



# DESAFIOS NA MODELAÇÃO E PREVISÃO DE SÉRIES TEMPORAIS NO SETOR ALIMENTAR/CADEIAS DISTRIBUIÇÃO/LOGÍSTICA: DAS METODOLOGIAS CLÁSSICAS AO DEEP LEARNING.

NUNO MOURA MONTEIRO

julho de 2024

**DESAFIOS NA MODELAÇÃO E PREVISÃO DE SÉRIES  
TEMPORAIS NO SETOR ALIMENTAR/CADEIAS  
DISTRIBUIÇÃO/LOGÍSTICA: DAS METODOLOGIAS  
CLÁSSICAS AO DEEP LEARNING**

**Nuno Moura Monteiro**

**Dissertação para obtenção do Grau de Mestre em  
Engenharia Mecânica, Área de Especialização em  
Gestão Industrial**

**Orientador: Doutora Marisa João Guerra Pereira De Oliveira**

**Coorientador: Doutora Maria Teresa Ribeiro Pereira**

**Júri:**

Presidente:

Luís Norberto Miranda Torres, professor adjunto do Instituto Superior de Engenharia do Porto

Vogais:

Filipe Roberto de Jesus Ramos, Professor Auxiliar Convidado da Faculdade de Ciências –  
Universidade de Lisboa

Marisa João Guerra Pereira de Oliveira, professora adjunta do Instituto Superior de  
Engenharia do Porto

Porto, Julho 2024



# Agradecimentos

Em primeiro lugar deixo o meu agradecimento à minha família que sempre me apoiou neste meu percurso académico. Aos meus amigos deixo um grande obrigado por todo o incentivo que me deram, por me terem sempre encorajado a ir mais além, fazendo com que superasse os meus próprios limites. Agradeço também, a todos os professores e docentes que me acompanharam ao longo desta jornada, em especial à Professora Doutora Marisa Oliveira por toda a disponibilidade e orientação nesta dissertação. Ao gabinete de relações externas deixo também a minha gratidão por todo o empenho que demonstraram, tornando possível um dos grandes planos que tinha do ponto de vista pessoal. Aos colegas que encontrei e conheci durante este período por o terem tornado mais leve e divertido. A todos vocês o meu muito obrigado.



# Resumo

O presente relatório tem por base a análise dos desafios na modelação e previsão de séries temporais referentes ao setor alimentar e às respetivas cadeias de distribuição, bem como, os devidos problemas que recaem na logística empresarial. Com o objetivo de obter uma melhor compreensão da temática em análise, foi realizada uma pesquisa bibliográfica onde são expostos diversos métodos de previsão clássicos, sendo eles os modelos qualitativos ou modelos quantitativos e as respetivas subdivisões. É apresentado também o modo de escolha do modelo adequado para a análise de um qualquer caso em estudo. A acrescentar à pesquisa bibliográfica dos referidos métodos de previsão, é também apresentada uma contextualização do tema deste relatório, com vista a uma melhor precessão do historial, desenvolvimentos passados e importância no mundo atual do referido tópico. Subsequentemente, é elaborado o procedimento de criação de uma série temporal, mostrando como se procede à recolha dos dados necessários à sua execução, assim como a exposição das dificuldades inerentes a todo o processo. É ainda descrito a estrutura e funcionamento dos algoritmos de *machine learning* a sua classificação, após a qual são apresentados os modelos de *deep learning*, bem como uma comparação com o anterior e a sua aplicabilidade no estudo de séries temporais. Termina-se com a exposição de um caso de estudo sobre o tema em análise, onde se analisa os índices de volumes de negócios no comércio a retalho nos quais são aplicados diversos modelos de previsão com o intuito de compreender se os modelos de inteligência artificial obtêm um melhor resultado do que as metodologias clássicas. Após a análise do caso de estudo concluiu-se que o modelo clássico utilizado obteve uma boa performance, baseado na métrica de erro MAPE. Dado que a série escolhida segue os parâmetros de normalidade com um forte ciclo de tendência e sazonalidade, foi possível obter em média um erro de 2.39%. Por seu lado, o modelo de inteligência artificial tirando partido do seu maior poder computacional consegue captar as dinâmicas da série em análise alcançando um bom ajuste e podendo assim devolver previsões adequadas. Desta forma obteve melhores resultados que os modelos anteriores, atingindo em média um erro de 1.86% ficando, assim, provado a sua eficácia e futura aplicabilidade.

**Palavras-chave:** Métodos de Previsão; Séries Temporais; Métodos Qualitativos; Métodos Quantitativos; *Deep Learning*.



# Abstract

This report is based on the analysis of the challenges in modeling and forecasting time series in the food sector and in the respective distribution channels, as well as the problems that fall within the scope of corporate logistics. In order to obtain a better understanding of the topic under analysis, a bibliographical analysis was conducted in which several classical forecasting methods are presented, being them qualitative models or quantitative models and their respective subdivisions. It is also presented how to choose the appropriate model for the analysis of any case under study. In addition to the bibliographical research of the referred prediction methods, a contextualization of the subject of this report is also presented in order to better precess the history, past developments and importance in today's world of the referred topic. Subsequently, the procedure for creating a time series is described, showing how to collect the data needed to implement it, as well as the difficulties associated with the whole process. The structure and functioning of machine learning algorithms and their classification are also described, after which, deep learning models are presented, as well as a comparison with the previous one and its applicability in the study of time series. It concludes with a case study on the subject under analysis, analysing turnover indices in the retail trade in which various forecasting models are applied in order to understand whether artificial intelligence models obtain better results than classical methodologies. After analysing the study case, it can be concluded that the classic model used obtained a good performance, based on the MAPE error metric. Given that the chosen series follows the parameters of normality with a strong trend cycle and seasonality, it was possible to obtain an average error of 2.39%. On the other hand, the artificial intelligence model, taking advantage of its greater computing power, manages to capture the dynamics of the series being analysed, achieving a good adjustment and thus being able to return adequate forecasts. This way, it obtained better results than the previous models, achieving an average error of 1.86%, therefore proving its effectiveness and future applicability

**KEYWORDS:** Forecasting Methods; Time Series; Qualitative Methods; Quantitative Methods; Deep Learning.



# Índice

Lista de Figuras.....	xi
Lista de Tabelas.....	xiii
Acrónimos.....	xv
1. INTRODUÇÃO.....	1
1.1. Contextualização.....	1
1.2. Objetivos.....	2
1.3. Metodologia.....	3
1.4. Estrutura do Relatório.....	3
2. REVISÃO BIBLIOGRÁFICA.....	5
2.1. Setor alimentar/cadeias de distribuição/logística.....	5
2.2. Métodos de previsão.....	9
2.2.1. Séries temporais.....	9
2.2.2. Modelos qualitativos.....	11
2.2.3. Modelos quantitativos.....	12
2.2.3.1. Métodos causais ou explicativos.....	12
2.2.3.2. Métodos extrapolativos ou estatísticos.....	13
2.3. Testes de adequação do modelo.....	15
3. DESENVOLVIMENTO DE SÉRIES TEMPORAIS.....	17
3.1. Criação de séries temporais.....	17
3.2. Dificuldades na criação de uma série temporal.....	18
3.3. Como ultrapassar problemas identificados.....	19
4. MACHINE LEARNING.....	21
4.1. Machine learning.....	21
4.1.1. Deep learning.....	22
4.2. Aplicabilidade no estudo de séries temporais.....	23
5. METODOLOGIA.....	25
5.1. Apresentação de dados.....	25
5.2. Implementação computacional.....	25
5.3. Previsão e avaliação do erro.....	28
6. CASO DE ESTUDO.....	29
6.1. Análise exploratória.....	29
6.2. Aplicação dos modelos.....	33
6.3. Comparação dos resultados.....	37
7. CONCLUSÃO.....	39

BIBLIOGRAFIA.....	41
DECLARAÇÃO DE INTEGRIDADE.....	45
ANEXOS .....	47
Anexo A.....	47
Anexo B .....	50

# Lista de Figuras

Figura 1 - Produção e ocupação dos solos (elaboração própria) .....	6
Figura 2 - Produtividade na agricultura (PIB/trabalhador) (elaboração própria) .....	6
Figura 3 - Valor da produção agrícola, 2022 (elaboração própria) .....	7
Figura 4 - Exportações produtos agroalimentares (elaboração própria).....	8
Figura 5 - Canais de distribuição (elaboração própria) .....	8
Figura 6 - Métodos de previsão (adaptado Gonçalves, 2010) .....	10
Figura 7 - Metodologia de implementação computacional dos modelos ETS.....	26
Figura 8 - Metodologia de implementação computacional dos modelos DNN.....	27
Figura 9 - Índices de volume de negócios no comercio a retalho (elaboração própria).....	31
Figura 10 - Gráfico box-plot anual.....	31
Figura 11 - Grafico box-plot mensal.....	32
Figura 12 - Análise dos diversos modelos .....	34
Figura 13 - Comparação dos modelos.....	34
Figura 14 - Resíduos do modelo (A,M).....	35
Figura 15 - Previsões modelo (A,M) .....	35
Figura 16 - Modelo BJ-DNN.....	36
Figura 17 - Previsão modelo BJ-DNN .....	36
Figura B.1 - Histograma com curva de densidade.....	50
Figura B.2 - Função cumulativa de distribuição .....	50
Figura B.3 - Decomposição da série (aditiva e multiplicativa) .....	51



# Lista de Tabelas

Tabela 1 - Índices de volume de negócios no comércio a retalho (Produtos alimentar) .....	29
Tabela 2 - Estatísticas descritivas .....	32
Tabela 3 - Testes de hipótese.....	33
Tabela 4 - Valores de força de tendência e sazonalidade.....	33
Tabela 5 - Nomenclatura dos modelos aplicados .....	34
Tabela 6 - Comparação dos resultados obtidos.....	37
Tabela A.1 - Índices de volume de negócios no comércio a retalho.....	47



# Acrónimos

## Lista de Acrónimos

ISEP	Instituto Superior de Engenharia do Porto
P.PORTO	Instituto Politécnico do Porto
Tt	Tendência
S <sub>T</sub>	Sazonalidade
C <sub>T</sub>	Ciclo
R <sub>t</sub>	Ruido Aleatório



# 1. INTRODUÇÃO

Neste capítulo é realizada uma introdução ao tema em análise “*Desafios Na Modelação E Previsão De Séries Temporais No Setor Alimentar/Cadeias Distribuição/Logística: Das Metodologias Clássicas Ao Deep Learning*”. Começando por uma contextualização do assunto, referindo a sua importância no mundo atual, é também apresentada a história e o desenvolvimento das séries temporais. Posteriormente são explicados os objetivos, que passam pela análise das metodologias de previsões temporais, visando os mais adequados para os respetivos problemas, seguindo-se a metodologia de trabalho que foi adotada para a realização deste relatório. Este capítulo é finalizado com a exposição da estrutura do relatório, onde se pode analisar o conteúdo de cada capítulo.

## 1.1. Contextualização

Um dos maiores problemas da gestão de empresas recai no momento da tomada de decisão sobre a estratégia futura a seguir, uma vez que existe sempre alguma incerteza sobre os períodos subsequentes devido aos vários parâmetros que os influenciam. Deste modo, são procuradas informações nas quais serão possíveis basear a escolha tomada, assim sendo, muitas vezes utiliza-se uma análise de dados obtidos em momentos anteriores, histórico, com o intuito de realizar uma previsão futura. Neste sentido, ao longo dos anos têm vindo a ser desenvolvidos inúmeros métodos de previsão, evoluindo-se dos modelos clássicos para aprendizagem máquina com o aparecimento da Inteligência Artificial, apresentando variados processos de obtenção dessa mesma previsão.

Numa fase inicial da análise das séries temporais é considerado que estas são compostas por quatro componentes não observáveis: a tendência ( $T_t$ ), a sazonalidade ( $S_t$ ), o ciclo ( $C_t$ ) e o ruído aleatório ( $R_t$ ). De acordo com este pressuposto as séries temporais são definidas como função destas componentes,  $Y_t = f(T_t, S_t, C_t, R_t)$ . Em 1938, H. O. Wold, (Wold, 1938) expôs as séries temporais como sendo representáveis por modelos autorregressivos e de médias móveis, sendo que a sua implementação só seria possível na década de 60 com o desenvolvimento de melhores *softwares* de cálculo, mais concretamente os computadores de segunda geração, (Murteira, Muller, & Turkman, 2000).

## Introdução

Na década de 60, com a computação, os métodos automáticos ganhavam força, tornando-se mais populares e mais utilizados, uma vez que, depois de devidamente programados, requeriam pouca intervenção por parte do analista, obtendo um melhor desempenho do que os restantes métodos à data. Baseando-se nos resultados obtidos por Wold (1938), em 1970 Box e Jenkins propuseram a sua própria metodologia (Box & Jenkins, 1976), na qual, após a utilização do processo de diferenciação para transformar a série numa série estacionária, são analisadas as funções de autocorrelação com a finalidade de determinar os parâmetros autorregressivos, os processos de diferenciação e os parâmetros de médias móveis. Este estudo revelou-se um dos mais importantes na área da previsão de séries temporais, tendo sido responsável pelo desenvolvimento e formalização da área.

Recentemente foram desenvolvidas diversas outras técnicas de previsão tais como o método de Forsys, método Ararma e a utilização de combinações de previsões, sendo que atualmente se tem dado grande ênfase à aprendizagem máquina da inteligência artificial: *Machine Learning* e dentro desta o *Deep Learning* e ao uso de inteligência artificial para a análise de séries temporais.

*“The ARARMA methodology of time series forecasting introduced by Parzen has compared well with longer established techniques such as Box and Jenkins ARIMA models (...) The results indicate that there are additional benefits to be gained from the ARARMA approach.”*, (Meade & Smith, 1985).

A análise destas séries temporais pode assumir vários objetivos, podendo estes passar pela descrição do comportamento da série ou pela compreensão dos diversos parâmetros que a compõem, chegando até à previsão de valores futuros. Esta análise representa uma área de estudo que tem vindo a ser objeto de grande estudo e desenvolvimento ao longo dos anos com uma vasta seleção de artigos científicos a serem redigidos por diversos investigadores, (Chatfield, 2016).

## 1.2. Objetivos

O objetivo principal deste trabalho prende-se com comparação entre as metodologias clássicas e os modelos de inteligência artificial, tentando concluir quais as mais adequadas para as previsões futuras dos setores em análise. O presente trabalho tem como objetivo complementar a análise das diversas metodologias de previsão de séries temporais que se encontram no setor alimentar, nas cadeias de distribuição e no setor da logística, verificando quais poderão ser aplicáveis e quais serão os pressupostos para a sua utilização. Com este propósito será redigida uma revisão bibliográfica sobre o setor em questão, de forma a melhor compreender o passado e quais os obstáculos que poderão surgir no futuro, com vista a

compreender o impacto de acontecimentos de ordem global, tais como a pandemia de COVID-19 ou a guerra na Ucrânia.

### 1.3. Metodologia

Para a realização desta dissertação a metodologia seguida começa por uma revisão de literatura sobre método de previsão, inicializando-se pelos modelos clássicos e mais recentemente os baseados em Inteligência Artificial assentes na aprendizagem máquina. Seguidamente, considerando o objetivo desta dissertação, procurou-se data sets, dados, quer públicos quer privados, para a realização de um caso de estudo no setor alimentar com vista à aplicação dos modelos de previsão identificados e à sua comparação em termos de erro, parâmetros de escolha e impacto no contexto dos resultados obtidos.

### 1.4. Estrutura do Relatório

Este relatório encontra-se dividido em sete capítulos. No primeiro capítulo é feita uma breve introdução do tema que será analisado neste trabalho, sendo esclarecido o contexto do tema no mundo atual, com a explicação dos objetivos, metodologia usada e terminando com a estrutura do relatório. Com vista a responder com a maior qualidade aos objetivos propostos na elaboração desta dissertação foi feita pesquisa bibliográfica sendo analisados diversos livros e artigos científicos sobre o tema, tendo o cuidado de utilizar apenas fontes credíveis e que tenham sido verificadas. Deste modo, foram utilizadas algumas bibliotecas online tais como o "b-on", "EBSCO eBook", "Google Scholar", "Science.gov", "SciELO" e Repositórios científicos como o RECIIP e RCAAP.

