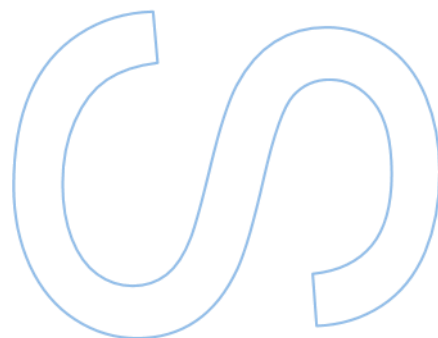
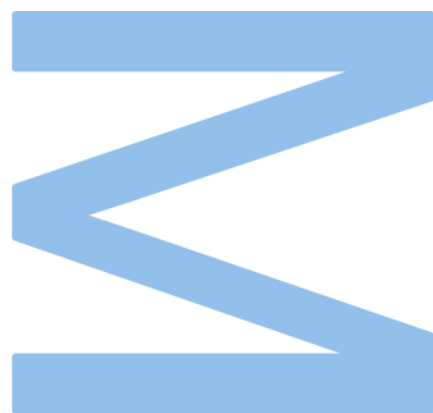


Otimização de Stocks para peças OEM

João Pedro Taveira Amaral
Mestrado em Engenharia Matemática
Departamento de Matemática
Faculdade de Ciências da Universidade do Porto
2026



Otimização de Stocks para peças OEM

João Pedro Taveira Amaral

Relatório de Estágio realizado no âmbito do
Mestrado em Engenharia Matemática

Departamento de Matemática
2026

Orientador

Manuel Cruz, Professor Adjunto, Instituto Superior de Engenharia do Porto

Coorientador

Ana Paula Rocha, Professora Associada, Faculdade de Ciências da Universidade do Porto

Supervisor

Pedro Louro, Gestor de Business Intelligence, NORS

Agradecimentos

Assim chega ao fim aquilo que é um percurso de cinco longos, mas curtíssimos anos, algo que, apesar de conseguido por mim, não podia deixar passar sem mencionar algumas pessoas que tornaram toda esta jornada possível.

Em primeiro lugar, aos meus pais e avós, por todo o sacrifício que fizeram para me poderem dar algo que eles nunca tiveram a oportunidade de ter. Sem vocês, nada disto era possível.

Ao quinto ano de MateC, que tornou o Porto casa. Vocês vão ser sempre a parte mais bonita de todo o meu percurso.

A toda a gente que teve de me ouvir durante horas intermináveis no planetário. Sem vocês, isto seria bastante aborrecido.

Ao meu orientador Manuel Cruz e ao meu supervisor Pedro Louro, pela oportunidade de poder trabalhar com vocês e por me terem guiado nesta última etapa.

Por fim, à minha coorientadora Ana Paula Rocha e a toda a gente que, de forma direta ou indireta, me ajudou a concluir o meu percurso.

Resumo

O presente trabalho foi desenvolvido no âmbito do Mestrado em Engenharia Matemática da Faculdade de Ciências da Universidade do Porto, em colaboração com o Grupo NORS. O principal objetivo consiste na otimização de stocks de peças OEM no setor agrícola brasileiro, problema caracterizado pela elevada variabilidade e intermitência da procura de peças de substituição, o que dificulta simultaneamente a previsão da procura e a definição de níveis adequados de stock de segurança.

Numa fase inicial, procedeu-se ao tratamento e análise de dados históricos de vendas e stocks, construindo séries temporais sobre os mesmos. Posteriormente, os produtos foram segmentados segundo diferentes características de procura, recorrendo a classificação por rotação e à classificação SBC.

Foram desenvolvidos e avaliados diferentes modelos de previsão da procura, incluindo métodos estatísticos, modelos específicos para procura intermitente como o Croston e o SBA, distribuições probabilísticas e abordagens de Machine Learning, com destaque para o LightGBM. Paralelamente, foram estudadas diferentes metodologias de cálculo de stock de segurança e pontos de encomenda, incluindo abordagens baseadas na distribuição normal, distribuições discretas e técnicas de *bootstrapping*.

O desempenho das diferentes metodologias foi avaliado através de simulações e indicadores de performance associados à precisão das previsões, níveis de serviço e eficiência global do sistema. Por fim, os modelos foram integrados na infraestrutura de dados empresarial e foi desenvolvida um dashboard de suporte à decisão para acompanhamento operacional.

Palavras-Chave: Gestão de stocks, previsão da procura, peças OEM, procura intermitente, stock de segurança, otimização

Abstract

This work was carried out as part of the Master's program in Mathematical Engineering at the Faculty of Sciences of the University of Porto, in collaboration with the NORS Group. The main objective is to optimize OEM parts inventories in the Brazilian agricultural sector, a problem characterized by high variability and intermittent demand for replacement parts, which simultaneously makes it difficult to forecast demand and define appropriate safety stock levels.

In an initial phase, historical sales and inventory data were processed and analyzed, constructing time series. Subsequently, products were segmented according to different demand characteristics, using turnover classification and the SBC classification.

Various demand forecasting models were developed and evaluated, including statistical methods, specific models for intermittent demand such as Croston and SBA, probabilistic distributions, and machine learning approaches, notably LightGBM. In parallel, different methodologies for calculating safety stock and reorder points were studied, including approaches based on the normal distribution, discrete distributions, and bootstrapping techniques.

The performance of the different methodologies was evaluated through simulations and performance indicators related to forecast accuracy, service levels, and overall system efficiency. Finally, the models were integrated into the enterprise data infrastructure, and a decision support dashboard was developed for operational monitoring.

Keywords: Inventory management, demand forecasting, OEM parts, intermittent demand, safety stock, optimization

Índice

Lista de Figuras	vi
Lista de Tabelas	vii
Lista de Abreviaturas	viii
1. Introdução	1
1.1. O Grupo NORS	1
1.2. Descrição do Problema.....	2
1.3. Objetivos	5
1.4. Ferramentas e Tecnologias.....	5
1.5. Estrutura do Relatório	6
2. Revisão de Literatura	7
3. Metodologia	12
3.1. Preparação e Tratamento dos Dados	12
3.1.1. Descrição e Estrutura das Bases de Dados.....	12
3.1.2. Preparação dos Dados e Construção das Séries Temporais.....	13
3.2. Segmentação de Produtos	14
3.2.1. Segmentação por Rotação dos Produtos	14
3.2.2. Segmentação por Classificação SBC.....	15
3.3. Modelos de Previsão	17
3.3.1. <i>Baseline Models</i>	17
3.3.1.1 <i>Zero Forecast</i>	18
3.3.1.2 <i>Naive Forecast</i>	18
3.3.2. Modelos Paramétricos	19
3.3.2.1 Médias Móveis	19
3.3.2.2 Suavização Exponencial Simples	19
3.3.2.3 Método de Croston	20
3.3.2.4 Método de Syntetos-Boylan Approximation	20
3.3.2.5 Método de Teunter-Syntetos-Babai	21
3.3.2.6 Método de Holt	21
3.3.3. Modelos Baseados em Distribuições	22
3.3.3.1 Distribuição de Poisson	22
3.3.3.2 Distribuição Binomial Negativa	23
3.3.3.3 <i>Zero-Inflated Poisson</i>	24
3.3.3.4 <i>Zero-Inflated Negative Binomial</i>	25
3.3.4. Modelos de Machine Learning	26
3.4. Stock de Segurança e Ponto de Encomenda	27
3.4.1. Stock de Segurança Dinâmico	28
3.4.2. Stock de Segurança com Base na Distribuição Normal	29
3.4.3. Stock de Segurança Baseado em Distribuições Discretas	30
3.4.4. <i>Bootstrapping</i> de Willemain	31
4. Metodologia Experimental e Estratégia de Avaliação	33

4.1.	Enquadramento Geral	33
4.2.	Estratégia Experimental e Divisão dos Dados	33
4.3.	Otimização de Hiperparâmetros	35
4.4.	Seleção dos Modelos	36
4.5.	Simulação e Validação dos Modelos de Gestão de Stocks.....	37
4.6.	Indicadores de Performance	38
4.6.1.	Performance das Previsões	38
4.6.2.	Performance das Políticas de Stock de Segurança	39
4.6.3.	Performance Global do Sistema	41
5.	Resultados e Discussão	43
5.1.	Resultados da Segmentação dos Produtos	43
5.1.1.	Segmentação por Rotação dos Produtos	43
5.1.2.	Segmentação por Classificação SBC.....	46
5.2.	Análise dos Modelos.....	49
5.2.1.	Análise por Métrica de Seleção	49
5.2.1.1	Análise dos Modelos selecionados pelo MSE	49
5.2.1.2	Análise dos Modelos selecionados pelo WSSE	52
5.2.1.3	Análise dos Modelos selecionados pelo MAE	54
5.2.2.	Performance das Políticas de Stock de Segurança	56
5.2.3.	Performance Global.....	58
6.	Integração dos Dados e Estrutura da Base de Dados	63
6.1.	Modelo Dimensional e Integridade Referencial	63
6.2.	Estrutura da Base de Dados Operacional	64
6.3.	Estrutura da Base de Dados Analítica	66
6.4.	Integração com SSAS e Power BI	67
6.4.1.	Dashboard de Suporte à Decisão.....	68
6.4.1.1	Análise de Vendas e Stocks	68
6.4.1.2	Sugestão de Compra	69
7.	Conclusões.....	71
	Referências	72

Lista de Figuras

3.1. Estrutura dos ficheiros gerados após a preparação e transformação dos dados (valores meramente ilustrativos do formato)	14
3.2. Matriz de classificação SBC	17
4.1. Estrutura temporal utilizada na estratégia experimental.	34
5.1. Distribuição da percentagem de períodos sem procura nas referências analisadas	44
5.2. Exemplo de série temporal de um produto de alta rotação	45
5.3. Exemplo de série temporal de um produto de média rotação	45
5.4. Exemplo de série temporal de um produto de baixa rotação	45
5.5. Distribuição das referências segundo a classificação SBC no plano $ADI-CV^2$	47
5.6. Exemplo de série temporal classificada como suave	48
5.7. Exemplo de série temporal classificada como intermitente	48
5.8. Exemplo de série temporal classificada como errático	49
5.9. Exemplo de série temporal classificada como irregular	49
5.10. Métrica MSE: Distribuição dos métodos selecionados segundo a classificação SBC	52
5.11. Métrica WSSE: Distribuição dos métodos selecionados segundo a classificação SBC	53
5.12. Métrica MAE: Distribuição dos métodos selecionados segundo a classificação SBC	55
5.13. Evolução do valor total em stock para as diferentes metodologias analisadas.	60
5.14. Evolução do stock total em unidades para as diferentes metodologias analisadas. .	60
6.1. Modelo relacional da base de dados operacional.	65
6.2. Modelo relacional da base de dados analítica e dos modelos.	66
6.3. Arquitetura da solução desenvolvida.	68
6.4. Separador de análise de vendas e stocks.	69
6.5. Painel de filtros do dashboard.	69
6.6. Separador de sugestões de compra.	70

Lista de Tabelas

3.1. Notação utilizada nos modelos de previsão	18
5.1. Distribuição das referências segundo a classificação por rotação.....	43
5.2. Distribuição das referências segundo a classificação SBC	46
5.3. Métrica MSE: Distribuição dos métodos selecionados	50
5.4. Métrica MSE: Resultados médios referentes aos KPI1, 2 e 3 durante os 12 meses de período simulado.....	50
5.5. Métrica WSSE: Distribuição dos métodos selecionados	52
5.6. Métrica WSSE: Resultados médios referentes aos KPI1, 2 e 3 durante os 12 meses de período simulado	53
5.7. Métrica MAE: Distribuição dos métodos selecionados	54
5.8. Métrica MAE: Resultados médios referentes aos KPI1, 2 e 3 durante os 12 meses de período simulado.....	54
5.9. Resultados obtidos para as diferentes políticas de stock de segurança.....	57
5.10. Resultados globais considerando o valor monetário do inventário.....	58
5.11. Resultados globais considerando o número de unidades em stock.	59

Lista de Abreviaturas

BI	Business Intelligence	1
FCUP	Faculdade de Ciências da Universidade do Porto	1
IBGE	Instituto Brasileiro de Geografia e Estatística	4
ISEP	Instituto Superior de Engenharia do Porto	1
LEMA	Laboratório de Engenharia Matemática	1
LSPA	Levantamento Sistemático da Produção Agrícola	4
MEIO	Multi-Echelon Inventory Optimization	3
OEM	Original Equipment Manufacturer	1
SS	Stock de Segurança	10

1. Introdução

Este relatório resume o trabalho desenvolvido durante o estágio curricular do Mestrado em Engenharia Matemática na Faculdade de Ciências da Universidade do Porto (FCUP). O projeto foi realizado no Departamento de Business Intelligence (BI) do Grupo NORS, em parceria com o Laboratório de Engenharia Matemática (LEMA) do Instituto Superior de Engenharia do Porto (ISEP)

O objetivo principal deste estágio foi melhorar a gestão de stocks de peças *Original Equipment Manufacturer (OEM)* na empresa NORS Equipamentos Agrícolas Brasil Centro-Oeste que opera na região do Mato Grosso. Na prática, o desafio é encontrar um equilíbrio: garantir que não faltam peças para os clientes, mas evitar ter demasiados recursos financeiros alocados ao armazém. Para resolver este problema, recorreu-se a modelos matemáticos que ajudam a prever as necessidades e a decidir as quantidades a encomendar, de forma a tornar a operação da NORS Equipamentos Agrícolas Brasil Centro-Oeste eficiente.

Por razões de confidencialidade, os valores apresentados neste relatório não correspondem aos valores reais observados pela empresa, tendo sido efetuadas transformações nos dados que preservam as conclusões e os resultados da análise.

Neste capítulo é feita uma introdução ao grupo NORS, enquadramento do problema, bem como os objetivos, software utilizado e metodologia.

1.1. O Grupo NORS

Com mais de 90 anos de história, a NORS é um grupo português que começa em 1933 quando Luiz Óscar Jervell tornou-se representante da marca VOLVO em Portugal. Em 1949 surge pela primeira vez a denominação Auto Sueco, quando Yngvar Poppe e Luiz Óscar Jervell decidem autonomizar o negócio. Desde então a empresa foi expandindo-se tanto nacional e internacionalmente.

Em 2013, o Grupo assume a sua vocação multinacional, com presença em 4 continentes e 24 mercados, o que dá origem a uma nova identidade e um novo nome para a *holding*: NORS.

No ano de 2023 foi adquirida uma nova empresa pelo grupo, a Agrofito Case Máquinas, posteriormente NORS Equipamentos Agrícolas Brasil Centro-Oeste, um concessionário de equipamentos agrícolas da marca CASE IH no Estado de Mato Grosso, Brasil.

Atualmente o Grupo NORS detém uma forte presença internacional, opera em mercados estratégicos de quatro continentes, incluindo países como Portugal, Canadá, Brasil e várias regiões do continente africano, atua nas seguintes áreas:

- **Trucks and Buses**

É a área de negócio dedicada às soluções de mobilidade e transporte, abrange a comercialização de camiões, autocarros, motores e geradores, bem como soluções de mobilidade adaptadas às necessidades do mercado.

- **Aftermarket**

É a área responsável pelos serviços de pós-venda e assistência técnica, inclui a comercialização de peças multimarcas, serviços de manutenção, reparação e suporte aos clientes ao longo do ciclo de vida dos equipamentos.

- **Construction Equipment**

É a área especializada na comercialização e suporte de máquinas e equipamentos destinados aos setores da construção, reciclagem e exploração mineira.

- **Agro**

É a área de negócio focada em soluções para o setor agrícola, disponibiliza máquinas e equipamentos utilizados em todas as fases da atividade agrícola, desde a preparação do solo, sementeiras entre outras.

- **Ventures**

É a divisão dedicada ao desenvolvimento de novos negócios e iniciativas inovadoras, atua em áreas como os seguros, sustentabilidade e soluções ambientais.

1.2. Descrição do Problema

A otimização de stocks constitui um dos desafios mais complexos e, ao mesmo tempo, uma das maiores oportunidades para a criação de valor nas empresas.

A previsão da procura é uma componente fundamental da gestão da cadeia de abastecimento e é uma preocupação central para inúmeras empresas, uma vez que constitui a base essencial para as atividades de planeamento, onde a obtenção de previsões rigorosas é crucial em todos os aspetos da gestão de inventário. (Syntetos e Boylan 2001).

De forma geral, a otimização de stocks procura encontrar um equilíbrio entre o investimento em inventário e o nível de serviço oferecido ao cliente, com o objetivo de maximizar a eficiência da cadeia de abastecimento. Neste contexto, a previsão da procura desempenha um papel essencial na redução da incerteza e no apoio à tomada de decisão ao longo da cadeia de abastecimento. Previsões mais precisas permitem reduzir custos associados ao excesso de stock, minimizar ruturas e melhorar os níveis de serviço ao cliente, contribuem simultaneamente para uma utilização mais eficiente dos recursos empresariais (Ghobbar e Friend 2003).

Tradicionalmente associada a funções de armazenamento e controlo físico, a otimização de stocks evoluiu para uma área de forte componente analítica e estratégica. Atualmente, as decisões relacionadas com níveis de inventário têm impacto direto não apenas na eficiência da operação, mas também na utilização do capital, na capacidade de resposta ao mercado e na sustentabilidade financeira das organizações. Em ambientes caracterizados por elevada volatilidade da procura e crescente complexidade das cadeias de abastecimento, a gestão eficiente de stocks assume um papel particularmente relevante, sendo frequentemente utilizada como indicador da qualidade da gestão operacional e do desempenho financeiro das empresas (Capkun, Hameri e Weiss 2009).

Segundo Gallagher, Mitchke e Rogers (2005), o mercado global de peças sobresselentes movimentava, em 2005, cerca de 400 mil milhões de dólares anuais, o que evidencia a elevada dimensão económica associada à gestão de inventário. Contudo, apesar do refinamento dos sistemas de gestão, a chamada distorção de inventário continua a representar um problema financeiro de enorme escala para as empresas. Este conceito refere-se ao desfasamento entre o inventário registado nos sistemas e o inventário efetivamente disponível, bem como situações de excesso de stock e ruturas ao longo da cadeia de abastecimento. Como consequência, as empresas enfrentam frequentemente capital imobilizado em produtos de baixa rotação, perda de vendas devido à indisponibilidade de artigos e aumento de custos operacionais associados a encomendas urgentes e reposições ineficientes. De acordo com o estudo *Retail Inventory Distortion*, o impacto económico global destas ineficiências atinge aproximadamente 1.77 biliões de dólares, refletindo o efeito significativo da má gestão de stocks na rentabilidade empresarial (IHL Group 2023).

Apesar do vasto número de estudos desenvolvidos na área da gestão de stocks e previsão da procura, continua a verificar-se um desfasamento significativo entre os modelos propostos na literatura e a sua aplicação efetiva em contexto empresarial (Bacchetti e Sacconi 2012). Embora metodologias avançadas, como a Multi-Echelon Inventory Optimization (MEIO), apresentem

resultados superiores em termos teóricos, a sua adoção prática nas organizações permanece ainda relativamente reduzida.

