010), muito utilizada em problemas de modelação e previsão, consiste em dividir o conjunto de treino em dois subconjuntos, um dos quais (conjunto de validação) é utilizado para avaliar a qualidade do treino e a capacidade de generalização no final

de cada época. O algoritmo termina quando o erro aumenta sucessivamente por um número definido de iterações, sendo a melhor solução a que apresentou menor erro durante o processo de treino.

2.3.3. Arquitecturas *Perceptron Multilayer* e *Long Short-Term Memory* ¹⁰

As arquiteturas *Multilayer Perceptron* (MLP) têm particular aplicação em problemas cujo objetivo passa pelo treino de um conjunto de dados de entrada, procurando captar a informação relevante (por exemplo, tendências), para conseguir uma generalização.

Embora a arquitetura MLP seja muito utilizada, esta apresenta menor robustez face a outras arquiteturas pois cada neurónio recebe apenas como input os dados de treino, não registando aprendizagem ao longo do tempo. Tal facto pode ser uma limitação quando os dados apresentam alguma sequência e/ou dependência temporal, limitação essa que se torna evidente no caso de séries temporais.

As arquiteturas *Long Short-Term Memory* (LSTM), tidas como redes de memória de longo prazo, são um tipo especial de RNN capazes de “aprender” dependências de longo prazo. Acrescendo alguns procedimentos internos aos mecanismos gerais de aprendizagem descritos nas RNN, estas ‘novas’ arquiteturas de rede neuronal conseguem selecionar informação, preservando não só as memórias antigas consideradas mais importantes (as que mais contribuíram para reduzir o erro), como também eliminar outras que se mostraram menos relevantes. Além de outras potencialidades, no caso específico do estudo de séries temporais, esta arquitetura apresenta melhor desempenho que as RNN (ou as MLP) na captura de dependências de longo prazo (Data Science Academy, 2019).

Deste modo, são reconhecidas vantagens aquando da modelação e previsão de séries temporais pela dupla capacidade da rede: (i) “lembrar” aprendizagens importante mais antigas (ou mesmo o ponto de partida), algo que nas RNN padrão vai sendo esquecido (apenas há input de aprendizagens recentes); (ii) “esquecer” a informação considerada não relevante para aprendizagem.

Assim, pode-se fazer uma breve comparação entre estas duas redes neuronais. Onde se verifica que as diferenças no processo de aprendizagem destas duas arquiteturas ocorrem essencialmente ao nível do fluxo de informação dentro de cada neurónio das camadas ocultas (célula).

¹⁰ Secção parcialmente adaptada de Ramos (2021).

2.3.4. Hybrid Model Box & Jenkins – DNN (BJ – DNN)

No estudo de Ramos *et al.* (2022) propõe-se uma abordagem híbrida entre a metodologia de *Box&Jenkins* e a de redes neurais com aprendizagem profunda (DNN), com recurso a metodologias de *cross-validation* mais robustas – Grupo *k-Fold*.

De acordo com os autores e de um modo sucinto, em termos de procedimentos metodológicos, foram consideradas três arquiteturas DNN (MLP, RNN e LSTM), combinadas com a metodologia *Box&Jenkins*, sendo proposto um modelo híbrido, *Box&Jenkins-DNN* (BJ-DNN), como se pode ver na Figura 2.2

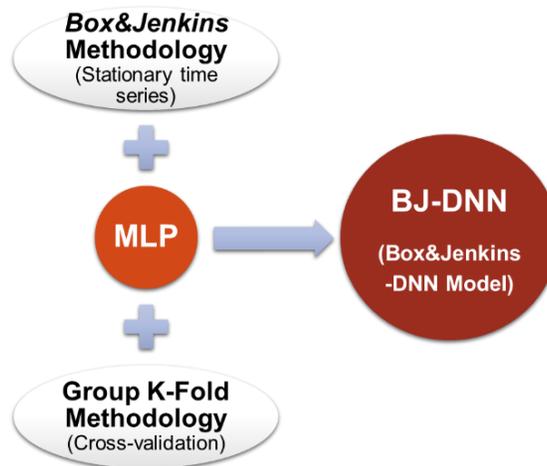


Figura 2.2 – Modelo BJ-DNN

Fonte: Ramos *et al.* (2022)

Pela importância da estacionaridade das séries temporais na aplicabilidade dos modelos ARMA, avaliaram-se os benefícios, em termos de pré-processamento dos dados, em considerar uma série temporal estacionária no processo de treino da rede neuronal. Mais, eliminando a presença de autocorrelação, foi viável a utilização da metodologia de *cross-validation Group k-Fold*, cuja bibliografia reconhece vantagens e potencialidades ¹¹.

Em relação às arquiteturas DNN, em Ramos *et al.* (2021) e Lopes *et al.* (2021) é mencionada a capacidade de memória de longo prazo das redes LSTM, mas destacam o alto custo computacional em relação ao MLP. Devido a este fato, ambas as redes foram consideradas na

¹¹ Consultar Ramos (2021)

investigação desenvolvida em Ramos *et al.* (2022). Através da construção de rotinas computacionais completas e automatizadas, o modelo BJ-DNN proposto é testado com a modelação de duas séries temporais financeiras com perturbações em seus dados históricos, esta abordagem é comparada a modelos de redes neurais com arquiteturas MLP e LSTM, tendo os autores concluído pela não existência de vantagens na utilização das arquiteturas LSTM. Os autores destacam ainda que, além de reduzir o tempo computacional implícito em 20%, o modelo BJ-DNN proposto mostra uma melhor qualidade preditiva (considerando o *Mean Absolute Percentage Error – MAPE*).

3. Metodologia

Neste capítulo demonstramos os dados que utilizamos para este estudo. Temos ainda uma secção onde explicamos como foi aplicada a implementação computacional que nos permitiu tratar os dados.

3.1. Apresentação dos Dados

Os dados utilizados no presente estudo foram recolhidos do site do Instituto Nacional de Estatística (INE, 2022) e referem-se ao número de dormidas em estabelecimentos de alojamento turístico em Portugal. Seleccionaram-se os documentos *XLSX* que continham os dados pretendidos e, de forma a ter uma massa de dados robusta, incluíram-se o número de dormidas mensais desde janeiro de 2002 até março de 2022. Dado que a informação estava dispersa, foi necessário desenhar um plano sistemático de recolha, tratamento e organização de dados numa folha de cálculo *Microsoft Excel*. Depois de termos o documento pronto, criamos um ficheiro *csv* que foi importado para um programa de código de leitura, previamente criado para o estudo de Ramos (2021) e disponível em Lopes e Ramos (2020).

Com o intuito de dar resposta aos objetivos traçados na investigação, consideramos duas séries temporais:

- 1) Série Total – Número total de dormidas em estabelecimentos de alojamento turístico em Portugal;
- 2) Série Residentes – Número de dormidas de residentes em Portugal em estabelecimentos de alojamento turístico em Portugal.

3.2. Implementação Computacional

Tal como no estudo de Ramos (2021), utilizamos o ambiente computacional *Jupyter Notebook*, com a linguagem de programação *Python* (em específico a versão 3.7.3). O código de leitura é simples e automatizado pois recorre a várias bibliotecas: *numpy* (que permite, de modo eficiente, tratar dados matemáticos); *pandas* (que permite estruturar e analisar dados de modo mais simples); *statsmodels* (que permite a implementação de diversos testes estatísticos);

`matplotlib` (associado a ferramentas de visualização) e, finalmente, `tensorflow` (que permite, de modo rápido e eficiente, implementar redes neurais).

Para desenvolver o estudo empírico, a primeira etapa passa pela análise e caracterização da série temporal. Essa fase envolve o recurso a representações gráficas, a análise de existência de quebras de estrutura, a determinação de medidas descritivas e a implementação de alguns testes estatísticos (normalidade, estacionariedade e independência).

Foram implementados três *notebooks*: *ExploratoryDataAnalysis.ipynb*, respeitante à análise exploratória (técnicas descritivas e inferenciais) das séries temporais em estudo; *ExponenTialSmoothing.ipynb*, relativo à implementação da família de modelos ETS; *DNN_OurApproach.ipynb*, que inclui as rotinas referentes aos modelos DNN.

3.2.1 Análise Exploratória de Séries Temporais

Para a implementação de alguns métodos de previsão, em particular os ETS, é útil reconhecer a presença de comportamentos de tendência e de sazonalidade e de qualquer “ruído” presente nas observações (Ramos, 2021). Recorremos ao *Hodrick–Prescott filter* (Hodrick & Prescott, 1997) para remover a componente ciclo de tendência, CT_t , a partir dos dados ‘brutos’, através da função `hpfilter` da biblioteca `statmodels` (*Python*).

Para estudar a existência de quebras de estrutura ou mudanças abruptas de regime das séries temporais, utilizamos o algoritmo CUSUM. Na implementação em *Python* foi usada a função `detect_cusum` desenvolvida por Duarte e Watanabe (2018). Em termos de análise inferencial, aplicamos testes de hipóteses relacionados com o estudo da normalidade (Teste de *Jarque-Bera*, Teste de *Skewness* e Teste de *Kurtosis*). A implementação em *Python* está na biblioteca `scipy`. Para o estudo da independência dos dados (Teste BDS), em *Python* obtém-se na biblioteca `statsmodels`.

No que respeita à avaliação da existência de quebras de estrutura, optamos pelo algoritmo CUSUM. Relativamente à implementação em *Python* foi usada a função `detect_cusum` desenvolvida por Duarte e Watanabe (2018).

3.2.2 Modelos de Previsão: Modelos de Alisamento Exponencial (ETS) e Modelos de Redes Neurais

No que respeita à parte computacional em *Python*, a implementação dos modelos ETS (*notebook ExponentialSmoothing.ipynb*) está na biblioteca `statsmodels` por meio das classes `SimpleExpSmoothing`, `Holt` e `ExponentialSmoothing`. Em termos de implementação computacional, consideráramos neste trabalho nove modelos ETS, como está descrito na

Tabela 3.1. Cada modelo é rotulado por um par de letras (T, S), de acordo com as componentes tendência (T) e Sazonalidade (S).

Tabela 3.1 – Classificação dos modelos ETS

Componente de tendência (T)	Componente sazonal (S)		
	Nenhuma (N)	Modelo Aditivo (A)	Modelo Multiplicativo (M)
Nenhuma (N)	(N, N)	(N, A)	(N, M)
Modelo Aditivo (A)	(A, N)	(A, A)	(A, M)
Modelo Aditivo amortecido (A_d)	(A_d, N)	(A_d, A)	(A_d, M)

Fonte: Ramos (2021)

De acordo com a

Tabela 3.1, identificam-se três grupos de modelos:

- *Single Exponential Smoothing* (N, N)
- *Double Exponential Smoothing* (A, N) e (A_d, N)
- *Triple Exponential Smoothing* (N, A), (A, A), (A_d, A), (N, M), (A, M) e (A_d, M).

Uma forma robusta e objetiva de obter o valor dos hiperparâmetros numéricos, é estimá-los a partir dos dados históricos, numa perspetiva de minimizar o erro Hyndman e Athanasopoulos (2018), para modelos onde se verifica convergência. De acordo com o estudo de Ramos (2021), provou-se que, em séries com comportamentos mais irregulares (que não apresentam padrões definidos de tendência e/ou sazonalidade, por exemplo) as metodologias clássicas não exibem boa qualidade preditiva. Porém, como o presente estudo trata séries com um padrão claro de tendência e/ou sazonalidade, os modelos ETS parecem conseguir captar alguma da dinâmica dos dados e apresentar cenários preditivos relativamente ajustados. São apontas, ainda, como vantagens destas

metodologias a pouca duração de implementação computacional e não ser necessário cuidados em termos de pré-processamento dos dados.

Relativamente aos modelos de redes neuronais, como referido na secção 2.3, foi utilizado o modelo desenvolvido em Ramos (2021) e Ramos *et al.* (2022). Em *Python*, a implementação deste modelo (*notebook DNN_OurApproach.ipynb*) é fornecida na biblioteca `tensorflow` por meio das classes `Dense`, `SimpleRNN` e `LSTM`. A metodologia implícita é descrita pela Figura 3.1.