No seguinte capítulo, dois, é feita uma revisão bibliográfica sobre o tema em questão, começando por uma abordagem ao setor alimentar, às cadeias de distribuição e à área da logística. Posteriormente são explicados os vários métodos de previsão, em concreto os métodos qualitativos e os métodos quantitativos.

No terceiro capítulo é aprofundada a temática das séries temporais, apresentando o seu processo de criação e as formas de obtenção dos dados necessários para o mesmo. São também expostos alguns problemas e dificuldades comuns no desenvolvimento e na aplicação das referidas séries temporais, bem como algumas formas de os ultrapassar e eliminar.

No capítulo posterior, quarto, é analisado o tópico da inteligência artificial, mais concretamente do *machine learning* e, nestes, do *deep learning*, mostrando quer o seu conceito quer os seus subconjuntos. O capítulo termina mostrando a sua aplicabilidade no estudo das séries temporais.

## Introdução

Nas últimas etapas desta dissertação, no capítulo cinco, é apresentada a metodologia seguida no caso de estudo, expondo a implementação do modelo clássico e de inteligência artificial utilizados, bem como a métrica de erro que será utilizada para os comparar. Posteriormente, no capítulo seis é apresentada a análise dos Índices de Volume de Negócios no Comércio a Retalho, com o propósito de prever valores futuros através da utilização de métodos de previsão. É realizada uma análise exploratória da série de dados recolhidos, na qual serão aplicados os modelos clássicos, analisando qual obteve o menor erro. Esse mesmo modelo será, de seguida, comparado com o modelo de inteligência artificial de modo a entender qual obteve o melhor resultado.

No capítulo sete é feita uma conclusão dos trabalhos desenvolvidos nesta dissertação e são feitas considerações sobre o trabalho futuro.

## **2. REVISÃO BIBLIOGRÁFICA**

Neste capítulo será realizada uma revisão bibliográfica sobre o tema a desenvolver nesta dissertação, visando a sua melhor compreensão. Começando por uma breve apresentação do setor alimentar, cadeias de distribuição e logística, passando depois à apresentação e exposição dos vários métodos de previsão, explicando as séries temporais e os respetivos métodos qualitativos e quantitativos. No último subcapítulo é explicada a metodologia de escolha entre os vários modelos de previsão.

### **2.1. Setor alimentar/cadeias de distribuição/logística**

Desde tempos antigos que a produção de alimentos toma uma elevada importância para o ser humano, pois dela necessita para a sua sobrevivência e foi através do bom planeamento da mesma que se tornou possível a criação e o desenvolvimento das primeiras cidades. A partir do momento em que esta necessidade se tornava satisfeita, estas cidades podiam então concentrar-se no seu desenvolvimento, criando novas infraestruturas e estabelecendo novas tecnologias, tornando assim a região mais desenvolvida e atrativa, fazendo com que a cidade crescesse.

Com a revolução industrial, o setor alimentar mudou de forma drástica, estando agora focado essencialmente em produzir alimentos que possuam um prazo de validade mais extenso, dado que cada vez menos pessoas tinham acesso direto aos produtos frescos que necessitam, devido ao grande aumento das zonas urbanas. Os alimentos deixaram de ser provenientes das próprias mãos dos cidadãos passando a serem fabricados e distribuídos por grandes produtores e corporações, resultando na descida dos preços dos alimentos e numa maior acessibilidade dos mesmos.

Mais recentemente, a introdução de fertilizantes, adubos e maquinaria mais avançada possibilitou o aumento considerável de bens alimentares, bem como uma maior velocidade de produção, colheita e processamento destes, tal como é possível observar através da figura 1, que retrata a produção de milho, comparando-a com a sua área utilizada (verifica-se que apesar da área utilizada permanecer relativamente constante, a produção destes bens tem aumentado consideravelmente). A posterior automatização que este setor teve levou a um aumento da

eficiência quer da sua produção quer da sua preparação, levando a que a produtividade deste setor crescesse (figura 2 que demonstra a evolução da produtividade na agricultura).

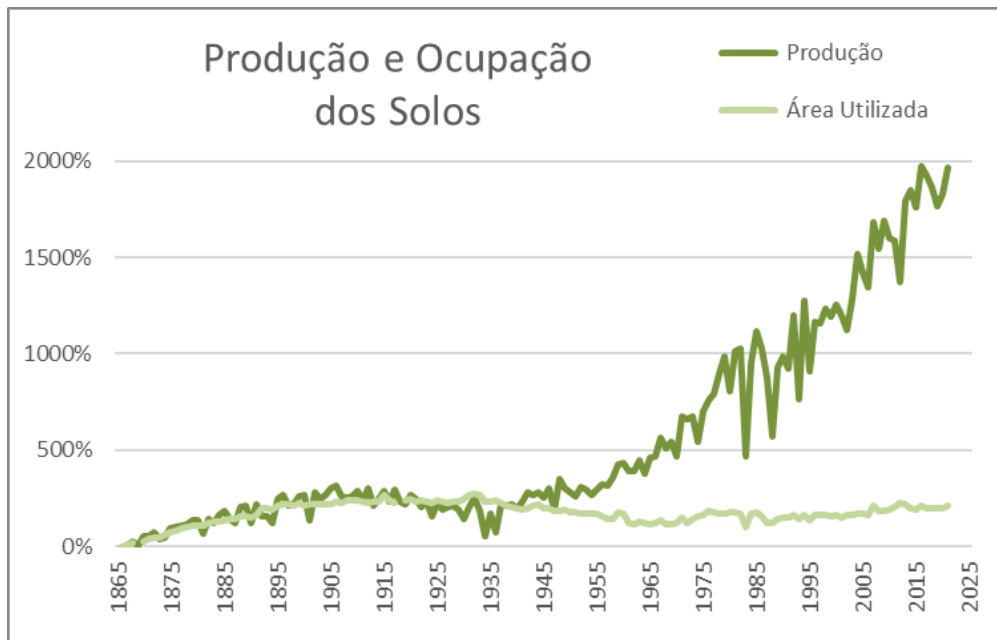


Figura 1 - Produção e ocupação dos solos (elaboração própria)

(Os valores medem a variação percentual da produção de milho e da utilização das terras nos Estados Unidos em relação ao primeiro ano da série cronológica)

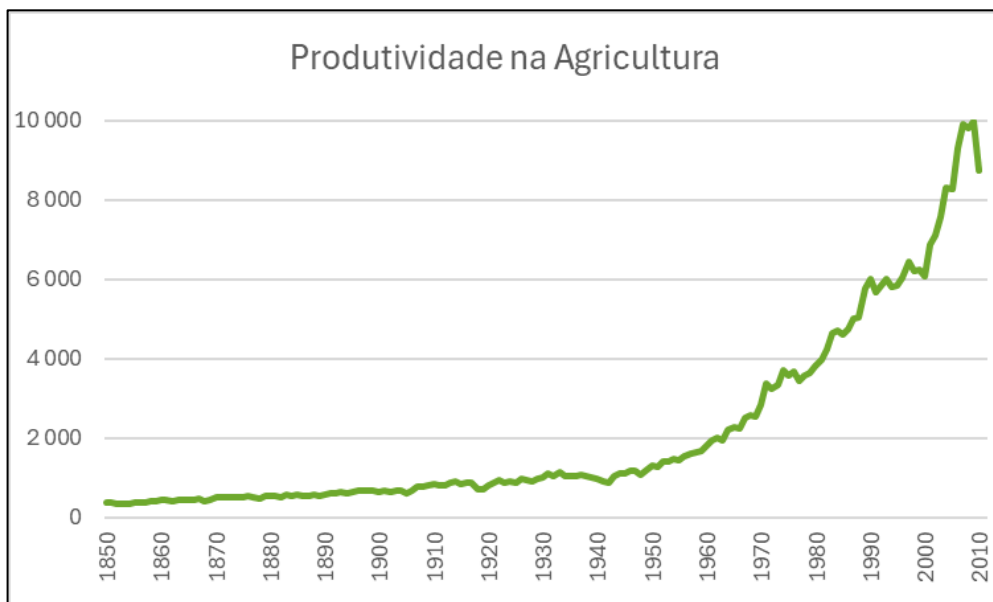


Figura 2 - Produtividade na agricultura (PIB/trabalhador) (elaboração própria)

(A produtividade do trabalho corresponde ao rácio entre o valor acrescentado na agricultura e o número de pessoas empregadas na agricultura, Suécia)

Juntamente com isto, os avanços na preservação destes alimentos deram a possibilidade à extensa gama de produtos disponíveis no mercado, indo desde os produtos frescos e naturais até aos mais processados, passando pelos congelados ou até mesmo os desidratados.

Atualmente este setor abrange uma vasta gama de atividades industriais, desde a produção de alimentos (como o setor agrícola, piscatório ou vinícola) à exportação dos mesmos (como é o caso do comércio ao retalho), passando pela confeção de bens para a preparação dos alimentos (caso do setor da moagem), até ao armazenamento destes (tais como o setor das conservas). Esta área envolve diversos processos complexo que são estritamente regularizados de forma a manter os níveis de higiene adequados e prevenir a contaminação do produto, de forma que o consumível final seja seguro e de qualidade.

É estimado que, a nível mundial, a produção alimentar atinja cerca de 4.2 triliões de euros até 2026, com um crescimento médio anual de 1,1%. Já na Europa prevê-se que as vendas relacionadas com artigos alimentares tenham um crescimento de 0,9% por ano, totalizando 1,2 triliões de euros, até à mesma data, (Linker, 2022).

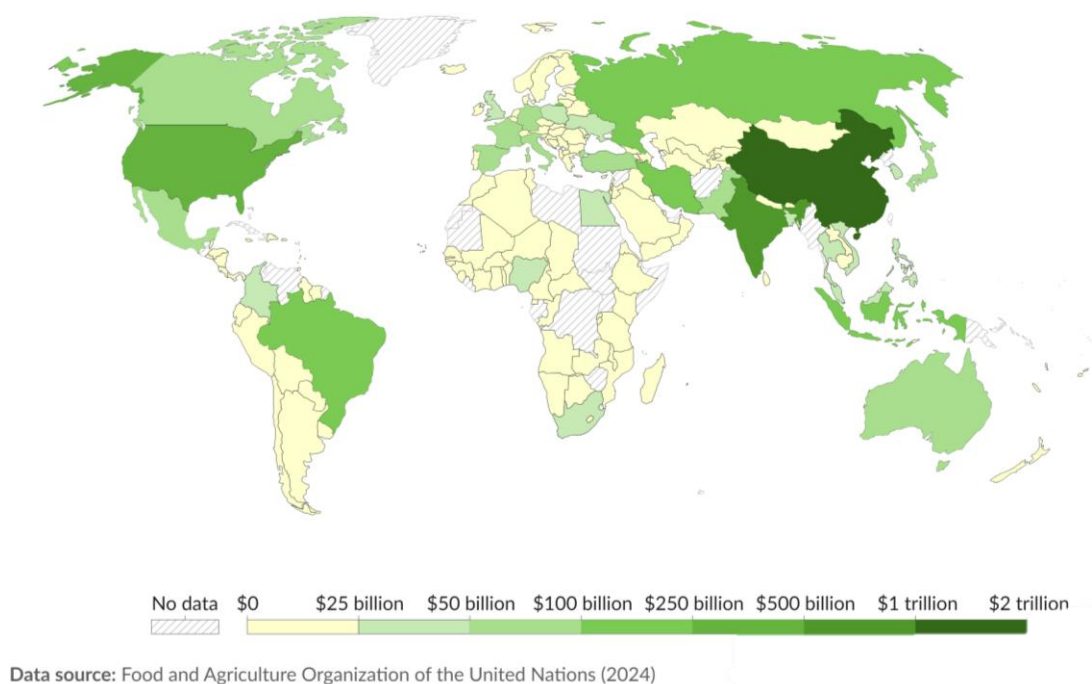


Figura 3 - Valor da produção agrícola, 2022 (elaboração própria)

Em Portugal, a indústria alimentar é um dos setores com maior peso na nossa economia, tendo-se vindo a observar um aumento considerável da mesma. De igual forma, tem-se constatado um crescimento das suas exportações, como se pode conferir no gráfico que se segue, onde se pode verificar que as exportações de produtos agroalimentares (milhões de euros) têm mantido um crescimento constante ao longo dos anos e consequentemente da sua percentagem na totalidade das exportações do nosso país, (INE, 2023).

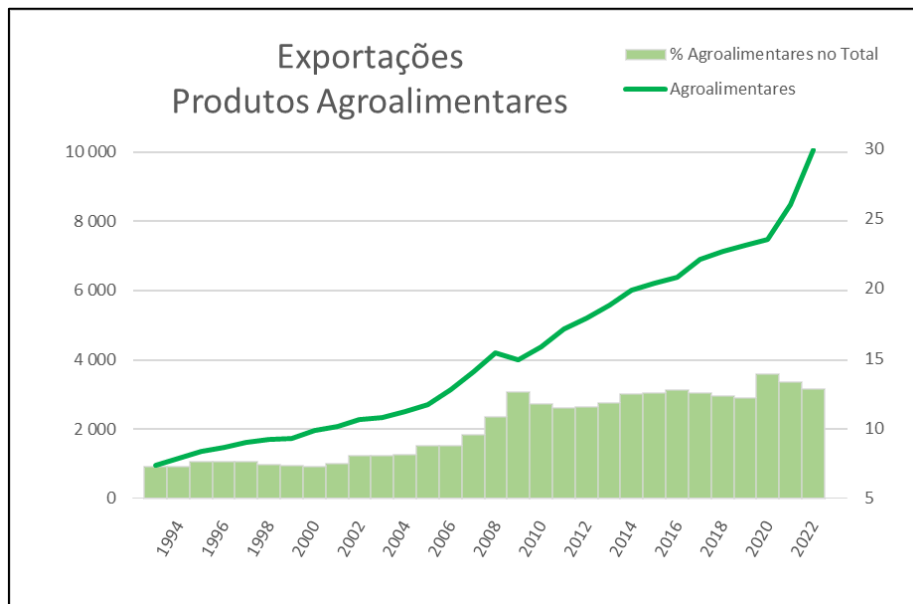


Figura 4 - Exportações produtos agroalimentares (elaboração própria)

Para que todo este setor funcione em pleno é necessário que as várias partes integradoras funcionem em concordância. Deste modo é essencial que as cadeias de distribuição sejam corretamente planeadas, ligando com sucesso os produtores aos consumidores.

As cadeias de distribuição são responsáveis por fazer chegar o produto ao consumidor final, servindo-se de redes de distribuidores, armazéns ou lojas físicas para o conseguirem. Podendo os produtores optarem por uma distribuição direta onde controlam todo o percurso até ao consumidor final ou então optar por uma distribuição indireta, envolvendo intermediários que se encarregam de fazer chegar o produto final ao seu destino.

Geralmente, a escolha de um canal de distribuição direto representa menores custos para o consumidor, uma vez que estes estarão a comprar diretamente aos produtores. Por seu lado, um canal de distribuição indireto passa por diversos intervenientes, começando no produtor o qual transfere o consumível para um vendedor grossista, que irá vender grandes quantidades do produto aos diferentes revendedores, os quais por fim vendem o produto ao cliente final.

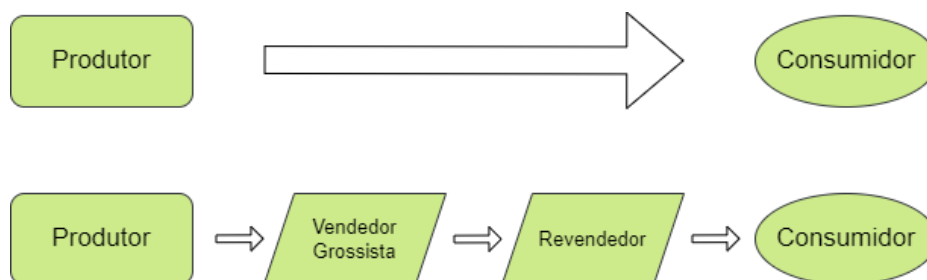


Figura 5 - Canais de distribuição (elaboração própria)

Estes canais de distribuição possuem um papel fundamental no desempenho de uma empresa ou organização, garantindo que os produtos produzidos serão transportados até ao consumidor final no momento correto. Um mau planeamento destes canais de distribuição pode levar a que o comprador final receba um excesso de produtos ou possivelmente a ter falta destes. Erros desta natureza podem deixar a empresa ou organização num momento menos proveitoso, aumentando a insatisfação dos clientes, podendo mesmo vir a perder alguns, reduzindo assim a sua cota de mercado e por consequente a sua capacidade de venda e lucro final.