Outro fator de complexidade resulta da própria natureza da procura de peças sobresselentes, frequentemente caracterizada por padrões irregulares, elevada variabilidade e procura intermitente, o que torna particularmente complexa a previsão da procura. Segundo Silver (1981), a procura intermitente caracteriza-se pela presença frequente de períodos sem consumo, originando séries temporais com uma elevada proporção de zeros. Para além disso, a reduzida disponibilidade de dados e a elevada variabilidade associada aos períodos com procura positiva tornam difícil a previsão rigorosa da procura. Este problema é especialmente relevante em setores como o das máquinas agrícolas e de construção, onde a necessidade de peças de substituição depende de condições operacionais difíceis de antecipar (Costantino, Di Gravio e Shaban 2018).

Adicionalmente, a procura de peças sobresselentes apresenta comportamentos irregulares e descontínuos, sem padrões evidentes de tendência ou sazonalidade, o que dificulta a aplicação de métodos tradicionais de previsão (Willemain, Smart e Schwarz 2004). Como consequência, a previsão de procura intermitente tem sido amplamente reconhecida na literatura como um dos principais desafios associados à gestão de inventário de peças de substituição.

Neste contexto, a gestão eficiente de peças sobresselentes assume particular importância em setores fortemente dependentes da disponibilidade contínua de equipamentos, como a agricultura. Em atividades agrícolas, a indisponibilidade de máquinas devido à falta de peças pode provocar atrasos significativos em períodos críticos de produção e colheita, tornando essencial garantir a rápida reposição das peças.

Este problema é bastante relevante no contexto brasileiro, em particular na região Centro-Oeste, responsável por cerca de 50% da produção agrícola nacional segundo dados do Levantamento Sistemático da Produção Agrícola (LSPA) divulgados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para o ano de 2023 (IBGE 2023).

Para além disso, o agronegócio brasileiro representa um dos setores mais relevantes da economia do país, correspondendo a cerca de 21.4% do PIB brasileiro em 2019. No mesmo ano, as exportações do setor atingiram aproximadamente 96.8 mil milhões de dólares, representando 43.2% do total exportado pelo Brasil (Christ et al. 2022).

1.3. Objetivos

O objetivo geral deste trabalho consiste em desenvolver um novo modelo de gestão de stocks para uma empresa do grupo NORS, focado no equilíbrio entre o nível de serviço e o investimento em stock minimizando as vendas perdidas através da utilização de modelos de previsão, bem como avaliar o desempenho com base em indicadores de performance.

Numa fase inicial, analisaram-se os dados disponibilizados pela empresa, estudou-se a estrutura das bases de dados, as variáveis existentes, a existência de valores omissos, inconsistências e outros aspetos relevantes para o tratamento dos dados e utilização futura no modelo.

Aproveitou-se esta fase inicial para segmentar os produtos comercializados com base nos padrões de procura, de forma a adaptar os modelos de previsão escolhidos às características de cada tipo de procura.

Modelaram-se também, os níveis de stock de segurança e os pontos de encomenda mais adequados através de diferentes políticas de stock de segurança, integrando as previsões dadas pelos modelos.

Por fim, comparou-se o desempenho dos diversos modelos com a política de compras atualmente utilizada, através simulações com diferentes níveis de serviço.

1.4. Ferramentas e Tecnologias

O trabalho realizado foi desenvolvido maioritariamente em Python 3.11 (2022), que serviu de base para o processamento dos dados, construção dos modelos de previsão e a implementação de diferentes políticas de stock de segurança, simulações, visualização e análise dos resultados. Para tal, recorreu-se a bibliotecas como Pandas e Numpy para o tratamento de dados, ao `scipy.stats` e `scikit-learn` para cálculos estatísticos e simulações de *Bootstrap*. Utilizaram-se ainda bibliotecas auxiliares como `regular expression`, `glob`, `os` para o tratamento de ficheiros e cadeias de caracteres. Para a visualização recorreu-se ao Matplotlib e Seaborn.

Após a criação do modelo, a integração com a infraestrutura de dados da empresa foi feita através da exportação da informação para o Microsoft SQL Server 2022 (2022), recorrendo a bibliotecas como o SQLAlchemy e o PyODBC. Os processos de ETL (*Extract, Transform and Load*) foram orquestrados recorrendo ao SQL Server Integration Services. Sobre esta base de dados recorreu-se ao SQL Server Analysis Services, para definir e aplicar modelos tabulares e métricas relevantes ao negócio.

A componente de visualização e suporte à decisão foi desenvolvida em Power BI, utilizando-se do modelo tabular e das métricas definidas anteriormente.

1.5. Estrutura do Relatório

O presente trabalho está organizado da seguinte forma no **Capítulo 2** é efetuada uma revisão da literatura relevante, aborda os principais conceitos associados à procura intermitente, gestão de stocks e políticas de stock de segurança.

No **Capítulo 3** é descrita a metodologia utilizada ao longo do trabalho, incluindo o tratamento de dados, a segmentação dos produtos, os modelos de previsão e as políticas de stock de segurança.

O **Capítulo 4** detalha as técnicas utilizadas no desenvolvimento dos modelos, como a escolha dos dados de treino, validação e teste, a otimização dos parâmetros, os critérios de seleção de modelos, a estratégia de simulação e os indicadores de performance utilizados.

No **Capítulo 5** são apresentados e discutidos os resultados obtidos, analisado o desempenho dos modelos de previsão e das diferentes políticas de stock de segurança. De seguida, no **Capítulo 6**, é apresentada a *dashboard* desenvolvida bem como o processo de integração nos sistemas da empresa e no último capítulo são apresentadas as principais conclusões, bem como possíveis orientações para a continuação do trabalho realizado no contexto aqui apresentado.

2. Revisão de Literatura

Grande parte da abordagem desenvolvida no presente trabalho foi inspirada no artigo “*Order and Stock Costs Optimization in an Automotive Spare Parts Wholesaler*” de Cruz et al. (2020). Este trabalho apresenta uma aplicação prática no contexto da otimização de stocks de uma empresa do grupo NORS. O artigo aborda a utilização de modelos de previsão da procura e políticas de stock de segurança com o objetivo de otimizar os níveis de inventário e reduzir os custos associados ao armazenamento e às encomendas.

A aplicação eficaz destes modelos pode beneficiar de uma segmentação prévia dos produtos, dado que diferentes produtos apresentam padrões de consumo distintos. Neste sentido, a classificação da procura constitui uma etapa importante na gestão de stocks e na seleção de modelos de previsão adequados. Uma das classificações mais utilizadas é a classificação ABC, baseada no princípio de Pareto, sendo frequentemente utilizada para segmentar produtos de acordo com a sua importância relativa no inventário. Assim, os produtos são divididos em três categorias com base no volume de vendas ou valor movimentado, o que permite definir estratégias de previsão e níveis de serviço distintos entre cada grupo (Syntetos et al. 2009).

Apesar de ser bastante utilizada, a classificação ABC apresenta algumas limitações quando aplicada no contexto da previsão de procura, uma vez que produtos com a classificação mais baixa podem apresentar padrões de procura distintos. Para corrigir este problema foram surgindo vários métodos alternativos que classificam com base na importância económica assim como nas características da procura. Syntetos, Boylan e Croston (2005) (SBC) propuseram uma classificação baseada no intervalo médio entre procura (*Average Demand Interval* -ADI) e no coeficiente de variabilidade da procura não nula (CV^2), permitindo assim segmentar em diferentes padrões de consumo. Com base nestes dois indicadores pode-se classificar em quatro categorias: suave, intermitente, errático e irregular. Esta classificação conhecida como SBC, rapidamente tornou-se uma das mais utilizadas na literatura, devido à sua simplicidade de implementação e capacidade de adaptação a diversos tipos de procura.

A previsão de procura intermitente tem vindo a tornar-se uma área de investigação bastante relevante, uma vez que os métodos tradicionais de séries temporais tendem a apresentar um desempenho pouco satisfatório quando encontram largos períodos sem procura. Um dos trabalhos mais influentes nesta área foi desenvolvido por Croston (1972), que propôs separar a série temporal em duas componentes distintas: a quantidade de procura e o intervalo entre ocorrências de procura. Cada uma destas componentes é prevista individualmente através de suavização

exponencial simples e posteriormente combinadas para produzir a previsão final. Apesar da sua simplicidade, estudos posteriores revelaram que o método apresenta algum enviesamento nas previsões, o que levou ao desenvolvimento de versões melhoradas destacando-se a de Syntetos e Boylan (2005), que se tornou a principal referência ao introduzir um fator de correção na previsão.

Uma abordagem alternativa consiste no desenvolvimento de estratégias de *bootstrapping*. De entre estas estratégias destaca-se Willemain, Smart e Schwarz (2004), que estima a distribuição da procura durante o *lead time* através de um processo com duas etapas. Primeiro, utiliza-se uma cadeia de Markov com dois estados para simular e capturar a autocorrelação da ocorrência de procura. De seguida, para os períodos em que se determina haver procura positiva, aplica-se *bootstrapping* para selecionar valores historicamente observados, aos quais se adicionam pequenas variações (*jittering*). Desta forma consegue-se representar a natureza irregular e esporádica da procura intermitente melhor que os métodos determinísticos.

Nos últimos anos, métodos de *ensemble learning* têm ganho notoriedade em problemas de previsão devido a sua capacidade de combinar vários modelos simples e produzir previsões mais robustas. Entre estas abordagens destaca-se o *AdaBoost* (*Adaptive Boosting*), um dos primeiros modelos de *boosting*, que constrói iterativamente uma sequência de modelos atribuindo maior peso às observações incorretamente classificadas nas iterações anteriores. Os princípios do *AdaBoost* serviram como base para o aparecimento de novos métodos de *boosting* (Freund e Schapire 1997; Zhang et al. 2020).

Uma evolução deste modelo é o *Gradient Boosting*, no qual cada novo modelo é ajustado aos erros residuais produzidos pelos modelos anteriores. Dentro dos modelos de *Gradient Boosting* destaca-se o *eXtreme Gradient Boosting* (*XGBoost*), que utiliza técnicas de regularização, amostragem e otimização (Chen e Guestrin 2016). Esta abordagem tem-se mostrado particularmente útil quando existem variáveis explicativas com relações não lineares (Zhang et al. 2020).

Mais recentemente em Ke et al. (2017) surgiu o *Light Gradient Boosting Machine* (*LightGBM*). Este modelo baseado em árvores de decisão foi concebido para reduzir os tempos de treino e o consumo de memória através de estratégias para controlar o crescimento das árvores e da discretização eficiente das variáveis contínuas, tornando-o assim adequado em conjuntos de dados com grandes dimensões e de elevada complexidade. De acordo com Shehadeh et al. (2021), tanto o *LightGBM* como o *XGBoost* são alternativas promissoras aos métodos estatísticos tradicionais de previsão, como por exemplo modelos autoregressivos (ARIMA).

Outra etapa fundamental no processo de modelação diz respeito à avaliação dos modelos preditivos, uma vez que permite quantificar a capacidade de previsão e comparar objetivamente diferentes modelos. A escolha de métricas adequadas assume um papel determinante na análise da qualidade das previsões, pois diferentes métricas capturam diferentes características. De acordo com Hyndman e Koehler (2006), não existe uma métrica universalmente superior para todos os problemas de previsão.

No contexto de previsão da procura, a utilização de múltiplas métricas torna-se ainda mais relevante devido à elevada variabilidade nas séries temporais e à possibilidade de ocorrência de valores extremos ou de procura reduzida. De entre as métricas mais utilizadas nestes problemas, encontram-se: *Mean Absolute Error* – MAE, *Root Mean Squared Error* – RMSE e o *Mean Absolute Percentage Error* – MAPE, uma vez que fornecem perspetivas complementares sobre a precisão das previsões produzidas pelos modelos (Hyndman e Koehler 2006). Enquanto o MAE mede a magnitude média dos erros sem penalizar desvios extremos, o RMSE penaliza mais os erros de grande dimensão devido à utilização do quadrado dos resíduos. Por sua vez, o MAPE expressa o erro em termos relativos, o que permite avaliar a precisão das previsões independentemente da escala dos dados e facilitando a comparação entre diferentes modelos e conjuntos de observações.

Em Cruz et al. (2020), foram comparados diversos métodos de previsão da procura, incluindo médias móveis, suavização exponencial e técnicas específicas para procura intermitente, como Croston, Syntetos–Boylan Approximation (SBA), entre outros. Os resultados obtidos foram utilizados para estimar níveis de stock de segurança com base nos erros das previsões, contribuindo assim para uma gestão mais eficiente do inventário e para a redução dos custos operacionais.

Para além das previsões, a distribuição dos erros das mesmas assume um papel fundamental na gestão de stocks. Muitos modelos de controlo de stocks assumem que os erros de previsão não são enviesados e seguem aproximadamente uma distribuição normal, fazendo com que a variabilidade da procura seja estimada através do desvio padrão dos erros. Nesta abordagem é comum utilizar um fator multiplicativo vindo da distribuição normal, para determinar a quantidade de stock adicional para mitigar o risco de ruturas (Barrow e Kourentzes 2016; Axsäter 2006). Segundo Barrow, Kourentzes e Sandberg (2017), o RMSE deve ser preferido ao desvio padrão dos erros no cálculo do stock de segurança baseado em previsões, uma vez que consegue capturar tanto a dispersão como eventuais enviesamentos.

No entanto, a hipótese de normalidade nem sempre é válida. Para ultrapassar esta limitação Strijbosch, Heuts e Schoot (2000) desenvolveram uma abordagem que utiliza distribuições como

a de Poisson e a Binomial Negativa, para modelar a frequência e a magnitude da procura.

Esta ligação entre previsão da procura e o cálculo de Stock de Segurança (SS) tem sido amplamente reconhecida na literatura, contudo vários autores destacam que estas áreas são frequentemente estudadas de forma separada, apesar de a qualidade das previsões influenciar diretamente as decisões de compra e os níveis de stock de segurança. Assim, a análise dos erros de previsão torna-se particularmente relevante, uma vez que a sua variabilidade influencia diretamente os níveis de stock de segurança necessários para alcançar o mesmo nível de serviço (Goltsos, Syntetos e Boylan 2022).

Podemos dividir as políticas de stock de segurança, com base no sistema de revisão: o sistema de revisão contínua, em que uma encomenda é desencadeada sempre que a posição de inventário atinge um determinado ponto de encomenda (*reorder point*), sendo o stock de segurança utilizado para proteger durante o tempo de reposição. Outra alternativa passa pelo sistema de revisão periódica, em que o inventário é analisado em intervalos regulares, pode exigir um maior stock de segurança, dependendo da periodicidade da reposição e consequente exposição à incerteza da procura (Silver, Pyke e Thomas 2017).

No contexto da crescente complexidade dos sistemas de informação empresariais, tem-se verificado a adoção de arquiteturas de *Business Intelligence* (BI), onde o principal objetivo consiste em transformar grandes volumes de dados em informação útil. Neste contexto, os *Data Warehouses* assumem um papel central ao integrar dados provenientes de diversas fontes (Golfarelli e Rizzi 2009).

Assim surgem os modelos semânticos, cuja função é disponibilizar uma representação analítica dos dados independente da sua estrutura física. Estes modelos servem para centralizar as relações entre entidades, as regras de negócio e os indicadores utilizados, bem como permitir que diferentes utilizadores acessem à mesma informação de forma consistente. Os modelos semânticos reduzem inconsistências entre relatórios e contribuem para um melhor desempenho das consultas, uma vez que os cálculos e agregações passam a ser efetuados sobre um modelo otimizado para análise multidimensional (Russo, Ferrari e Webb 2012; Microsoft 2025).

No contexto da gestão de stocks, estas arquiteturas assumem especial importância devido à necessidade de integrar informação proveniente de vendas, inventário, compras e previsões da procura.

A plataforma Microsoft tem um papel de destaque neste domínio, ao disponibilizar um ecossistema integrado de ferramentas para armazenamento, processamento, modelação e visualização

de dados. Destacam-se tecnologias como o SQL Server, utilizado para a implementação do *data warehouse*, o SQL Server Integration Services (SSIS), responsável pela orquestração dos processos de ETL, o SQL Server Analysis Services (SSAS), utilizado na construção do modelo tabular e na definição das métricas analíticas, e o Power BI, responsável pela visualização e exploração dos resultados. Atualmente, este conjunto de ferramentas é amplamente utilizado a nível mundial, sendo uma das soluções mais adotadas pelas organizações para análise de dados e suporte à decisão (Zierz et al. 2026).

3. Metodologia

Neste capítulo é apresentada a metodologia utilizada ao longo do trabalho, bem como os principais conceitos teóricos associados às diferentes etapas da análise. A metodologia adotada procura integrar técnicas de tratamento de dados, métodos de segmentação de produtos, modelos de previsão da procura e políticas de stock de segurança.

3.1. Preparação e Tratamento dos Dados

3.1.1. Descrição e Estrutura das Bases de Dados

Neste trabalho foram utilizados dados disponibilizados pela empresa contendo informação histórica relativa às vendas e aos stocks dos produtos. Uma base de dados contém registos das vendas desde 1 de janeiro de 2021 até 30 de junho de 2024 em que cada linha corresponde aos movimentos diários (entradas e saídas) referentes ao período em análise. Esta base de dados era constituída por: data do movimento, o código do produto, o preço, informações fiscais associadas à transação, número da nota fiscal, vendedor, canal de compra e tipo de compra, entre outras variáveis utilizadas internamente pela empresa. Para além desta tabela principal, existiam também outras tabelas auxiliares relacionadas com produtos, vendedores, compradores, entre outras, estando assim a base de dados já organizada segundo uma estrutura relacional normalizada.

Apesar da boa organização da informação, uma das primeiras dificuldades encontradas consistiu em perceber exatamente quais os registos que representavam vendas efetivas relevantes para o estudo, uma vez que a base de dados incluía diferentes tipos de movimentos, como devoluções, trocas, peças cobertas por garantia e operações internas, pelo que foi necessário compreender o significado das variáveis e dos códigos existentes no sistema antes de selecionar os dados realmente associados a vendas.

Não foram encontrados, nesta base de dados, valores duplicados nem valores omissos, pelo que não foi necessário realizar procedimentos de imputação ou limpeza adicional, foi apenas aplicada, por segurança, uma verificação e remoção de eventuais espaços em branco no início e fim de cada célula. Da mesma maneira, optou-se por não se remover nenhum *outlier*, uma vez que eles representam situações reais de procura, logo fazem parte dos padrões que se pretendem analisar e prever.

Foi também utilizada uma base de dados relacionada com os stocks dos produtos relativa ao

mesmo período da base de dados anterior. A sua estrutura é bastante semelhante à base de dados das vendas, embora mais focada para a informação interna do inventário, uma vez que existiam várias colunas associadas a códigos e estados internos de stock.

Posteriormente, foram utilizadas ainda duas bases de dados adicionais relativas ao período de 1 de julho de 2024 até 31 de dezembro de 2025, com informação semelhante às inicialmente analisadas, mas com uma estrutura de extração diferente. Ao contrário das anteriores, que continham a informação toda num único ficheiro, estas novas extrações foram realizadas mensalmente, sendo cada ficheiro composto por informação correspondente a um período dos últimos três meses.