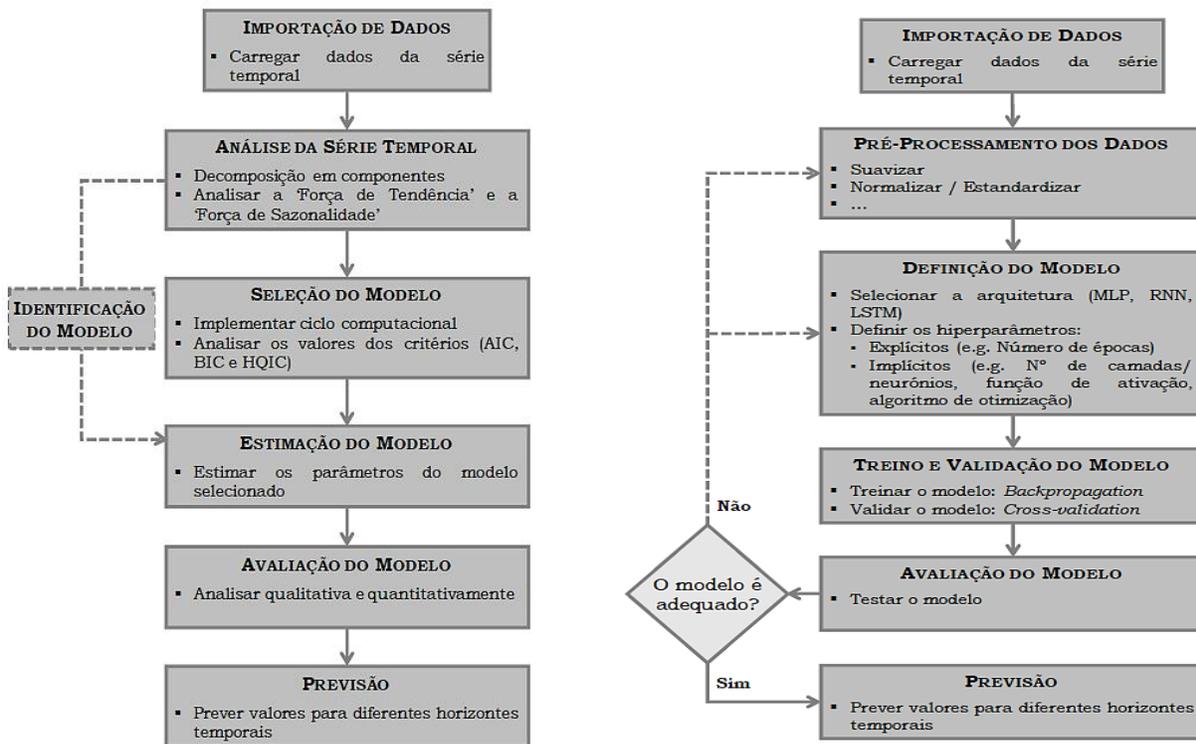


Figura 3.1 – Metodologia de implementação computacional dos modelos: ETS (à esquerda) e DNN (à direita)

Fonte: Ramos (2021)

3.2.3 *Forecasting* e Avaliação do Erro (Métricas)

Para se obter uma previsão, ou seja, ao ajustar uma série de forma a projetá-la para além do período da amostra, surge o conceito de ‘erro de previsão’ (Ramos, 2021). Existem dois tipos de erros: (i) aleatórios, devido ao desconhecimento de variações futuras cujos fatores não são contemplados no modelo; e (ii) sistemáticos, cometidos consistentemente devido, por exemplo, à seleção de relações matemáticas incorretas entre as variáveis, ou a diferenças entre os parâmetros verdadeiros e as suas

estimativas. Ambos contribuem para o erro de previsão e o melhor modelo será aquele que permite minimizar as medidas do erro sistemático (Hamilton, 1994).

Assim, sendo y_{t+h} o valor desconhecido no período futuro $t+h$ e \hat{y}_{t+h} a sua previsão obtida com base na informação disponível até ao momento t , o erro de previsão corresponde à diferença entre esses dois valores (14):

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h} \quad (14)$$

Sendo o propósito avaliar o desempenho global de um determinado modelo de previsão, torna-se essencial analisar os valores de indicadores capazes de traduzir, em termos numéricos, a medida de erro. De entre as métricas de erro comumente referidas na bibliografia, destacamos as seguintes, onde s denota o número de previsões a efetuar (janela de previsão)¹²:

Mean Error (ME)

$$ME = \frac{\sum_{i=1}^s e^i}{s} \quad (15)$$

Mean Absolute Error (MAE)

$$MAE = \frac{\sum_{i=1}^s |e^i|}{s} \quad (16)$$

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{\sum_{i=1}^s \left| \frac{y_{t+i} - \hat{y}_{t+i}}{y_{t+i}} \right|}{s} \times 100 \quad (17)$$

De acordo com Ramos (2021), ao comparar as duas métricas, MAE e MAPE, verifica-se que o MAE é uma ‘grandeza dimensional’ por ser expresso na unidade de medida dos dados, enquanto o MAPE é uma ‘grandeza adimensional’ por avaliar a dimensão do erro em termos percentuais.

¹² Na Literatura verifica-se a aplicação de mais medidas, contudo apenas se apresentam as que foram utilizadas no presente estudo. Para mais detalhes consultar Ramos (2021).

Esta característica apresenta vantagens, não só pela simplificação da interpretação, mas também por permitir comparar os erros de séries de dados com unidades de medida distintas ¹³.

¹³ Importa, contudo, salvaguardar que o MAPE não pode ser usado quando existem valores nulos para y_{t+i} . Além de ser uma medida sensível, que não deve ser utilizada quando se trabalha com poucos dados históricos (Ramos, 2021).

4. Resultados

Seguidamente apresentamos os resultados do nosso estudo. Começamos por expor o que observamos relativamente às séries temporais, e seguidamente apresentamos os resultados obtidos nos métodos seleccionados: Alisamento Exponencial (ETS) e *Deep Learning*. Por fim temos uma secção onde é apresentada uma discussão dos resultados.

4.1. Apresentação de Resultados

4.1.1 Séries Temporais

Os dados históricos relativos às séries em estudo, (1) Série Total – Número total de dormidas ($\times 10^7$) em estabelecimentos de alojamento turístico em Portugal, e (2) Série Residentes – Número de dormidas de residentes em Portugal em estabelecimentos de alojamento turístico em Portugal, apresentam-se nas Figura 4.1 e Figura 4.2. Das representações gráficas depreende-se uma tendência crescente com evidências de um comportamento sazonal (ao longo do ano), verificando-se uma clara mudança de regime em 2020 e período subsequente.

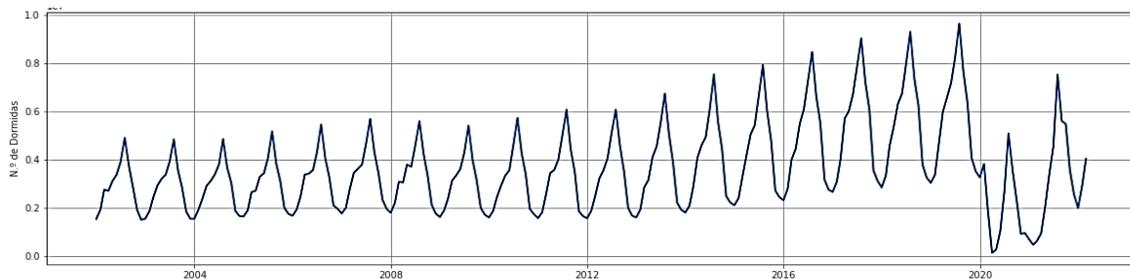


Figura 4.1 – Série Dormidas Total ($\times 10^7$) – Representação gráfica (por anos)

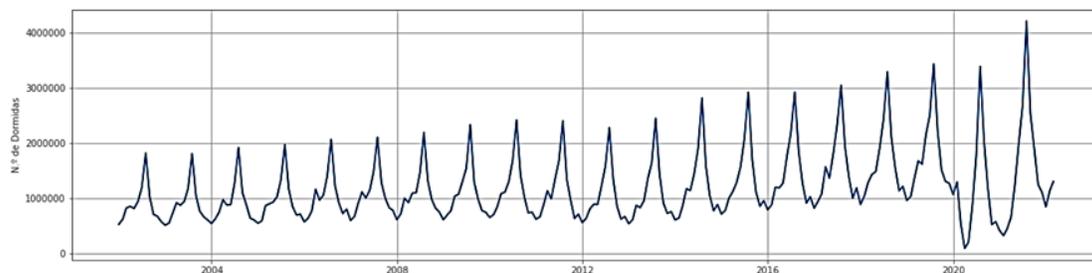


Figura 4.2 – Série Dormidas Residentes – Representação gráfica (por anos)

Através da representação de *box-plots* anuais agrupados (Figura 4.3 e Figura 4.4), verifica-se um aumento gradual da amplitude amostral e interquartil (nomeadamente desde 2011 até 2019), com a existência de observações *outliers* na série Residentes (mês de férias de Verão). Em 2020 devido à pandemia, com o fecho de fronteiras por todo o mundo, as implicações no fluxo de Turismo foram evidentes. Assim, pode-se verificar nas Figura 4.3 e Figura 4.4 uma quebra abrupta em 2020 e a existência de indícios de uma recuperação em 2021 e nos primeiros três meses de 2022 (de reforçar que em 2022 só foram considerados os primeiros três meses do ano).

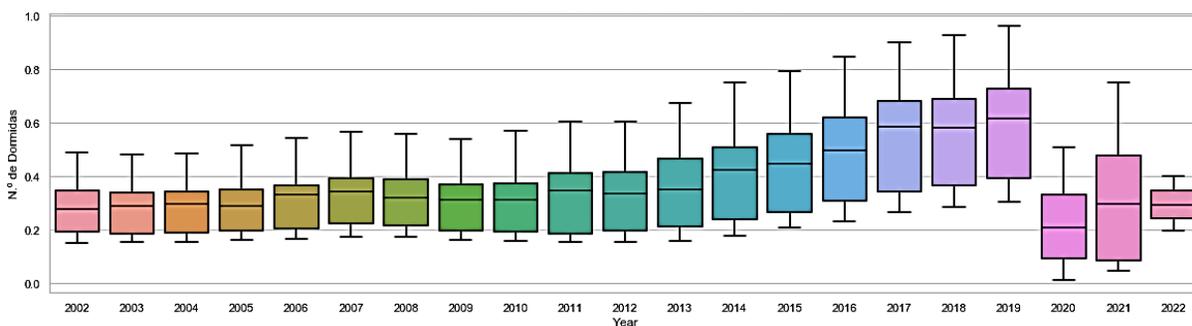


Figura 4.3 – Série Dormidas Total: Representação gráfica dos *box-plots* anuais

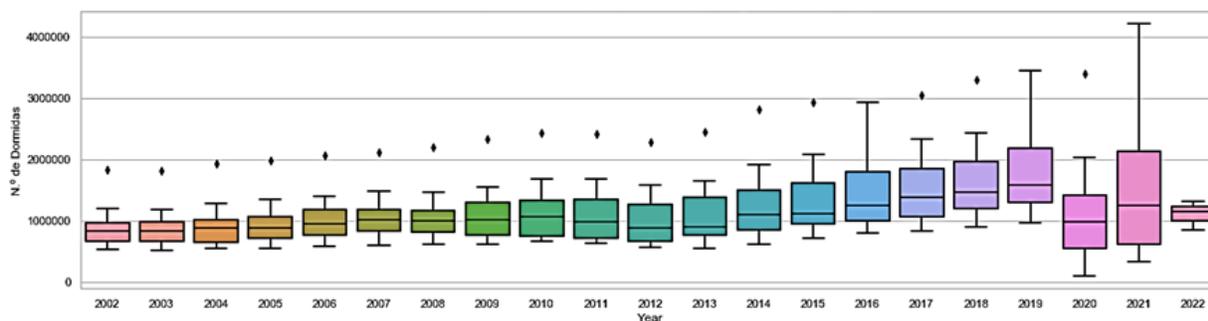


Figura 4.4 – Série Dormidas Residentes: Representação gráfica dos *box-plots* anuais

Pela representação gráfica dos *box-plots* mensais (Figura 4.5 e Figura 4.6) é evidente o comportamento sazonal das séries temporais: o mês com maior número de dormidas é agosto, sendo os mais baixos os meses de janeiro e dezembro. De referir, ainda, a existência de observações *outliers* nos cinco primeiros meses, parte delas consequentes dos confinamentos e das restrições de circulação resultantes da pandemia (anos de 2020 e 2021).