Todos estes problemas recaem na logística da empresa, onde terão de ser escolhidos os processos e métodos mais eficientes para a sua realização. Como se pode imaginar este pode ser um problema de grande dimensão, pois apesar dos destinatários finais permanecerem geograficamente estáticos, as quantidades que são requeridas e, por consequente, expedidas poderão variar com grande facilidade. Deste modo, é necessária alguma metodologia que ajude a prever os referidos envios e é neste momento que entram na equação os métodos de previsão, (Vriens & Versteijnen, 2016).

## **2.2. Métodos de previsão**

Métodos de previsão são procedimentos pré-estabelecidos e guiados por uma metodologia que permite prever de forma mais exata o futuro de uma sequência temporal. Assim, através destas metodologias é possível visualizar, analisar e possivelmente prever o padrão futuro da referida série temporal, (Brockwell & Davis, 1991).

Por vezes, a incerteza, associada aos vários parâmetros que influenciam um fenómeno temporal, não permite obter o conhecimento completo da série temporal em estudo, devido a este problema é necessário realizar previsões com o objetivo de tentar determinar o seguimento da série. Com isto, a empresa ou organização consegue ter uma base para se sustentar na escolha do melhor percurso e estratégia a tomar.

As análises podem ser classificadas segundo a sua hierarquia temporal, podendo ser previsões a curto, médio e longo prazo. As previsões a curto prazo tendem a ser utilizadas quando é necessário um maior rigor da previsão e quando uma atuação mais breve é fundamental, deste modo este género de previsões é frequentemente utilizado na área financeira e logística quando se tem por objetivo a determinação da procura de determinados processos. Por seu lado, as previsões a longo prazo são geralmente aplicadas na tomada de decisões estratégicas, onde se determina o rumo a seguir e os objetivos a alcançar, (Gonçalves, 2010).

### **2.2.1. Séries temporais**

Uma série temporal é definida como um conjunto de observações contínuas ordenadas temporalmente, podendo estas ser medidas em intervalos de tempo iguais ou aleatórios, apresentando uma dependência entre elas. A análise destas constitui uma área que recentemente tem vindo a ser objeto de grande estudo e desenvolvimento, com vários artigos e livros a serem publicados sobre o tema, (Wei, 2006).

São vários os motivos pelos quais se analisa uma série temporal, desde a compreensão dos vários parâmetros que a influenciam até à geração de previsões futuras. Ambos estes objetivos requerem que o padrão da referida série seja identificado e descrito para que com isto seja possível interpretar ou, até mesmo, para que seja possível integrar com outros dados adquiridos noutra local ou numa faixa temporal distinta, (Chatfield, 2016).

A análise das séries temporais é aplicável numa variedade de disciplinas como administração de empresas, economia, finanças públicas, engenharia, estatísticas e matemática, nas múltiplas áreas das ciências sociais e em qualquer área em que o desenvolvimento e análise de séries temporais possa ser relevante.

*“Time series occurs in a variety of fields. In agriculture we observe annual crop production and prices. In business and economics, we observe daily closing prices (...) In engineering we observe sound, electric signals and voltage.”*, (Wei, 2006).

Os modelos de previsão encontram-se divididos em dois grandes grupos: os métodos subjetivos ou qualitativos e os métodos quantitativos. Os primeiros podem recair sobre o método de *Delphi*, o método de júri de executivos, a análise de mercado ou o modelo de decisão de base. Por sua vez, os métodos quantitativos podem assumir duas formas sendo elas os modelos causais ou explicativos ou então modelos extrapolativos ou estatísticos, apresentando estes também as suas subdivisões, tal como se pode conferir na figura 6, (Gonçalves, 2010).

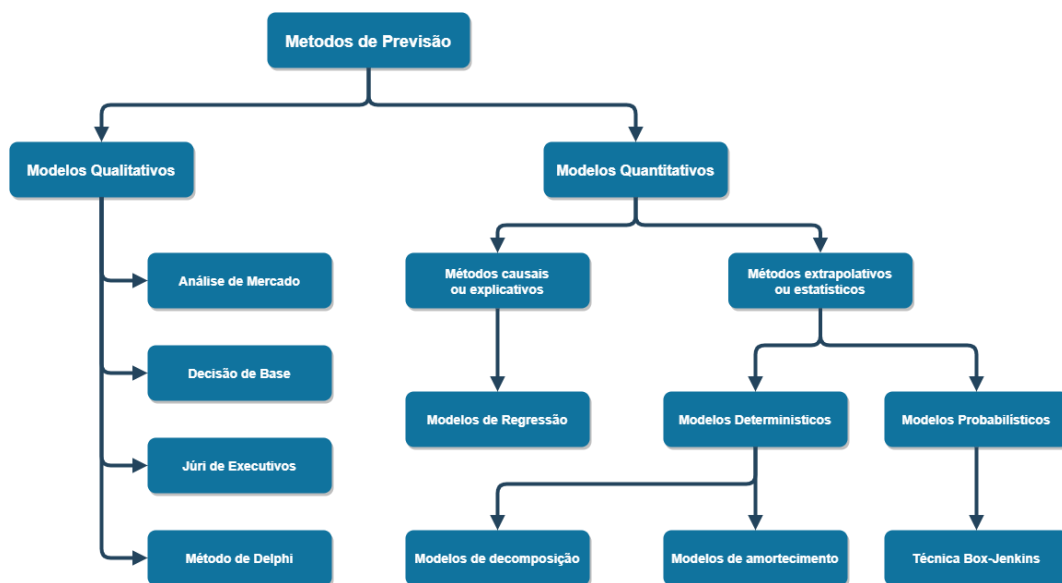


Figura 6 - Métodos de previsão (adaptado Gonçalves, 2010)

### 2.2.2. Modelos qualitativos

Os modelos qualitativos tomam por base o julgamento, intuição, opinião e experiência de um ou vários especialistas no assunto para chegar a uma decisão, tentando prever eventos futuros. Pelo modo como são obtidas as decisões, estes modelos são considerados modelos subjetivos, sendo que são utilizados quando não há dados históricos ou caso existam e não sejam descritivos do futuro. Estes podem assumir variadas metodologias com diversos procedimentos, sendo que entre os mais utilizados encontram-se o método de Delphi, o método de júri de executivos, o método de análise de mercado e o método de decisão de base ou força de vendas, (Corrêa & Corrêa, 2012).

No método Delphi, através de um questionário, é considerada a opinião de especialistas sobre o tema, tentando chegar a um consenso sobre eventos futuros ou que direção e estratégia seguir. Este método começa com cada um dos especialistas a preencher um questionário sobre o tema em análise. De seguida, o moderador recolhe e analisa as respostas, redigindo um novo questionário no qual as interações da ronda anterior foram tomadas em conta, incentivando assim a uma reflexão sobre as respostas anteriores, quer suas quer dos seus colegas. O mesmo procedimento é levado a cabo até que seja obtido um consenso ou assim que um critério de termino seja alcançado (número de rondas, por exemplo), (Dalkey, 2003).

No método júri de executivos, os executivos de diversos departamentos são reunidos com o intuito de analisar a sua opinião. Assim sendo é formado um júri composto por vários executivos das diversas áreas da empresa (planeamento, produção, finanças, marketing, compras, ...) onde cada um é encorajado a partilhar o seu parecer e previsão, fazendo com que a sua opinião pese no julgamento dos restantes ao mesmo tempo que analisa o juízo dos seus pares, (Chambers, Mullick, & Smith, 1971).

No método de análise de mercado é analisado o mercado no qual a empresa ou organização se integra com o objetivo de compreender as tendências deste. Este procedimento pode ser realizado através de um questionário orientado diretamente aos clientes finais, tentando obter os seus requisitos, condições, exigências e formalidades. Nesta metodologia deve-se ter o cuidado de garantir a aleatoriedade da amostra recolhida nomeadamente em termos de demografia e regiões analisadas. Este género de metodologias são benéficas quando se pretende divulgar novos produtos no mercado ou para analisar produtos com pouco ou nenhum histórico de vendas.

No método de decisão de base ou força de vendas é tomada em conta a opinião e apreciação dos que se encontram na base da hierarquia empresarial com o intuito de prever eventos futuros e definir o melhor rumo a traçar. Geralmente o departamento de vendas encontra-se nesta posição e, sendo os vendedores aqueles que têm um contacto direto e continuo com os clientes, estes possuem conhecimentos que possivelmente serão úteis e valiosos nos momentos em que for necessário estimar eventos posteriores.

Este grupo de metodologias, apesar de bastante útil, apresenta alguns problemas, entre eles o facto de por vezes ser impossível de chegar a um consenso caso as opiniões dos especialistas

sejam diversas ou caso eventualmente a opinião destes seja tendenciosa, fazendo com que estes métodos sejam bastante subjetivos e com um baixo grau de precisão. (Goodwin & Wright, 1994)

### **2.2.3. Modelos quantitativos**

Os modelos quantitativos tomam por base dados históricos como fundamento das previsões futuras. Estes tentam identificar padrões dentro das séries, transpondo-os para os restantes períodos temporais. Nos seguintes subcapítulos serão analisados os diversos subgrupos de modelos quantitativos, entre eles métodos causais ou explicativos e os métodos extrapolativos ou estatísticos. Será atribuído um especial destaque a este último conjunto de modelos pois este trabalho terá um método quantitativo de alisamento exponencial no seu caso de estudo.

#### **2.2.3.1. Métodos causais ou explicativos**

Os métodos causais ou explicativos procuram determinar possíveis causas e efeitos de um padrão na série temporal, tentando explicar as suas relações, com esse intuito são geralmente utilizados modelos de regressão. Estes conjuntos de modelos estabelecem uma relação entre uma variável dependente e uma ou mais variáveis independentes. A regressão pode ser considerada simples caso descreva uma relação entre duas variáveis quantitativas, uma independente e outra dependente, por outro lado, esta pode ser considerada regressão múltipla no caso da variável dependente ser função de duas ou mais variáveis independentes. Estes modelos podem também ser catalogados em relação à sua ordem, sendo classificados como regressão linear caso a equação da solução seja de primeira ordem ou regressão polinomial caso seja de segunda ordem ou superior, (Ballou, 2004).

A Regressão linear representa a forma mais simples de regressão e analisa a influência das variáveis independentes ( $a, b_1, b_2, \dots, b_n$ ) na variável em estudo ( $Y$ ), este processo pode ser obtido através da seguinte equação:

$$Y = a + b_1 \times X_1 + b_2 \times X_2 + \dots + b_n \times X_n \quad (1)$$

A regressão polinomial toma por base os conceitos da regressão linear, aplicando-os a casos nos quais se verificam não linearidades, deste modo é utilizada uma equação polinomial para melhor definir a relação entre a variável em estudo e os parâmetros que a influenciam.

$$Y = a + b_1 \times X_j + b_2 \times X_j^2 + \dots + b_n \times X_j^p \quad (2)$$

Este género de análise pode ter como objetivo a compreensão das relações entre os diversos parâmetros e o objeto em análise podendo, por outro lado, ter como objetivo a previsão de séries futuras procurando uma base para sustentar uma escolha de estratégia, (Bai & Perron, 1998).

### 2.2.3.2. Métodos extrapolativos ou estatísticos

Estes métodos encontram-se geralmente divididos em métodos determinísticos ou probabilísticos. Os primeiros são compostos principalmente por modelos de decomposição por média móvel ou modelos de amortecimento enquanto os segundos são representados pela técnica Box-Jenkins, (Box & Jenkins, 1976).

O modelo de decomposição de médias móveis é um método que tenta prever valores futuros, utilizando os valores mais recentes, reduzindo assim a importância de dados mais longínquos. Deste modo, este modelo calcula uma média utilizando os valores mais atuais da série temporal para realizar a previsão de períodos futuros, possibilitando também o alisamento de flutuações aleatórias. Neste grupo, os modelos podem ser classificados como médias móveis simples caso todos os valores utilizados no cálculo da média tenham o mesmo peso, como médias móveis centradas caso o primeiro e o último termos tenham metade do peso dos restantes ou podem ser classificadas como médias móveis ponderadas caso as parcelas tenham ponderações distintas, (Cowpertwait & Metcalfe, 2009).

Os modelos de amortecimento exponencial, por seu lado, utilizam ponderações de dados anteriores com o intuito de prever valores futuros, sendo que estes podem ser divididos em modelos de amortecimento simples, duplo e triplo. O primeiro representa o modelo mais simples de alisamento exponencial para uma série temporal, sendo descrito pela equação 3. O termo  $\hat{y}_{(t+1|t)}$  representa a previsão correspondente ao instante  $t+1$ , estando o coeficiente de amortecimento simbolizado por  $\alpha$ . Este pode variar entre 0 e 1, favorecendo valores mais antigos caso esteja mais próximo de 0 obtendo assim previsões mais estáveis ou, por outro lado, favorecendo valores mais recentes caso esteja mais próximo de 1 tornando-se mais sensível a oscilações recentes dos dados. Tendo este género de modelo apenas um único hiperparâmetro ( $\alpha$ ) torna-o indicado para a realização de previsões de séries que não apresentem tendência ou sazonalidade.

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha) \hat{y}_{t|t-1}, \quad \forall t \in T \quad (3)$$

No caso do modelo de amortecimento exponencial duplo, este é indicado para séries que possuam uma tendência com efeito aditivo, ou seja, caso a média não se apresente constante, mas sim com uma tendência linear crescente ou decrescente (Holt, 1957). De modo a contabilizar esta tendência este modelo utiliza duas constantes de amortecimento,  $\alpha$  e  $\beta$ , com valores entre 0 e 1, estando definido pelas equações abaixo apresentadas. Para além da equação de previsão (4), este modelo requer duas equações de suavização, uma para o nível (5) e outra para a tendência (6).

$$\hat{y}_{t+h|t} = \ell_t + h t_t, \quad \forall t \in T \quad (4)$$

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + t_{t-1}), \quad \forall t \in T \quad (5)$$

$$t_t = \beta(\ell_t + \ell_{t-1}) + (1 - \beta) t_{t-1}, \quad \forall t \in T \quad (6)$$

Podendo existir séries temporais que apresentem padrões de tendência linear bem como sazonalidade, para efetuar previsões dos mesmos deve ser utilizado o modelo de amortecimento exponencial triplo, definida pelas equações abaixo. A primeira corresponde à equação de previsão sendo as seguintes responsáveis pela suavização do nível, tendência e componente sazonal.

$$\hat{y}_{t+h|t} = \ell_t + h t_t + s_{t+h-m(k+1)}, \quad \forall t \in T \quad (7)$$

$$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + t_{t-1}), \quad \forall t \in T \quad (8)$$

$$t_t = \beta(\ell_t + \ell_{t-1}) + (1 - \beta) t_{t-1}, \quad \forall t \in T \quad (9)$$

$$s_t = \gamma(\ell_t + \ell_t) + (1 - \gamma) s_{t-m}, \quad \forall t \in T \quad (10)$$

Na qual  $\alpha$ ,  $\beta$  e  $\gamma$  representam os parâmetros de suavização, sendo  $m$  usado para indicar a frequência da sazonalidade (por exemplo o número de partes do ano,  $m=12$  para dados mensais e  $m = 4$  para dados trimestrais).

A metodologia Box-Jenkins aplica-se a séries estocásticas, ou seja, a séries na qual a média e a variância não se alteram com o tempo. Esta metodologia tem como objetivo a determinação dos parâmetros que constituem a estrutura da série, sendo eles os parâmetros autorregressivos ( $p$ ), os processos de diferenciação ( $d$ ) e os parâmetros de médias móveis ( $q$ ), formando assim uma estrutura ARIMA ( $p,d,q$ ). Deste modo, segundo Box e Jenkins, a série original será diferenciada com o intuito de tornar a série estacionária. De seguida, através da análise das funções de autocorrelação são identificados os valores de  $p$  e  $q$ . Estes valores permitem determinar o número de parâmetros autorregressivos e de médias móveis que deverão ser escolhidos de forma que a estrutura que se ajuste aos dados fornecidos seja a mais reduzida possível. Com isto, é possível realizar as previsões de valores posteriores, (Box & Jenkins, 1976).