A estrutura relacional destas novas bases de dados manteve-se muito semelhante à da base inicial, no entanto alguns dos códigos internos utilizados pela empresa sofreram alterações tendo sido necessário proceder ao mapeamento dos mesmos, de forma a garantir a consistência da informação ao longo do tempo.

Estas novas extrações apresentam uma estrutura mais simples, uma vez que várias colunas relacionadas com a informação interna da empresa deixaram de estar incluídas nos ficheiros disponibilizados. Ainda assim, as variáveis essenciais para a análise, previsão da procura e análise dos stocks continuaram disponíveis. Como as extrações eram realizadas mensalmente e cada ficheiro incluía informação relativa aos três meses anteriores, foi necessário remover linhas duplicadas e em casos de observações diferentes para o mesmo mês em ficheiros diferentes optou-se por manter a observação que correspondia à extração mais recente.

3.1.2. Preparação dos Dados e Construção das Séries Temporais

Numa fase inicial da preparação dos dados, foram removidas várias colunas que não apresentavam relevância para a análise da procura. Foram igualmente excluídos produtos não operacionais, assim como referências com menos de dois anos de histórico de vendas, sobrando assim 15542 referências diferentes. De seguida, foi necessário identificar os diferentes tipos de movimentos presentes na base de dados. Para simplificar o tratamento da informação, foi atribuído o valor 1 aos movimentos correspondentes a vendas e o valor -1 aos movimentos associados a devoluções, com os movimentos identificados e após calcular o balanço diário de cada produto, foi construída uma matriz em que cada coluna corresponde a um produto diferente e cada linha a um dia sendo esta matriz preenchida com o balanço final do respetivo dia.

Foi ainda implementado um tratamento adicional de valores negativos nas séries temporais.

Sempre que num determinado dia o balanço era negativo, esse valor era redistribuído retroativamente pelos dias anteriores com vendas positivas, subtraindo progressivamente essas quantidades até eliminar o saldo negativo registado.

Optou-se também por trabalhar com os dados das vendas diárias por permitir manter o maior nível de detalhe nos dados e por ser sempre possível agregar em outros períodos.

No caso da informação relativa aos stocks, foram igualmente construídas séries temporais seguindo uma estrutura semelhante. Neste caso não existia a necessidade de atribuir sinais positivos ou negativos aos movimentos, uma vez que os dados já representavam os valores existentes no inventário naquele período.

Após a construção das séries temporais, os dados foram guardados em ficheiros Excel num formato estruturado semelhante ao apresentado na Figura 3.1. Isto permitiu organizar melhor a informação e facilitar o processamento dos dados ao longo do trabalho, sobretudo devido à dimensão das matrizes temporais utilizadas.

Código do Produto	Data	Qtd_pm
0000901127	2025-10-01 00:00:00	15
0000923547	2024-07-09 00:00:00	10
0000923547	2024-08-15 00:00:00	-10
0000924530	2024-07-09 00:00:00	1
0000924530	2024-08-15 00:00:00	-1
0000934542	2025-02-20 00:00:00	8
0000934542	2025-05-15 00:00:00	6
0000934542	2025-07-09 00:00:00	2
0000934542	2025-08-28 00:00:00	2
0000934542	2025-08-29 00:00:00	4

(a) Estrutura da base de dados de vendas

Referencia	Data	EEPE	Custo Medio
000011	2021-01-31 00:00:00	69	17,07893
000011	2021-02-28 00:00:00	69	17,07893
000011	2021-03-31 00:00:00	69	17,07893
000011	2021-04-30 00:00:00	69	17,07893
000011	2021-05-31 00:00:00	69	17,07893
000011	2021-06-30 00:00:00	69	17,07893
000011	2021-07-31 00:00:00	69	17,07893
000011	2021-08-31 00:00:00	69	17,07893
000011	2021-09-30 00:00:00	105	17,07893
000011	2021-10-31 00:00:00	105	17,07893

(b) Estrutura da base de dados de stocks

Figura 3.1 - Estrutura dos ficheiros gerados após a preparação e transformação dos dados (valores meramente ilustrativos do formato)

3.2. Segmentação de Produtos

3.2.1. Segmentação por Rotação dos Produtos

A segmentação dos produtos com base no seu nível de rotação permite distinguir referências com diferentes padrões de consumo e frequência de vendas (Syntetos, Boylan e Croston 2005). Para esta classificação, considerou-se a proporção de períodos com procura nula ao longo do horizonte temporal analisado, permitindo diferenciar produtos com diferentes níveis de procura.

Os produtos são então segmentados de acordo com a frequência com que registam períodos sem procura:

- **Alta Rotação:** Produtos com vendas frequentes e menos de 30% de períodos sem procura.
- **Média Rotação:** Produtos com um comportamento intermédio, apresentando períodos com e sem vendas.
- **Baixa Rotação:** Produtos com procura intermitente, registando mais de 70% de períodos sem vendas.

A percentagem de períodos sem procura para cada produto é:

$$Z_i = \frac{N_{0,i}}{N_i} \times 100 \quad (3.1)$$

Onde,

- Z_i representa a percentagem de períodos sem vendas do produto i ;
- $N_{0,i}$ corresponde ao número de períodos com procura nula;
- N_i representa o número total de períodos observados para esse produto.

3.2.2. Segmentação por Classificação SBC

Para complementar a segmentação baseada na rotação dos produtos, foi utilizada a classificação SBC (Syntetos, Boylan e Croston 2005), o que permite caracterizar de forma mais detalhada o comportamento da procura de cada produto. Esta abordagem considera não apenas a frequência com que ocorrem vendas, mas também a variabilidade associada aos valores de procura.

A classificação SBC baseia-se em dois indicadores principais:

- **ADI (*Average Demand Interval*):** mede o intervalo médio entre períodos com procura diferente de zero, sendo calculado através da expressão:

$$ADI_i = \frac{N_i}{N_{z,i}} \quad (3.2)$$

- CV_i^2 (**Squared Coefficient of Variation**): mede a variabilidade da procura através do coeficiente de variação quadrático, sendo definido por:

$$CV_i^2 = \left(\frac{\sigma_i}{\mu_i} \right)^2 \quad (3.3)$$

Onde,

- $N_{z,i}$ representa o número de períodos com procura diferente de zero;
- μ_i corresponde à média da procura do produto i ;
- σ_i representa o desvio padrão da procura do produto i .

Com base nos valores de ADI_i e CV_i^2 , os produtos são classificados em diferentes categorias de procura:

- **Suave (Smooth)**: Produtos com vendas frequentes e valores de procura relativamente estáveis ao longo do tempo;
- **Intermitente (Intermittent)**: Produtos em que a procura ocorre de forma espaçada, embora os valores registados apresentem pouca variação;
- **Errático (Erratic)**: Produtos com vendas frequentes, mas com grandes oscilações nas quantidades procuradas;
- **Irregular (Lumpy)**: Produtos caracterizados por longos períodos sem procura e por fortes variações nas quantidades vendidas quando ocorre consumo.

A distinção entre as categorias tem por base os valores de corte tradicionalmente utilizados na literatura, correspondendo a $ADI = 1.32$ e $CV^2 = 0.49$. A Figura 3.2 apresenta a matriz de classificação.

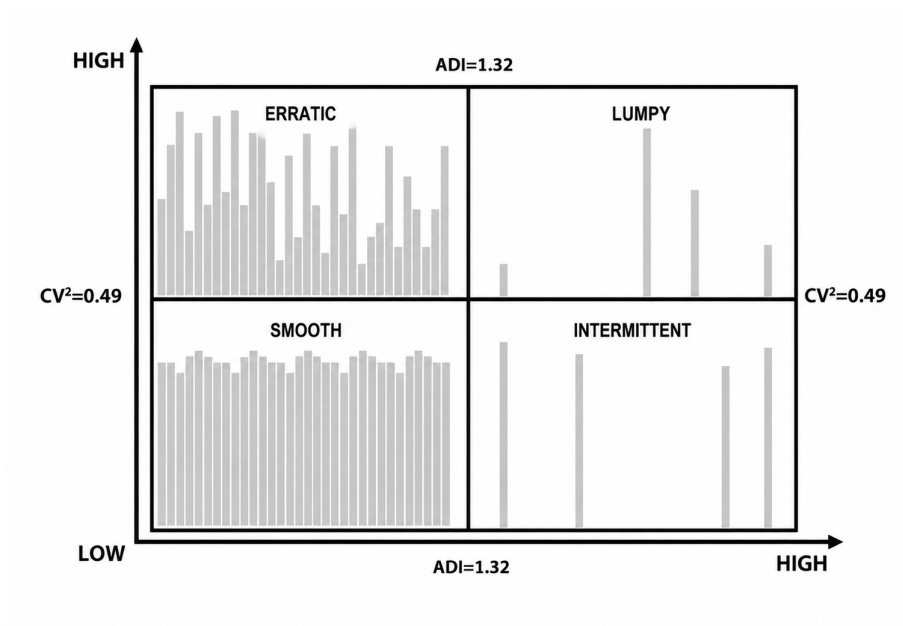


Figura 3.2 - Matriz de classificação SBC

3.3. Modelos de Previsão

Nesta secção são apresentados e discutidos os métodos de previsão utilizados para a previsão da procura. A seleção destes métodos teve por base os trabalhos de: Cruz et al. (2020) e Romeijnnders, Teunter e Jaarsveld (2012).

Na Tabela 3.1 encontra-se a notação utilizada ao longo desta secção.

3.3.1. *Baseline Models*

Os *Baseline Models* correspondem a modelos de previsão simples, utilizados frequentemente como referência para comparação com modelos mais complexos (Hyndman e Athanasopoulos 2018). Apesar da sua simplicidade, estes modelos assumem especial importância na previsão de procura de peças sobresselentes, uma vez que este tipo de procura é caracterizado por longos períodos sem consumo. Nestas situações, modelos simples podem, em determinados casos, apresentar desempenhos superiores a modelos mais sofisticados.

Neste trabalho foram utilizados dois *Baseline Models*: o *Zero Forecast* e o *Naive Forecast*.

Notação	Descrição
d_t	Procura observada no período t
$\hat{X}_t$	Previsão da procura para o período t
$\hat{S}_t$	Previsão da procura positiva no período t
$\hat{K}_t$	Previsão do intervalo entre procuras positivas
k_t	Número de períodos desde a última procura positiva
$\hat{p}_t$	Previsão da probabilidade de ocorrência de procura positiva
p_t	Variável binária indicadora de procura positiva
a_t	Nível da série temporal no período t
b_t	Tendência da série temporal no período t
α, β	Constantes de suavização ($0 \leq \alpha, \beta \leq 1$)
λ, σ	Média e desvio padrão da distribuição
n, π	Parâmetros da distribuição Binomial Negativa
ϕ	Parâmetro de dispersão da distribuição Binomial Negativa
ψ	Probabilidade de ocorrência de zeros estruturais
M	Número total de árvores do modelo
$f_m(x_t)$	Previsão produzida pela árvore m no período t
$\mathcal{L}$	Função de perda do modelo
$L(d_t, \hat{X}_t)$	Erro de previsão entre d_t e $\hat{X}_t$ segundo a métrica L
$\Omega(f_m)$	Termo de regularização aplicado à árvore m

Tabela 3.1 - Notação utilizada nos modelos de previsão

3.3.1.1. Zero Forecast

O modelo *Zero Forecast* (ZF) assume que a previsão para o período seguinte é igual a zero (Romeijnders, Teunter e Jaarsveld 2012). Este método revela-se particularmente adequado para séries temporais com elevada intermitência, nas quais predominam períodos sem procura.

A previsão é dada por:

$$\hat{X}_{t+1} = 0 \quad (3.4)$$

3.3.1.2. Naive Forecast

O modelo *Naive Forecast* (NF) utiliza como previsão para o período seguinte o valor observado no período atual (Romeijnders, Teunter e Jaarsveld 2012).

A previsão é definida da seguinte forma:

$$\hat{X}_{t+1} = d_t \quad (3.5)$$

3.3.2. Modelos Paramétricos

Os modelos paramétricos assumem que a estrutura da série temporal pode ser descrita através de um conjunto reduzido de parâmetros estimados ao longo do tempo. Em métodos de previsão baseados em suavização exponencial, estes parâmetros controlam a forma como as observações mais recentes influenciam as previsões futuras. Devido à sua simplicidade, baixo custo computacional e facilidade de implementação, os modelos paramétricos continuam a ser amplamente utilizados em problemas de previsão da procura e gestão de stocks (Hyndman e Athanasopoulos 2018).

Neste trabalho foram considerados seis modelos paramétricos: Médias Móveis, Suavização Exponencial Simples, Método de Croston, Syntetos-Boylan Approximation, Método de Teunter-Syntetos-Babai e Método de Holt.

3.3.2.1. Médias Móveis

O modelo das Médias Móveis (MM) é um dos mais famosos para lidar com séries temporais (Silver, Pyke e Thomas 2017). Neste modelo a previsão das vendas para o instante futuro é determinada pela média dos valores observados nos últimos N instantes. Embora possa ser possível a atribuição de pesos diferentes às observações (Média Móvel Ponderada), neste trabalho será utilizado um peso uniforme de $\frac{1}{N}$ a cada observação. Em que N é um parâmetro a definir, consoante as particularidades do processo a modelar.

$$\hat{X}_{t+1} = \frac{1}{N} \sum_{i=0}^N d_{t-N+i} \quad (3.6)$$

3.3.2.2. Suavização Exponencial Simples

A Suavização Exponencial Simples (SES) é um dos métodos de previsão mais utilizados em séries temporais, sendo frequentemente aplicada em problemas de previsão da procura e gestão de stocks (Silver, Pyke e Thomas 2017).

Ao contrário do método das Médias Móveis, a SES atribui maior importância às observações mais recentes através de pesos exponencialmente decrescentes ao longo do tempo. Desta forma, as previsões adaptam-se de forma mais rápida a alterações no comportamento da procura.

A previsão para o período seguinte é obtida através de uma combinação entre a previsão do período atual e o valor observado nesse período, sendo definida por:

$$\hat{X}_{t+1} = (1 - \alpha)\hat{X}_t + \alpha d_t, \quad 0 < \alpha < 1 \quad (3.7)$$

3.3.2.3. Método de Croston

O Método de Croston (CR), proposto por Croston (1972), baseia-se na decomposição da série temporal em duas componentes: a magnitude da procura positiva e o intervalo entre ocorrências de procura. Cada componente é atualizada recorrendo a suavização exponencial simples apenas quando ocorre procura positiva.

A previsão para a magnitude da procura é dada por:

$$\hat{S}_{t+1} = \begin{cases} \hat{S}_t, & \text{se } d_t = 0 \\ (1 - \alpha)\hat{S}_t + \alpha d_t, & \text{se } d_t > 0 \end{cases} \quad (3.8)$$

A previsão para o intervalo entre procuras positivas é definida por:

$$\hat{K}_{t+1} = \begin{cases} \hat{K}_t, & \text{se } d_t = 0 \\ (1 - \alpha)\hat{K}_t + \alpha k_t, & \text{se } d_t > 0 \end{cases} \quad (3.9)$$

A previsão final da procura é então obtida por:

$$\hat{X}_{t+1} = \frac{\hat{S}_{t+1}}{\hat{K}_{t+1}} \quad (3.10)$$

3.3.2.4. Método de Syntetos-Boylan Approximation

O método *Syntetos-Boylan Approximation* (SBA), proposto por Syntetos e Boylan (2005), corresponde a uma modificação do Método de Croston desenvolvida com o objetivo de corrigir o enviesamento das previsões.

O método mantém a mesma estrutura do Método de Croston, utiliza as previsões para a magnitude da procura positiva e para o intervalo entre procuras positivas anteriormente definidas nas

Equações (3.8) e (3.9). A principal diferença consiste na introdução de um fator de correção $(1 - \frac{\alpha}{2})$ na previsão final.

Assim, a previsão da procura é dada por:

$$\hat{X}_{t+1} = \left(1 - \frac{\alpha}{2}\right) \frac{\hat{S}_{t+1}}{\hat{K}_{t+1}} \quad (3.11)$$

3.3.2.5. Método de Teunter-Syntetos-Babai

O método de Teunter-Syntetos-Babai (TSB), proposto por Teunter, Syntetos e Babai (2011), corresponde a uma modificação do método de Croston desenhada para mitigar o problema da obsolescência. O TSB não se foca na correção do viés da previsão uma vez que atualiza diretamente a probabilidade de ocorrência de procura positiva em cada período, o que permite uma adaptação mais rápida do que a do método de Croston em situações de redução gradual da procura.

Tal como no Método de Croston, a atualização da magnitude da procura positiva é dada pela Equação (3.8).

A probabilidade de ocorrência de procura positiva é atualizada através de:

$$\hat{p}_{t+1} = (1 - \beta)\hat{p}_t + \beta p_t \quad (3.12)$$

A previsão final da procura é então obtida por:

$$\hat{X}_{t+1} = \hat{p}_{t+1}\hat{S}_{t+1} \quad (3.13)$$

3.3.2.6. Método de Holt

O método de Holt, também conhecido por Suavização Exponencial Dupla é uma extensão da Suavização Exponencial Simples adequada para séries temporais que apresentam tendência (Hyndman e Athanasopoulos 2018).

O modelo utiliza duas componentes: o nível da série temporal e a tendência, atualizadas recursivamente através de Suavização Exponencial Simples.

A atualização do nível da série é dada por:

$$a_t = \alpha d_t + (1 - \alpha)(a_{t-1} + b_{t-1}) \quad (3.14)$$

A atualização da tendência é definida por:

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \quad (3.15)$$

A previsão para o período seguinte é então obtida através de:

$$\hat{X}_{t+1} = a_t + b_t \quad (3.16)$$

3.3.3. Modelos Baseados em Distribuições

Adicionalmente aos modelos de previsão anteriormente apresentados, foram também considerados modelos baseados em distribuições probabilísticas. Nestes modelos assume-se que a procura pode ser descrita através de uma variável aleatória associada a uma distribuição teórica ajustada aos dados históricos.

A previsão é obtida através da utilização dos quantis da função distribuição, o que permite determinar diretamente níveis de procura associados a diferentes probabilidades de ocorrência. Tal como nos modelos paramétricos, foi considerado um parâmetro α , neste caso associado ao quantil da distribuição utilizado na previsão.

Uma vez que as distribuições consideradas neste trabalho são discretas, foi utilizada interpolação linear sobre a função de distribuição entre valores consecutivos da variável aleatória, permitindo obter previsões contínuas e reduzir os efeitos de descontinuidade associados às previsões inteiras.

Neste trabalho foram consideradas as distribuições de Poisson, Binomial Negativa, *Zero-Inflated Poisson* e *Zero-Inflated Negative Binomial*.

3.3.3.1. Distribuição de Poisson

A distribuição de Poisson é frequentemente utilizada para modelar o número de ocorrências de um determinado evento num intervalo fixo de tempo, assumindo que esses eventos ocorrem de forma aleatória e independente. No contexto deste trabalho, o evento considerado corresponde à procura observada para uma determinada referência num dado período.

Seja X_i a variável aleatória correspondente ao número de unidades procuradas da referência i .