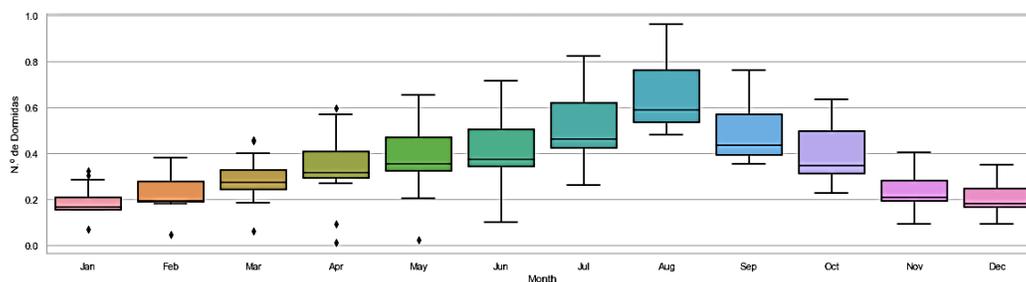


Figura 4.5 – Série Dormidas Total: Representação gráfica dos *box-plots* mensais

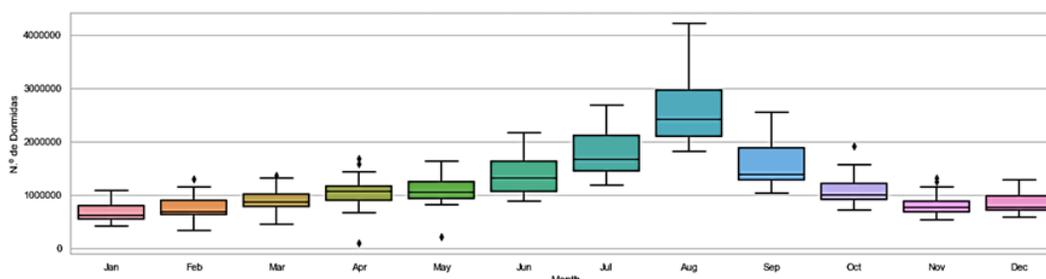


Figura 4.6 – Série Dormidas Residentes: Representação gráfica dos *box-plots* mensais

As principais estatísticas descritivas das séries temporais são apresentadas na Tabela 4.1, onde se pode verificar que, em ambas as séries, o valor máximo corresponde às dormidas em agosto de 2019 e o valor mínimo reflete as dormidas em abril de 2020.

Tabela 4.1 – Estatísticas descritivas

	<i>Count</i>	<i>Mean</i>	<i>Std</i>	<i>Min</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Max</i>	<i>Kurtosis</i>	<i>Skewness</i>
Total	243	3644629	1817023	133212	2208905	3365240	4583488	9633427	0,4438	0,8370
Residentes	243	1201920	625718	96673	77319,5	1035100	1400719	4214577	3,2374	1,6136

As duas séries têm uma média superior à mediana, impacto esperado no valor da assimetria (*skewness*) que é positivo – distribuição assimétrica à direita. Relativamente à curtose (*kurtosis*) verifica-se em ambos os casos uma distribuição leptocúrtica do número de dormidas. Estas informações reforçam algumas características observáveis na curva de densidade resultante da distribuição do número de dormidas (ver Figura 7.2 – Anexo B e Figura 7.5 – Anexo C).

Na Tabela 4.2 apresentam-se os resultados dos testes de hipóteses referentes à normalidade, estacionariedade e independência (sob hipótese nula de que os dados são independentes e identicamente distribuídos) para cada uma das séries em estudo. Para os níveis de significância usuais, e para ambas as séries, é rejeitada a normalidade, a estacionariedade e a independência dos dados relativos às séries estudadas.

Tabela 4.2 – Testes de normalidade, estacionariedade e independência

	Testes de Normalidade			Testes de Raiz Unitária/ Estacionariedade		Teste de Independência
	<i>Kurtosis</i>	<i>Skewness</i>	<i>Jarque-Bera</i>	ADF	KPSS	BDS (Dim. 2 – Dim. 6)
Série Dormidas Total						
<i>statistic</i>	1.4453	4.8491	30.3689	-2.8783	0.5982	18.8298 – 22.1105
<i>p-value</i>	0.1484*	0.0000*	0.0000*	0.0479	-----	0.0000*
Série Dormidas Residentes						
<i>statistic</i>	4.8882	7.7890	211.56449	-1.3561	1.0849	5.0778 – 9.7241
<i>p-value</i>	0.0000*	0.0000*	0.0000*	0.6031	-----	0.0000*

*Rejeita-se H0 para os níveis de significância de 1%, 5% e 10%

Alinhado com a literatura, face à importância que a presença de quebras de estrutura pode ter no processo de modelação e previsão, foi analisada a presença de quebras de estrutura com a aplicação do algoritmo CUSUM, onde identificamos quebras de estrutura constantes que resultam, na sua maioria, das alterações de comportamento sazonal, conforme representado na Figura 4.7. Porém, no seguimento do já exposto, de salientar, uma vez mais, a mudança de regime observada em cada uma das séries (embora com impactos distintos) na sequência da pandemia em 2020.

Outras informações complementares sobre as séries temporais, nomeadamente, sua decomposição (ciclo de tendência, sazonalidade e componente aleatória), são apresentadas na Figura 7.1 – Anexos B e Figura 7.4 – Anexo C.

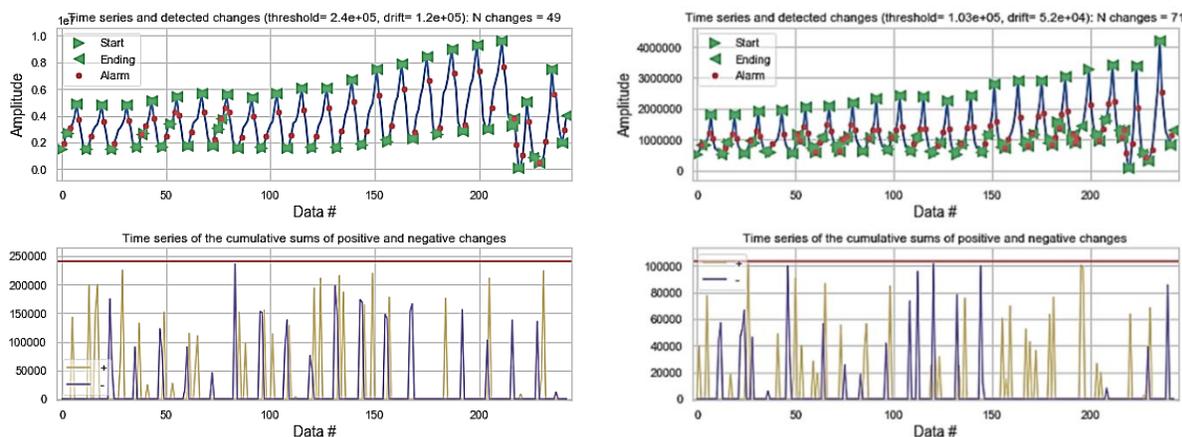


Figura 4.7 – Mudanças/quebras estruturais (série Total à esquerda, série Residentes à direita)

4.1.2 Previsão: Alisamento Exponencial (ETS)

Para a modelação e previsão usando metodologia ETS, considerou-se útil o cálculo da “força de tendência” (\mathcal{F}_T) e “força de sazonalidade” (\mathcal{F}_S). Conforme descrito na secção 2.1.2, os valores observados evidenciam que ambas as séries apresentam uma forte tendência (valores \mathcal{F}_T iguais a 1 numa vertente multiplicativa) e uma forte sazonalidade (valores de \mathcal{F}_S próximos de 1), factos que confirmam o referido anteriormente: da análise da representação gráfica de cada uma das séries temporais (Figura 4.1 e Figura 4.2) depreendem-se evidencias tendência e sazonalidade.

Tabela 4.3 – Valores da “força de tendência” e da “força de sazonalidade”

	Série Dormidas Total		Série Dormidas Residentes	
	\mathcal{F}_T	\mathcal{F}_S	\mathcal{F}_T	\mathcal{F}_S
Modelo Aditivo	0.7232	0.8536	0.1882	0.8247
Modelo Multiplicativo	1.0000	0.8141	1.0000	0.8706

Seguindo os procedimentos de implementação computacional descritos na secção 3.2.2., para determinar qual o modelo ETS mais apropriado para efeitos de previsão (de entre os modelos que apresentam uma convergência dos parâmetros a usar no modelo), de acordo com os critérios de seleção dos modelos de ETS, para a série Dormidas Total, os mais adequados são o TS (Ad, M) e TS (Ad, A), considerando tendência aditiva e sazonalidade multiplicativa e aditiva, respetivamente

(visivelmente, pela Figura 4.8, verifica-se que são as barras mais pequenas). Por outro lado, para a série Dormidas Residentes, selecionamos o modelo TS (N, N), de entre os possíveis modelos (os que apresentam uma convergência dos parâmetros), que não conta com tendência nem sazonalidade da série.¹⁴

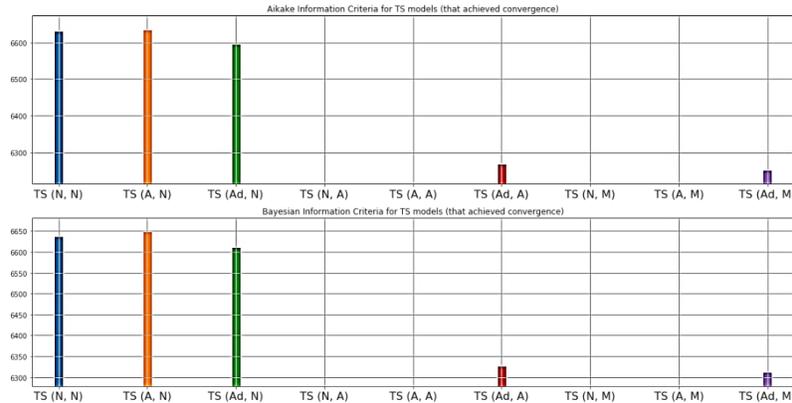


Figura 4.8 – Critérios de Informação para modelos ETS da série Dormidas Total

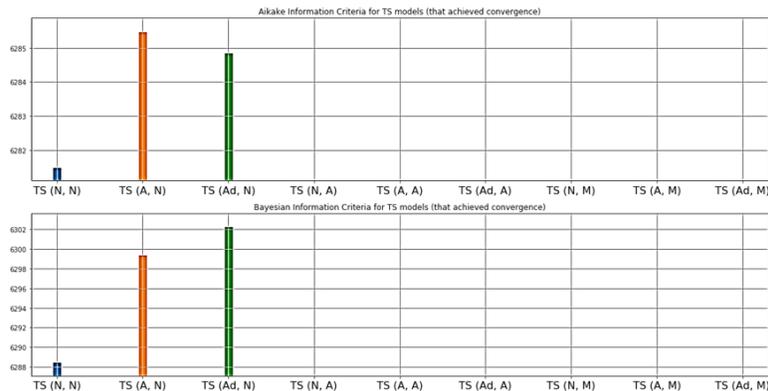


Figura 4.9 – Critérios de Informação para modelos ETS da série Dormidas Residentes

Relativamente à série Dormidas Total, verificamos na Figura 4.10 que existe um ajuste bastante satisfatório entre a linha de previsão de ambos os modelos considerados e a flutuação global dos dados reais (linha preta), dado que a linha de previsão acompanha, em geral, a tendência da série. Globalmente, pode verificar-se que a linha do modelo TS (Ad, M), a linha cor de laranja, está mais próxima dos valores reais.

¹⁴ Estas características contradizem o comportamento da série, facto que irá comprometer a qualidade preditiva, como iremos ver adiante.

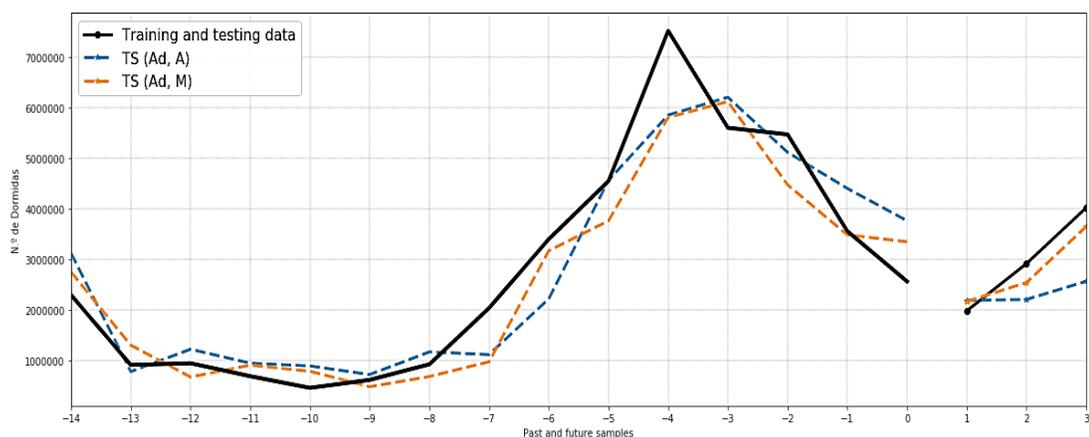


Figura 4.10 – Modelos ETS (*fitting e forecasting*) da série Dormidas Total

Na série Dormidas Residentes, dado que foi utilizado um modelo sem tendência e/ou sazonalidade, verifica-se na Figura 4.11 que a previsão é praticamente um valor constante.