### 2.3. Testes de adequação do modelo

Um dos problemas que residem com o uso de modelos de previsão consiste na determinação do método mais apropriado para o caso em análise. Uma vez que a escolha de um modelo incorreto para estudo de uma série temporal levará a previsões incorretas enquanto, por outro lado, a escolha correta do modelo de previsão trará previsões mais adequadas, levando, possivelmente, a uma vantagem competitiva da empresa ou organização sobre os seus concorrentes.

Para este efeito é útil utilizar-se a decomposição de séries temporais, pois para além de poder ser necessária na modelação da mesma, pode ainda ajudar na sua compreensão, aumentando assim a precisão das previsões. (Makridakis et al. 1998). A decomposição temporal pode ser descrita através da equação 3, onde  $CT_t$  representa a componente cíclica da tendência,  $S_t$  representa a componente sazonal e  $R_t$  a componente residual da série.

$$y_t = CT_t \times S_t \times R_t \quad (11)$$

Esta decomposição pode ser utilizada com o intuito de medir a força da tendência ( $F_t$ ) e a força da sazonalidade ( $F_s$ ) presentes numa série temporal. Deste modo, para se medir a primeira utiliza-se o quociente entre a variância da componente residual,  $Var(R_t)$ , e a variância da soma da componente residual com a componente cíclica da tendência,  $Var(CT_t + R_t)$ , se o objetivo for medir a força da sazonalidade, é utilizado o mesmo raciocínio mas utiliza-se a componente sazonal em vez da componente cíclica da tendência,  $Var(S_t + R_t)$ , tal como é definido pelas equações (4) e (5). Uma série que apresente um valor de  $F_t$  ou  $F_s$  igual a 0 será uma série sem tendência ou sazonalidade, por outro lado caso apresente um valor próximo de 1, é evidencia de uma forte tendência ou sazonalidade.

$$F_t = \max \left( 0, 1 - \frac{Var(R_t)}{Var(CT_t + R_t)} \right) \quad (12)$$

$$F_s = \max \left( 0, 1 - \frac{Var(R_t)}{Var(S_t + R_t)} \right) \quad (13)$$

Com estas informações são vários os testes disponíveis para verificar a adequação do modelo escolhido para a análise da série temporal. Box e Jenkins sugerem a utilização do teste de sobrefixação e do teste dos resíduos gerados, (Box & Jenkins, 1976).

No teste da sobrefixação elabora-se um modelo com um número de parâmetros superiores ao do caso em análise, verificando se estes parâmetros extras assumem significância no referido modelo. Caso estes novos parâmetros sejam significativos, isto significaria que o método estaria

## Revisão Bibliográfica

subidentificado. No teste dos resíduos gerados analisa-se se o ruído que foi gerado por este modelo corresponde a um ruído branco, ou seja, se o comportamento da série se encontra corretamente descrito, tal que este ruído não tenha interferência ou correlação com a série ou parâmetros em estudo, (Bai & Perron, 1998)

## 3. DESENVOLVIMENTO DE SÉRIES TEMPORAIS

Neste capítulo é exposto o processo de construção de uma série temporal, começando pela sua criação e os diversos métodos de recolha de dados necessários para a sua realização, passando depois por apresentar as dificuldades que poderão surgir nesses momentos. Este capítulo finaliza com a apresentação de algumas possíveis soluções e métodos para ultrapassar as referidas adversidades.

### 3.1. Criação de séries temporais

O processo de criação de uma série temporal deverá começar com a recolha de dados sobre os parâmetros ou ações que queremos analisar. A recolha de dados é um processo de aquisição de informações com o intuito de serem estudadas de forma que posteriormente possam servir como base em tomadas de decisões, podendo estas ser aplicadas em artigos de pesquisa, planeamentos estratégicos, decisões de negócios, entre outros. Esta recolha de informações pode ser adquirida de diversas maneiras, dependendo principalmente da natureza dos dados a serem analisados, podendo estes ser agrupados em duas grandes classes: os dados qualitativos e os quantitativos, (Guo, Williams, & Smith, 2007).

Os dados qualitativos são dados não numéricos, que geralmente medem atributos, qualidades ou características de um certo produto ou serviço prestado. Estes podem ser obtidos através de grupos focais (“focus group”), entrevistas, questionários (quer sejam pessoalmente, por telefone, email ou outro meio alternativo) ou por observação direta dos participantes num estudo. A vantagem destas metodologias de obtenção de informação é que proporciona um contacto mais direto com o consumidor final, dando voz à experiência vivida por estes, apesar de lhes ser inerente um elevado nível de subjetividade e, por consequente, um baixo grau de precisão.

Como se pode compreender, esta análise é de grande importância no setor alimentar, uma vez que o sucesso de um produto e, por consequente de uma empresa, depende em grande parte da opinião do público e do consumidor final.

Os dados quantitativos, por outro lado, medem valores numéricos, ou seja, medidas concretas que expressam uma quantidade ou intervalo. Deste modo é possível que lhe sejam aplicados modelos mais complexos e por consequente gerar previsão com um maior grau de exatidão. Outra grande vantagem deste género de dados é o nível de automatização que pode ser aplicado na sua obtenção.

No caso do setor alimentar, durante todo o processo industrial existem diversos sensores que medem, pesam e controlam continuamente os seus produtos e caso surja alguma anormalidade, esta será imediatamente detetada.

Relativamente à gestão da própria empresa, os sistemas de informação tecnológica (IT systems) conseguem obter informações, tais como, sobre as transações que são efetuadas, levando assim a determinação da frequência que um certo artigo é comprado ou à localização geográfica onde o referido produto será mais procurado.

### **3.2. Dificuldades na criação de uma série temporal**

Quando se realiza um projeto desta natureza, o primeiro passo deve ser a determinação do objetivo final, ou seja, o propósito do estudo da referida série temporal, pois é a partir deste que serão tomadas as decisões para o atingir. Na maioria dos casos o objetivo final assenta nestes dois pontos: na sua análise, de forma a compreender com mais exatidão os parâmetros que a influenciam e a importância que cada um deles pode tomar ou poderá assentar na previsão de valores futuros, com o intuito de estimar e antever o desenvolvimento da série temporal.

A determinação incorreta destes pontos levará a alguns problemas futuros, gerando dados irrelevantes ou em excesso que não serão necessários para a sua finalidade. Esta recolha excessiva trará custos acrescidos para a empresa, uma vez que serão necessários mais instrumentos, dispositivos e mecanismos de medição acrescentando-se ainda a possibilidade de estes aumentarem o tempo de confeção do produto. Isto também tornará mais trabalhosa a sua análise, quer na preparação dos dados quer na criação do modelo a aplicar, levando a um maior tempo computacional devido a uma maior entrada de dados, (Government, 2022).

De mesmo modo, caso a série a ser analisada possua uma componente sazonal, que afetará os dados da mesma, esta pode ser determinada, com maior ou menor precisão através dos dados históricos que já tenham sido obtidos. Como se pode compreender, aqui reside uma das grandes limitações e dificuldades da análise de séries temporais, uma vez que caso não exista um conjunto significativo de valores à priori, a referida componente sazonal dificilmente será obtida.

Noutros casos, mesmo aquando da existência de dados históricos, poderão aparecer alguns problemas na análise das séries temporais devido à fraca qualidade dos valores recolhidos. Estes podem ser provenientes de diversos fatores tais como: valores em falta, fazendo com que

a série não seja contínua ou que possua grandes lacunas; formatação incorreta dos dados obtidos, levando a que a sua análise seja mais demorada e complicada; valores inconsistentes, possivelmente devido a uma calibração incorreta ou má utilização dos instrumentos de medição, (Rasdorf, Hummer, & Vereen, 2015).

Outra das grandes dificuldades na criação e na consequente análise de séries temporais advém da complexidade e imprevisibilidade dos dados obtidos, pois como se pode compreender, com um maior volume de dados, a sua análise será mais complexa e tornar-se-á mais difícil de identificar os dados que realmente serão importantes para o seu objetivo final. Um bom exemplo desta situação é o caso do mercado financeiro onde grandes volumes de informações são gerados num curto espaço de tempo, surgindo frequentemente oscilações aleatórias que dificultam a sua análise, diminuindo consideravelmente a precisão dos modelos utilizados.

Mesmo estando ultrapassadas estas adversidades, a previsão final poderá ainda estar incorreta caso o modelo escolhido não seja o indicado. Devido à sua formulação alguns modelos não contabilizam com parâmetros como a tendência ou a sazonalidade e caso a série a ser analisada possua tais características a sua previsão terá uma exatidão reduzida, (Svoboda, Kotik, & Platos, 2021).

### **3.3. Como ultrapassar problemas identificados**

Como é de fácil compreensão, os pontos apresentados no subcapítulo anterior não são bem-vindos aquando da criação e análise de séries temporais, deste modo as empresas e organizações empregam diversas metodologias para reduzir o impacto destes problemas ou até mesmo eliminando-os completamente.

Assim sendo, primeiramente, o objetivo final da análise de uma série temporal deve ser claro e transmitido a todas as partes necessárias para a sua realização. De seguida é recomendável que se determine onde as informações e dados necessários serão recolhidos, focando-se apenas nestas e evitando a recolha de valores desnecessários à sua finalidade.

No caso de haver uma escassa coleção de dados históricos ou mesmo de não existirem de todo, uma das estratégias que uma empresa pode tomar será a de consultar e analisar o mercado na qual se encontra, com o intuito de compreender as tendências que este possui e dessa forma prever as suas vendas de uma forma mais eficaz. Outra estratégia poderá passar pela criação de alguns cenários: um mais pessimista, outro mais otimista e um cenário realista que servirá de base para as previsões e que será ajustando no decorrer do tempo.

A qualidade dos dados recolhidos é de grande importância no mundo atual, pois para muitas empresas estes funcionam como ponto de partida na tomada de certas decisões. Assim as mesmas devem confirmar que as informações que recolhem são fidedignas, especialmente se tais valores forem recolhidos de terceiros. Caso seja a própria organização a obter essa informação, esta deve ter o cuidado de utilizar os aparelhos de medida corretamente,

## Desenvolvimento de Séries Temporais

mantendo-os calibrados, no caso de se tratar de uma linha de montagem, ou no caso de estes serem obtidos de forma digital tal como o registo de vendas, essa informação deve ser tornada acessíveis às pessoas necessárias e apresentada na formatação pretendida.

Relativamente à aplicação de um modelo a uma série temporal, podemos verificar se a sua utilização é correta através de testes de adequação do modelo. De acordo com Box e Jenkins, os testes que devem ser realizados serão o teste de sobrefixação e do teste dos resíduos gerados, os quais já foram expostos anteriormente no subcapítulo 2.3. De modo a determinar qual o modelo com maior precisão, de entre os que se mostram adequados, é frequente utilizar-se o erro médio de previsão ou o desvio médio absoluto, (Bai & Perron, 1998).

## 4. MACHINE LEARNING

Neste capítulo é abordado o tópico da inteligência artificial com vista a uma melhor compreensão do tema. Começando por expor os conceitos de *Machine Learning* e *Deep Learning*, explicando em que consistem e o seu funcionamento, bem como as suas subdivisões. No último subcapítulo é apresentada a sua aplicabilidade no estudo de séries temporais.

### 4.1. Machine learning

*Machine Learning* é uma subdivisão da inteligência artificial, à qual são aplicados algoritmos capazes de analisar e estudar os dados que lhes são fornecidos, tomando posteriormente decisões baseadas no que aprenderam. Os modelos de *machine learning* utilizam na sua base codificação e matemática complexa, conseguindo com isto detetar e identificar padrões nos dados e informações que lhes são concedidas. Isto faz com que seja possível a realização de diversas tarefas automatizadas nas mais variadas áreas, desde a área financeira que procura indicadores favoráveis para emitir alertas de compra e venda, a uma empresa de ciber segurança que tenta encontrar vírus e *malwares* nos softwares.

Os modelos de *machine learning* podem ser classificados através da sua construção, sendo esta em função do género de dados que lhe será fornecido. Deste modo, as respetivas divisões podem ser apresentadas como: *Supervised learning*, *Unsupervised learning* e *Reinforcement learning*, (Russell, Norvig, & Davis, 2010).

O primeiro subconjunto de *machine learning* é o que requer mais supervisão. O computador recebe dados de treino e um modelo para ajustar os dados. À medida que novas informações lhe são introduzidas, o operador supervisiona e verifica o processo, confirmando as respostas exatas do computador e corrigindo as respostas incorretas do mesmo. *Supervised learning* envolve dar ao modelo todas as "respostas corretas" devidamente catalogadas com o objetivo de treinar e ensinar o modelo a reconhecer os restantes dados que lhe serão fornecidos.

No grupo de *Unsupervised learning*, os dados fornecidos não são rotulados, deste modo o modelo tenta encontrar pontos comuns entre os valores de entrada, extrapolando de forma a conseguir reconhecer padrões e relações entre estes. Uma vez que não é necessário o trabalho

prévio de catalogar os dados, esta categoria consegue ser de mais fácil abordagem e utilização, dando maior possibilidade à sua automatização. Este subconjunto de algoritmos é particularmente valioso pois da generalidade dos dados os não catalogados são os mais habituais, (Marcus & Davis, 2019).

O último género de algoritmos funciona de modo ligeiramente diferente dos apresentados anteriormente, guiando-se por uma metodologia de tentativa e erro. Modelos esses que são recompensados sempre que atingem um certo objetivo ou cumprem determinada tarefa, tendo assim o algoritmo como foco a maximização da sua recompensa. Com isto, o algoritmo procurará determinar o melhor caminho possível que deve tomar numa certa situação. Este subconjunto denominado de *Reinforcement learning* é bastante útil quando não estão disponíveis quaisquer dados de treino que possam servir de ferramenta de aprendizagem para os mesmos. Assim, estes algoritmos vão simplesmente explorando, criando diversas tentativas, aprendendo através dos seus erros e das recompensas que lhes são atribuídas, (Sutton & Barto, 2018).

#### **4.1.1. Deep learning**

*Deep Learning* é um ramo do *machine learning*, que estrutura os seus algoritmos em camadas de forma a criar uma rede neural artificial capaz de aprender por si própria e de tomar decisões inteligentes de forma autónoma. Os modelos de *deep learning* são desenhados e concebidos de forma a analisarem continuamente toda a informação que lhes é fornecida, para isso utilizam estruturas lógicas com várias camadas de neurónios interligados, possibilitando que os vários sistemas trabalhem em conjunto, semelhante à forma como um cérebro humano usaria para obter conclusões.

Quando estes modelos recebem dados e informação estas circulam por todas as diversas camadas de sistema, fazendo com que o mesmo aprenda progressivamente. Desta forma, estes modelos são capazes de aprender através do seu próprio código de computação, (Wani et al., 2019).

Os modelos de *Deep learning* têm normalmente um melhor desempenho do que os modelos de *machine learning* em problemas complexos ou conjuntos de dados de grandes dimensões. Por outro lado, os primeiros geralmente necessitam de um grande volume de dados de treino, levando por isso uma maior porção de tempo até atingirem o seu estado final.

Apesar de terem um modo de funcionamento semelhante, estes diferem em alguns aspetos, sendo o principal o nível de intervenção humana que necessita. No caso do *machine learning*, apesar de conseguir captar e assimilar informação de forma autónoma, se este devolver uma resposta incorreta, não será capaz de a reconhecer como tal, necessitando de intervenção humana de forma a ser ajustada. No caso do *deep learning* estes algoritmos são capazes de determinar se a resposta obtida será válida e aceite através da sua própria rede neural, não requerendo qualquer intervenção humana, (Bengiolan, Goodfellow, & Courville, 2015).

## 4.2. Aplicabilidade no estudo de séries temporais

A previsão de séries temporais é uma área importante da inteligência artificial, sendo estes métodos capazes de analisar uma maior quantidade de dados em relação às metodologias clássicas, possibilitando assim uma previsão com um maior grau de exatidão. As redes neurais que constituem os algoritmos de inteligência artificial são, por norma, mais robustas em relação ao ruído de entrada, podendo até serem capazes de assumir valores em falta.