Assume-se então que:

$$X_i \sim \text{Poisson}(\lambda_i) \quad (3.17)$$

Em que λ_i pode ser estimado através de:

$$\lambda_i = \frac{1}{N} \sum_{t=1}^N d_t \quad (3.18)$$

A função de probabilidade da distribuição de Poisson é dada por:

$$P(X_i = x) = \frac{e^{-\lambda_i} \lambda_i^x}{x!}, \quad x = 0, 1, 2, \dots \quad (3.19)$$

A previsão da procura é então obtida através do quantil associado ao nível de probabilidade definido pelo parâmetro α , isto é, através do menor valor $\hat{X}_{t+1}$ tal que:

$$P(X_i \leq \hat{X}_{t+1}) \geq 1 - \alpha \quad (3.20)$$

3.3.3.2. Distribuição Binomial Negativa

A distribuição Binomial Negativa é frequentemente utilizada para modelar variáveis discretas com sobredispersão, isto é, situações em que a variância observada é superior à média (Greene 1994). Esta característica torna a distribuição particularmente adequada para séries temporais com elevada variabilidade da procura.

Seja X_i a variável aleatória correspondente ao número de unidades procuradas da referência i .

Assume-se então que:

$$X_i \sim NB(n_i, \pi_i) \quad (3.21)$$

A média da distribuição é estimada através de (3.18) e a variância através de:

$$\sigma_i^2 = \frac{1}{N} \sum_{t=1}^N (d_t - \lambda_i)^2 \quad (3.22)$$

O parâmetro de dispersão é então obtido por:

$$\phi_i = \max\left(\frac{\sigma_i^2 - \lambda_i}{\lambda_i^2}, 0\right) \quad (3.23)$$

A partir deste parâmetro são calculados os parâmetros da distribuição:

$$\pi_i = \frac{1}{1 + \phi_i \lambda_i} \quad (3.24)$$

$$n_i = \frac{1}{\phi_i} \quad (3.25)$$

Quando $\phi_i = 0$, assume-se um valor elevado para n_i , o que aproxima a distribuição Binomial Negativa à distribuição de Poisson. A função de probabilidade da distribuição Binomial Negativa é dada por:

$$P(X_i = x) = \frac{\Gamma(x + n_i)}{\Gamma(n_i)x!} \pi_i^{n_i} (1 - \pi_i)^x, \quad x = 0, 1, 2, \dots \quad (3.26)$$

Tal como na distribuição de Poisson, a previsão da procura é obtida através de (3.20).

3.3.3.3. Zero-Inflated Poisson

A distribuição *Zero-Inflated Poisson* (ZIP) é uma extensão da distribuição de Poisson desenvolvida para modelar séries temporais com excesso de zeros relativamente ao esperado numa distribuição de Poisson tradicional (Lambert 1992). Este tipo de comportamento é frequente em problemas de procura intermitente, nos quais existe uma elevada proporção de períodos sem procura.

A distribuição ZIP assume que a procura pode resultar de dois processos distintos: um processo responsável por zeros estruturais, correspondentes a períodos em que não existe possibilidade de ocorrência de procura, e um segundo processo descrito por uma distribuição de Poisson (Lambert 1992).

Seja X_i a variável aleatória correspondente à procura da referência i . Assume-se então que:

$$X_i \sim ZIP(\lambda_i, \psi_i) \quad (3.27)$$

Tal como na distribuição de Poisson, a média da distribuição é estimada através da Equação (3.18).

O parâmetro ψ_i é estimado através da diferença entre a proporção observada de zeros e a proporção de zeros esperada pela distribuição de Poisson:

$$\mathbb{E}[N_{0,i}] = Ne^{-\lambda_i} \quad (3.28)$$

$$\psi_i = \frac{\hat{N}_{0,i} - Ne^{-\lambda_i}}{N} \quad (3.29)$$

A função de probabilidade da distribuição ZIP é dada por:

$$P(X_i = 0) = \psi_i + (1 - \psi_i)e^{-\lambda_i} \quad (3.30)$$

e, para $x > 0$:

$$P(X_i = x) = (1 - \psi_i) \frac{e^{-\lambda_i} \lambda_i^x}{x!} \quad (3.31)$$

A previsão da procura é obtida através de (3.20).

3.3.3.4. Zero-Inflated Negative Binomial

A distribuição *Zero-Inflated Negative Binomial* (ZINB) corresponde a uma extensão da distribuição Binomial Negativa para séries temporais com excesso de zeros. Tal como na distribuição ZIP, assume-se a existência de dois processos distintos: um processo responsável por zeros estruturais e um segundo processo descrito por uma distribuição Binomial Negativa (Greene 1994).

Seja X_i a variável aleatória correspondente à procura da referência i . Assume-se então que:

$$X_i \sim ZINB(n_i, \pi_i, \psi_i) \quad (3.32)$$

Tal como na distribuição Binomial Negativa, os parâmetros da distribuição são estimados através das Equações (3.23)–(3.25).

O parâmetro ψ_i é estimado através da diferença entre a proporção observada de zeros e a proporção de zeros esperada pela distribuição Binomial Negativa:

$$\mathbb{E}[N_{0,i}] = N\pi_i^{n_i} \quad (3.33)$$

$$\psi_i = \frac{\hat{N}_{0,i} - N\pi_i^{n_i}}{N} \quad (3.34)$$

A função de probabilidade da distribuição ZINB é dada por:

$$P(X_i = 0) = \psi_i + (1 - \psi_i)\pi_i^{n_i} \quad (3.35)$$

e, para $x > 0$:

$$P(X_i = x) = (1 - \psi_i) \frac{\Gamma(x + n_i)}{\Gamma(n_i)x!} \pi_i^{n_i} (1 - \pi_i)^x \quad (3.36)$$

A previsão da procura é obtida através da Equação (3.20).

3.3.4. Modelos de Machine Learning

Foram também considerados modelos de *Machine Learning* aplicados à previsão da procura, os quais se distinguem pela capacidade de identificar relações não lineares e padrões complexos diretamente a partir dos dados, sem assumir explicitamente uma estrutura estatística fixa para a série temporal. Nos últimos anos, estes modelos têm vindo a ganhar relevância em problemas de previsão da procura, particularmente em contextos caracterizados por elevada variabilidade e grande volume de dados (Ke et al. 2017). Neste trabalho foi utilizado o modelo *Light Gradient Boosting Machine* (LightGBM), desenvolvido por Ke et al. (2017), devido à sua elevada capacidade preditiva e eficiência computacional.

O LightGBM é um modelo de *Machine Learning* baseado na técnica de *Gradient Boosting* (Friedman 2001), que recorre a várias árvores de decisão construídas de forma sequencial com o objetivo de melhorar progressivamente a capacidade preditiva do modelo.

A previsão final resulta da combinação das previsões produzidas por todas as árvores construídas ao longo do processo iterativo, pode ser representada por:

$$\hat{X}_{t+1} = \sum_{m=1}^M f_m(x_t) \quad (3.37)$$

Em cada iteração, o algoritmo adiciona uma nova árvore ajustada aos erros das previsões anteriores, para minimizar uma função de perda. De forma geral, o problema consiste em minimizar:

$$\mathcal{L} = \sum_{t=1}^n L(d_t, \hat{X}_t) + \sum_{m=1}^M \Omega(f_m) \quad (3.38)$$

O LightGBM distingue-se de outros métodos de *Gradient Boosting* pela forma como as árvores são construídas. Em vez de expandir todos os nós de um determinado nível da árvore, o algoritmo utiliza uma abordagem baseada em folhas (*leaf-wise tree growth*), ou seja expande prioritariamente o nó que proporciona a maior redução da função de perda. Esta estratégia permite, obter melhores níveis de precisão com menor custo computacional (Ke et al. 2017).

3.4. Stock de Segurança e Ponto de Encomenda

Para além da previsão da procura, é igualmente necessário determinar níveis adequados de stock de segurança, de forma a proteger contra a incerteza associada à procura futura. O stock de segurança corresponde a uma quantidade adicional de inventário mantida para acomodar consumos superiores aos previstos, reduzindo assim o risco de ruturas durante o período de reposição (Axsäter 2006). Assim, assume um papel importante na gestão de stocks, uma vez que influencia ao mesmo tempo os níveis de serviço e os custos de inventário.

O ponto de encomenda é determinado através da combinação entre a procura esperada durante o período de reposição (*Lead Time*) e o stock de segurança necessário para acomodar possíveis desvios da procura real face à previsão.

O cálculo do stock de segurança encontra-se diretamente relacionado com a qualidade das previsões e com o comportamento dos respetivos erros de previsão. Como referido anteriormente, muitos modelos assumem que os erros seguem aproximadamente uma distribuição normal, recorrendo à variabilidade residual para estimar a quantidade adicional de stock necessária (Barrow e Kourentzes 2016). Contudo, diferentes abordagens podem ser utilizadas, particularmente em contextos de procura intermitente ou discreta, onde a hipótese de normalidade pode não ser adequada (Strijbosch, Heuts e Schoot 2000). No restante da secção serão apresentados diferentes métodos de cálculo do stock de segurança.

3.4.1. Stock de Segurança Dinâmico

Com base nos resultados obtidos durante a fase de avaliação dos modelos de previsão, foi também considerada uma abordagem dinâmica para o cálculo do stock de segurança. Esta metodologia parte do princípio de que a principal função do stock de segurança consiste em proteger o sistema contra situações em que a procura real excede a procura prevista, ou seja, nos períodos em que ocorre subestimação da procura por parte do modelo de previsão (Cruz et al. 2020).

Neste contexto, em vez de utilizar toda a distribuição dos erros de previsão, o cálculo do stock de segurança é efetuado apenas com base nos períodos em que a procura observada foi superior à previsão realizada. Esta abordagem permite focar especificamente nos cenários que efetivamente originam risco de rutura de stock, tornando o cálculo do stock de segurança mais parecido ao comportamento real da procura.

Assim, começa-se por definir o conjunto dos períodos em que se verifica:

$$A_t = \{i = 1, \dots, n : d_{t-i} > \hat{X}_{t-i}\} \quad (3.39)$$

O erro médio associado às situações de subestimação da procura é dado por:

$$\text{Error}_{med,t} = \frac{1}{\#A_t} \sum_{i \in A_t} (d_{t-i} - \hat{X}_{t-i}) \quad (3.40)$$

O stock de segurança é então calculado através da multiplicação deste erro médio por um fator de risco $q_{1-\alpha/2}$, que provém de uma distribuição normal padrão $N(0, 1)$ e está associado ao nível de serviço pretendido:

$$\text{SS}_t = q_{1-\alpha/2} \times \text{Error}_{med,t} \quad (3.41)$$

Este fator permite ajustar o stock de segurança em função do risco que se pretende assumir relativamente à ocorrência de ruturas de stock. Segundo Cruz et al. (2020), a parametrização deste fator pode ainda incorporar conhecimento técnico da empresa, nomeadamente a criticidade de cada referência ou produto.

Por fim, o ponto de encomenda é obtido através da combinação entre a procura prevista durante o tempo de reposição e o stock de segurança calculado:

$$OP_t = \hat{X}_t \times LT + SS_t \times LT \quad (3.42)$$

3.4.2. Stock de Segurança com Base na Distribuição Normal

Uma das abordagens mais utilizadas para o cálculo do stock de segurança assume que os erros de previsão seguem aproximadamente uma distribuição normal. Neste contexto, o stock de segurança é determinado a partir da variabilidade da procura ou dos erros de previsão, sendo ajustado através de um fator associado ao nível de serviço pretendido (Axsäter 2006; Silver, Pyke e Thomas 2017).

Assumindo que os erros de previsão são independentes e seguem uma distribuição normal com média nula e variância constante, o stock de segurança pode ser obtido através de:

$$SS_t = z_\alpha \sqrt{\sigma_t^2} \quad (3.43)$$

onde z_α representa o quantil da distribuição normal associado ao nível de serviço pretendido e σ_t^2 corresponde à variância da procura durante o período de proteção.

Quando a variabilidade é estimada a partir dos erros de previsão, é comum utilizar o desvio padrão residual ou o RMSE como medida de dispersão (Barrow e Kourentzes 2016; Barrow, Kourentzes e Sandberg 2017).

O RMSE encontra-se relacionado com a variância dos erros através da seguinte fórmula:

$$RMSE = \sqrt{\sigma^2 + \mu^2} \quad (3.44)$$

Desta forma, o RMSE incorpora simultaneamente a dispersão e o enviesamento das previsões, o que leva a uma medida mais adequada para o cálculo do stock de segurança (Barrow, Kourentzes e Sandberg 2017).

O ponto de encomenda é então definido através da soma entre a procura esperada durante o período de reposição e o stock de segurança ajustado ao período de reposição:

$$OP_t = \hat{X}_t \times LT + SS_t \times \sqrt{LT} \quad (3.45)$$

3.4.3. Stock de Segurança Baseado em Distribuições Discretas

Para além das abordagens tradicionais, foi também desenvolvida uma metodologia própria para o cálculo do stock de segurança. Esta abordagem procura representar de forma mais realista o comportamento da procura de peças sobresselentes, caracterizado frequentemente por elevados níveis de intermitência e pela presença recorrente de períodos sem consumo (Syntetos, Boylan e Croston 2005; Teunter, Syntetos e Babai 2011).

A metodologia proposta baseia-se na utilização de três distribuições probabilísticas distintas para modelar a procura durante o período de reposição: uma distribuição de Poisson parametrizada diretamente pela procura prevista, uma distribuição de Poisson parametrizada através dos erros históricos observados e uma distribuição Binomial Negativa. A seleção da distribuição é realizada dinamicamente em função das características estatísticas observadas em cada referência, incorpora ainda a previsão da procura e a variabilidade histórica associada aos erros de previsão.

Quando existe procura prevista positiva e a variância observada é aproximadamente semelhante à média, assume-se uma distribuição de Poisson tradicional. Neste caso, o parâmetro da distribuição é estimado diretamente através da procura prevista durante o período de reposição:

$$\lambda_t = \hat{X}_t \times LT \quad (3.46)$$

assumindo-se então:

$$D_t \sim \text{Poisson}(\lambda_t) \quad (3.47)$$

Contudo, em muitas referências de procura intermitente, a previsão pode assumir valores nulos ou bastante reduzidos, apesar de ocorrerem consumos ocasionais. Nestes casos, a utilização direta da distribuição de Poisson leva frequentemente a pontos de encomenda iguais a zero. Ainda, para incorporar a incerteza associada aos erros de previsão, é calculado adicionalmente um segundo ponto de encomenda através de uma distribuição de Poisson que utiliza como parâmetro o desvio-padrão dos erros de previsão, ajustado ao tempo de reposição. O valor final corresponde ao maior dos dois pontos de encomenda obtidos, garantindo-se assim um nível mínimo de proteção mesmo quando a procura prevista é muito reduzida.

De outra forma, quando se observa sobredispersão, a distribuição de Poisson tende a subestimar a probabilidade de ocorrências extremas, o que conduz a níveis insuficientes de stock de segurança. Para lidar com esta variabilidade adicional, o modelo recorre à distribuição Binomial Negativa.

A parametrização da distribuição Binomial Negativa é realizada com base nas equações (3.22)–(3.25), novamente ajustadas ao tempo de reposição.

Após a seleção e parametrização da distribuição mais adequada para cada referência, o ponto de encomenda é determinado através do quantil associado ao nível de serviço pretendido. Assim, considerando um nível de serviço β , o ponto de encomenda corresponde ao menor valor OP_t tal que:

$$P(D_t \leq OP_t) \geq \beta \quad (3.48)$$

O stock de segurança é então obtido implicitamente através da diferença entre o ponto de encomenda calculado e a procura média prevista durante o período de reposição:

$$SS_t = OP_t - \lambda_t \quad (3.49)$$

3.4.4. *Bootstrapping* de Willemain

Uma abordagem alternativa para o cálculo do ponto de encomenda consiste na utilização de métodos de *bootstrapping* para estimar diretamente a distribuição da procura durante o período de reposição. Ao contrário das abordagens paramétricas tradicionais, este tipo de metodologia não assume explicitamente uma distribuição teórica para a procura, visa antes reproduzir empiricamente o comportamento observado nas séries históricas. Neste contexto, Willemain, Smart e Schwarz (2004) propõem um método de *bootstrapping* especificamente desenvolvido para procura intermitente, capaz de estimar diretamente a distribuição da procura durante o tempo de reposição.

Esta abordagem, em vez de utilizar a previsão da procura e calcular o stock de segurança, constrói diretamente uma distribuição empírica da procura durante o tempo de reposição, a partir da qual é possível determinar o ponto de encomenda associado ao nível de serviço desejado.

Começa-se pela modelação da ocorrência de procura através de uma cadeia de Markov com dois estados, correspondentes à presença ou ausência de procura em cada período. Desta

forma, procura-se capturar a dependência temporal frequentemente observada em séries de procura intermitente, caracterizadas por sequências consecutivas de períodos com ou sem procura (Willemain, Smart e Schwarz 2004).

Após gerar a sequência de estados para o horizonte correspondente ao tempo de reposição, os períodos classificados como procura positiva recebem valores positivos obtidos através de um processo de *bootstrapping* com reposição a partir das observações históricas não nulas, ou seja:

$$\mathcal{D}^+ = \{d_i : d_i > 0\} \quad (3.50)$$

o conjunto das observações históricas positivas da procura. Para cada período em que ocorre procura, seleciona-se aleatoriamente um valor pertencente a $\mathcal{D}^+$:

$$d_t^* \sim \mathcal{D}^+ \quad (3.51)$$

De forma a evitar que a distribuição simulada reproduza apenas valores previamente observados, Willemain, Smart e Schwarz (2004) introduzem adicionalmente um mecanismo de perturbações, responsável por adicionar pequenas perturbações aleatórias aos valores amostrados. Assim, cada valor *bootstrap* é ajustado segundo:

$$d_t^{boot} = 1 + \text{INT}(d_t^* + Z\sqrt{d_t^*}) \quad (3.52)$$

onde $Z \sim N(0, 1)$ representa uma variável aleatória normal padrão e $\text{INT}(\cdot)$ corresponde à parte inteira do valor obtido.

Este processo é repetido múltiplas vezes, o que gera uma distribuição empírica da procura durante o período de reposição. O ponto de encomenda é posteriormente determinado através do quantil associado ao nível de serviço pretendido β :

$$OP_t = F^{-1}(\beta) \quad (3.53)$$

onde $F^{-1}(\beta)$ representa o quantil empírico da distribuição simulada da procura acumulada durante o *lead time*.

4. Metodologia Experimental e Estratégia de Avaliação

Neste capítulo é apresentada a metodologia experimental utilizada ao longo do trabalho, incluindo os procedimentos de divisão dos dados, otimização de hiperparâmetros, seleção dos modelos e avaliação da qualidade das previsões. É também descrita a forma como as previsões obtidas são posteriormente utilizadas nos modelos de gestão de stocks.

4.1. Enquadramento Geral

A segmentação dos produtos apresentada no capítulo anterior teve como principal objetivo confirmar as características da procura no conjunto de dados, como a existência de procura intermitente e produtos de baixa rotação. No entanto, essa segmentação não foi utilizada para definir formas distintas de previsão entre segmentos, uma vez que, após discussão com a equipa responsável pela realização do trabalho, optou-se por utilizar um modelo que comparasse os diversos métodos disponíveis e escolhesse, para cada referência, o que apresenta melhor desempenho na previsão segundo determinadas métricas.