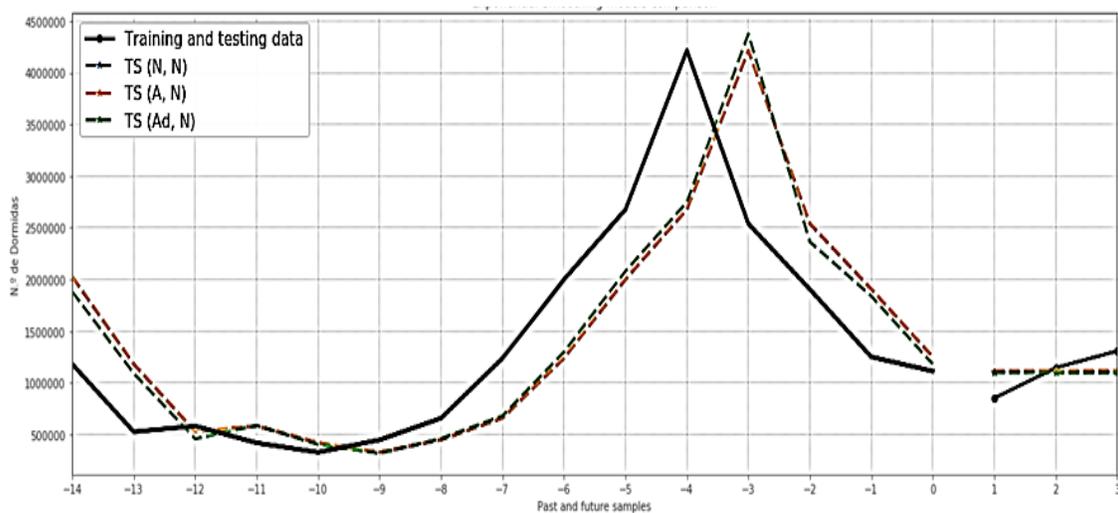


Figura 4.11 – Modelos ETS (*fitting e forecasting*) da série Dormidas Residentes

Para efeitos de previsão da série Total, o modelos ETS selecionado, TS (Ad, M), corresponde de forma positiva, embora com algum erro (Tabela 4.4), sendo que a métrica de erro utilizada na avaliação foi *Mean Absolute Percentage Error* – MAPE. Já o mesmo não se verificou para a série Residentes. Nesta série, por ter uma dinâmica recente e completamente atípica, verifica-se que os modelos ETS potencialmente mais adequados (modelos com tendência e sazonalidade), não

apresentam, em geral, uma convergência dos respectivos parâmetros para efeitos de previsão. Os únicos modelos que apresentam convergência são modelos sem componente sazonal, o que não é concordante com o histórico de dados, comprometendo, assim, a qualidade preditiva. Apresenta-se, igualmente na Tabela 4.4 o MAPE referente ao modelo TS (N, N).

Tabela 4.4 – Previsões segundo os modelos ETS (MAPE)

	Série Total (Ad, M)			Série Residentes (N, N)		
	janeiro	fevereiro	março	janeiro	fevereiro	Março
MAPE	9.29%	11.05%	10.44%	30.81%	16.8%	16.15%

Correspondente ao modelo selecionado para cada caso (ver Figura 7.3 e Figura 7.6 do Anexo D), na

Figura 4.12 (série Total) identificamos, globalmente, um ruído branco, existindo um pico que foge ao padrão no ano 2020. Contudo, na Figura 4.13 (série Residentes), verifica-se um evidente comportamento sazonal nos resíduos, uma vez que, como já referido, o modelo selecionado não capta esta dinâmica (evidenciando limitações das metodologias ETS).

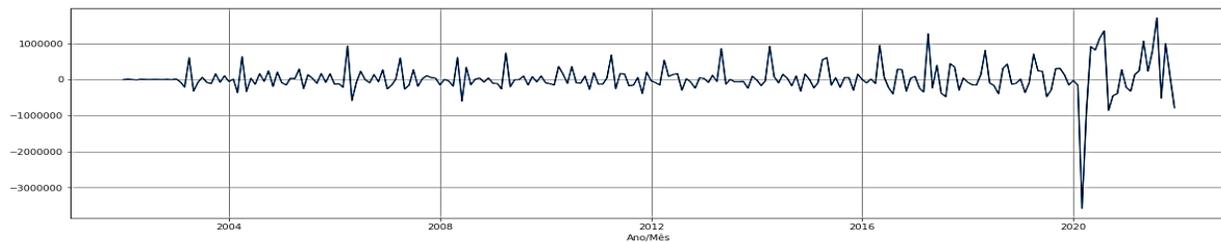


Figura 4.12 – Resíduos de modelos ETS da série Dormidas Total

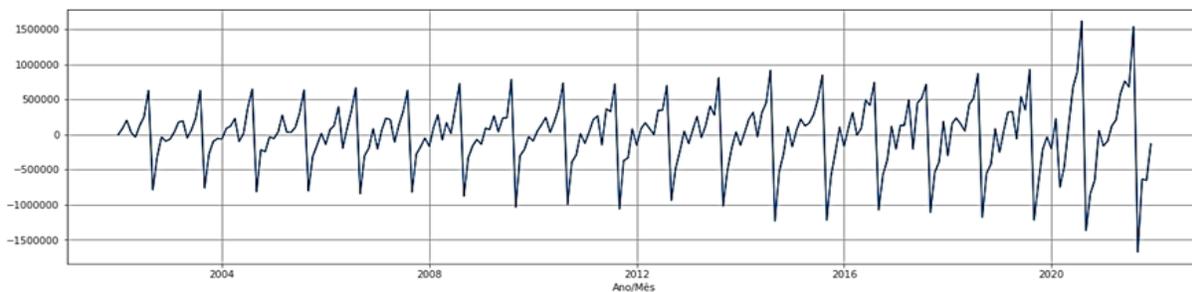


Figura 4.13 – Resíduos de modelos ETS da série Dormidas Residentes

4.1.3 Previsão: *Deep Learning*

Em termos de previsão usando redes neurais, foi considerado o modelo BJ-DNN (secção 2.3.4). Nas Figura 4.14 e Figura 4.15, para cada uma das séries verifica-se: (1) a janela de dados usada para treino (zona sombreada); (2) as previsões *in-sample* (linha cor de laranja) que nos dão conta do *fitting* de cada modelo e (3) finalmente, ajustado o modelo, a partir da amostra de dados destacada a azul, são apresentados os dados das previsões *out-of-sample* (linha a azul), a qual pode ser comparada com os dados reais (linha a preto).

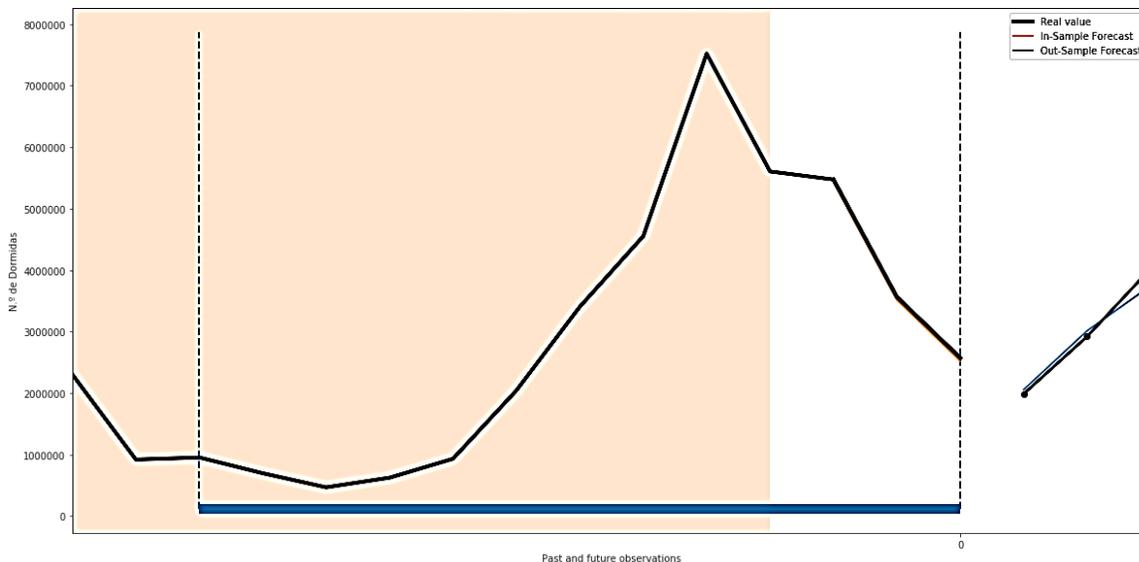


Figura 4.14 – Modelos BJ-DNN: LSTM da série Dormidas Total

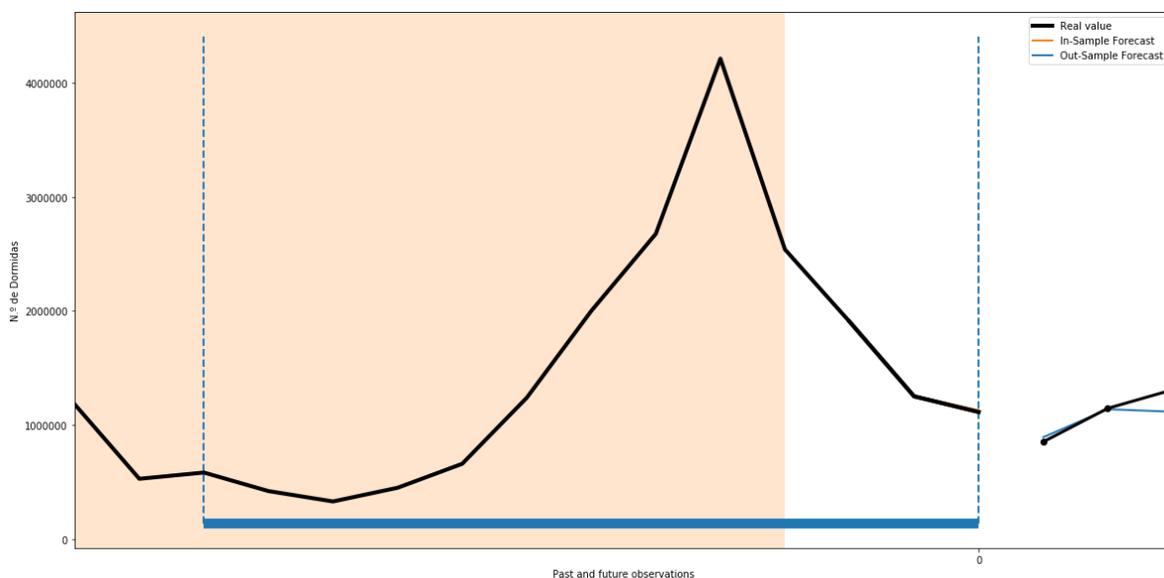


Figura 4.15 – Modelos BJ-DNN: LSTM da série Dormidas Residentes

Da análise das representações gráficas, verificamos que, em ambas as séries, o ajuste da linha de previsão à flutuação dos dados reais é bom (muito bom na série Total). No caso da previsão *in-sample*, verifica-se que as linhas de previsão são praticamente coincidentes com os dados reais. Já as linhas de previsão *out-of-sample*, acompanham, em geral, a tendência das séries. Os valores do MAPE, para cada caso, são apresentados na Tabela 4.5 (ver Figura 7.8 e Figura 7.10 – Anexo D).

Tabela 4.5 – Previsões segundo os modelos BJ-DNN (MAPE)

	Série Total			Série Residentes		
	janeiro	fevereiro	março	janeiro	fevereiro	Março
MAPE	3.41 %	3.29 %	4.5 %	4.74 %	2.72 %	6.71 %

4.2. Discussão dos Resultados

Do que verificamos da Apresentação dos Resultados da secção anterior, avaliamos como existe um bom ajuste global dos dois modelos na série total, o que não se verifica com tanto destaque na série residentes. Para melhor compreensão e comparação dos resultados, elaboramos a Tabela 4.6 onde é possível verificar os valores reais de dormidas para cada uma das séries nos três primeiros meses de 2022, e ainda compará-los com os valores previstos tanto com os modelos ETS como BJ-DNN.