A previsão de séries temporais utilizando estas metodologias possui uma elevada eficácia na identificação dos padrões presentes nas sequências de dados que lhes são fornecidos, conseguindo até detetar comportamentos que poderão ser demasiado complexos por observação humana. Simultaneamente estes modelos extraem e identificam de forma automática as características que possam ser úteis e relevantes para o problema que está a ser analisado.

As relações que são apreendidas entre as diferentes variáveis podem alterar-se durante o tempo, tornando assim a análise mais dinâmica e fazendo com que esta se consiga adaptar mais rapidamente às mudanças que irão aparecendo. Isto faz com que este género de modelos seja de grande utilidade nos dias de hoje e neste mundo em rápida transição, adicionando valor à empresa e possibilitando o aumento das receitas e consequentemente o desempenho da mesma, (Firat, 2008; Sohrabpour et al., 2021; Hu et al., 2020).



## 5. METODOLOGIA

Neste capítulo serão apresentados os dados necessários ao caso de estudo e que servirão de base para a execução dos modelos de previsão, sendo posteriormente exposta a implementação computacional dos mesmos. No final do capítulo salienta-se algumas das métricas de erro mais comuns, evidenciando a escolhida.

### 5.1. Apresentação de dados

Os dados utilizados para a realização do presente relatório foram retirados a partir do Instituto Nacional de Estatística (INE), dizendo respeito aos índices de volumes de negócios no comércio a retalho, procurando assim usar dados disponíveis a toda a comunidade científica que possibilitem posteriores estudos comparativos. Estes dados foram agrupados num ficheiro XLSX e, de forma a formarem um volume de dados significativo, foram selecionados os dados referentes aos índices mensais desde janeiro de 2015 até outubro de 2022. A grande vantagem destes dados terem sido retirados do INE, está no facto de estes serem previamente tratados, levando a que não haja erros de formatação, valores em falta ou mesmo duplicados, entre outros.

### 5.2. Implementação computacional

A implementação computacional necessária para a realização do caso de estudo presente neste relatório teve por base o estudo realizado por Ramos em 2021, tendo sido os algoritmos previamente criados para este mesmo estudo. Foi utilizado o ambiente computacional *Jupyter Notebook* com a linguagem de programação *Python*, pois esta possui uma vasta informação disponível (facilitando a aprendizagem e compreensão), uma simplicidade de utilização e uma grande comunidade de utilizadores que facilita o acesso a soluções para os problemas encontrados e a bibliotecas abertas, (Ramos, 2021).

De modo a desenvolver o estudo empírico, o primeiro passo passou pela análise e caracterização da série temporal, usando para isso o recurso a representações gráficas, estatísticas descritivas e testes de hipótese. As representações gráficas prendem-se com o gráfico cronológico dos dados, bem como a utilização de gráficos box-plots anuais e mensais. À série em análise serão aplicados os testes normalidade Kurtosis, Skewness e Jarque-Bera, bem como os testes de estacionariedade ADF e KPSS, sendo a independência medida através do teste BDS.

De seguida, de forma a serem implementados os modelos ETS, foi utilizada a biblioteca statsmodels por meio das classes SimpleExpSmoothing, Holt e ExponentialSmoothing. Como descrito por Ramos (2021), serão assim aplicados os vários modelos de amortecimento exponencial com as diversas componentes de tendência e sazonalidade, podendo estes serem divididos em três grupos: os de amortecimento exponencial simples que contempla o modelo sem tendência e sazonalidade; os de amortecimento exponencial duplo englobando os modelos sem sazonalidade e de tendência aditiva ou aditiva amortecida; os de amortecimento exponencial triplo compostos pelos restantes modelos. Tendo isto por base foi aplicada a metodologia exposta na figura 7.

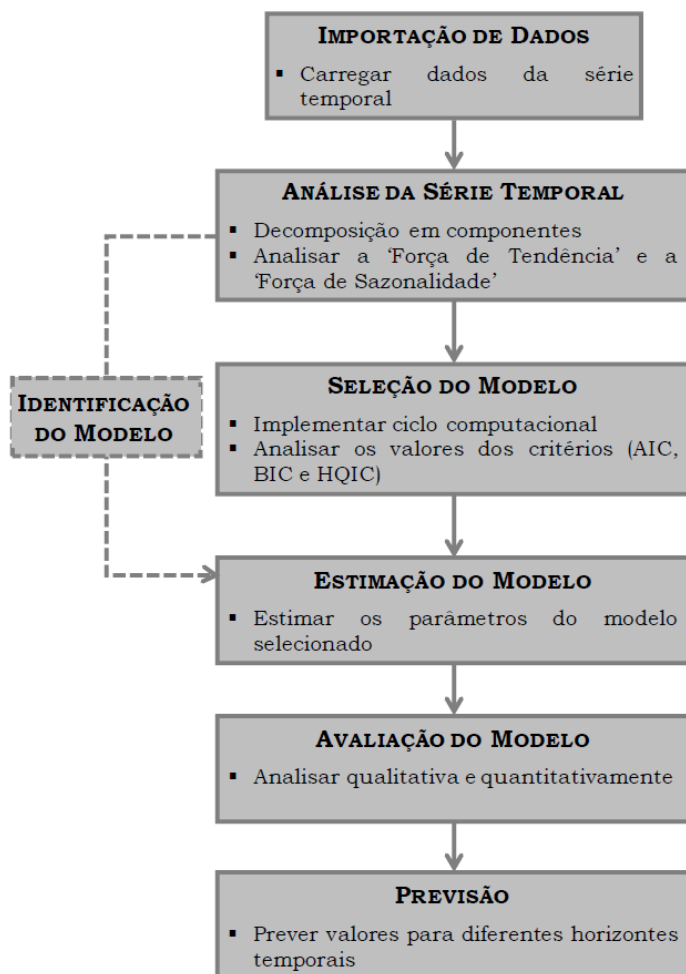


Figura 7 - Metodologia de implementação computacional dos modelos ETS

Fonte: Ramos (2021)

De modo a escolher o melhor modelo entre os implementados, Hyndman e Athanasopoulos (2018) sugerem, para os modelos onde se verifica convergência, que uma forma robusta e objetiva de obter o valor dos hiperparâmetros numéricos será estimá-los a partir dos dados históricos, Optou-se assim por analisar todos os modelos em paralelo com a informação obtida pelos critérios AIC e BIC, sendo o melhor modelo aquele que minimiza o erro.

Na parte que diz respeito aos modelos de redes neurais, foi utilizado o modelo desenvolvido em Ramos (2021), sendo utilizada a biblioteca *tensorflow* por meio das classes *Dense*, *SimpleRNN* e *LSTM*, seguindo a metodologia apresentada na figura 8.

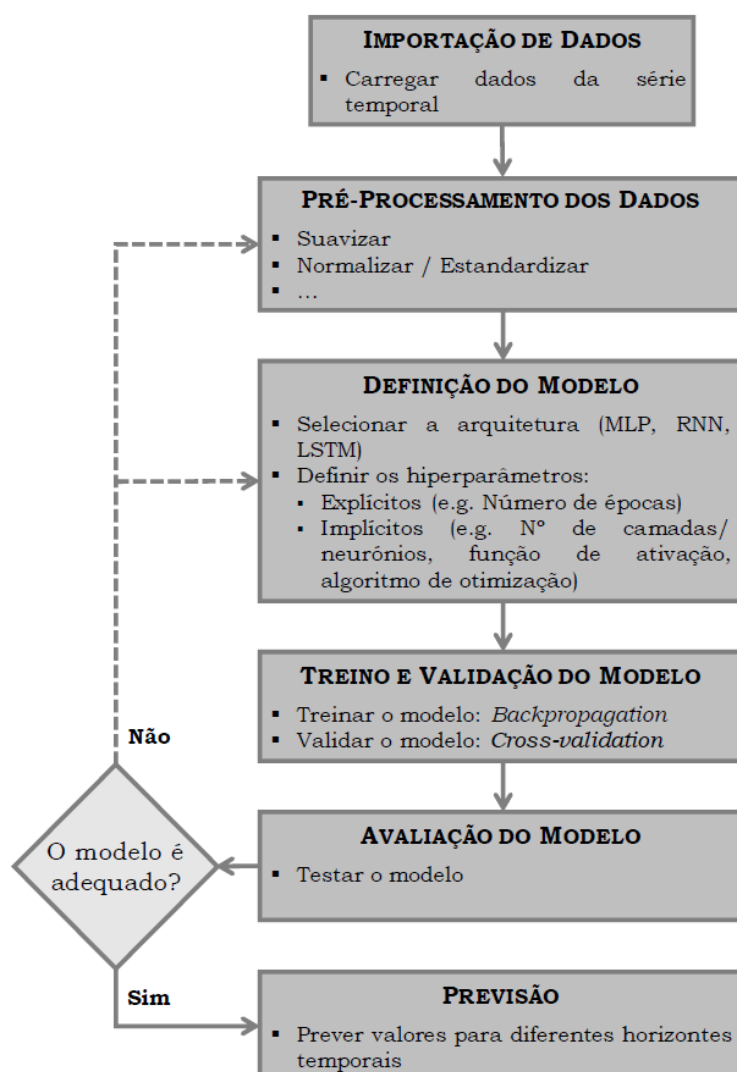


Figura 8 - Metodologia de implementação computacional dos modelos DNN

Fonte: Ramos (2021)

### 5.3. Previsão e avaliação do erro

Projetar a série para valores além do período amostral é o principal objetivo do ajustamento da série a este género de modelos, trazendo consigo o conceito de erro de previsão, caracterizado pela diferença entre os valores previstos e os valores reais. Segundo Hamilton (1994), existem dois tipos de erros: os primeiros, aleatórios, acontecem devido ao desconhecimento de variações futuras cujos fatores não são contemplados no modelo, sendo os segundos denominados de erros sistemáticos que são cometidos de forma consistente devido à seleção de relações matemáticas incorretas entre as variáveis ou a diferenças entre os parâmetros verdadeiros e as suas estimativas. Ambas as formas contribuem para o erro de previsão, sendo o melhor modelo escolhido pelo que minimiza o erro sistemático. Deste modo, para um qualquer valor  $y_{t+h}$ , no período  $t+h$ , e sendo a sua previsão representada por  $\hat{y}_{t+h}$  (obtida com base nos dados disponíveis até ao período  $t$ ) então o seu erro de previsão pode ser definido pela diferença entre estes valores.

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

Este erro torna-se útil para se poder avaliar o desempenho de um determinado modelo de previsão, podendo utilizar para isto algumas métricas de erro comumente referidas na bibliografia, de onde são destacadas as seguintes:

*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)*

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

De entre os trabalhos científicos é possível identificar diversas ocasiões em que são utilizados quaisquer uma das métricas, podendo a opção passar qualquer uma delas. Contudo, as métricas MAE e MAPE tem vindo a ser as mais utilizadas. Comparando as mesmas, verifica-se que MAE é uma grandeza dimensional expressa na unidade de medida dos dados, enquanto MAPE é uma grandeza adimensional, representando a dimensão do erro de forma percentual. Esta particularidade revela-se vantajosa pois permite uma simplicidade de interpretação dos resultados, bem como a fácil comparação do erro com outros conjuntos de dados de diferentes unidades de medida. Tendo sido esta a métrica utilizada na avaliação do erro

## 6. CASO DE ESTUDO

Neste capítulo será exposta uma análise ao setor alimentar em Portugal com o intuito de compreender os acontecimentos passados bem como tentar prever o destino desta área de negócios. Com vista a este objetivo, foram analisados os Índices de Volume de Negócios no Comércio a Retalho, sendo sobre estes utilizados modelos de previsão de séries temporais clássicos, bem como modelos de inteligência artificial, tentando determinar quais levam mais vantagem e quais obtêm um melhor desempenho.

### 6.1. Análise exploratória

Primeiramente foi necessário efetuar a recolha de valores sobre os quais seja possível debruçar a análise pretendida, deste modo, e sendo os dados referentes ao nosso próprio país, recorreu-se ao Instituto Nacional de Estatística (INE). Após a exploração de qual o conjunto de valores mais significativos para a análise em curso, foram retirados os dados expostos na tabela A.1 do anexo A, os quais apresentam os índices de volumes de negócios no comércio a retalho separados pelas respetivas categorias. Estes dados foram posteriormente filtrados de forma a apenas serem apresentados os índices referentes aos produtos alimentares, criando assim a tabela 1

Tabela 1 - Índices de volume de negócios no comércio a retalho (Produtos alimentar)

Data	Índice de Volume	Data	Índice de Volume
<b>jan/15</b>	90,2	out/15	100,8
fev/15	84,8	nov/15	94,8
mar/15	94,9	dez/15	119,1
abr/15	95,6	<b>jan/16</b>	92,0
mai/15	100,4	fev/16	87,9
jun/15	97,9	mar/16	98,7
jul/15	108,0	abr/16	97,2
ago/15	114,1	mai/16	99,8
set/15	99,6	jun/16	102,9

## Caso de Estudo

Data	Índice de Volume	Data	Índice de Volume
jul/16	115,9	set/19	116,4
ago/16	118,7	out/19	115,6
set/16	105,1	nov/19	115,0
out/16	104,6	dez/19	138,9
nov/16	99,5	<b>jan/20</b>	108,5
dez/16	124,0	fev/20	108,2
<b>jan/17</b>	94,3	mar/20	121,7
fev/17	90,2	abr/20	106,8
mar/17	101,5	mai/20	117,2
abr/17	106,6	jun/20	112,4
mai/17	104,0	jul/20	125,1
jun/17	106,8	ago/20	130,2
jul/17	117,3	set/20	117,8
ago/17	120,7	out/20	121,3
set/17	109,4	nov/20	115,4
out/17	107,2	dez/20	142,6
nov/17	106,8	<b>jan/21</b>	111,2
dez/17	132,0	fev/21	106,3
<b>jan/18</b>	98,3	mar/21	123,3
fev/18	93,8	abr/21	118,2
mar/18	113,4	mai/21	120,9
abr/18	102,5	jun/21	120,3
mai/18	111,7	jul/21	132,6
jun/18	112,0	ago/21	137,3
jul/18	121,1	set/21	123,8
ago/18	129,9	out/21	128,9
set/18	115,0	nov/21	126,6
out/18	113,5	dez/21	156,5
nov/18	111,4	<b>jan/22</b>	116,9
dez/18	137,7	fev/22	112,2
<b>jan/19</b>	102,9	mar/22	130,2
fev/19	96,8	abr/22	133,3
mar/19	112,6	mai/22	135,1
abr/19	112,3	jun/22	134,2
mai/19	115,7	jul/22	151,5
jun/19	116,1	*ago-22	155,1
jul/19	124,6	*set-22	139,5
ago/19	132,7	*out/22	145,1

Sendo este conjunto de dados relativamente extensos, a sua análise imediata torna-se algo complicada, assim sendo e de forma a melhor compreender a sua evolução construiu-se o gráfico exposto na figura 9.