De forma a avaliar a precisão das previsões, é importante ter em conta os erros das mesmas, e como elas afetam os modelos de gestão de stocks. Para cada referência, os modelos são avaliados segundo três métricas distintas: *Mean Squared Error* (MSE), *Weighted Sum of Squared Error* (WSSE) e *Mean Absolute Error* (MAE). Como consequência, uma mesma referência pode ter três modelos selecionados diferentes, dependendo da métrica utilizada como critério de minimização do erro.

As previsões obtidas e os respetivos erros são posteriormente utilizados no cálculo do stock de segurança e do ponto de encomenda segundo as diferentes metodologias consideradas. Adicionalmente, é ainda considerado um modelo baseado em *Bootstrap*, descrito na Secção 3.4.4. Totalizando assim dez metodologias distintas para o cálculo do stock de segurança e do ponto de encomenda. Assim, para além de avaliar a qualidade preditiva dos modelos, torna-se também possível analisar o impacto das diferentes abordagens no desempenho operacional da gestão de stocks, nomeadamente na qualidade dos níveis de stock de segurança e na capacidade de resposta do sistema de reposição.

4.2. Estratégia Experimental e Divisão dos Dados

Os dados, agora agregados mensalmente, foram divididos cronologicamente em diferentes janelas temporais, de forma a preservar a estrutura temporal das séries e evitar a fuga de informação,

isto é, a utilização de informação futura durante o processo de validação e avaliação, que conduz a estimativas excessivamente otimistas do seu desempenho (Hyndman e Athanasopoulos 2018).

A divisão dos dados adotada consiste em cinco etapas principais: inicialização dos modelos, otimização de hiperparâmetros, seleção dos modelos, simulação para validação das políticas de stocks e validação futura dos resultados obtidos.

Para a primeira fase foi considerado um período inicial de 3 meses de dados, utilizados para inicializar os modelos e evitar instabilidade e para reduzir futuros erros associados aos valores iniciais. De seguida, foram considerados 12 meses de dados para validação, período no qual é realizada a otimização dos hiperparâmetros através uma pesquisa em grelha e de validação cruzada temporal.

Após a otimização dos hiperparâmetros, os modelos são avaliados num período de teste de 6 meses, utilizado exclusivamente para realizar a seleção de modelos e identificar, para cada referência, o modelo com melhor desempenho preditivo segundo cada métrica considerada.

Uma vez selecionado o modelo final para cada referência, são considerados 12 meses adicionais de dados para simular o comportamento das diferentes metodologias de stock de segurança.

Finalmente, são ainda utilizados 2 meses adicionais de dados para previsão futura e para testes extra na passagem dos dados para o SQL Server.

A Figura 4.1 apresenta de forma resumida a estrutura temporal considerada, que corresponde a um horizonte total de 35 meses.

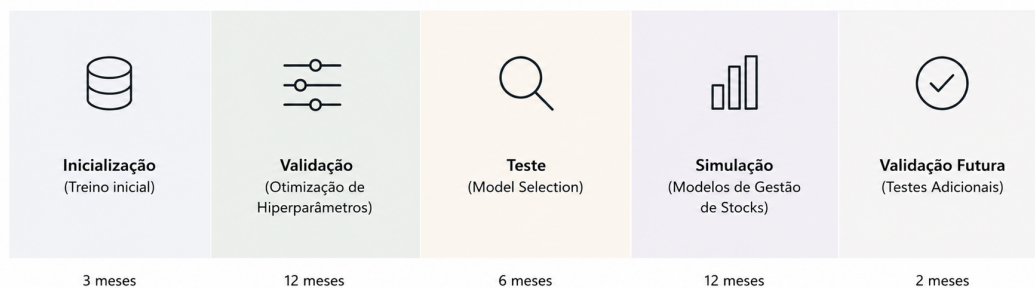


Figura 4.1 - Estrutura temporal utilizada na estratégia experimental.

4.3. Otimização de Hiperparâmetros

A otimização dos hiperparâmetros dos modelos foi realizada através de um procedimento de pesquisa em grelha. Para os hiperparâmetros contínuos, uma vez que se encontravam entre 0 e 1, foram considerados incrementos de 0.01, sendo avaliadas múltiplas combinações de parâmetros de forma automática, para o modelo de LightGBM foi utilizada a biblioteca Optuna do Python 3.11 2022, uma vez que neste modelo podem ser considerados bastantes parâmetros diferentes.

Para cada combinação de hiperparâmetros considerada uma estratégia de *Rolling Window Cross-Validation* em que o modelo é inicialmente treinado com os dados de treino. De seguida, na fase de validação, é realizada uma previsão *one-step-ahead* para a observação imediatamente a seguir, sendo calculado o respetivo erro de previsão segundo o MSE.

Após cada previsão, é considerado um novo mês do conjunto de validação, sendo a janela temporal deslocada ao longo do horizonte de validação. O processo é repetido iterativamente ao longo de todo o conjunto de validação, sendo no final calculado o valor do MSE para cada combinação de hiperparâmetros considerada. A combinação que minimiza o MSE obtido no conjunto de validação é selecionada como configuração final do modelo.

A utilização exclusiva do MSE durante a fase de otimização dos hiperparâmetros deveu-se ao facto de esta métrica apresentar uma maior estabilidade matemática e uma elevada sensibilidade a erros de maior magnitude. Como o principal objetivo desta etapa é identificar a combinação de hiperparâmetros mais adequada para cada modelo, a utilização de uma única métrica permite tornar o processo de otimização computacionalmente mais eficiente. Além disso, por penalizar de forma mais acentuada previsões com erros elevados, o MSE facilita a distinção entre diferentes combinações de hiperparâmetros ao longo do processo de *Grid Search*.

De forma simplificada, o procedimento de *Rolling Window Cross-Validation* pode ser representado da seguinte forma (Hyndman e Athanasopoulos 2018):

$$\mathcal{D}_{\text{train}}^{(1)} = \{d_1, d_2, \dots, d_t\} \quad (4.1)$$

$$\hat{X}_{t+1} = f(d_1, d_2, \dots, d_t) \quad (4.2)$$

$$L(d_{t+1}, \hat{X}_{t+1}) = (d_{t+1} - \hat{X}_{t+1})^2 \quad (4.3)$$

Após realizar a previsão para o período seguinte, a nova observação é incorporada no conjunto de treino:

$$\mathcal{D}_{\text{train}}^{(2)} = \{d_2, \dots, d_t, d_{t+1}\} \quad (4.4)$$

De seguida, é realizada uma nova previsão *one-step-ahead*:

$$\hat{X}_{t+2} = f(d_2, \dots, d_t, d_{t+1}) \quad (4.5)$$

$$L(d_{t+2}, \hat{X}_{t+2}) = (d_{t+2} - \hat{X}_{t+2})^2 \quad (4.6)$$

O processo é repetido iterativamente ao longo de toda a janela de validação, sendo no final calculada a média dos erros obtidos:

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^N L(d_t, \hat{X}_t) \quad (4.7)$$

Os hiperparâmetros selecionados são posteriormente utilizados na fase de teste, onde é realizada a *model selection*.

4.4. Seleção dos Modelos

Após a otimização dos hiperparâmetros, procedeu-se à comparação dos diferentes modelos de previsão no conjunto de teste. Este passo tem como objetivo identificar, para cada referência, o modelo com melhor capacidade preditiva.

As previsões foram realizadas através de uma estratégia de *rolling window* com uma janela temporal de 12 meses, sendo realizada uma previsão *one-step-ahead* para o período seguinte. Este procedimento, para cada referência é repetido ao longo do conjunto de teste para todos os modelos considerados.

Após as previsões, são calculadas as diferentes métricas de erro para cada modelo e para cada referência, podendo assim existir modelos distintos para a mesma referência dependendo da métrica utilizada.

A utilização do MSE como métrica de seleção segue a mesma motivação apresentada na secção anterior. Já o WSSE foi incluído por permitir atribuir diferentes pesos aos erros de previsão, o que torna a avaliação mais sensível a períodos mais recentes, abordagem inspirada em Cruz et al. (2020).

Por outro lado, o MAE foi considerado devido à sua interpretação direta nas unidades originais da variável analisada. Esta métrica mede o erro médio absoluto das previsões e apresenta menor

sensibilidade a erros extremos quando com o MSE. Como consequência, a seleção baseada no MAE tende a privilegiar modelos que minimizam o erro médio global.

4.5. Simulação e Validação dos Modelos de Gestão de Stocks

Após a seleção dos modelos de previsão para cada uma das métricas consideradas, realizou-se uma simulação para validar o cálculo do stock de segurança e do ponto de encomenda segundo as metodologias apresentadas na Secção 3.4. Para cada uma das três métricas de previsão selecionadas, foram calculados o stock de segurança dinâmico, o stock de segurança com base na distribuição normal e o stock de segurança baseado em distribuições discretas.

Foi também considerado o modelo de *Bootstrapping* de Willemain apresentado na Secção 3.4.4. No total, foram assim avaliadas dez metodologias distintas de gestão de stocks para cada referência.

Inicialmente, para cada metodologia, foram calculados o stock de segurança e o ponto de encomenda com base no nível de stock efetivamente observado na empresa no instante inicial da análise. A partir desses valores iniciou-se uma simulação com duração de 12 meses, com o objetivo de avaliar o comportamento operacional de cada metodologia em condições semelhantes ao que acontece na realidade.

De uma forma geral a simulação realiza uma previsão da procura para o período seguinte através do modelo selecionado para cada referência. Com base nessa previsão, no stock disponível nesse instante e no ponto de encomenda calculado, verifica-se se existe a necessidade de realizar novas encomendas. Sempre que o stock disponível se encontra abaixo do ponto de encomenda, é gerada uma nova compra correspondente à quantidade necessária para repor o nível de stock definido pela metodologia considerada.

Nesta simulação, assume-se que as encomendas chegam imediatamente após serem realizadas. Apesar de esta hipótese não refletir totalmente o comportamento real, a sua utilização justifica-se pelo facto de não existir informação relativa às encomendas que já se encontravam em trânsito no momento inicial da simulação.

As previsões são realizadas apenas no final de cada mês e permanecem fixas ao longo de todo o mês seguinte. Durante esse período, o stock é atualizado diariamente através da subtração das vendas observadas. Sempre que o stock disponível não é suficiente para satisfazer a procura observada, considera-se a existência de vendas perdidas. Ao contrário da realidade empresarial, nesta simulação não são realizadas compras de emergência nem ajustes extraordinários ao

sistema de reposição.

No final de cada mês, o stock remanescente resultante da simulação diária passa a constituir o stock inicial do mês seguinte. A partir desse valor são novamente realizadas previsões, calculados os pontos de encomenda e determinadas as quantidades a encomendar segundo cada metodologia considerada.

Este procedimento permite simular o comportamento que o sistema de gestão de stocks da empresa apresentaria caso cada uma das metodologias avaliadas fosse implementada na totalidade, ao considerar um processo de reposição mensal sem compras de emergência e com atualização contínua do stock ao longo do tempo.

4.6. Indicadores de Performance

A avaliação das metodologias consideradas ao longo do trabalho não depende apenas da qualidade das previsões. Como um dos objetivos consiste na utilização deste modelo numa situação real, torna-se igualmente importante analisar a capacidade das diferentes abordagens para evitar ruturas de stock e o impacto global que estas têm nos níveis de serviço e no nível do inventário da empresa.

Assim, foram definidos diferentes grupos de Indicadores Globais de Desempenho ou *Key Performance Indicators* (KPIs), organizados de acordo com o objetivo da análise. Dada a especificidade deste trabalho, utilizou-se a seguinte divisão: performance das previsões, performance das políticas de stock de segurança e performance global do sistema de gestão de stocks.

4.6.1. Performance das Previsões

Para avaliar a qualidade das previsões produzidas pelos diferentes modelos, em cada métrica, foram utilizados os seguintes indicadores:

- **KPI 1:** Percentagem de referências com previsão correta.

$$KPI_1 = \frac{\sum_{t=1}^N \mathbb{I}(\hat{X}_t = d_t)}{N} \times 100 \quad (4.8)$$

onde N representa o número total de referências e $\mathbb{I}(\cdot)$ é a função indicatriz definida por:

$$\mathbb{I}(\hat{X}_t = d_t) = \begin{cases} 1, & \text{se } \hat{X}_t = d_t, \\ 0, & \text{caso contrário.} \end{cases}$$

- **KPI 2:** Percentagem de referências com previsão superior às vendas observadas.

$$KPI_2 = \frac{\sum_{t=1}^N \mathbb{I}(\hat{X}_t > d_t)}{N} \times 100 \quad (4.9)$$

onde:

$$\mathbb{I}(\hat{X}_t > d_t) = \begin{cases} 1, & \text{se } \hat{X}_t > d_t, \\ 0, & \text{caso contrário.} \end{cases}$$

- **KPI 3:** Percentagem de referências com previsão inferior às vendas observadas.

$$KPI_3 = \frac{\sum_{t=1}^N \mathbb{I}(\hat{X}_t < d_t)}{N} \times 100 \quad (4.10)$$

onde:

$$\mathbb{I}(\hat{X}_t < d_t) = \begin{cases} 1, & \text{se } \hat{X}_t < d_t, \\ 0, & \text{caso contrário.} \end{cases}$$

Os KPIs são calculados em cada instante de previsão e, no final, é apresentada a média dos valores obtidos ao longo de todo o período analisado. Os indicadores foram avaliados tanto ao nível do número de referências como a nível do valor monetário associado às vendas, através da razão entre o valor correspondente a cada KPI analisado e o valor total das vendas observadas.

4.6.2. Performance das Políticas de Stock de Segurança

Para avaliar se os níveis de stock de segurança calculados pelas diferentes metodologias são suficientes para satisfazer a procura ao longo de cada instante de previsão, foram considerados os seguintes indicadores:

- **KPI 4:** Percentagem de referências em que a procura observada é satisfeita pela soma da procura prevista com o stock de segurança.

$$KPI_4 = \frac{\sum_{t=1}^N \mathbb{I}(\hat{X}_t + SS_t \geq d_t)}{N} \times 100 \quad (4.11)$$

- **KPI 4.1:** Percentagem de peças cuja procura é satisfeita pela soma da procura prevista com o stock de segurança.

$$KPI_{4.1} = \frac{\sum_{t=1}^N d_t \cdot \mathbb{I}(\hat{X}_t + SS_t \geq d_t)}{\sum_{t=1}^N d_t} \times 100 \quad (4.12)$$

- **KPI 4.2:** Percentagem do valor monetário das vendas satisfeitas pela soma da procura prevista com o stock de segurança.

$$KPI_{4.2} = \frac{\sum_{t=1}^N c_t \cdot d_t \cdot \mathbb{I}(\hat{X}_t + SS_t \geq d_t)}{\sum_{t=1}^N c_t \cdot d_t} \times 100 \quad (4.13)$$

onde c_t representa o custo unitário associado à referência no período t .

- **KPI 5:** Percentagem de referências em situação de rutura.

$$KPI_5 = 100 - KPI_4 \quad (4.14)$$

- **KPI 5.1:** Percentagem de peças em situação de rutura.

$$KPI_{5.1} = 100 - KPI_{4.1} \quad (4.15)$$

- **KPI 5.2:** Percentagem do valor monetário das vendas em situação de rutura.

$$KPI_{5.2} = 100 - KPI_{4.2} \quad (4.16)$$

- **KPI 6:** Número médio de dias necessários para escoar o excesso de stock das compras realizadas.

Para cada instante temporal, considera-se inicialmente o excesso de stock correspondente à diferença entre as compras realizadas e as vendas observadas no mês seguinte:

$$Excesso_t = \max(0, Q_t - d_{t+1}) \quad (4.17)$$

onde Q_t representa a quantidade comprada no período t .

Após determinar o excesso de stock, é calculado o número de dias necessários para que esse excesso seja totalmente absorvido pelas vendas futuras. O KPI corresponde assim ao número médio de dias necessários para escoar o excesso de stock acumulado:

$$KPI_6 = \frac{\sum_{t=1}^N Dias(Excesso_t)}{N} \quad (4.18)$$

Os KPIs são calculados em cada instante temporal da simulação e, no final, é apresentada a média dos valores obtidos ao longo de todo o horizonte analisado.

4.6.3. Performance Global do Sistema

Para avaliar o impacto global das diferentes metodologias de gestão de stocks no desempenho operacional da empresa foram considerados os seguintes indicadores.

- **KPI 9:** Percentagem de vendas perdidas devido à inexistência de stock disponível.

$$KPI_9 = \frac{\sum_{t=1}^N \max(0, d_t - S_t)}{\sum_{t=1}^N d_t} \times 100 \quad (4.19)$$

onde S_t representa o stock disponível no período t .

- **KPI 10:** Percentagem de stock excedente face às vendas observadas.

O KPI é então calculado através da razão entre o excesso total acumulado e o total de vendas observadas:

$$KPI_{10} = \frac{\sum_{t=1}^N Excesso_t}{\sum_{t=1}^N d_t} \times 100 \quad (4.20)$$

- **KPI 11:** Percentagem de variação do stock ao longo do horizonte temporal de análise, medida através da variação percentual do stock total entre o início e o final do período considerado.

$$KPI_{11} = \frac{S_{t-N} - S_t}{S_{t-N}} \times 100 \quad (4.21)$$

- **KPI 12:** Rotação do stock (*Inventory Turnover*).

Este indicador mede a relação entre as vendas realizadas e o stock médio disponível ao longo do período considerado:

$$KPI_{12} = \frac{d_t}{\bar{S}_t} \quad (4.22)$$

onde $\bar{S}_t$ representa o stock médio disponível no período t .

Os indicadores 9 e 10 foram avaliados tanto ao nível do número de referências como ao nível do valor monetário associado às vendas, através da razão entre o valor correspondente a cada KPI analisado e o valor total das vendas observadas. Já o KPI 11 avalia a variação percentual entre

o stock inicial e o stock final do período de simulação, enquanto o KPI 12 mede a relação entre as vendas totais observadas e o stock médio disponível ao longo do período considerado.

Importa salientar que a interpretação do indicador 11 difere dos restantes indicadores analisados. Neste caso, valores positivos correspondem a uma diminuição do stock ao longo do horizonte temporal considerado, o que reflete uma menor quantidade de capital imobilizado em inventário. Por outro lado, valores negativos indicam um aumento do stock total.

5. Resultados e Discussão

Neste capítulo são apresentados e analisados os principais resultados obtidos ao longo do trabalho. Na primeira secção, são discutidos os resultados da segmentação dos produtos. De seguida, é avaliado o desempenho dos modelos de previsão para cada métrica. São também analisados os resultados das diferentes políticas de stock de segurança nos níveis de serviço e no investimento em inventário, tendo por base os indicadores de performance apresentados no capítulo anterior.

5.1. Resultados da Segmentação dos Produtos

5.1.1. Segmentação por Rotação dos Produtos

Tal como referido no Capítulo 3, a segmentação dos produtos foi realizada com o objetivo de distinguir diferentes padrões de procura.