Tabela 4.6 – Comparações entre valores reais e previsões

	Dormidas Total			Dormidas Residentes		
	Real	Modelos ETS	Modelos DNN	Real	Modelos ETS	Modelos DNN
Janeiro	1988869	2173621	2056734	851424	1113740	891739
Fevereiro	2920471	2641112	3013021	1145829	998627	1137803
Março	4023327	3713308	3877065	1307984	1145101	1227788

Em termos da metodologia usada, a literatura aponta para o sucesso das metodologias ETS em séries temporais com presença de um ciclo de tendência e clara sazonalidade, que são características presentes nas séries em estudo. Contudo, a dificuldade destas metodologias em lidar com quebras abruptas de estrutura, no histórico dos dados, fica patente neste estudo. Não obstante de previsões mais ou menos ajustadas, o destaque vai para a não convergência dos parâmetros (para efeitos de previsão) de alguns modelos que, em certa medida, se poderiam ajustar melhor ao histórico de dados. Neste sentido, impôs-se a procura de metodologias alternativas capazes de responder positivamente aos desafios identificados. O recurso às metodologias de *Deep Learning* (*Deep Neural Networks*) é apontado na literatura como promissor. Tal facto ficou comprovado no nosso estudo, a avaliar pelos valores do MAPE, que são inferiores ambos os casos.

Analisando os gráficos das séries temporais podemos observar como o Turismo interno teve um *boom* no período pós 2020, sendo responsável por salvar, em parte, o Turismo a nível nacional.

5. Conclusão

Serve este capítulo para sintetizar os pontos centrais do estudo, através da resposta às questões de investigação levantadas no início do mesmo, que serviram de linha condutora para o raciocínio e tomadas de decisão feitas ao longo desta investigação. Assim, iremos apresentar cada pergunta e a respetiva resposta, dado que resumem e realçam os pontos fundamentais de todo o estudo.

Seguidamente, iremos também referir limitações sentidas no desenvolvimento do trabalho bem como sugestões de investigações futuras e, ainda, as contribuições deste estudo para a comunidade científica.

Quais as alterações na dinâmica do fluxo de Turismo decorrente da pandemia?

De acordo com os resultados e com a Revisão da Literatura, demonstramos como, para suportar os negócios associados ao Turismo, Portugal estava muito dependente do Turismo externo. Com a pandemia, esta situação alterou-se, e como todos os países fecharam as fronteiras, tornamo-nos dependentes de movimentações internas, sendo os próprios portugueses quem segurou alguma atividade empresarial.

Torna-se evidente a necessidade de uma gestão estratégica capaz de fomentar o Turismo interno, devido à importância que os ‘turistas residentes’ tiveram no período pandémico, se quisermos criar mecanismos de ‘não dependência’ externa. Realçamos assim o interesse em não esquecer e apostar em políticas como o “Vá para fora cá dentro”.

Que potencialidades/limitações encontramos nas metodologias clássicas e de *Machine Learning* na previsão do fluxo de Turismo?

A previsão é uma excelente ferramenta de apoio à tomada de decisão, como ficou claro pela Revisão da Literatura. Por isso interessa encontrar metodologias de previsão robustas e capazes de dar resposta às oscilações bruscas ocorrentes nos dados históricos.

As metodologias clássicas são as mais conhecidas, pela sua facilidade de compreensão e implementação (nomeadamente metodologias ETS), contudo o nosso estudo confirma que estas metodologias apresentam limitações perante quebras de estrutura. Por isso, as metodologias de *Machine Learning* surgem como uma boa alternativa aos obstáculos existentes nas clássicas, apesar do maior custo computacional e complexidade de compreensão e implementação.

Qual a importância da utilização da informação obtida na modelação e previsão, para as empresas afetas ao setor do Turismo?

A modelação e previsão não tinha como propósito prever dados a longo prazo (tal que não o fizemos), mas dar indicação de modelos que as empresas afetas ao Turismo possam usar para obter resultados robustos que lhes permitam definir estratégias de atuação futura.

Que diretrizes podem ser apontadas, numa perspetiva de Gestão Estratégica, para desenvolver o Turismo interno?

Começando pela previsão, caso os dados apresentem quebras, temos de olhar com ponderação para as metodologias clássicas. Relativamente à Gestão Estratégica, é importante que, em situações adversas, existam alternativas para evitar falência e declínio dos negócios. Verificamos ainda que o fluxo de Turismo interno, apesar da dimensão da população portuguesa, foi fundamental para evitar uma situação (ainda) mais drástica do que aquela que se verificou nas atividades afetas ao Turismo (falências, despedimentos em massa, entre outros). Por isso, dada a importância em evitar dependência do Turismo externo, consideramos imperioso implementar mecanismos de captação de turistas internos.

Para concluir, em termos de limitações na investigação traçada, reconhecemos o aspeto redutor de se analisar o fluxo do Turismo com base no “número de dormidas em estabelecimentos de alojamento turístico em Portugal”. Porém, sendo um fenómeno difícil de quantificar, impunha-se a definição de uma estratégia de análise que, em certa medida, tornasse o fenómeno quantificável.

Consideramos ainda como possibilidade de investigação futura, a comparação do impacto do turismo interno no contexto pandémico em diferentes países localizados próximos de Portugal, como por exemplo Espanha, França e Itália, por suspeitarmos que terão uma dependência do turismo externo semelhante à de Portugal. Assim poderíamos discutir de forma informada, as vantagens e limitações de um maior foco no turismo interno.

Por fim apresentamos as contribuições deste estudo, dado que serviu, em parte, para a elaboração de um manuscrito que foi aceite e apresentado na Conferência Internacional de Gestão e Negócios (ICABM2022). Este está publicado no *e-book* disponibilizado aos participantes do evento, e posteriormente, submetido para indexação na *Web of Science (Clarivate Analytics)* e

SCOPUS (*Elsevier*). No seguimento da comunicação, está em processo de submissão um artigo completo para a *special issue of European Journal of Applied Business and Management* (EJABM).

6. Referências Bibliográficas

- Ahmed, N. K., Atiya, A. F., el Gayar, N., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5), 594–621.
- Akaike, H. (1974). A New Look at the Statistical Model Identification. In *Selected Papers of Hirotugu Akaike* (pp. 215–222). New York: Springer
- Antonova, A. (2014). Emerging technologies and organizational transformation. *Technology, Innovation and Enterprise Transformation*, 20–34.
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40–79.
- Arruda, E. F., Ferreira, R. T., & Castelar, I. (2011). Modelos lineares e não lineares da curva de phillips para previsão da taxa de inflação no Brasil. *Revista Brasileira de Economia*, 65(3), 237–252.
- Atkinson, R. (2016). “It’s Going to Kill Us!” and Other Myths About the Future of Artificial Intelligence. *Information Technology & Innovation Foundation*, 1–49.
- BCcampus. (2008). *Introduction to Tourism and Hospitality in BC*. Retrieved from <https://opentextbc.ca/introtourism2e/chapter/what-is-tourism/>
- Botelho, N. (2019). *Mais turistas que habitantes? Portugal faz parte de clube restrito*. Retrieved from <https://visao.sapo.pt/atualidade/economia/2019-02-08-Mais-turistas-que-habitantes-Portugal-faz-parte-de-clube-restrito/>
- BPI Research (2022). *Portugal - Setor do Turismo: situações e perspetivas*. Retrieved from https://www.bancobpi.pt/contentservice/getContent?documentName=PR_WCS01_UCM01212077
- Brandon, C., Jarrett, J. E., & Khumawala, S. B. A. (1987). Comparative Study of the Forecasting Accuracy of Holt-Winters and Economic Indicator Models of Earnings Per Share For Financial Decision Making. *Managerial Finance*, 13(2), 10–15.
- Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society*, 37(2), 149–192.
- Brownlee, J. (2017). *Introduction to Time Series Forecasting With Python*. Machine Learning Master Pty. Retrieved from <https://machinelearningmastery.com/introduction-to-time-series-forecasting-with-python/>

- Camara, I., Monteiro, J. E., & Santos, G. O. (2021). Fatores determinantes da demanda turística internacional para o Rio de Janeiro: evidências baseadas em modelos de regressão linear. *Revista Turismo Em Análise*, 32(1), 100–119.
- Casanova, S. (2012). Previsão da Demanda Turística da Cidade de Foz do Iguaçu: Uma Aplicação com os Modelos ARIMA. *Revista Turismo, Visão e Ação*, 14(3), 366–385.
- Clement, E. P. (2014). Using Normalized Bayesian Information Criterion (Bic) to Improve Box - Jenkins Model Building. *American Journal of Mathematics and Statistics*, 4(5), 214–221.
- Costa, A., Ramos, F., Mendes, D., & Mendes, V. (2019). *Forecasting financial time series using deep learning techniques*. In IO 2019 - XX Congresso da APDIO 2019, Instituto Politécnico de Tomar.
- Cristina, S. (2017). New Perspectives of the Tourism and Air Travel Relationship. *Cactus Tourism Journal*, 15(2), 24–32.
- Data Science Academy. (2019). *Deep Learning Book*. Retrieved from <http://deeplearningbook.com.br/>
- Duarte, M., & Watanabe, R. N. (2018). *Notes on Scientific Computing for Biomechanics and Motor Control*. GitHub. Retrieved from <https://github.com/BMClab/BMC>
- Egg, A. (1962). *The Travelling Companions*. Retrieved from https://commons.wikimedia.org/wiki/File:Augustus_Leopold_Egg_-_The_Travelling_Companions_-_Google_Art_Project.jpg
- Faulkner, D. O., & Campbell, A. (2003). *The Oxford handbook of Strategy: a strategy overview and competitive strategy*. Oxford: Oxford University Press.
- Ghalekhondabi, I., Ardjmand, E., Young, W., & Weckman, G. (2019). A review of demand forecasting models and methodological developments within tourism and passenger transportation industry. *Journal of Tourism Futures*, 5(1), 75-93.
- Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, 5, 499–510.
- Gonçalves, E., & Lopes, N. M. (2008). *Séries Temporais - Modelações lineares e não lineares*. Portugal: Sociedade Portuguesa de Estatística.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2015). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232.
- Gunter, U., & Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management*, 46, 123–135.

- Gupta, A. (2018, June). *Introduction to deep learning: Part 1*. Retrieved from <https://www.aiche.org/resources/publications/cep/2018/june/introduction-deep-learning-part-1>
- Hamilton, J. D. (1994). *Time series analysis: Econometric Theory*. New Jersey: Princeton University Press.
- Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics.
- Hawkins, D. M. (2004). The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*. American Chemical Society.
- Haykin, S. (2009). *Neural Networks and Learning Machines* (3rd ed.). Upper Saddle River, NJ: Pearson.
- Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1), 1–16.
- Holt, C. (1957). *Forecasting seasonals and trends by exponentially weighted moving averages*. Pittsburgh Pa.: Carnegie Institute of Technology Graduate school of Industrial Administration.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). Melbourne, Australia: OTexts.
- INE. (2021). *Atividade Turística*. Retrieved from https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=472742885&DESTAQUESmodo=2&xlang=pt
- INE. (2022). *Estatísticas do Turismo*. Retrieved from https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpagenumber=1&PUBLICACOEScolecao=107668&PUBLICACOESTipo=ea&selTab=tab0
- Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., & Callot, L. (2020). Criteria for classifying forecasting methods. *International Journal of Forecasting*, 36(1), 167–177.
- Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015). An Empirical Exploration of Recurrent Network Architectures. In *ICML - International Conference on Machine Learning*.
- Kirovska, Z. (2011). Strategic management within the Tourism and the world globalization. *Journal of Economics*, 2(1), 69–76.
- Kitagawa, G. (2010). *Introduction to time series modeling*. New York: Chapman and Hall/CRC.