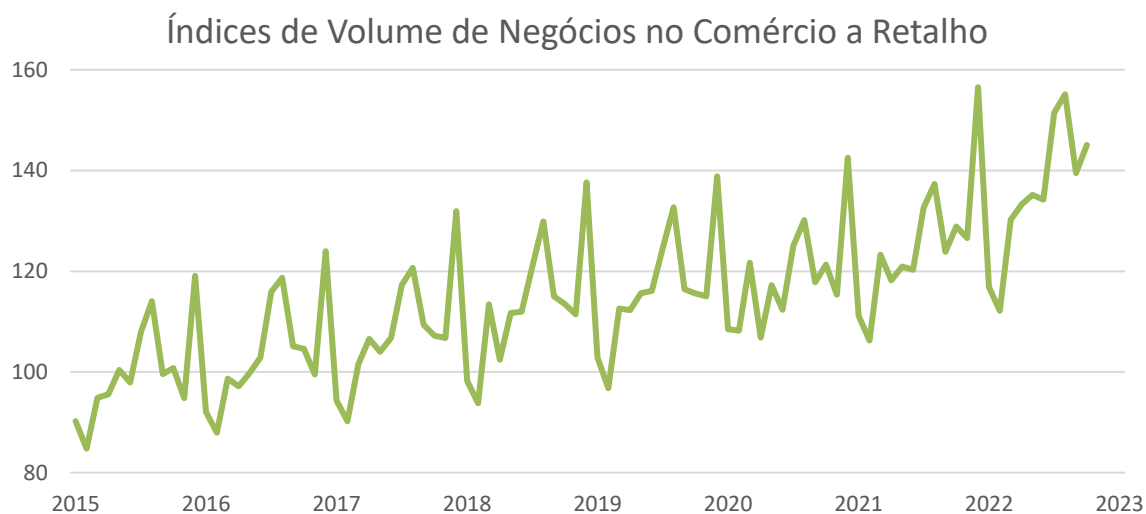


Figura 9 - Índices de volume de negócios no comércio a retalho (elaboração própria)

Através da apresentação do gráfico *box-plots* anual, ver figura 10, é possível verificar que as linhas do primeiro e terceiro quartil se encontram relativamente próximas, o que significa que a maioria dos valores de cada ano estão próximos entre si. Para além disso é possível observar que a série temporal possui uma tendência crescente ao longo dos anos.

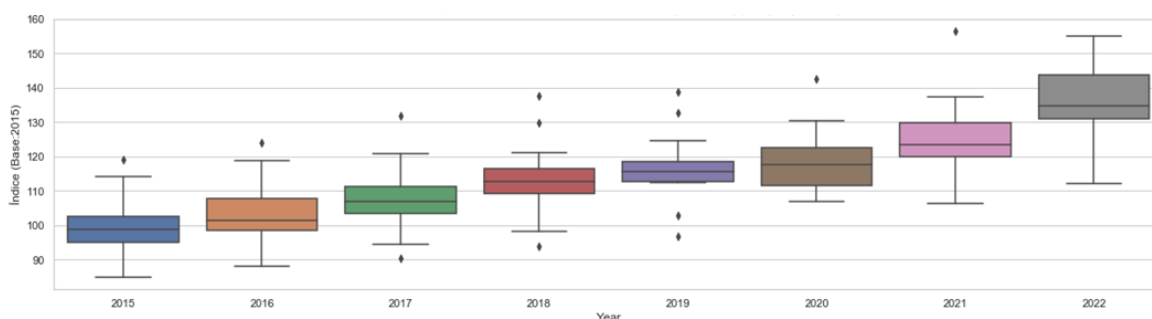


Figura 10 - Gráfico box-plot anual

Com a o gráfico *box-plots* mensais, ver figura 11, torna-se perceptível que esta série de dados possui uma clara sazonalidade, podendo ser observado no seguinte gráfico de caixa a sua variação ao longo dos vários meses do ciclo anual. Neste pode-se verificar que o mês de maior índice corresponde a dezembro e os menores dizem respeito aos meses de janeiro e fevereiro, isto poderá dever-se à época natalícia que por norma há um maior consumo, sendo os meses seguintes de maior retração. Também é possível observar o aumento dos meses de julho e agosto, devido à época de férias.

## Caso de Estudo

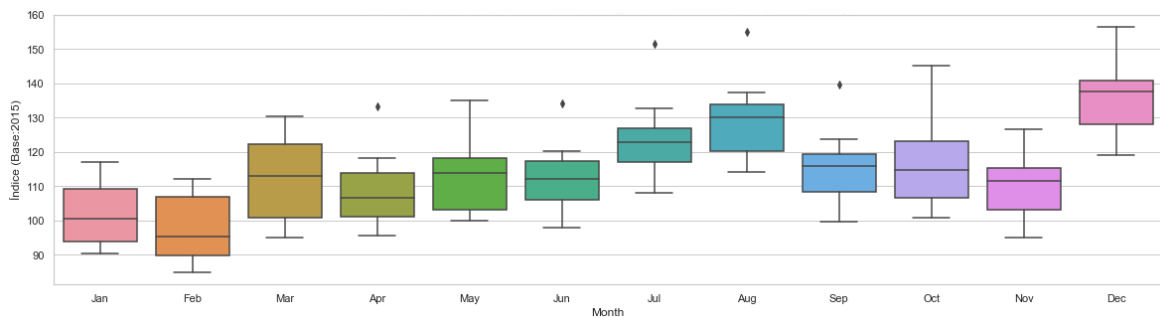


Figura 11 - Grafico box-plot mensal

De seguida apresenta-se na tabela 2 as principais estatísticas descritivas da série temporal, onde se verifica que o valor mais baixo corresponde ao mês de fevereiro de 2015 e que o valor mais alto diz respeito a dezembro de 2021. É também possível de verificar através do valor positivo da assimetria *Skewness* que a série possui uma distribuição assimétrica à direita, o que também pode ser verificado pelo facto da média ser superior à mediana.

Tabela 2 - 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>
Índice (Base: 2015)	94	114.77	15.36	84.80	103.18	113.80	123.65	156.50	-0.042	0.492

Sobre estes dados foram realizados os testes estatísticos de normalidade, estacionariedade e independência, estando os resultados obtidos expostos na tabela 3. Tendo em conta os resultados podemos verificar que para os níveis de significância de 5 e 10% a hipótese nula do teste de *Skewness* é rejeitada, significando que “a assimetria da população, de onde a amostra foi retirada, não é a mesma de uma distribuição normal”. No caso dos testes de normalidade de *Kurtosis* e *Jarque-Bera*, a hipótese  $H_0$  não é rejeitada, sugerindo que os dados da série se ajustam a uma distribuição normal, ver figura B.1 e B.2 – Anexo B.

Relativamente à estacionariedade, uma vez que não há evidência para rejeitar a existência de uma raiz unitária (teste *ADF*) e estando o valor do teste *KPSS* superior aos valores críticos, conclui-se a não estacionariedade da série. Quanto ao teste de independência, verifica-se que a hipótese nula é rejeitada para qualquer nível de significância, concluindo que existirá algum tipo de dependência nos dados. Outras informações complementares como o ciclo de tendência, sazonalidade e componente aleatória podem ser observados na figura B.3 do anexo B.

Tabela 3 - Testes de hipótese

		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)
Frete/ Volume	<i>statistic</i>	0.2372	1.9912	3.8032	2.3973	0.8403	7.7946 – 8.1024
	<i>p-value</i>	0.8125	0.0465**	0.1493	0.9990	-----	0.0000*

\*Rejeita-se  $H_0$  para os níveis de significância de 1%, 5% e 10%

\*\*Rejeita-se  $H_0$  para os níveis de significância de 5% e 10%

## 6.2. Aplicação dos modelos

De modo a escolher a modelação mais acertada foi realizado o cálculo da força da tendência (Ft) e força da sazonalidade (Fs). Conforme exposto no subcapítulo 2.3, os resultados obtidos na tabela 4 encontram-se próximos de 1, comprovando assim a forte tendência e sazonalidade da série em análise.

Tabela 4 - Valores de força de tendência e sazonalidade

	Força Tendência	Força Sazonalidade
Modelo Aditivo	0.9324	0.9430
Modelo Multiplicativo	1.0000	0.9435

Deste modo e observando os modelos clássicos aplicáveis a séries temporais com estas características, foram escolhidos os modelos de amortecimento exponencial para a sua análise. Na nomenclatura aplicada a tendência e a sazonalidade da série temporal são descritas através de duas letras separadas por vírgula contidas entre parenteses. A primeira letra diz respeito á tendência sendo a segunda utilizada para descrever a sazonalidade, assim sendo uma série temporal que não possua tendência e tenha uma sazonalidade multiplicativa possuirá a nomenclatura (N,M), ver Tabela 5.

Caso de Estudo

Tabela 5 - Nomenclatura dos modelos aplicados

Tendência	Sazonalidade		
	Nenhum	Aditiva	Multiplicativa
Nenhuma	(N,N)	(N,A)	(N,M)
Aditiva	(A,N)	(A,A)	(A,M)
Aditiva Amortizada	(Ad,N)	(Ad,A)	(A,M)

Após a aplicação destes modelos, no gráfico apresentado na figura 12 estão expostos os que obtiveram convergência, estando no gráfico da figura 13, ver figura 13, a representação dos seus valores, onde se verifica que vários modelos obterem uma boa aproximação. Para o estudo em questão foi escolhido o modelo de tendência aditiva e de sazonalidade multiplicativa (A,M), uma vez que este apresentou o menor valor de resíduos, ver figura 14.

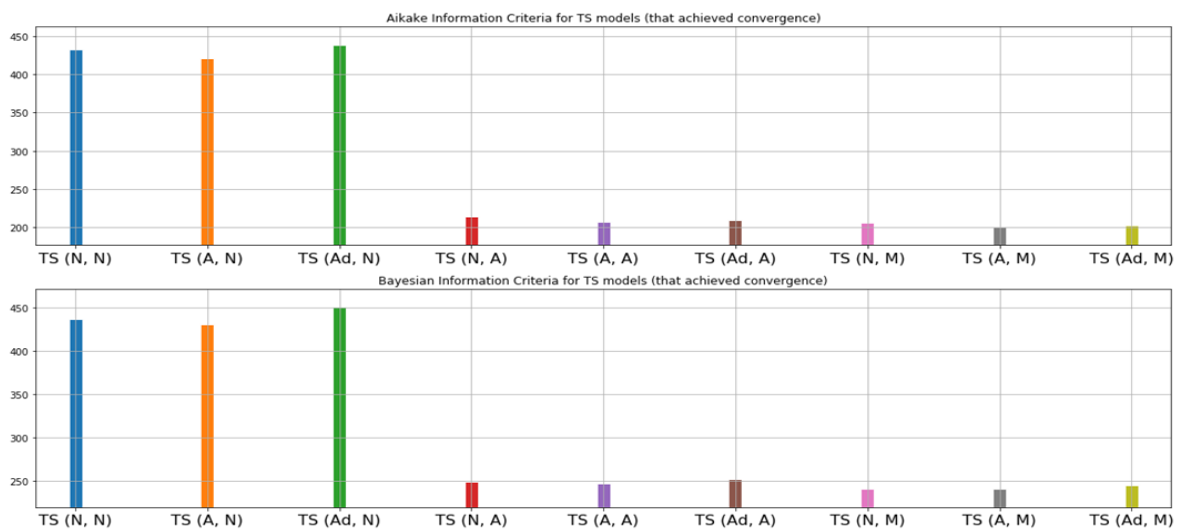


Figura 12 - Análise dos diversos modelos

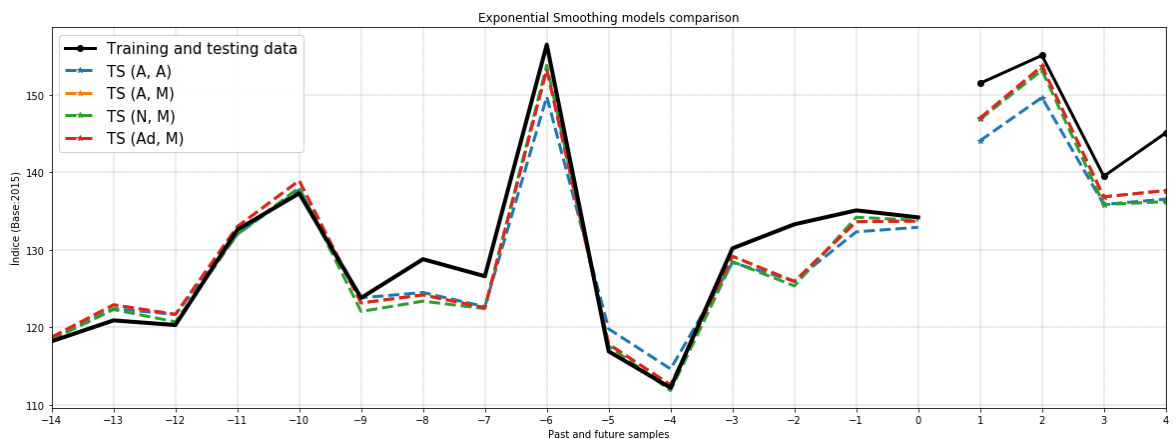


Figura 13 - Comparação dos modelos

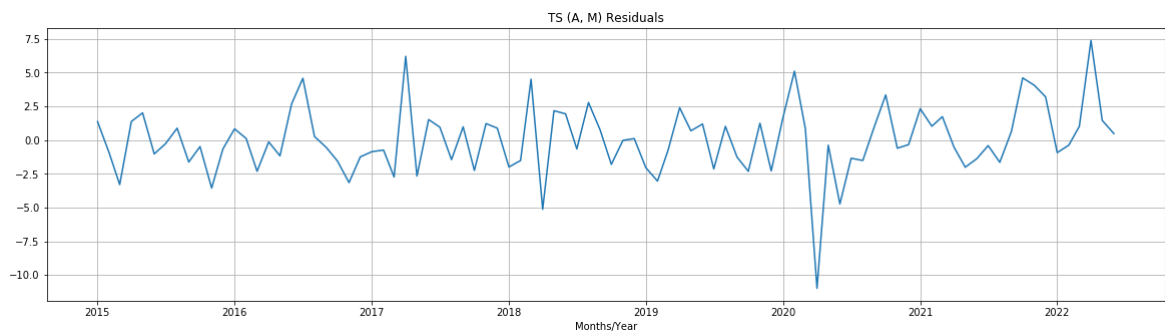


Figura 14 - Resíduos do modelo (A,M)

Na figura 14 é apresentado o ruído branco ou resíduo do modelo de tendência aditiva e sazonalidade multiplicativa (A,M), no qual se torna evidente o pico no início do ano de 2020, que pode ser explicado devido a pandemia SARS-CoV 2 (Covid 19) e ao consequente confinamento que gerou. É possível também verificar um ponto que foge á norma no início de 2022 possivelmente devido ao início da guerra na Ucrânia a 24 de fevereiro.

Com o modelo anteriormente selecionado foi possível executar uma previsão (linha azul) de modo a ser viável conferir com os dados reais já obtidos pelo INE (linha preta), ilustrado no gráfico da figura 15. Na parte inferior da referida figura é exposto o erro desta previsão onde se verifica que para o primeiro mês de previsão se obteve um erro de valor 4.5, tendo este diminuído para 1.4 e 2.7 no segundo e terceiro mês respetivamente, atingindo o máximo de 7.4 no quarto mês de previsão. Desta forma torna-se possível comparar a qualidade dos valores obtidos com os resultados que irão ser devolvidos pelos modelos de inteligência artificial.

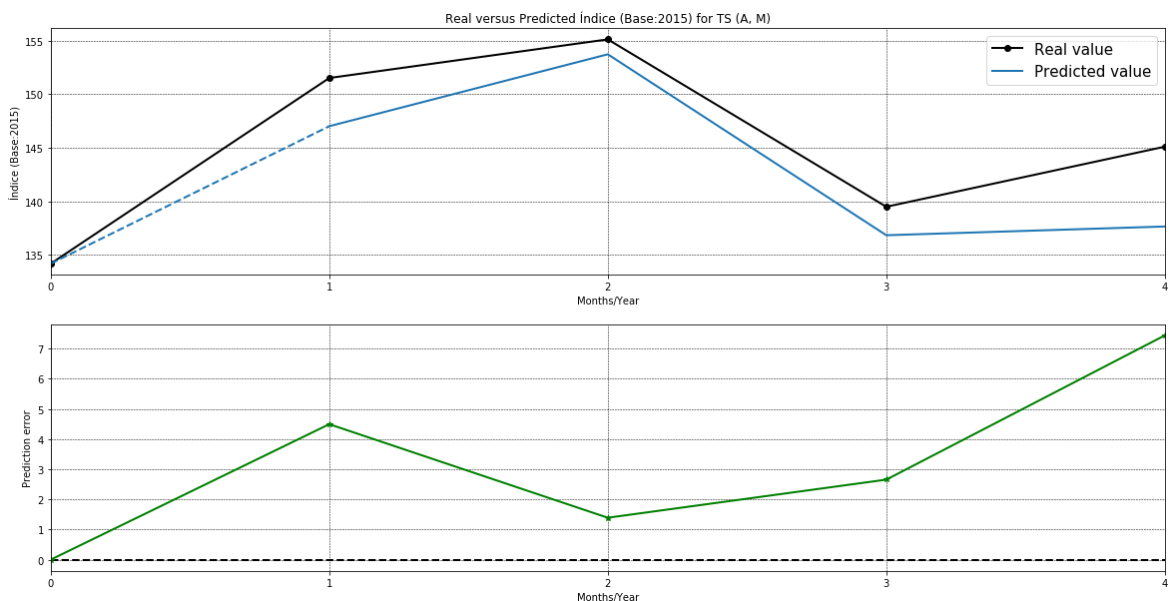


Figura 15 - Previsões modelo (A,M)

## Caso de Estudo

Relativamente aos modelos utilizando redes neurais, foi utilizado o modelo híbrido *Box&Jenkins-DNN*, representado na figura 16. Nesta é possível verificar as previsões efetuadas *in-sample* (previsões que produzem resultados dentro do intervalo temporal da série) representadas pela secção laranja, bem como a janela de dados utilizada para o treino deste modelo (faixa azul). Na parte direita do gráfico são apresentados os valores das previsões *out-sample* representados pela linha azul, que podem ser comparados com os valores reais (linha preta). Na figura 17 é apresentada em maior detalhe a zona de previsão descrita, bem como o erro relativamente ao valor real.