Inicialmente, as referências foram classificadas de acordo com o seu nível de rotação, utilizou-se a proporção de períodos sem procura ao longo do horizonte temporal analisado. Esta abordagem permitiu segmentar os produtos em três grupos principais: alta rotação, média rotação e baixa rotação.

Os produtos de alta rotação caracterizam-se por apresentarem vendas frequentes e menos de 30% de períodos sem procura, enquanto os produtos de baixa rotação apresentam procura intermitente e mais de 70% de períodos sem vendas ao longo do horizonte temporal analisado. Entre estes dois extremos encontram-se os produtos de média rotação, que evidenciam um comportamento intermédio, apresentando entre 30% e 70% de períodos sem procura.

A Tabela 5.1 apresenta a distribuição das referências pelas diferentes categorias de rotação, tanto em número absoluto como em percentagem relativamente ao total de referências analisadas.

Categoria	Número de Referências	Percentagem (%)
Alta Rotação	765	4.92
Média Rotação	2379	15.31
Baixa Rotação	12398	79.77

Tabela 5.1 - Distribuição das referências segundo a classificação por rotação

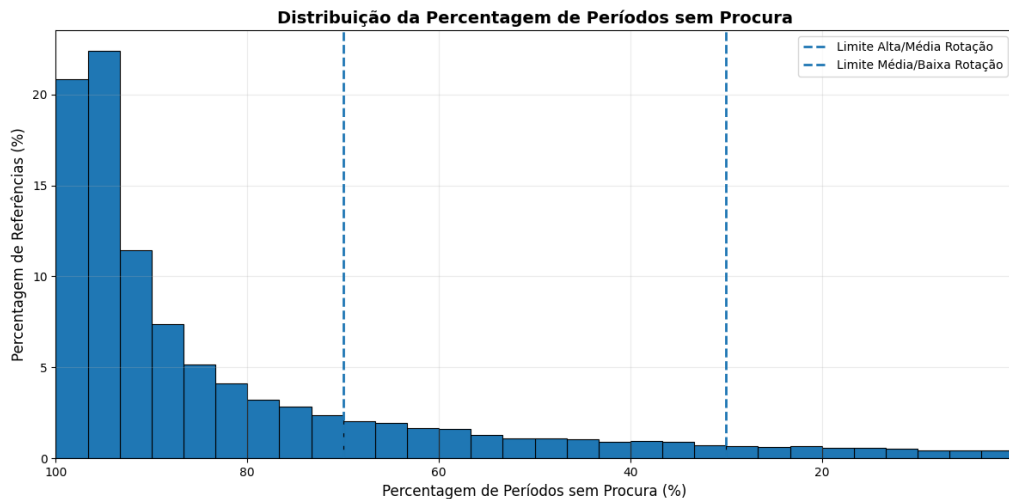


Figura 5.1 - Distribuição da percentagem de períodos sem procura nas referências analisadas

Os resultados mostram que existe uma forte presença de produtos de baixa rotação no conjunto de referências analisadas. Como pode ser visto na Tabela 5.1, cerca de 79.8% das referências apresentam mais de 70% de períodos sem procura, enquanto que apenas aproximadamente 4.9% correspondem a produtos de alta rotação. Isto significa que a maioria dos produtos passa longos períodos sem vendas, algo que vai ao encontro das características associadas ao setor agrícola onde os principais clientes desenvolvem a sua atividade predominantemente.

A Figura 5.1 permite observar de forma mais detalhada a distribuição da percentagem de períodos sem procura. Verifica-se uma distribuição fortemente assimétrica, caracterizada por uma cauda longa e pesada associada a valores elevados de períodos com vendas. Pode-se também ver que mais de 50% das referências apresentam mais de 90% dos períodos sem vendas. Isto pode traduzir-se em peças que ficam mais de um ano sem serem vendidas, originando investimento que poderá ficar parado por igual período de tempo

Apesar de representarem apenas uma pequena parte do total de referências, os produtos de alta rotação são aqueles que representam a maior proporção da receita mensal. A análise da receita total por categoria mostra que aproximadamente 5% das referências são responsáveis por cerca de 38% da receita total ao longo de todo o período em análise. Já os produtos de média rotação, que representam aproximadamente 15% das referências, contribuem com cerca de 30% da receita. Estes resultados mostram que uma pequena parte dos produtos concentra uma grande parte da receita da empresa, num comportamento que se assemelha ao princípio de Pareto.

Embora a maioria das referências apresente uma baixa procura, uma parte significativa da receita está concentrada em produtos com procura mais frequente e relativamente mais estável. Por esse motivo, faz sentido utilizar métodos tradicionais de previsão, como Médias Móveis e Suavização Exponencial Simples. Ao mesmo tempo, a elevada quantidade de referências com longos períodos sem procura justifica também a utilização de métodos específicos para procura intermitente, como Croston, SBA, TSB.

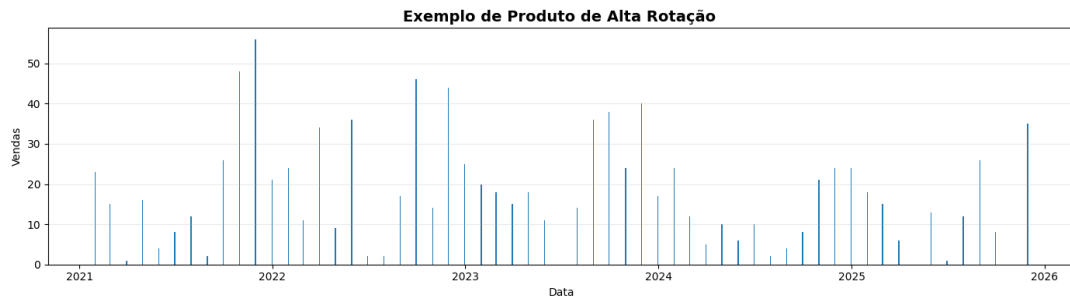


Figura 5.2 - Exemplo de série temporal de um produto de alta rotação

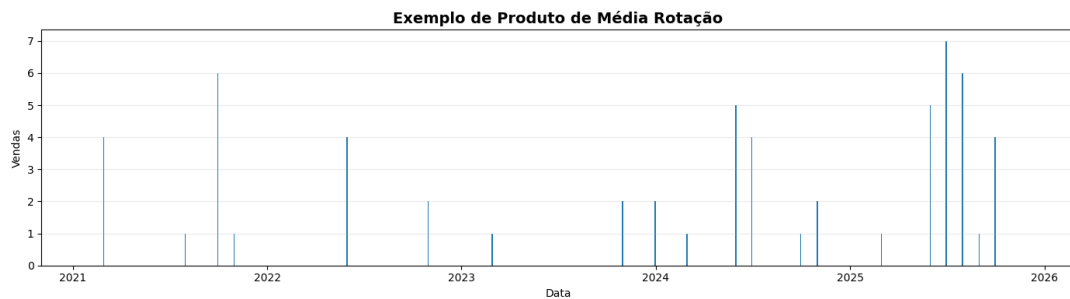


Figura 5.3 - Exemplo de série temporal de um produto de média rotação

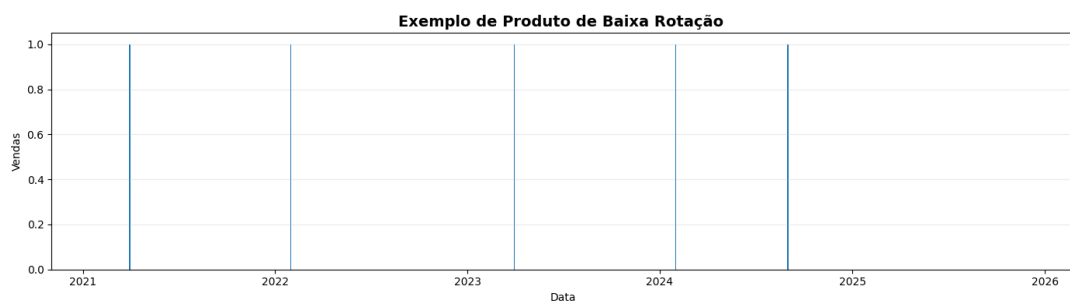


Figura 5.4 - Exemplo de série temporal de um produto de baixa rotação

5.1.2. Segmentação por Classificação SBC

A classificação SBC permite complementar a segmentação por rotação e introduz também uma caracterização mais detalhada do comportamento da procura das diferentes referências. Enquanto a segmentação por rotação considera apenas a proporção de períodos sem procura, a classificação SBC incorpora simultaneamente o intervalo médio entre procuras positivas (ADI) e a variabilidade da procura observada através do coeficiente de variação quadrático (CV^2).

Produtos classificados como suave apresentam procura frequente e reduzida variabilidade, correspondem a valores baixos de ADI e CV^2 . Os produtos classificados como intermitente caracterizam-se por longos intervalos sem procura, mas com quantidades relativamente estáveis quando ocorre procura, apresentam valores elevados de ADI e baixos de CV^2 . Já os produtos classificados como errático apresentam vendas frequentes, mas com elevada variabilidade nas quantidades procuradas, que resulta em baixos valores de ADI e elevados de CV^2 . Por fim, os produtos irregular apresentam simultaneamente longos períodos sem procura e elevada variabilidade nos consumos observados.

A Tabela 5.2 apresenta a distribuição das referências segundo cada categoria SBC, a Figura 5.5 representa dispersão das referências no plano definido pelo ADI e CV^2 , o que permite visualizar de forma mais clara a separação entre os diferentes padrões de procura identificados.

Categoria SBC	Número de Referências	Percentagem (%)
Suave (<i>Smooth</i>)	167	1.07
Intermitente (<i>Intermittent</i>)	6634	42.68
Errático (<i>Erratic</i>)	430	2.77
Irregular (<i>Lumpy</i>)	8311	53.47

Tabela 5.2 - Distribuição das referências segundo a classificação SBC

Os resultados da classificação SBC mostram quase a totalidade de produtos classificados como irregular e intermitente. Como pode ser observado na Tabela 5.2, aproximadamente 53.5% das referências pertencem à categoria irregular, enquanto cerca de 42.7% são classificadas como intermitente. Em contraste, os produtos suave e errático representam apenas uma pequena fração do total de referências analisadas.

A predominância de produtos classificados como irregular e intermitente ajuda a justificar a necessidade de utilizar modelos especificamente desenvolvidos para procura intermitente, tal como

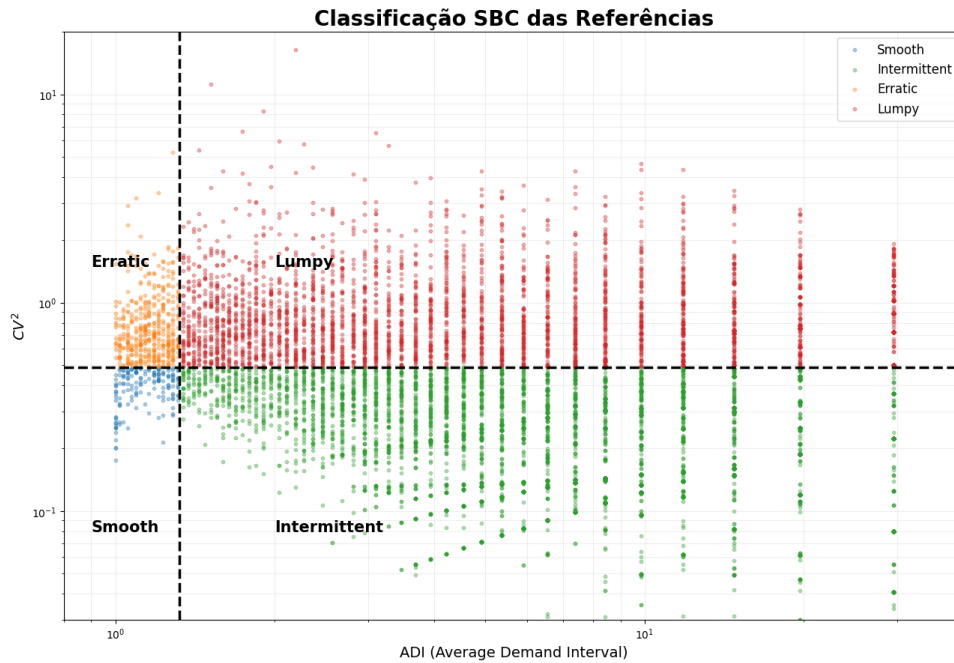


Figura 5.5 - Distribuição das referências segundo a classificação SBC no plano ADI- CV^2

foi discutido nos capítulos anteriores. No caso dos produtos irregular, a dificuldade não está apenas associada aos longos períodos sem procura, mas também à elevada variabilidade nas quantidades vendidas quando ocorre consumo. Neste tipo de séries temporais, métodos tradicionais como Médias Móveis ou Suavização Exponencial Simples tendem a apresentar maiores dificuldades, uma vez que a informação histórica contém poucos padrões consistentes e uma elevada quantidade de zeros consecutivos.

Já os produtos intermitente, apesar de apresentarem igualmente muitos períodos sem vendas, tendem a possuir quantidades procuradas relativamente mais estáveis quando ocorre consumo. Este comportamento favorece parcialmente modelos como Croston, SBA e TSB, que procuram separar a modelação dos intervalos entre procuras da modelação das quantidades procuradas. A elevada concentração de referências nestas duas categorias mostra que grande parte do problema estudado neste trabalho, como era de esperar, está diretamente associado à previsão de procura intermitente, e não apenas à previsão tradicional de séries temporais.

Por outro lado, os produtos suave, embora pouco em número, correspondem a séries temporais bastante mais regulares, com vendas frequentes e pouca variabilidade. Nestes casos os modelos tradicionais baseados em suavização tendem a funcionar de forma mais eficaz, uma vez que conseguem capturar melhor a dinâmica temporal da procura. Já os produtos errático representam um comportamento intermédio, no qual existe frequência relativamente elevada de vendas,

mas com elevada instabilidade nas quantidades procuradas, o que torna a previsão mais sensível a flutuações abruptas no histórico. Esta heterogeneidade entre padrões de procura reforça a motivação para utilizar diferentes modelos ao longo deste trabalho.

A análise da receita total por categoria SBC revela igualmente resultados bastante relevantes. Apesar de os produtos suave representarem uma pequena parte do número total de referências, estes contribuem com aproximadamente 14% da receita total observada. Os produtos intermitente representam cerca de 38% da receita, tornando-se a categoria com maior impacto financeiro para a empresa. Por outro lado, os produtos irregular, apesar de representarem mais de metade das referências analisadas, contribuem com aproximadamente 30% da receita total. Por fim, os produtos errático, apesar de corresponderem a apenas 2.77% das referências, contribuem com aproximadamente 18% da receita total, o que reflete a elevada concentração de valor nesta categoria, associada a vendas frequentes de produtos de valor unitário mais elevado.

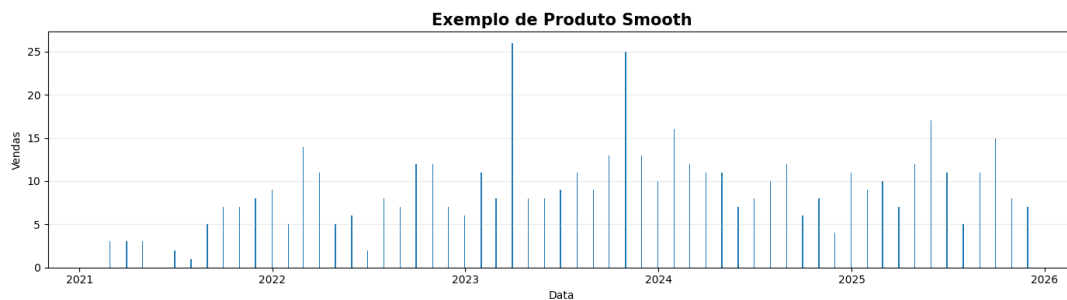


Figura 5.6 - Exemplo de série temporal classificada como suave

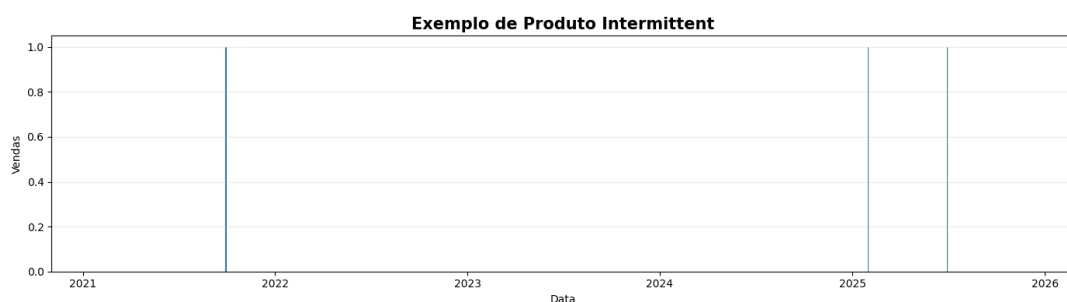


Figura 5.7 - Exemplo de série temporal classificada como intermitente

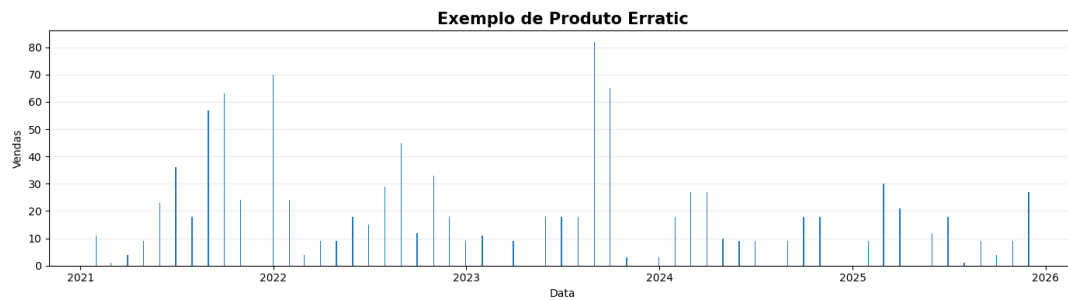


Figura 5.8 - Exemplo de série temporal classificada como errático

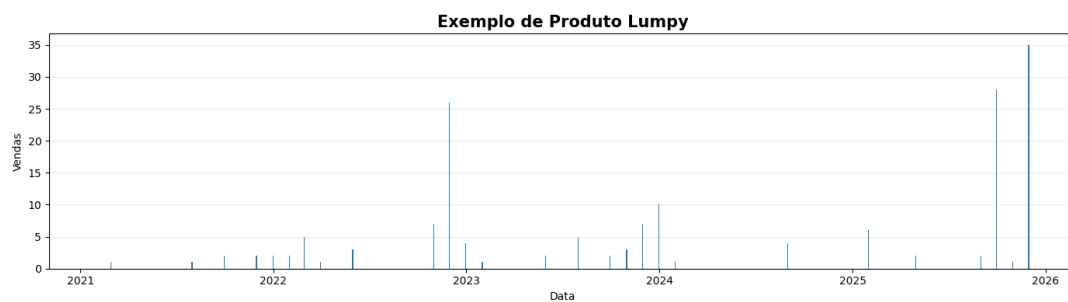


Figura 5.9 - Exemplo de série temporal classificada como irregular

5.2. Análise dos Modelos

De forma a garantir comparabilidade entre as diferentes abordagens, no cálculo dos níveis do stock de segurança foi utilizado o valor de 0.05 para o parâmetro α da suavização.