- Koutník, J., Greff, K., Gomez, F., & Schmidhuber, J. (2014). A Clockwork RNN. *31st International Conference on Machine Learning, ICML 2014*, 5, 3881–3889.
- Larguesa, A. (2017). *Portugal entra no top 10 mundial do Turismo de negócios*. Retrieved from <https://www.jornaldenegocios.pt/empresas/Turismo---lazer/detalhe/portugal-entra-no-top-10-mundial-do-Turismo-de-negocios>
- Lopes, D. R., Ramos, F. R., Mendes, D. & Costa, A. (2021). *Forecasting models for time-series: a comparative study between classical methodologies and Deep Learning*. SPE 2021 – XXV Congresso da Sociedade Portuguesa de Estatística. Évora.
- Lopes, D. & Ramos, F. (2020). *Univariate Time Series Forecast*. Retrieved from <https://github.com/DidierRLopes/UnivariateTimeSeriesForecast>
- López, A. M. (2021). *Travel and tourism in Portugal - statistics & facts*. Retrieved from <https://www.statista.com/topics/6863/travel-and-tourism-in-portugal/>
- Luna, I., Ballini, R., & Soares, S. (2006). Técnica de identificação de modelos lineares e não-lineares de séries temporais. *Sba: Controle & Automação Sociedade Brasileira de Automatica*, 17(3), 245–256.
- Magano, J., & Cunha, M. N. (2020). Digital marketing impact on tourism in Portugal: A quantitative study. *African Journal of Hospitality, Tourism and Leisure*, 9(1), 1–19.
- Makridakis, S. G., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications* (3rd ed.). New York, USA: Wiley.
- Marques, A. (2017). *Desafios na Previsão de Séries Temporais Financeiras: O caso da Taxa de Câmbio EUR/USD*. Tese de Mestrado, Faculdade de Ciências da Universidade de Lisboa.
- Mathieson, A., & Wall, G. (1982). *Tourism: Economic, Physical and Social Impacts*. London: Longman.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133.
- Minsky, M., & Papert, S. (1969). *Perceptrons; an introduction to computational geometry*. MIT Press.
- OMTNU. (2020). *International tourism and COVID-19*. Retrieved from <https://www.unwto.org/international-tourism-and-covid-19>
- ONU. (2021). *World Tourism Day*. Retrieved from <https://www.un.org/en/page/world-tourism-day-background>
- Ord, K. (2004). Charles Holt's report on exponentially weighted moving averages: an introduction and appreciation. *International Journal of Forecasting*, 20(1), 1–3.

- PORDATA. (2022). *Alojamentos Turísticos: estabelecimento*. Retrieved from <https://www.pordata.pt/DB/Portugal/Ambiente+de+Consulta/Gráfico>
- Porter, M. (1985). *Competitive advantage*. Retrieved from Free Press.
- Porto, B. M., Philippi, D. A., & Leite, V. A. W. (2020). Previsão dos modelos univariados e rede neural da demanda turística do estado de Mato Grosso do Sul. *Caderno Virtual de Turismo*, 19(3).
- Ramos, D. M., & Costa, C. M. (2017). Turismo: Tendências de Evolução. *PRACS: Revista Eletrônica de Humanidades Do Curso de Ciências Sociais Da UNIFAP*, 10, 21–33.
- Ramos, F. R. (2021). *Data Science na Modelação e Previsão de Séries Económico-financeiras: das Metodologias Clássicas ao Deep Learning*. Tese de Doutoramento, Instituto Universitário de Lisboa - ISCTE Business School.
- Ramos, F. R., Costa, A., Mendes, D. & Lopes, D. R. (2021). *Explorando o poder da memória das redes neuronais LSTM na modelação e previsão do PSI 20*. SPE 2021 – XXV Congresso da Sociedade Portuguesa de Estatística. Évora.
- Ramos, F. R., Lopes, D. R. & Pratas, T. E. (2022). *Deep Neural Networks: A Hybrid Approach Using Box&Jenkins Methodology*. In: Innovations in Mechatronics Engineering II. *icieng 2022. Lecture Notes in Mechanical Engineering*. Springer, Cham.
- Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain. *Psychological Review*, 65-386.
- Roser, M. (2019). *International tourist: Number of arrivals*. Retrieved from <https://ourworldindata.org/tourism>
- RTP. (2021). *RTP*. Retrieved from <https://www.rtp.pt/programa/tv/p40952>
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464.
- Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). *Financial time series forecasting with deep learning: A systematic literature review: 2005–2019*. Retrieved from <https://arxiv.org/pdf/1911.13288.pdf>
- Silva, I. (2021). *SLRMss: um pacote em R para a correção de testes em modelos de regressão linear simétricos*. Tese de Mestrado, Universidade Federal do Rio Grande do Norte.
- Simões, P., Costa, J., Provenza, M., Xavier, V., & Goulart, J. (2019). Modelos De Previsão Para Temperatura E Salinidade No Fenômeno De Ressurgência: Análise Dos Dados Da Boia 19°00's34°00'w No Período Entre 2005 E 2014. *XIX Simpósio de Pesquisa Operacional e Logística Da Marinha*, 1–23.

- Song, H., & Li, G. (2021). Editorial: Tourism forecasting competition in the time of COVID-19. In *Annals of Tourism Research*, 88, [103198].
- Song, H., Qiu, R., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, 75(3), 338–362.
- Tealab, A. (2018). Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*, 3, 334–340.
- Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, 38, 788–804.
- Turismo de Portugal (2013). *Vá para fora cá dentro*. Retrieved from <https://www.facebook.com/Vaparaforacadentro/>
- Turismo de Portugal (2017). *Estratégia Turismo 2027: Liderar o Turismo do Futuro*. Retrieved from <http://www.turismodeportugal.pt/SiteCollectionDocuments/Estrategia/estrategia-turismo-2027.pdf>
- Turismo de Portugal (2021a). *Prémios e distinções atribuídos ao Turismo de Portugal*. Retrieved from http://www.turismodeportugal.pt/pt/quem_somos/Organizacao/Premios_Distincoes/Paginas/default.aspx
- Turismo de Portugal (2021b). *Turismo de Portugal lança Programa Empresas Turismo 360*. Retrieved from <http://www.turismodeportugal.pt/pt/Noticias/Paginas/Turismo-de-portugal-lanca-programa-empresas-Turismo-360.aspx>
- Turismo de Portugal (2022). *Organização e Parcerias*. Retrieved from http://www.turismodeportugal.pt/pt/Turismo_Portugal/Organizacao_Parceiros/Paginas/default.aspx
- TV Guia (2018). *TV Guia*. Retrieved from <https://www.pressreader.com/portugal/tv-guia/20180824/281779924982552>
- Var, T., & Gunn, C. (2002). *Tourism Planning: Basics, Concepts, Cases* (4rd ed). UK: Routledge.
- Vieira, C. S. (2021). *Turismo de negócios: e se Bill Gates tem razão?* Retrieved from <https://www.dn.pt/opiniao/Turismo-de-negocios-e-se-bill-gates-tem-razao-1-13464921.html>
- Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, 13(3), 335–364.

- Winters, P. R. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science*, 6(3), 324–342.
- Zhang, H., Song, H., Wen, L., & Liu, C. (2021). Forecasting tourism recovery amid COVID-19. *Annals of Tourism Research*, 87, 1-10.
- Zhang, Y., Guo, Q., & Wang, J. (2017). Big data analysis using neural networks, 49, 9–18.

7. Anexos

7.1 Anexo A

Prémios Turismo de Portugal (Turismo de Portugal, 2021)

2021

World Travel Awards 2021 – edição Europa:

- Turismo de Portugal eleito Melhor Organismo Oficial de Turismo Europeu (*Europe's Leading Tourist Board*)

Prémios Marketing Meios & Publicidade:

- Turismo de Portugal conquistou o Ouro com a campanha *Can't Skip Hope*, na categoria Viagens & Turismo dos Prémios Marketing Meios & Publicidade, prata na categoria Digital e Redes Sociais e Bronze em Mercados Internacionais
- A campanha *Read Portugal* foi também distinguida com Prata na categoria Mercados Internacionais e Bronze na categoria Viagens e Turismo
- “Tu Podes” recebeu Prata nas categorias Viagens e Turismo e Digital e Redes Sociais
- A campanha *Wine Tourism* foi distinguida com Prata nas categorias Viagens e Turismo e Mercados Internacionais

5ª Edição dos Prémios Índice da Excelência 2020:

- Turismo de Portugal vence na categoria Setor Público e conquista o 8.º lugar das Grandes Empresas, passando a integrar o Top 10 deste ranking

2020

World Travel Awards:

- Visitportugal.com - Melhor Site Oficial de Turismo Mundial, na categoria *Travel Technology*
- Portugal foi eleito, pelo quarto ano consecutivo, o Melhor Destino da Europa

Country Brand Awards:

- Portugal lidera o ranking Europeu dos melhores países na promoção do Turismo e conquista o terceiro lugar mundial (bronze)

Prémios Eficácia 2020:

- O Turismo de Portugal foi distinguido com ouro pelo filme *#CantSkipHope* na categoria “Transportes, Viagens e Turismo”

WTTC:

- Portugal é o primeiro país Europeu a receber o Selo ‘*Safe Travels*’ do *World Travel & Tourism Council*

12.ª edição dos Prémios *Marketeer*:

- Turismo de Portugal ganha a categoria Turismo Destino

Prémios Marketing Meios & Publicidade:

- Turismo de Portugal - Ouro nas categorias Viagens & Turismo e Mercados Internacionais, com a campanha *#Brelcome* e «Onda *Times Square*»

2019

Reconhecimento PRÁTICAS RS E SUSTENTABILIDADE:

- Turismo de Portugal - vence a categoria Comunidade da 5.ª Edição - 2019 do prémio dinamizado pela Associação Portuguesa de Ética Empresarial, com o projeto “Educar para a responsabilidade”, reconhecendo as ações de responsabilidade social e sustentabilidade dinamizadas pela rede de Escolas do Turismo de Portugal.

***World Travel Awards* 2019:**

- Turismo de Portugal - Melhor Organismo Oficial de Turismo do Mundo, pelo terceiro ano consecutivo;
- Portugal - Melhor Destino Turístico do Mundo, pelo terceiro ano consecutivo.

Prémios Eficácia 2019:

- Turismo de Portugal - ouro pela «Onda em *Times Square*» nas categorias Utilização Criativa de Meio e Transportes e Viagens e Turismo

***Tourism Video Competition* da Organização Mundial do Turismo 2019:**

- Turismo de Portugal - a campanha internacional «Portugal. *The Summer*» venceu a 4.ª edição da competição

Prémios Criatividade Meios & Publicidade 2019:

- Turismo de Portugal – Anunciante do Ano
- Turismo de Portugal - «Onda *Times Square*» - Grande Prémio na categoria “Formato Especial – Digital” e Ouro nas categorias “Comunicação de Produto e Serviço” e “Mercados Internacionais”
- Turismo de Portugal - Campanha *#Brelcome* - Prata nas categorias “Filme”, “Redes Sociais” e “Mercados Internacionais”

***World Travel Awards* 2019 - edição Europa:**

- Turismo de Portugal - Melhor Organismo Oficial de Turismo da Europa, pelo sexto ano consecutivo.
- Portugal - Melhor Destino Europeu, pelo terceiro ano consecutivo.

11.ª edição dos Prémios *Marketeer*:

- Turismo de Portugal ganha a categoria Turismo Destino

XXI Festival do Clube de Criativos de Portugal:

- Turismo de Portugal - Anunciante do Ano
- Turismo de Portugal - Grande Prémio Jornalistas

- Turismo de Portugal - Melhor Cliente nas categorias Ativação de Marca e Publicidade, pela campanha “Onda gigante da Nazaré em Nova Iorque”.

Prémios Marketing Meios & Publicidade 2019 | 5 galardões:

- Turismo de Portugal - Ouro com a campanha internacional *Can't Skip Portugal*, na categoria Viagens & Turismo
- Turismo de Portugal - Prata com a campanha Um Plano Pelo Centro, nas categorias Viagens & Turismo e *Brand Content*.
- Turismo de Portugal - Prata com a campanha internacional *Can't Skip Portugal* nas categorias Digital & Redes Sociais e Mercados Internacionais.