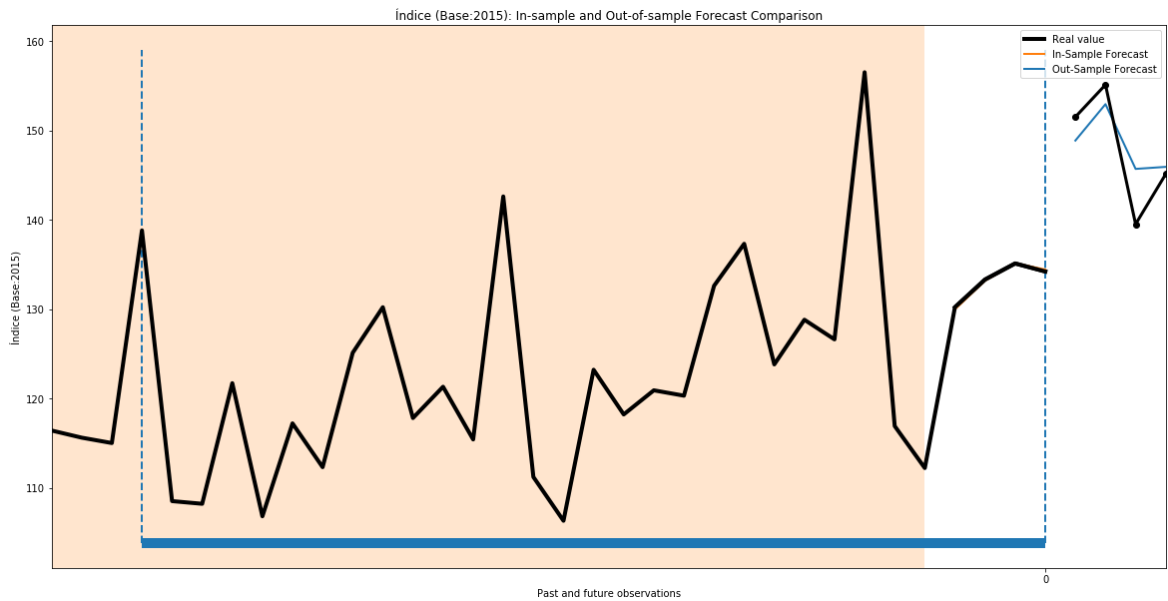


Figura 17 - Modelo BJ-DNN

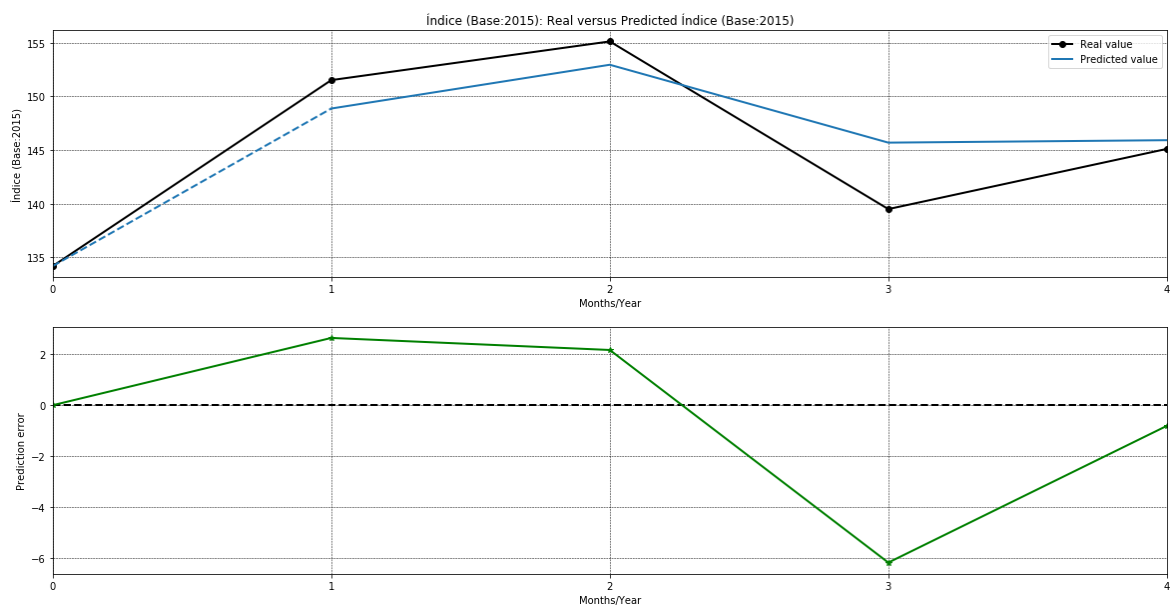


Figura 16 - Previsão modelo BJ-DNN

### 6.3. Comparação dos resultados

Tendo obtido os resultados para ambos os modelos, é agora possível compará-los e tentar compreender qual obteve melhor resultado, conforme o objetivo do relatório. Com vista a este objetivo foi criada a tabela abaixo, a qual apresenta os erros de previsão (MAPE) dos meses de junho a outubro de 2023, devolvidos pelo modelo ETS e pelo modelo DNN. Observando os gráficos das figuras 15 e 17 e tendo em conta os resultados apresentados na tabela 6, pode-se afirmar que ambos os modelos apresentam erros relativamente baixos, possuindo um bom ajuste à série em estudo.

A série em análise apresenta um claro ciclo tendência e sazonalidade tendo por isso o modelo ETS conseguido recolher a dinâmica dos dados e devolver previsões ajustadas, conseguindo assim obter um bom desempenho, conforme é apontado na literatura, isto deveu-se também ao facto desta mesma série estar dentro da normalidade, não tendo valores dispersos ou fora dos parâmetros expectáveis.

O modelo de inteligência artificial obtém um melhor resultado, estando na maioria dos casos com o seu intervalo de erro abaixo do erro do modelo clássico. Poderia ser esperado que os modelos DNN obtivessem uma maior vantagem sobre os ETS, mas sendo que os dados retirados do INE seguem a normalidade, as metodologias clássicas conseguem obter um desempenho bastante próximo das primeiras. Estando os valores do modelo ETS dentro ou bastante próximo dos intervalos devolvidos pelo modelo BJ-DNN, pode-se afirmar que ambos conseguiram resultados semelhantes levando possivelmente vantagem o modelo de inteligência artificial. Contudo o modelo clássico possui a vantagem de uma implementação mais simples sem ser necessário cuidados extras no pré-processamento dos dados. Desta forma consegue-se comprovar que os modelos de inteligência artificial são tidos como promissores, tal como é apontado na literatura científica.

Tabela 6 - Comparação dos resultados obtidos

	Julho	Agosto	Setembro	Outubro
Modelo ETS	2.97%	1.93%	1.92%	2.72%
Modelo BJ-DNN (*)	1.33% – 2.15%	1.09% – 2.05%	1.58% – 2.63%	1.46% – 2.60%

(\*) Valores mínimos e máximos (aparados em 5%) obtidos num total de 60 runs



## 7. CONCLUSÃO

Tendo em conta os objetivos propostos no início do relatório é possível afirmar que estes foram alcançados, tendo o presente trabalho levado a um maior conhecimento da história do setor alimentar, dos seus maiores produtores a nível mundial e da influência que este possui sobre as exportações nacionais. Um outro objetivo passava pela análise e exposição das diversas metodologias deste as mais clássicas até às de inteligência artificial, apresentando as várias divisões e o funcionamento de cada uma. Um último objetivo prendia-se com a comparação entre estas diferentes metodologias, para o qual foi utilizada como base a série de dados retirados do INE com vista a lhe serem aplicados os diferentes modelos.

Conforme exposto neste caso de estudo pode-se verificar que ambas as metodologias são capazes de alcançar bons resultados, uma vez que a série em análise é demonstrativa de uma série normal na qual não existem muitos pontos que fujam ao esperado. O modelo clássico obteve um erro médio de previsão de 2,39% comparativamente com 1,86% obtidos pelo modelo de inteligência artificial. Pode-se confirmar assim que os modelos de inteligência artificial demonstram ser bastante promissores, levando inclusive vantagem neste caso.

Este estudo demonstra que ambas as metodologias conseguem obter resultados bastante precisos, podendo as suas previsões ser tomadas em conta no momento de decisão, devendo ter em consideração que estas são suscetíveis a imprevistos e acontecimentos de ordem global, como referido anteriormente nos casos da pandemia covid-19 e da guerra da Ucrânia.

Ficou assim provado que é benéfica a utilização das metodologias clássicas em particular da ETS no setor alimentar, gerando previsões com qualidade, e levando como grandes vantagens a simplicidade de implementação, a minimização do armazenamento de dados e a não necessidade de pré-processamento dos dados. Verifica-se também incapacidade dos mesmos de lidar com alterações ou imprevistos que fujam aos seus padrões. Nestes casos, os modelos de inteligência artificial surgem como uma boa alternativa, pois apesar do maior custo computacional, estes conseguem obter uma maior compreensão das séries em análise.

De modo a concluir, referente às limitações efetuadas na investigação, é reconhecido o aspeto diminutivo de analisar o setor alimentar apenas com base nos índices de volume de negócios no comércio a retalho. Contudo, sendo este setor complexo e diverso, impunha-se a seleção de um conjunto de dados que representassem a globalidade do mesmo, tendo assim a escolha recaído sobre os dados referidos inicialmente. Deste modo, uma perspetiva de trabalho futuro

## Conclusão

poderia recair sobre a utilização de um conjunto de dados mais específicos ou até mesmo mais extensos, focando atenção nas temáticas atuais relacionadas com o Big Data.

Como possibilidade de investigação futura propõem-se, também, a análise do impacto da pandemia SARS-CoV 2 (Covid 19) e da guerra da Ucrânia neste setor e a comparação com países próximos de Portugal (Espanha ou França). Com isto seria possível averiguar a resiliência (ou falta da mesma) do setor alimentar nacional e tentar determinar a necessidade de mercados complementares alternativos aos europeus.

# BIBLIOGRAFIA

- Aertsen (2006). Trouble between Sales and Supply Chain: Planning with an unambiguous number gives transparency. Supply Chain Magazine
- Andrews, D. W. K. (2003). Tests for Parameter Instability and Structural Change with Unknown Change Point: A Corrigendum. *Econometrica*.
- Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*.
- Ballou, R. H. (2004). Business logistics and supply chain management. Pearson PrenticeHall.
- Bengiolan, Y., Goodfellow, I., & Courville, A. (2015). DeepLearning. The MIT Publisher.
- Box, G., & Jenkins, G. (1976). Time Series Analysis, Forecasting and Control. Holden-Day.
- Brockwell, P., & Davis, R. (1991). Time Series: Theory and Methods. Springer-Verlag.
- Brown, R. (1959). Statistical Forecasting for Inventory Control. New York: McGraw-Hill.
- Brown, R. (1964). Smoothing, Forecasting, and prediction of Discrete Time series. Prentice-Hall.
- Camil Gemael. (1994). Introdução ao Ajustamento de Observações, Editora UFPR.
- Chambers, J. C., Mullick, S. K., & Smith, D. D. (1971). How to Choose the Right Forecasting Technique. Harvard Business Review.
- Chatfield, C. (2016). The Analysis of Time Series: an introduction (6th ed.). Chapman and Hall/CRC.
- Chatfield, C., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2001). A New Look at Models for Exponential Smoothing. Journal of the Royal Statistical Society. WileyRoyal Statistical Society.
- Claeskens, G. & Hjort, N. L. (2008). Model selection and model averaging. Model Selection and Model Averaging. Cambridge University Press.
- Corrêa, H. L., & Corrêa, C. A. (2012). Administração de produção e operações. Editora Atlas.
- Cowpertwait, P., & Metcalfe, A. (2009). Introductory Time Series with R. Springer.
- Dalkey, N. C. (2003). The Delphy Methodology.
- Egrioglu, E., Bas, E., Yolcu, U. & Chen, M. Y. (2020) Picture fuzzy time series: Defining, modeling and creating a new forecasting method. Engineering Applications of Artificial Intelligence.

## Bibliografia

- Firat, M. (2008). Comparison of Artificial Intelligence Techniques for River Flow Forecasting. *Hydrology and Earth System Sciences*.
- Gonçalves, J. F. (2010). *Gestão de Aproveitamento – Stocks, Previsão, Compras*. Publindústria.
- Goodwin, P., & Wright, G. (1994). Heuristics, Biases and Improvement Strategies in Judgmental Time Series Forecasting. *Omega*.
- Government, V. S. (2022). Data collection challenges and improvements. <https://www.vic.gov.au/victorian-family-violence-data-collection-framework/data-collection-challenges-and-improvements>
- Gupta & Maranas (2003). Managing demand uncertainty in supply chain planning. *Computers and Chemical engineering*
- Guo, J., Williams, B. M., & Smith, B. L. (2007). Data Collection Time Intervals for Stochastic Short-Term Traffic Flow Forecasting. *Sage journals*.
- Hamilton, J. D. (1994). *Time series analysis: Econometric Theory*. New Jersey: Princeton University Press.
- Hannan, E. J. & Quinn, B. G. (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society*. WileyRoyal Statistical Society.
- Holt, C. (1957). *Forecasting seasonals and trends by exponentially weighted moving averages*. Pittsburgh Pa.: Carnegie Institute of Technology Graduate school of Industrial Administration.
- Hu, Z., Ge, Q., Li, S., & Xiong, M. (2020). Artificial Intelligence Forecasting of Covid-19 in China *International Journal of Educational Excellence*.
- Huang, X., Jin, G. & Ruan, W. (2012) *Machine Learning Basics*. Springer
- INE. (2023). *Estatísticas do Comércio Internacional de Bens*.
- Johnston, J. & DiNardo, J. (2001). *Métodos Econométricos (4th ed.)*. (M. Hiil, F. Ferrão, & R. Menezes, Trans.). McGraw-Hill.
- Kacmary, P., Rosova, A., Sofranko, M., Bindzar, P., Saderova, J. & Kovac, J. (2021) Creation of Annual Order Forecast for the Production of Beverage Cans—The Case Study. *Sustainability*.
- Linker, R. (2022). *Food Manufacturing Market Outlook 2022 - 2026*. <https://www.reportlinker.com/clp/global/5781>
- Makridakis, S. G., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications (3rd ed.)*. New York, USA: Wiley.
- Marina, D. & Pereira, L. (2022). *Template e regras na elaboração de dissertações MEGI*. Porto: ISEP
- Marcus, G. & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Knopf Doubleday Publishing Group.
- Meade, N. & Smith, I. (1985) ARARMA vs ARIMA - A study of the benefits of a new approach to forecasting. *Omega*.