5.2.1. Análise por Métrica de Seleção

Nesta secção são apresentados e analisados os resultados obtidos através do processo de seleção de modelos descrito no capítulo anterior. Para cada referência, foram avaliados diferentes métodos de previsão, sendo selecionado o modelo com melhor desempenho segundo as métricas de erro consideradas.

Em particular, serão analisados os resultados obtidos para os indicadores KPI 1, KPI 2 e KPI 3, que permite avaliar a qualidade das previsões.

5.2.1.1. Análise dos Modelos selecionados pelo MSE

A Tabela 5.3 e 5.4 apresenta a distribuição dos modelos selecionados ao minimizar o MSE e os KPI indicados.

Método	Número de Referências	Percentagem (%)
Zero	11211	72.13
Negative Binomial	692	4.45
SBA	687	4.42
Croston	650	4.18
LightGBM	456	2.93
Poisson	428	2.75
Exponential Smoothing	326	2.10
Moving Averages	271	1.74
Naive	205	1.32
TSB	183	1.18
Holt	162	1.04
Zero-Inflated Poisson	138	0.89
Zero-Inflated Negative Binomial	133	0.86

Tabela 5.3 - Métrica MSE: Distribuição dos métodos selecionados

KPI	Valor Médio (%)	Peso no Custo Total (%)
KPI 1	71	10.23
KPI 2	19	10.98
KPI 3	10	78.79

Tabela 5.4 - Métrica MSE: Resultados médios referentes aos KPI1, 2 e 3 durante os 12 meses de período simulado

Os resultados obtidos mostram que o modelo *Zero Forecast* foi responsável por mais de 72% das escolhas realizadas ao minimizar o MSE. Como observado, a maioria das referências apresenta longos períodos sem procura, sendo por isso expectável que modelos conservadores consigam minimizar o erro quadrático médio ao prever repetidamente zero.

Este resultado mostra também uma limitação do MSE como critério de seleção. Uma vez que esta métrica penaliza fortemente erros de grande magnitude, modelos que evitam previsões elevadas tendem a ser favorecidos. Assim, prever zero em séries fortemente intermitentes acaba frequentemente por produzir menor erro médio do que tentar antecipar ocorrências raras de procura elevada.

Observa-se igualmente uma presença relevante de métodos especificamente desenvolvidos para procura intermitente, como SBA, Croston.

Os modelos zero-inflated apresentaram uma frequência de seleção relativamente reduzida. Embora teoricamente adequados para séries com elevada quantidade de zeros, o seu desempenho poderá ter sido limitado pela elevada irregularidade simultânea das quantidades procuradas.

Os modelos de Exponential Smoothing, Holt e Moving Averages, apresentaram uma baixa frequência de seleção global. Ainda assim, estes métodos continuam a revelar utilidade em séries

mais regulares.

Relativamente aos KPI analisados, os resultados mostram que aproximadamente 71% das previsões realizadas coincidem com o comportamento observado, traduzindo-se num custo associado de apenas cerca de 10.23% do valor total analisado. Isto indica que, embora exista uma elevada taxa de acerto global, grande parte dessas previsões corretas está associada a referências de reduzido impacto económico, frequentemente caracterizadas por ausência de procura.

Por outro lado, observa-se que apenas cerca de 10% das previsões ficaram abaixo da procura observada, mas estes casos representam aproximadamente 80% do valor total. Este comportamento encontra-se diretamente relacionado com a própria natureza dos modelos utilizados. Em muitos casos, os métodos acabam por prever valores próximos da média histórica ou mesmo ausência de procura, o que reduz o erro médio global mas falha em prever eventos pontuais de maior magnitude. Como foi anteriormente observado através da classificação SBC, existe uma elevada imprevisibilidade simultaneamente ao nível do momento da procura e da quantidade procurada. Assim, torna-se fundamental a utilização de stock de segurança, uma vez que a previsão dificilmente consegue antecipar todos os eventos de procura sem introduzir níveis excessivos de erro ou excesso de stock.

A Figura 5.10 mostra que a distribuição dos modelos seleccionados varia significativamente consoante a categoria SBC considerada. Nos produtos classificados como intermitente e irregular, observa-se uma forte predominância do modelo *Zero Forecast*, que representa aproximadamente 74% e 76% das escolhas, respetivamente. Este comportamento encontra-se de acordo com as características destas séries temporais.

Por outro lado, os produtos classificados como suave e errático apresentam um comportamento bastante distinto. Nestas categorias, o modelo mais frequentemente seleccionado foi o SBA, seguido pelo LightGBM e por métodos tradicionais como Moving Averages, Holt e Exponential Smoothing. Este resultado é particularmente interessante, uma vez que demonstra que, mesmo em séries relativamente regulares, métodos desenvolvidos para procura intermitente continuam ter bons desempenhos. A escolha do LightGBM nestas categorias sugere também que modelos de *machine learning* conseguem beneficiar da maior estabilidade temporal presente nestas séries, e conseguem identificar padrões mais consistentes.

De forma geral, os resultados apresentados mostram a forte relação entre o padrão de procura

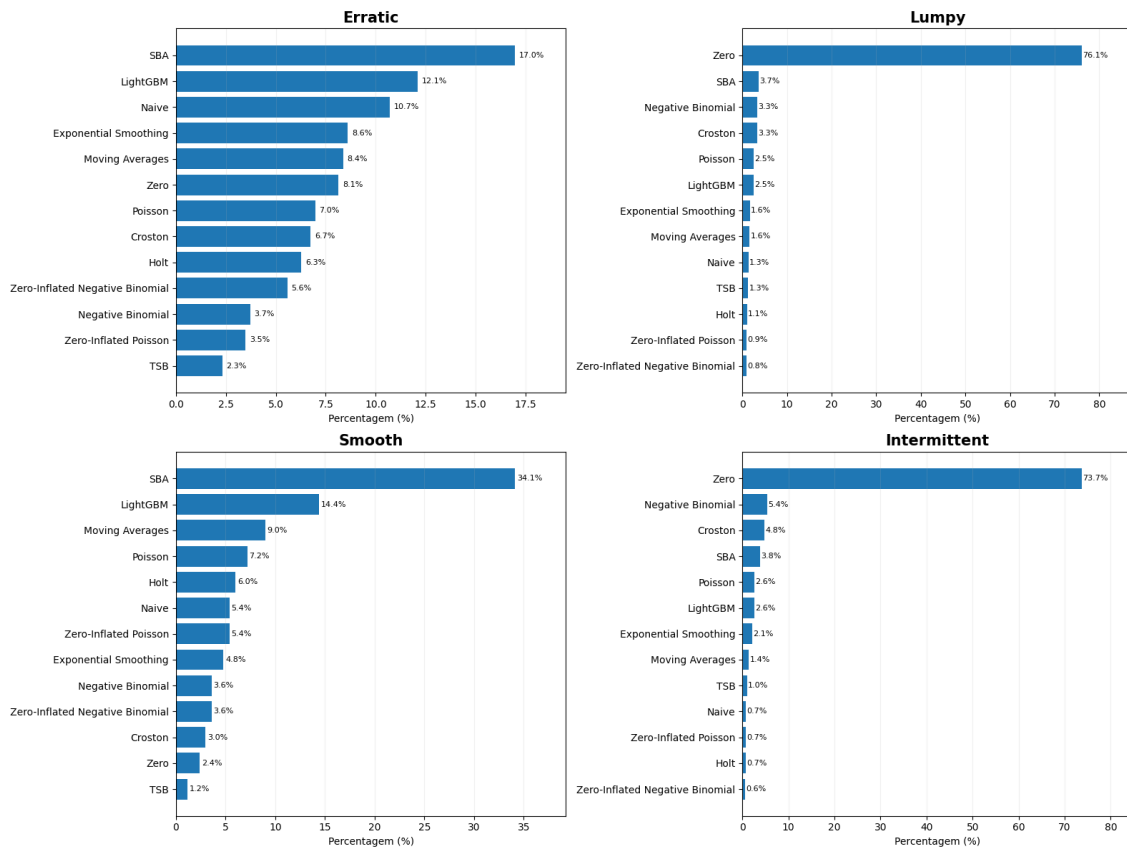


Figura 5.10 - Métrica MSE: Distribuição dos métodos selecionados segundo a classificação SBC

e o desempenho relativo dos diferentes modelos de previsão. Observa-se que séries mais intermitentes favorecem abordagens extremamente conservadoras ou modelos especificamente desenvolvidos para lidar com zeros consecutivos, enquanto séries mais regulares permitem maior competitividade entre métodos tradicionais, probabilísticos e modelos de *machine learning*.

5.2.1.2. Análise dos Modelos selecionados pelo WSSE

A Tabela 5.5 e 5.6 apresenta a distribuição dos modelos selecionados ao minimizar o WSSE e os KPI indicados.

Método	Número de Referências	Porcentagem (%)
Zero	15502	99.74
SBA	9	0.06
LightGBM	8	0.05
Exponential Smoothing	8	0.05
Moving Averages	7	0.05
Negative Binomial	4	0.03
Zero-Inflated Negative Binomial	3	0.02
Naive	1	0.01

Tabela 5.5 - Métrica WSSE: Distribuição dos métodos selecionados

KPI	Valor Médio (%)	Peso no Custo Total (%)
KPI 1	85	0.13
KPI 2	0	1.02
KPI 3	15	98.84

Tabela 5.6 - Métrica WSSE: Resultados médios referentes aos KPI1, 2 e 3 durante os 12 meses de período simulado

Os resultados obtidos mostram um comportamento mais extremo relativamente ao observado anteriormente para o MSE. Ao utilizar o WSSE como critério de seleção, verifica-se que o modelo *Zero Forecast* foi selecionado em praticamente todas as referências analisadas, representando cerca de 99.7% das escolhas realizadas.

Ao contrário do observado no MSE, praticamente todos os restantes modelos deixam de apresentar relevância estatística na seleção final. Métodos especificamente desenvolvidos para procura intermitente, como SBA, bem como modelos probabilísticos e métodos de *machine learning*, apresentaram frequências residuais de seleção.

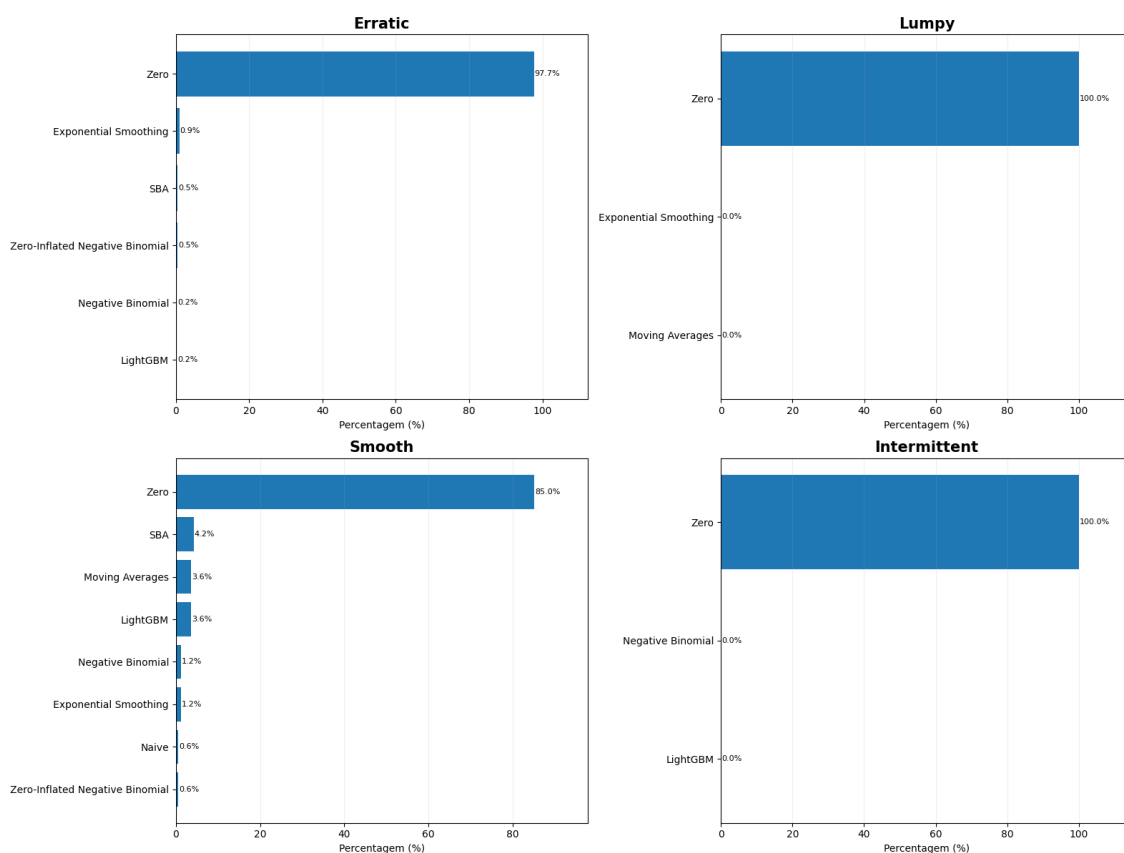


Figura 5.11 - Métrica WSSE: Distribuição dos métodos selecionados segundo a classificação SBC

Ao atribuir maior peso às observações mais recentes através do WSSE, os modelos acabam por ser fortemente pressionados a minimizar os erros nas previsões mais próximas no tempo. Este comportamento pode ser observado no KPI 1, onde foi obtida uma taxa de acerto de aproximadamente 85%. No entanto, este resultado traduz-se em pouco valor monetário.

Por outro lado, apesar de apenas cerca de 15% das previsões ficarem abaixo da procura observada, estes casos representam praticamente 99% do valor total associado.

A Figura 5.11 mostra que a predominância do modelo *Zero Forecast* ocorre de forma praticamente transversal a todas as categorias SBC consideradas. Mesmo em séries classificadas como suave e errático, onde anteriormente existia maior diversidade de modelos.

5.2.1.3. Análise dos Modelos selecionados pelo MAE

A Tabela 5.7 e 5.8 apresenta a distribuição dos modelos selecionados ao minimizar o MAE e os KPI indicados.

Método	Número de Referências	Percentagem (%)
Zero	13634	87.72
Exponential Smoothing	315	2.03
SBA	264	1.70
Naive	234	1.51
Poisson	214	1.38
Negative Binomial	144	0.93
Croston	138	0.89
LightGBM	128	0.82
Holt	112	0.72
Moving Averages	106	0.68
Zero-Inflated Negative Binomial	106	0.68
TSB	89	0.57
Zero-Inflated Poisson	58	0.37

Tabela 5.7 - Métrica MAE: Distribuição dos métodos selecionados

KPI	Valor Médio (%)	Peso no Custo Total (%)
KPI 1	81	4.74
KPI 2	7	8.09
KPI 3	12	87.17

Tabela 5.8 - Métrica MAE: Resultados médios referentes aos KPI1, 2 e 3 durante os 12 meses de período simulado

Os resultados obtidos mostram um comportamento intermédio relativamente ao observado anteriormente para o MSE e WSSE. Embora o modelo *Zero Forecast* continue a apresentar uma forte predominância, que representa aproximadamente 88% das escolhas realizadas, observa-se agora uma maior diversidade de modelos seleccionados.

Ao contrário do WSSE, onde praticamente todas as referências convergiam para o modelo Zero, o MAE permite uma presença relativamente mais significativa de métodos tradicionais e modelos especificamente desenvolvidos para procura intermitente. Em particular, modelos como Exponential Smoothing, SBA, Croston e Naive apresentam frequências de seleção superiores às observadas anteriormente no WSSE, mas menores que no MSE. Este comportamento encontra-se relacionado com a própria natureza do MAE, que penaliza os erros de forma linear.

Os modelos probabilísticos, como Poisson e Negative Binomial, continuam também a apresentar alguma relevância, embora significativamente inferior à observada no MSE. Por outro lado, modelos de *machine learning*, como LightGBM, continuam a apresentar uma frequência relativamente reduzida de seleção, sugerindo novamente dificuldades em capturar padrões robustos em séries temporais altamente esparsas e intermitentes.

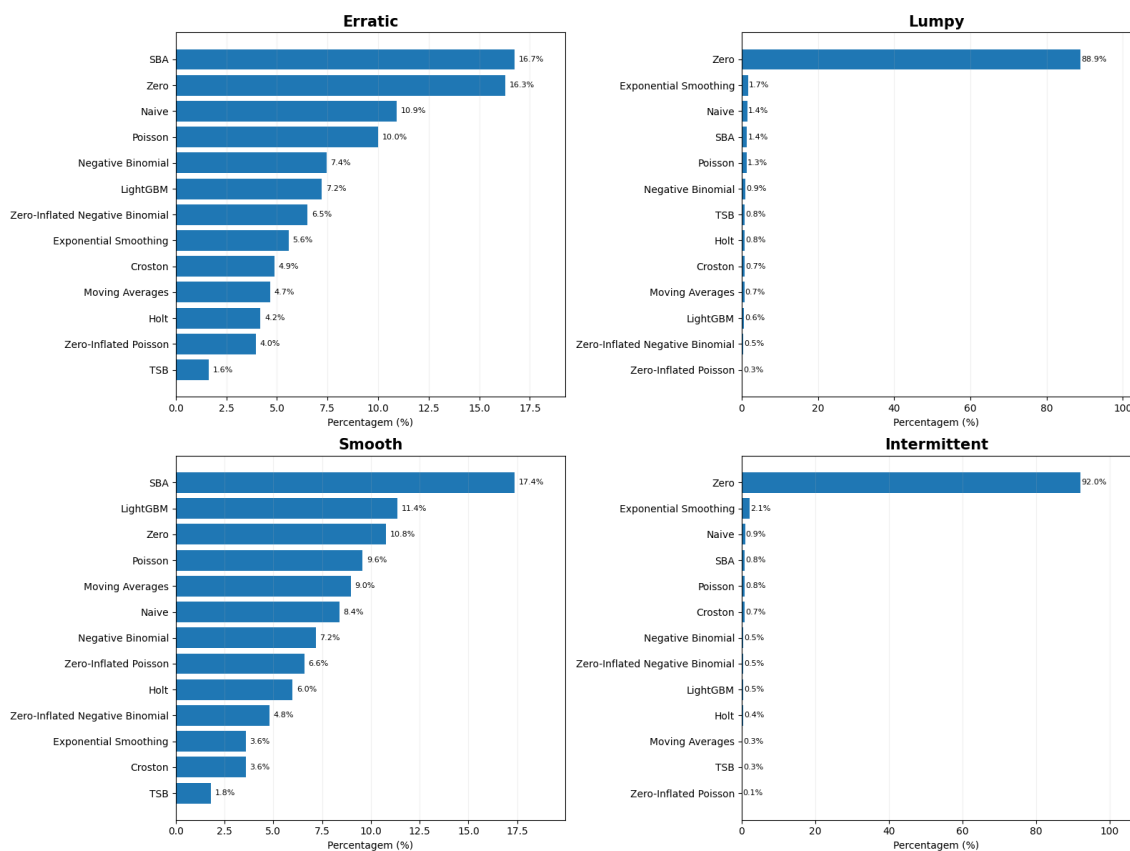


Figura 5.12 - Métrica MAE: Distribuição dos métodos seleccionados segundo a classificação SBC

Relativamente aos KPI obtidos, observa-se uma taxa de acerto de aproximadamente 81%, superior à observada no MSE. No entanto, este desempenho continua a traduzir-se num impacto monetário relativamente reduzido, representando apenas cerca de 4.74% do valor total analisado. Tal como observado anteriormente, uma parte significativa destes acertos encontra-se associada à previsão correta de ausência de procura em períodos dominados por zeros consecutivos.