2018

World Travel Awards 2018:

- Turismo de Portugal - Melhor Organismo Oficial de Turismo Mundial
- Visitportugal.com - Melhor Site Oficial de Turismo Mundial
- Portugal - Melhor Destino Turístico do Mundo e Melhor Destino Turístico Europeu

World Golf Awards 2018:

- Portugal Melhor Destino Mundial e Europeu de Golfe

Prémio ACEPI Navegantes XXI (Associação da Economia Digital - ACEPI) 2018:

- Plataforma *Portuguese Trails*

Prémios Meios & Publicidade 2018:

- Marca do Ano: Turismo de Portugal
- Personalidade de Marketing: Lúcia Monteiro (Diretora Coordenadora da Direção de Apoio à Venda)

2017

14.ª edição dos prémios UNWTO 2017:

- 1.º lugar na categoria Inovação e Políticas Públicas com o projeto formativo *Tourism Training Talent* das Escolas do Turismo de Portugal

World Travel Awards 2017:

- Turismo de Portugal - Melhor Organismo Oficial de Turismo Mundial
- Visitportugal.com - Melhor Site Oficial de Turismo Mundial
- Portugal - Melhor Destino Turístico do Mundo e Melhor Destino Turístico Europeu

World Golf Awards 2017:

- Portugal Melhor Destino Mundial e Europeu de Golfe

Prémios à Eficácia da Comunicação da Associação Portuguesa de Anunciantes – APAN:

- 3 galardões Prata nas categorias de Internacionalização, Comunicação Tática e *Low Budget* com a campanha internacional de Turismo “*Can't Skip Portugal*” e a ativação da marca durante o Euro2016

2008 – 2016

World Travel Awards 2014, 2015 e 2016:

- Melhor Organismo Oficial de Turismo da Europa - medalha de ouro

EIBTM Barcelona 2014:

- *Best Stand Design Overall*

Prémio Meios&Publicidade 2014:

- Marca do Ano

World Travel Market 2010:

- Melhor Stand

FITUR 2008, 2011, 2012, 2013 e 2015:

- Melhor Stand

Prémio Marketeer "Turismo" 2009 e 2014.

Ulysses 2009:

- Atribuído pela Organização Mundial do Turismo na categoria "Inovação e Aplicação do Conhecimento em *Governance*".

Melhor Organismo Oficial do Setor 2008 e 2009:

- "*Best Tourist Office*" nos *Danish Travel Awards* - prémios mais importantes de Turismo na Dinamarca

7.2 Anexo B

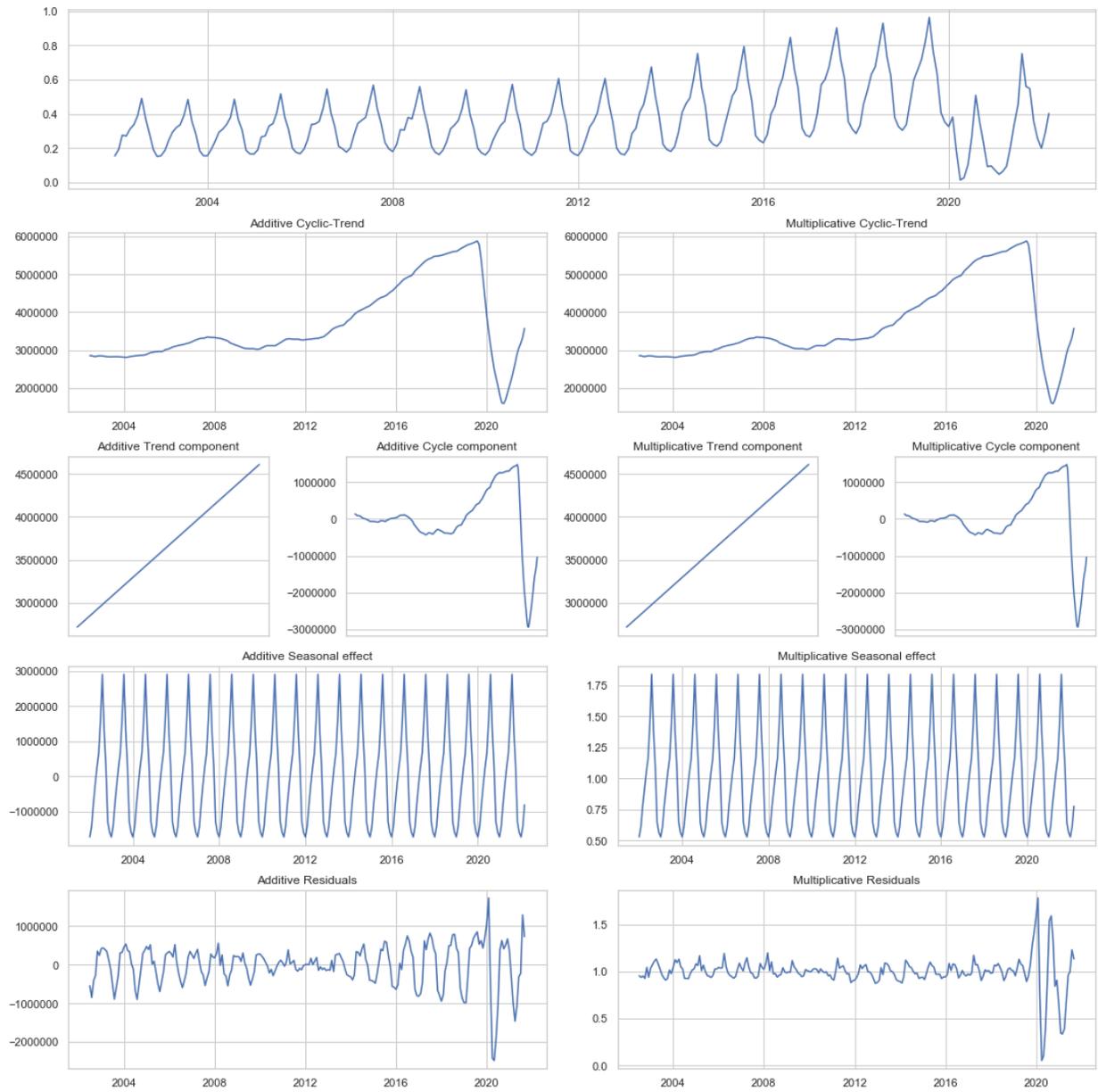


Figura 7.1 – Série Total: Decomposição (aditiva e multiplicativa)

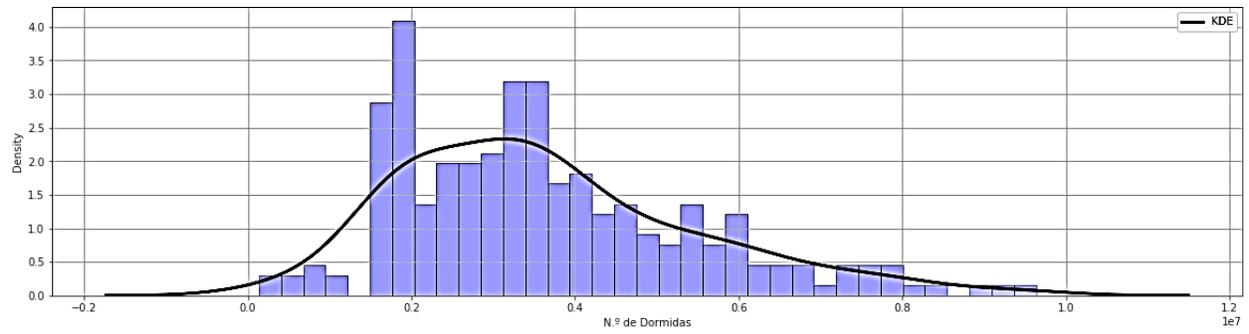


Figura 7.2 – Número Total: Histograma (com curva de densidade)

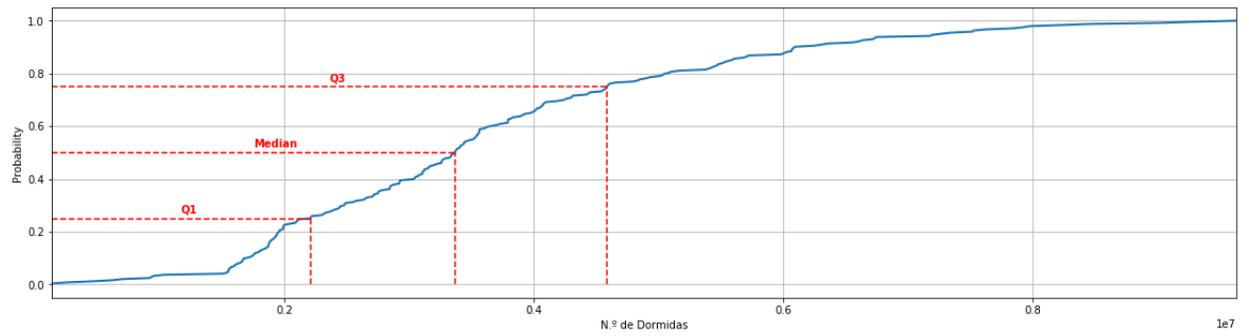


Figura 7.3 – Série Total – Função cumulativa de distribuição

7.3 Anexo C

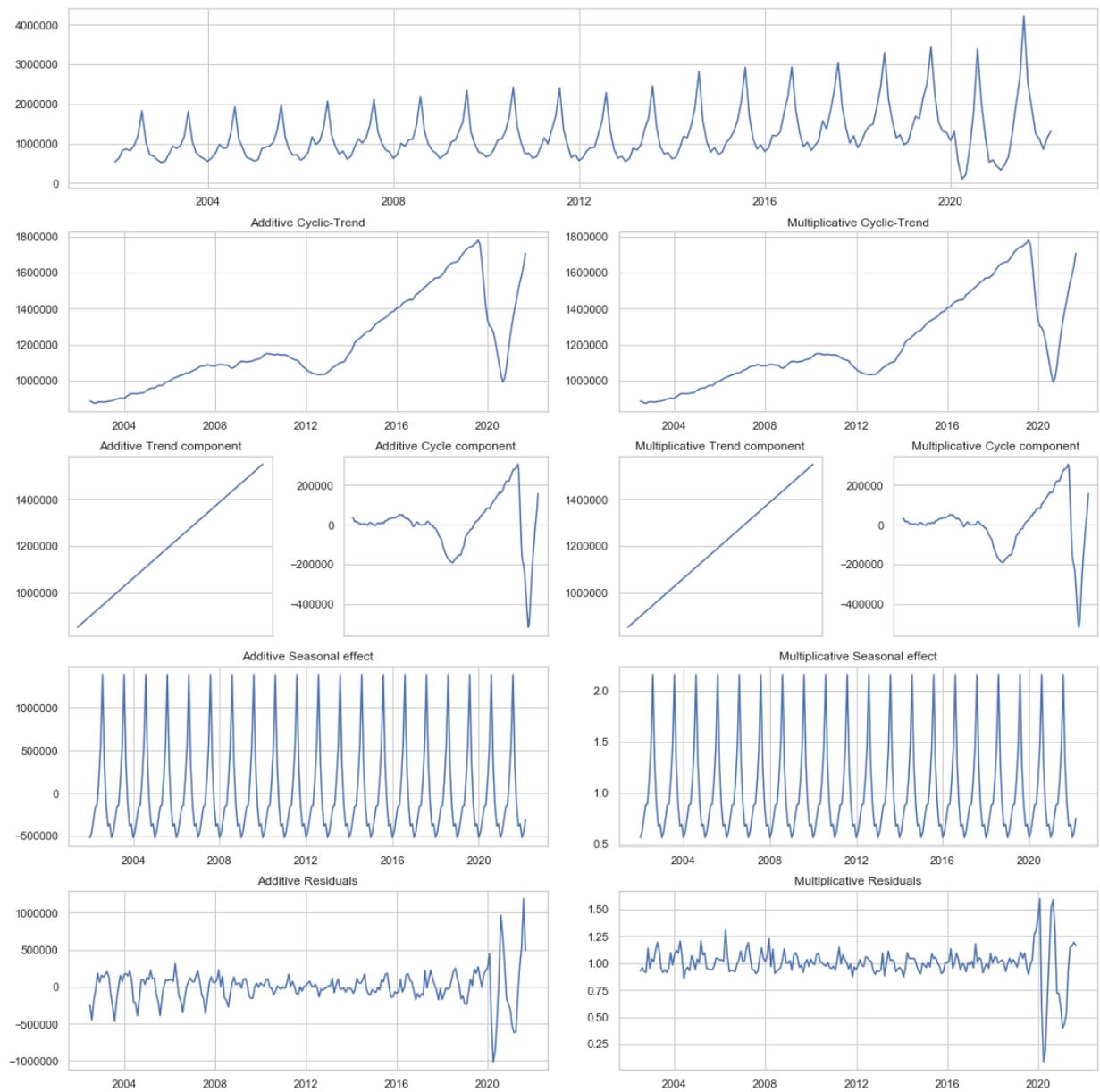


Figura 7.4 – Série Residentes: Decomposição (aditiva e multiplicativa)