- Monfared, M., Rastegar, H. & Kojabadi, H. M. (2009) A new strategy for wind speed forecasting using artificial intelligent methods. *Renewable Energy*.
- Murteira, B., Muller, D., & Turkman, K. (2000). *Análise de sucessões cronológicas*. McGrawHill.
- Nunes, C. (2012). *Probabilidades & Estatística*. Escolar Editora
- Oliveira, L. A. (2011). *Dissertação e Tese em Ciências e Tecnologia segundo Bolonha*. Lisboa: Lidel.
- Ord, K. (2004). Charles Holt's report on exponentially weighted moving averages: an introduction and appreciation. *International Journal of Forecasting*
- Phillips, B. Y. P. C. B. (1988). Testing for a unit root in time series regression. *Biometrika*.
- Qin, T. (2020) *Deep Learning Basics*. Springer.
- Ramos. R. F. (2021). *Data Science na modelação e previsão de séries económico-financeiras: das metodologias clássicas ao Deep Learning*. ISCTE. Lisboa.
- Rasdorf, W., Hummer, J. E., & Vereen, S. C. (2015). Data Collection Opportunities and Challenges for Skilled Construction Labor Demand Forecast Modeling. *Public Works Management & Policy*.
- Reis, F. L. dos. (2010). *Como Elaborar uma Dissertação de Mestrado segundo Bolonha*. Lisboa: Pactor.
- Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall.
- Serrano, P. (2004). *Redacção e Apresentação de Trabalhos Científicos (2a ed.)*. Lisboa: Relógio D'Água Editores.
- Sohrabpour, V., Oghazi, P., Toorajipour, R., & Nazarpour, A. (2021). Export sales forecasting using artificial intelligence. *Technological Forecasting and Social Change*, 163.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning An Introduction*. The MIT Press.
- Svoboda, R., Kotik, V., & Platos, J. (2021). Short-term natural gas consumption forecasting from long-term data collection. *Energy*.
- Vriens, A., & Versteijnen, E. (2016). *Forecasting & Planning in the Food Industry*. EyeOn.
- Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I. (2019). *Basics of Supervised Deep Learning*. Springer.
- Wei, W. W. S. (2006). *Time Series Analysis (2nd ed.)*. Pearson Addison Wesley.
- Wold, H. O. A. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksell.



# DECLARAÇÃO DE INTEGRIDADE

Declaro ter conduzido este trabalho académico com integridade. Não plagiei ou apliquei qualquer forma de uso indevido de informações ou falsificação de resultados ao longo do processo que levou à sua elaboração.

Declaro que o trabalho apresentado neste documento é original e de minha autoria, não tendo sido utilizado anteriormente para nenhum outro fim.

Declaro ainda que tenho pleno conhecimento do Código de Conduta Ética do P.PORTO.

NOME: Nuno Moura Monteiro

ISEP, Porto, 14 de junho de 2024



# ANEXOS

## Anexo A

Tabela A.1 - Índices de volume de negócios no comércio a retalho

CAE REV. 3	TOTAL	TOTAL exceto combustível	Produtos alimentares, bebidas e tabaco		
			Total	Em estabelecimentos não especializados	Em estabelecimentos especializados
<b>PONDERADOR</b>	<b>100,00</b>	<b>86,28</b>	<b>44,84</b>	<b>40,36</b>	<b>4,48</b>
<b>Média Anual</b>					
<b>2015</b>	100,0	100,0	100,0	100,0	100,0
<b>2016</b>	102,3	102,5	103,9	104,2	101,2
<b>2017</b>	108,1	107,1	108,0	108,4	104,8
<b>2018</b>	113,4	111,9	113,3	114,4	103,5
<b>2019</b>	116,8	115,5	116,6	117,8	105,7
<b>2020</b>	111,1	112,5	118,9	121,4	96,3
<b>2021</b>	119,2	119,4	125,5	128,1	102,2
<b>jan/15</b>	90,0	90,1	90,2	90,7	86,0
fev/15	84,0	83,6	84,8	85,0	83,1
mar/15	94,6	93,6	94,9	95,1	92,3
abr/15	94,4	93,4	95,6	95,1	99,6
mai/15	99,1	97,8	100,4	100,3	101,1
jun/15	98,4	97,5	97,9	97,6	101,2
jul/15	108,3	108,1	108,0	107,6	111,7
ago/15	109,1	108,8	114,1	114,2	112,6
set/15	100,4	100,5	99,6	99,1	103,8
out/15	101,5	101,9	100,8	100,4	104,3
nov/15	97,7	98,5	94,8	95,4	89,4
dez/15	122,5	126,3	119,1	119,6	115,0
<b>jan/16</b>	90,5	91,2	92,0	92,8	85,6
fev/16	86,3	86,7	87,9	88,3	84,6
mar/16	96,4	96,6	98,7	98,9	96,9

CAE REV. 3	TOTAL	TOTAL exceto combustível	Produtos alimentares, bebidas e tabaco		
			Total	Em estabelecimentos não especializados	Em estabelecimentos especializados
PONDERADOR	100,00	86,28	44,84	40,36	4,48
abr/16	94,4	94,0	97,2	97,3	95,7
mai/16	97,8	97,2	99,8	99,8	99,6
jun/16	100,9	100,7	102,9	102,8	103,4
jul/16	111,1	111,3	115,9	116,3	112,1
ago/16	112,0	111,7	118,7	119,1	115,7
set/16	103,6	103,3	105,1	104,8	108,5
out/16	104,2	104,1	104,6	105,1	99,5
nov/16	103,1	103,0	99,5	100,1	94,1
dez/16	127,9	130,1	124,0	124,6	118,6
<b>jan/17</b>	96,1	94,3	94,3	95,0	88,2
fev/17	90,4	89,0	90,2	90,7	86,1
mar/17	101,9	100,3	101,5	101,4	102,9
abr/17	102,7	101,4	106,6	106,7	104,8
mai/17	105,1	103,4	104,0	103,7	106,7
jun/17	106,2	105,0	106,8	106,6	108,2
jul/17	116,0	115,4	117,3	117,7	113,4
ago/17	116,5	115,2	120,7	121,0	118,1
set/17	109,2	108,2	109,4	109,4	109,1
out/17	107,3	106,0	107,2	107,7	102,2
nov/17	110,6	110,0	106,8	107,7	98,4
dez/17	135,2	137,1	132,0	133,3	119,5
<b>jan/18</b>	101,7	99,8	98,3	99,1	90,8
fev/18	94,5	92,5	93,8	94,6	86,6
mar/18	109,3	108,1	113,4	114,1	107,6
abr/18	102,3	100,3	102,5	103,3	95,0
mai/18	112,9	110,4	111,7	112,2	107,3
jun/18	112,5	110,7	112,0	112,3	108,7
jul/18	121,2	119,5	121,1	122,2	111,4
ago/18	123,2	120,8	129,9	131,1	118,5
set/18	112,4	110,0	115,0	116,5	102,2
out/18	115,8	113,7	113,5	114,6	103,4
nov/18	116,1	115,2	111,4	113,3	94,4
dez/18	139,0	141,3	137,7	140,0	116,2
<b>jan/19</b>	105,7	103,9	102,9	104,1	91,8
fev/19	98,4	96,7	96,8	97,9	86,7
mar/19	111,2	109,7	112,6	114,1	99,3
abr/19	111,3	108,8	112,3	113,1	105,0
mai/19	116,9	114,6	115,7	116,4	109,0
jun/19	114,4	112,6	116,1	117,4	104,4
jul/19	125,1	123,3	124,6	125,6	115,7
ago/19	125,8	124,0	132,7	133,8	123,0

CAE REV. 3	TOTAL	TOTAL exceto combustível	Produtos alimentares, bebidas e tabaco		
			Total	Em estabelecimentos não especializados	Em estabelecimentos especializados
PONDERADOR	100,00	86,28	44,84	40,36	4,48
set/19	113,9	112,7	116,4	117,8	104,1
out/19	117,6	116,3	115,6	116,4	108,3
nov/19	119,4	119,2	115,0	116,9	97,8
dez/19	142,3	144,5	138,9	140,6	122,7
<b>jan/20</b>	110,7	108,8	108,5	109,8	96,5
fev/20	106,9	105,4	108,2	109,6	95,2
mar/20	105,5	106,5	121,7	124,3	98,1
abr/20	85,6	88,8	106,8	109,5	82,7
mai/20	99,7	101,8	117,2	120,8	85,3
jun/20	106,5	108,0	112,4	114,9	89,6
jul/20	120,4	121,5	125,1	127,7	101,5
ago/20	119,9	120,4	130,2	133,2	103,1
set/20	113,9	114,9	117,8	120,0	98,1
out/20	115,7	117,6	121,3	123,8	98,7
nov/20	112,4	115,1	115,4	118,2	89,6
dez/20	136,3	141,6	142,6	145,4	117,4
<b>jan/21</b>	99,0	100,5	111,2	114,1	84,8
fev/21	91,7	92,8	106,3	109,2	79,8
mar/21	109,1	109,8	123,3	126,4	95,2
abr/21	110,0	109,8	118,2	120,7	95,7
mai/21	119,4	118,8	120,9	123,2	100,6
jun/21	117,4	116,7	120,3	122,3	101,6
jul/21	127,5	127,1	132,6	135,2	109,0
ago/21	128,3	127,1	137,3	139,9	114,7
set/21	121,0	120,0	123,8	125,7	106,8
out/21	125,2	123,8	128,9	131,1	108,3
nov/21	130,1	130,4	126,6	129,2	103,0
dez/21	151,9	156,3	156,5	159,8	127,1
<b>jan/22</b>	114,3	114,1	116,9	119,8	90,3
fev/22	112,4	111,1	112,2	114,3	92,8
mar/22	129,6	126,8	130,2	132,7	107,9
abr/22	127,4	127,0	133,3	135,5	113,6
mai/22	134,4	132,3	135,1	137,1	117,3
jun/22	132,3	130,2	134,2	135,7	121,3
jul/22	144,8	144,1	151,5	153,9	130,3
*ago-22	145,4	144,4	155,1	157,1	137,5
*set-22	135,3	135,3	139,5	141,0	126,1
*out-22	139,0	138,5	145,1	147,7	122,0

## Anexo B

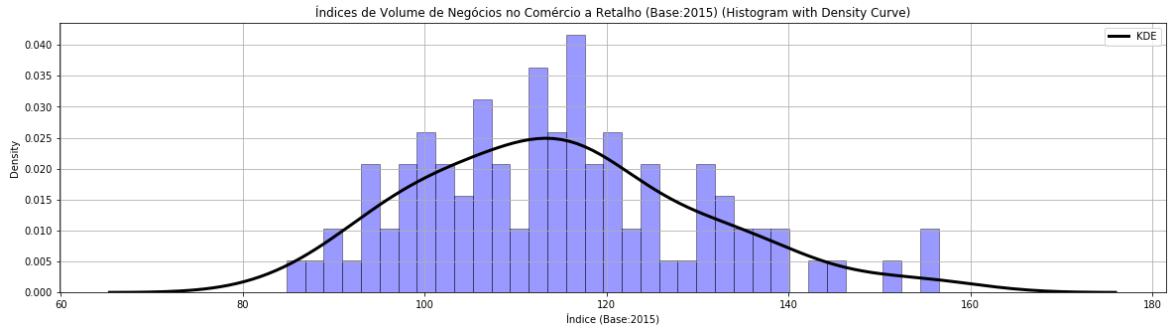


Figura 18 - Histograma com curva de densidade

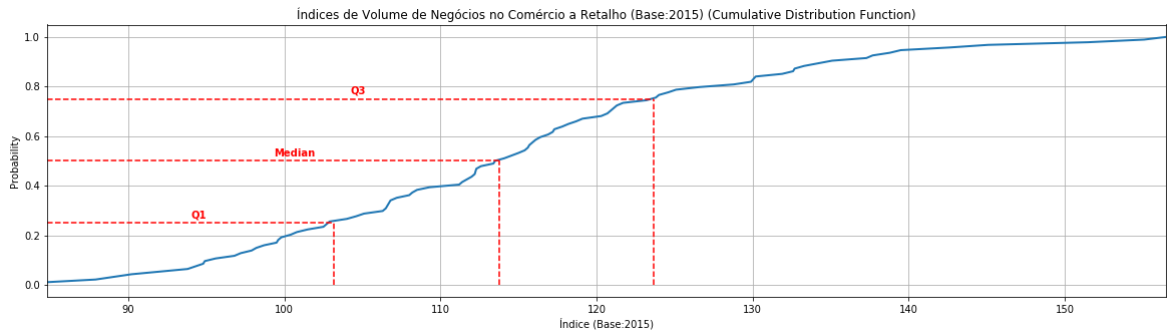


Figura 19 - Função cumulativa de distribuição

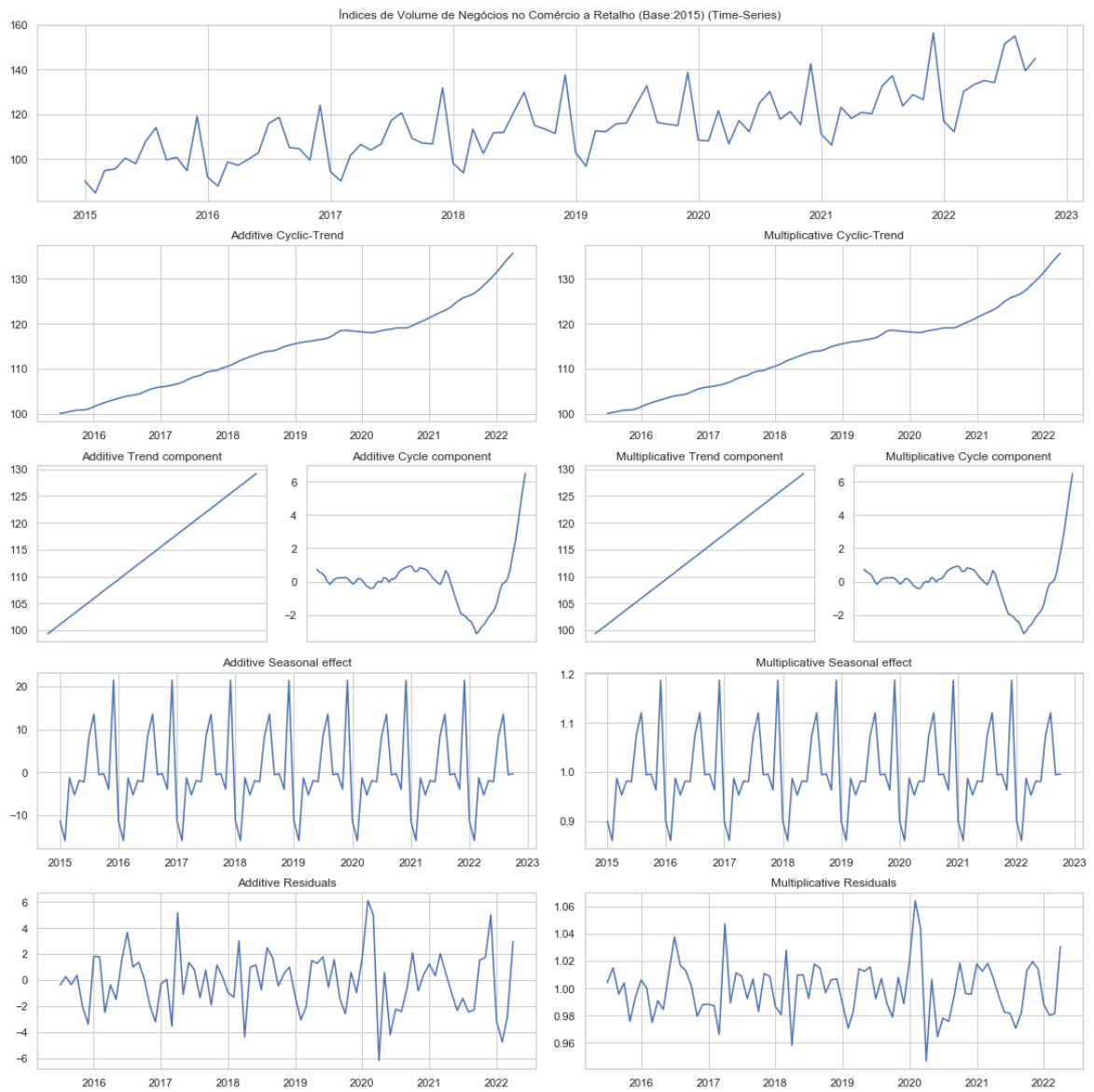


Figura 20 - Decomposição da série (aditiva e multiplicativa)