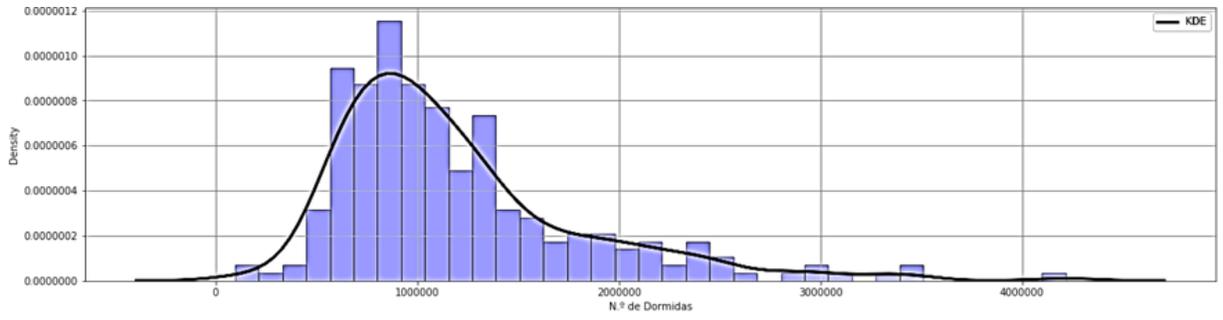


Figura 7.5 – Número Residentes: Histograma (com curva de densidade)

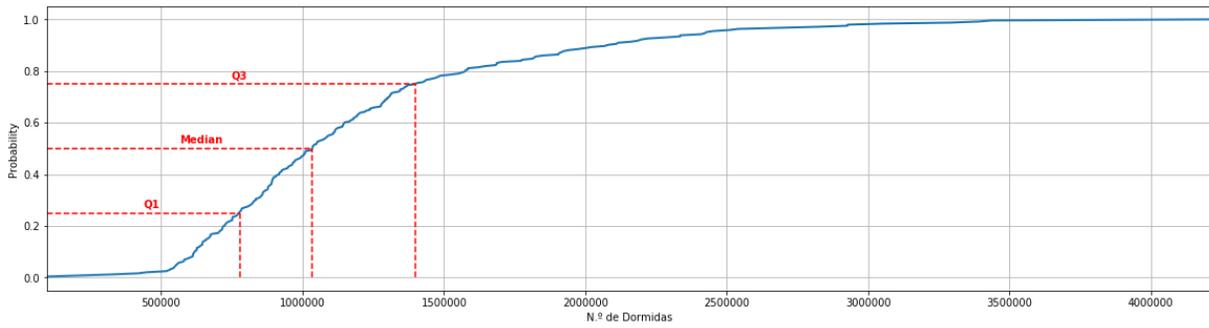


Figura 7.6 – Série Residentes: Função cumulativa de distribuição

7.4 Anexo D

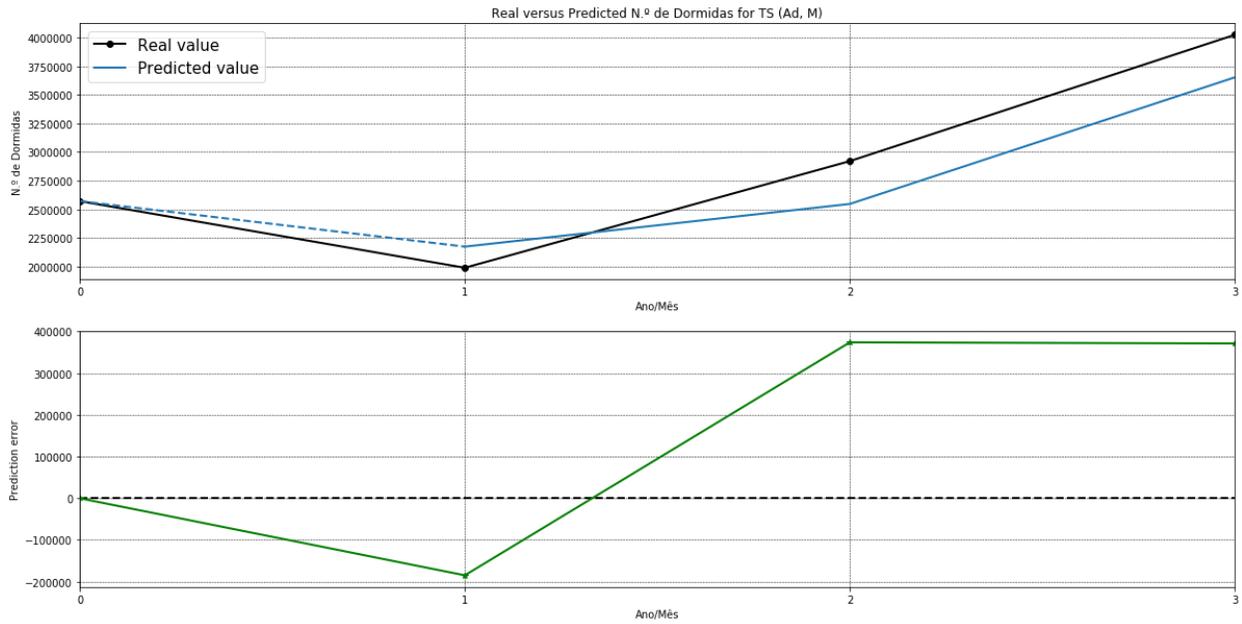


Figura 7.7 – Modelo ETS da Série Total

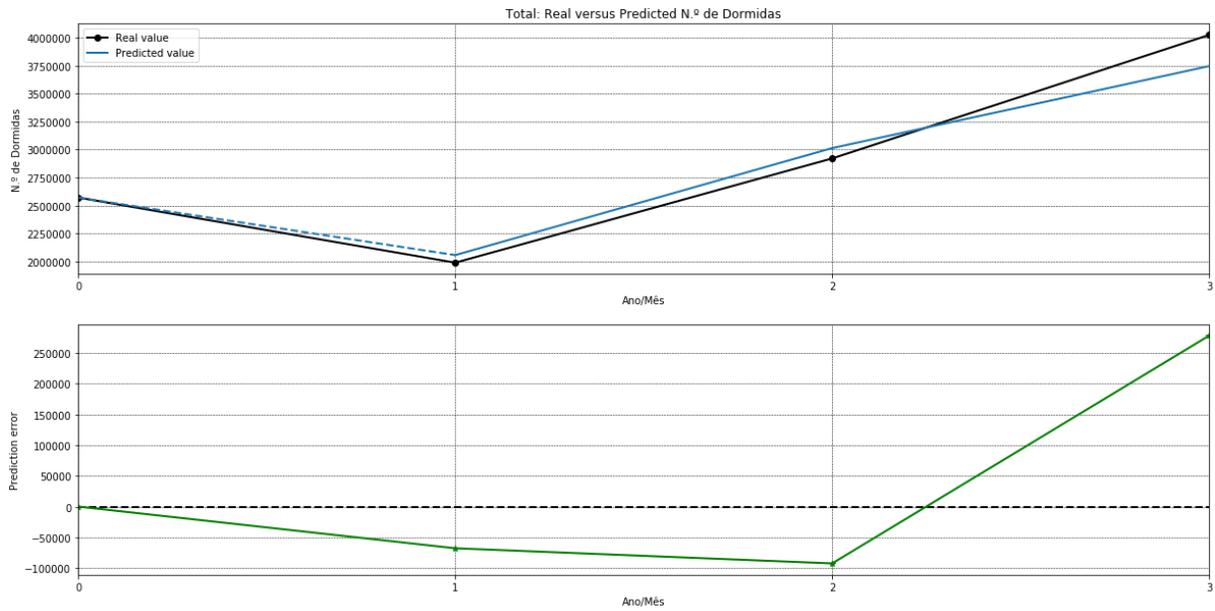


Figura 7.8 – Modelo BJ-DNN da Série Total

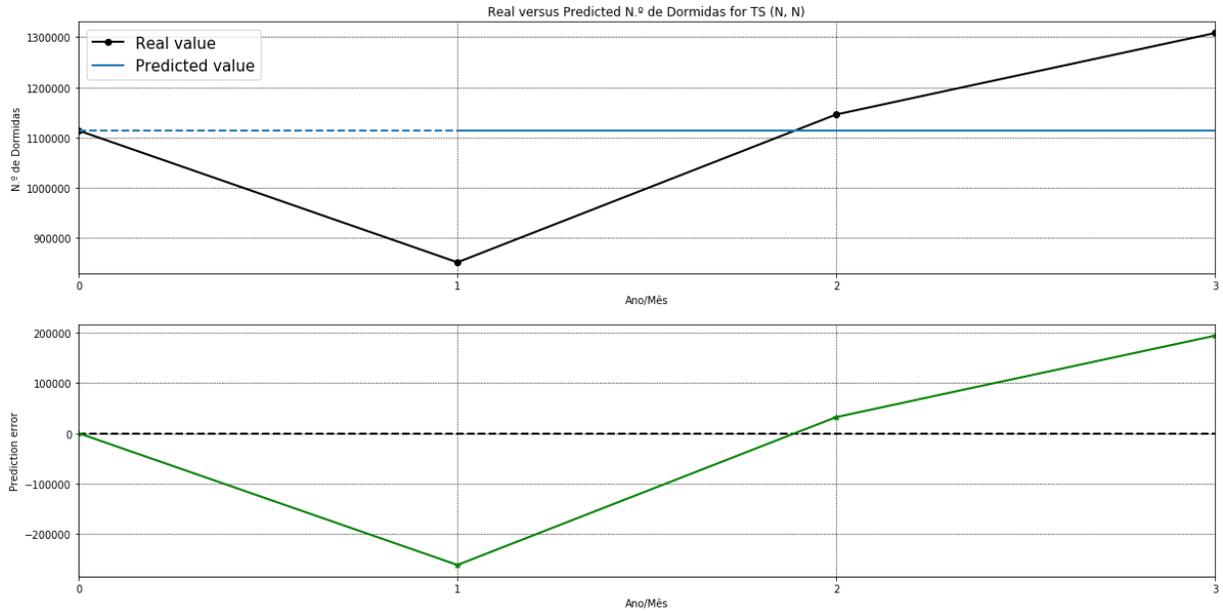


Figura 7.9 – Modelo ETS da Série Residentes

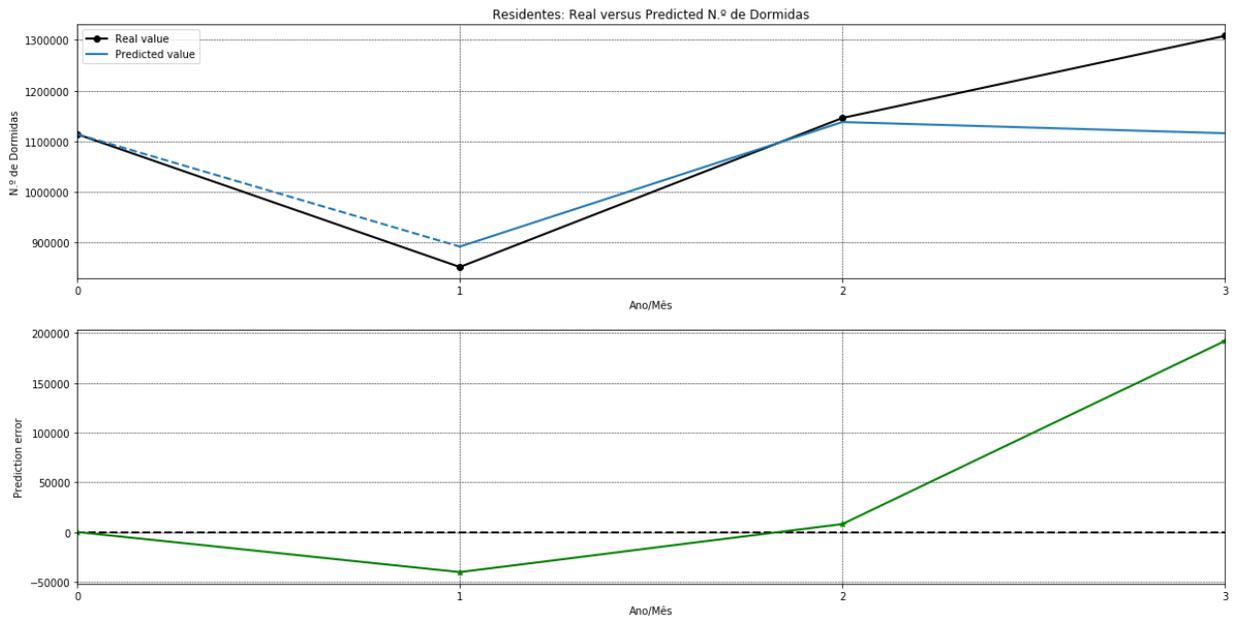


Figura 7.10 – Modelo BJ-DNN da Série Residentes