Por outro lado, apesar de apenas cerca de 12% das previsões ficarem abaixo da procura observada, estes casos representam aproximadamente 87% do valor total associado. Este resultado mostra uma vez mais que os erros mais relevantes encontram-se concentrados precisamente nos períodos onde ocorre procura efetiva, novamente salientando a importância de um stock de segurança.

A Figura 5.12 mostra que a predominância do modelo *Zero Forecast* continua a verificar-se sobretudo nas categorias intermitente e irregular, embora de forma menos extrema relativamente ao observado no WSSE. Já nas séries classificadas como suave e errático, observa-se novamente maior diversidade de modelos selecionados, particularmente métodos tradicionais de suavização, SBA e alguns modelos mais tradicionais.

5.2.2. Performance das Políticas de Stock de Segurança

Para avaliar o desempenho das diferentes políticas de stock de segurança, foram considerados indicadores relacionados com o nível de serviço obtido. Importa referir que estes indicadores de serviço são binários, uma vez que apenas verificam se a previsão mais o stock de segurança foi suficiente para satisfazer a procura observada em cada período. Os resultados devem ser interpretados com algum cuidado, uma vez que funcionam como uma medida mínima de desempenho garantido e não como uma avaliação completa do nível de serviço. Mais à frente, os níveis de serviço serão analisados através de métricas mais tradicionais utilizadas na literatura de gestão de stocks.

Foi também analisado o número médio de dias para escoar o excesso de compras, de forma a avaliar o impacto do excesso de compras nos níveis de inventário.

A Tabela 5.9 apresenta os resultados obtidos para as diferentes combinações entre métricas de seleção dos modelos e políticas de cálculo de stock de segurança.

Os resultados mostram que a metodologia de Bootstrap apresenta níveis de serviço mais elevados em todos os indicadores. Esta abordagem atingiu praticamente 100% de cobertura tanto ao nível das referências como das peças e do valor monetário. No entanto, este desempenho

Simulação	KPI4 (%)	KPI4.1 (%)	KPI4.2 (%)	KPI6 (dias)
MAE + Stock Dinâmico	98.91	71.95	84.30	6.02
MAE + Normal	97.75	65.10	74.48	3.26
MAE + Binomial Negativa/Poisson	98.04	60.75	77.92	4.71
MSE + Stock Dinâmico	98.93	76.13	85.59	6.81
MSE + Normal	97.86	66.54	75.38	4.02
MSE + Binomial Negativa/Poisson	98.38	68.75	81.12	5.55
WSSE + Stock Dinâmico	98.89	70.94	83.30	5.51
WSSE + Normal	97.72	64.25	74.05	2.66
WSSE + Binomial Negativa/Poisson	97.29	36.03	63.54	3.93
Bootstrap	99.98	98.57	99.77	10.33

Tabela 5.9 - Resultados obtidos para as diferentes políticas de stock de segurança.

encontra-se associado a um aumento significativo do excesso de stock, refletido no KPI6, que apresenta o valor mais elevado entre todas as metodologias.

As abordagens baseadas em stock dinâmico apresentaram também resultados bastante consistentes. Independentemente da métrica utilizada na seleção dos modelos de previsão, estas metodologias conseguiram níveis de cobertura superiores a 98% nas referências e valores entre 83% e 86% no KPI associado ao valor monetário satisfeito. Entre estas, destaca-se a combinação com o MSE, que obteve o melhor desempenho ao nível da percentagem de peças satisfeitas.

Por outro lado, as metodologias baseadas na distribuição normal apresentaram níveis de serviço mais reduzidos, embora com menores valores de excesso de stock. Este comportamento sugere uma política que apresenta um menor stock de segurança no geral, logo mais exposta ao risco de rutura mas com potencial para baixar mais o nível de stock.

A abordagem baseada em distribuições discretas apresentou resultados intermédios na maioria das situações. Contudo, a combinação entre WSSE e a abordagem Binomial Negativa/Poisson revelou um desempenho significativamente inferior nos indicadores associados às peças e ao valor satisfeito, o que pode indicar dificuldades desta metodologia em capturar a variabilidade da procura em grande parte do horizonte analisado.

Metodologias mais agressivas no cálculo do stock de segurança, como o Bootstrap e o Stock Dinâmico, conseguem garantir um maior nível de serviço, mas à custa de maiores níveis de inventário. Por outro lado, abordagens mais conservadoras reduzem o excesso de stock, embora apresentem menor capacidade de resposta perante variações da procura.

5.2.3. Performance Global

Após a análise individual das políticas de stock de segurança, é igualmente relevante avaliar o impacto global das diferentes abordagens na manutenção do inventário. Assim, foram considerados alguns KPIs globais definidos anteriormente, com especial foco no KPI9 e no KPI11, uma vez que estes indicadores permitem avaliar simultaneamente o nível de vendas perdidas e a evolução global do stock.

A Tabela 5.10 apresenta os resultados obtidos considerando o valor monetário do inventário.

Simulação	KPI9 (%)	KPI10 (%)	KPI11 (%)	KPI12
MAE + Stock Dinâmico	13.87	30.22	0.18	1.49
MAE + Normal	15.96	14.49	16.91	1.83
MAE + Binomial Negativa/Poisson	15.06	22.91	8.02	1.61
MSE + Stock Dinâmico	13.62	33.78	-3.47	1.45
MSE + Normal	15.65	19.89	10.76	1.73
MSE + Binomial Negativa/Poisson	14.32	28.48	0.60	1.52
WSSE + Stock Dinâmico	13.87	25.05	3.13	1.52
WSSE + Normal	16.03	8.51	21.11	1.88
WSSE + Binomial Negativa/Poisson	18.54	14.74	16.27	1.74
Bootstrap	8.79	43.71	-4.57	1.18

Tabela 5.10 - Resultados globais considerando o valor monetário do inventário.

Os resultados mostram que a metodologia de Bootstrap apresentou o menor nível de vendas perdidas, apenas 8.79% no KPI9. Este comportamento confirma os resultados anteriormente observados nos indicadores de nível de serviço, evidencia uma elevada capacidade de resposta perante a procura. Contudo, este desempenho encontra-se associado ao valor mais elevado no KPI10, que indica um aumento do nível de stock.

As abordagens baseadas em Stock Dinâmico apresentaram bons resultados. A combinação com MSE revela uma redução de vendas perdidas, mas um crescimento do inventário. Este resultado mostra que não foi possível reduzir o stock total ao longo do horizonte temporal sem comprometer significativamente o nível de serviço.

Por outro lado, as abordagens baseadas na distribuição normal apresentaram, de forma geral, maiores níveis de vendas perdidas e reduções mais significativas do inventário ao longo do tempo. Este comportamento torna-se particularmente evidente na combinação WSSE + Normal, que apresentou um dos maiores valores positivos no KPI11, a redução do inventário pode ter sido conseguida à custa de um aumento das vendas perdidas e de uma menor capacidade de

resposta perante a procura.

A Tabela 5.11 apresenta os resultados globais considerando o número de unidades em stock.

Simulação	KPI9 (%)	KPI10 (%)	KPI11 (%)	KPI12
MAE + Stock Dinâmico	13.26	21.98	19.59	1.52
MAE + Normal	13.38	15.33	26.86	1.68
MAE + Binomial Negativa/Poisson	14.81	17.13	28.17	1.69
MSE + Stock Dinâmico	11.96	27.48	14.10	1.45
MSE + Normal	12.59	22.92	21.92	1.57
MSE + Binomial Negativa/Poisson	12.70	25.67	19.94	1.53
WSSE + Stock Dinâmico	13.26	12.55	23.25	1.59
WSSE + Normal	13.41	6.24	31.34	1.74
WSSE + Binomial Negativa/Poisson	21.52	2.64	39.75	1.98
Bootstrap	8.94	15.21	16.86	1.42

Tabela 5.11 - Resultados globais considerando o número de unidades em stock.

Os resultados obtidos em unidades confirmam, em grande medida, as conclusões observadas na análise em valor. Novamente, a metodologia de Bootstrap apresentou o menor nível de vendas perdidas, reforçando a sua elevada capacidade de cobertura da procura.

Destaca-se também o desempenho da abordagem MSE + Stock Dinâmico, que apresentou um dos melhores equilíbrios globais entre redução de vendas perdidas e controlo do crescimento do stock. Em contraste, a combinação WSSE + Binomial Negativa/Poisson revelou um desempenho bastante inferior.

As Figuras 5.13 e 5.14 apresentam a evolução temporal do stock total em valor monetário e em unidades, respetivamente.

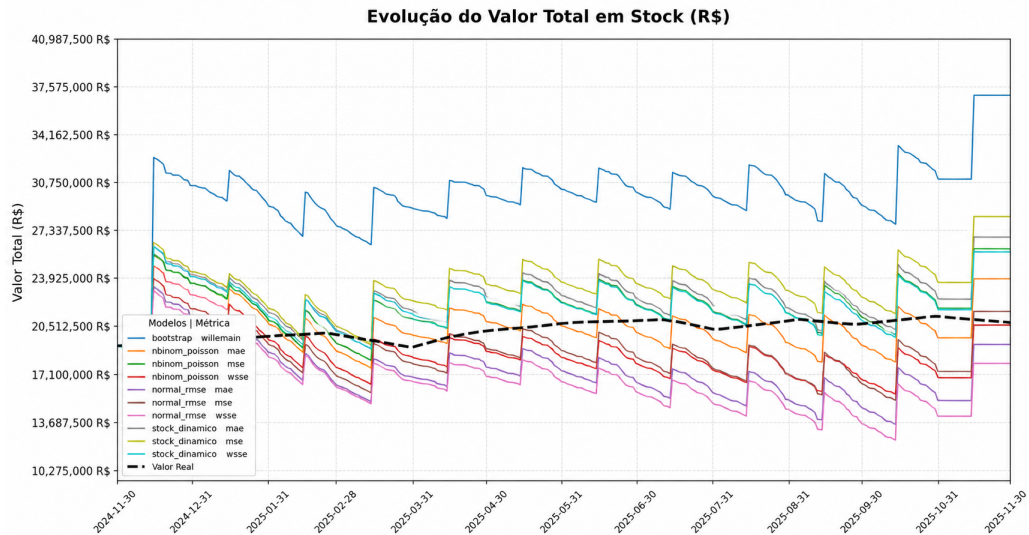


Figura 5.13 - Evolução do valor total em stock para as diferentes metodologias analisadas.

Observa-se que as metodologias associadas ao Bootstrap mantêm sistematicamente níveis de stock superiores ao stock real ao longo de praticamente todo o horizonte temporal. Este comportamento explica os níveis reduzidos de vendas perdidas anteriormente observados, embora implique um maior investimento em inventário, o que leva a que na realidade não possa ser utilizada, dado o seu alto nível de investimento.

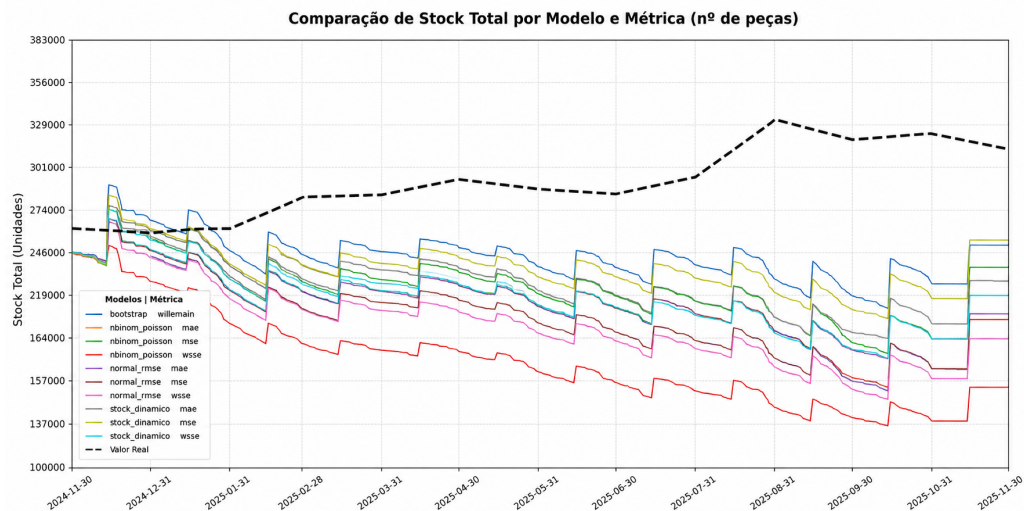


Figura 5.14 - Evolução do stock total em unidades para as diferentes metodologias analisadas.

De forma geral, a metodologia de Bootstrap apresentou os melhores resultados ao nível dos indicadores de serviço, registando os menores níveis de vendas perdidas. No entanto, este

desempenho encontra-se diretamente associado a níveis de stock significativamente superiores aos restantes modelos. Apesar da elevada capacidade de cobertura da procura, a necessidade de um investimento inicial muito elevado em inventário torna esta abordagem pouco viável do ponto de vista operacional e financeiro no contexto da empresa.

Relativamente à metodologia desenvolvida no âmbito deste trabalho baseada em distribuições discretas, os resultados obtidos ficaram, abaixo do esperado. Embora em algumas situações tenha apresentado desempenhos intermédios, a abordagem revelou dificuldades em manter simultaneamente baixos níveis de vendas perdidas e uma evolução controlada do stock, sobretudo nas combinações associadas ao WSSE.

As metodologias baseadas em Stock Dinâmico apresentaram um comportamento bastante consistente independentemente da métrica utilizada. Estas abordagens conseguiram manter bons níveis de serviço e baixos níveis de vendas perdidas, embora apresentem menor capacidade na redução do stock total, tanto em valor monetário como em unidades. Em contraste, as políticas baseadas na distribuição normal apresentaram níveis de serviço ligeiramente inferiores, mas conseguiram reduções mais significativas do inventário, o que evidencia um melhor equilíbrio entre vendas perdidas e investimento em stock.

Verificou-se que o stock real da empresa apresentou um crescimento significativo ao longo do horizonte temporal considerado, o que sugere a existência de fatores externos à modelação da procura e das políticas de stock que contribuíram para essa tendência de aumento. Neste contexto, a eliminação total desse crescimento revela-se particularmente difícil. Ainda assim, os resultados obtidos demonstram que as metodologias propostas limitaram esse aumento de forma substancial, que levou a um crescimento do stock significativamente inferior ao observado na empresa.

É importante ter em consideração uma limitação do processo de simulação adotado. Na operação real, a empresa pode recorrer a compras de emergência para responder a compras inesperadas. No entanto, esta possibilidade não foi incorporada nos modelos desenvolvidos. Assim, algumas situações que seriam resolvidas através deste tipo de compras acabam por se traduzir no modelo em vendas perdidas. Como consequência, os níveis de serviço obtidos tendem a ser mais baixos do que os observados na realidade.

Com base nos resultados obtidos, foram selecionadas três combinações de modelos para discussão com a empresa: MAE + Stock Dinâmico, WSSE + Stock Dinâmico e MSE + Normal.

Nessa discussão foram explicados os prós e contras de cada um dos modelos, sendo obviamente a sua seleção uma opção de gestão com base nos objetivos principais da empresa bem como restrições e/ou objetivos associados à parte financeira e comercial. Estas abordagens destacaram-se por apresentarem o melhor compromisso entre nível de serviço, controlo do stock e viabilidade operacional. Optou-se ainda por utilizar o mesmo valor do parâmetro α nas diferentes metodologias, uma vez que este demonstrou um comportamento equilibrado entre risco de rutura e investimento adicional em inventário.

6. Integração dos Dados e Estrutura da Base de Dados

Neste capítulo é descrita a solução de *Business Intelligence* desenvolvida, com foco na estruturação da Base de Dados e fluxos de integração nos sistemas em uso pela NORIS. O principal objetivo foi garantir que os dados fornecidos pela empresa fossem tratados de forma automática, consistente e compatível com os modelos de previsão, simulação e visualização desenvolvidos.

Tal como referido na Secção 3.1, os dados disponibilizados pela empresa apresentavam diferentes estruturas, o que levou a processos adicionais de transformação e validação antes da sua utilização. Assim, foi desenvolvida uma pipeline de ETL (*Extract, Transform and Load*), totalmente automatizada, responsável pela leitura dos ficheiros, transformação da informação e carregamento das tabelas finais para o Microsoft SQL Server. Uma descrição mais detalhada sobre processos de ETL pode ser encontrada em Golfarelli e Rizzi (2009).

Toda a lógica de transformação foi implementada em Python, em que se recorreu a bibliotecas como *Pandas*, *NumPy*, *SQLAlchemy* e *PyODBC*. O *SQL Server Integration Services* (SSIS) foi utilizado essencialmente como ferramenta de orquestração, através da execução encadeada de scripts Python responsáveis pelas diferentes etapas do processo.

6.1. Modelo Dimensional e Integridade Referencial

Após a preparação dos dados, foi criada uma base de dados relacional organizada segundo uma arquitetura dimensional do tipo *Star Schema*. Este tipo de estrutura é utilizado em sistemas de *Business Intelligence* e *Data Warehousing*, devido à sua simplicidade, elevada performance e facilidade de integração com ferramentas de visualização e processamento analítico (Golfarelli e Rizzi 2009).

Num modelo dimensional do tipo *Star Schema*, os dados são organizados em torno de tabelas de factos (*Fact Tables*) ligadas a tabelas auxiliares de dimensão (*Dimension Tables*). As tabelas de factos guardam valores quantitativos e métricas do negócio, enquanto as tabelas de dimensão fornecem o contexto necessário.

Nas tabelas desenvolvidas foram utilizados dois prefixos principais:

- **FACT:** tabelas de factos;
- **LK:** tabelas de dimensão.

Como os dados fornecidos eram de diferentes sistemas, existia o risco de inconsistências entre códigos de produtos, lojas, participantes, vendedores e datas. Para evitar estes problemas, foram implementadas validações automáticas antes do carregamento final dos dados para SQL Server.

Estas validações incluíam:

- Verificação da existência prévia de todas as chaves primárias nas tabelas de dimensão;
- Validação das relações entre tabelas de factos e dimensões;
- Controlo de chaves duplicadas;
- Uniformização de nomenclaturas.

Sempre que uma inconsistência era encontrada, o carregamento da tabela era interrompido e o erro registado automaticamente nos ficheiros de *log*.

Desta forma, foi possível garantir que todas as tabelas utilizadas nos modelos analíticos e no dashboard mantivessem integridade referencial.

6.2. Estrutura da Base de Dados Operacional

A Figura 6.1 apresenta o modelo relacional correspondente aos dados fornecidos pela empresa.

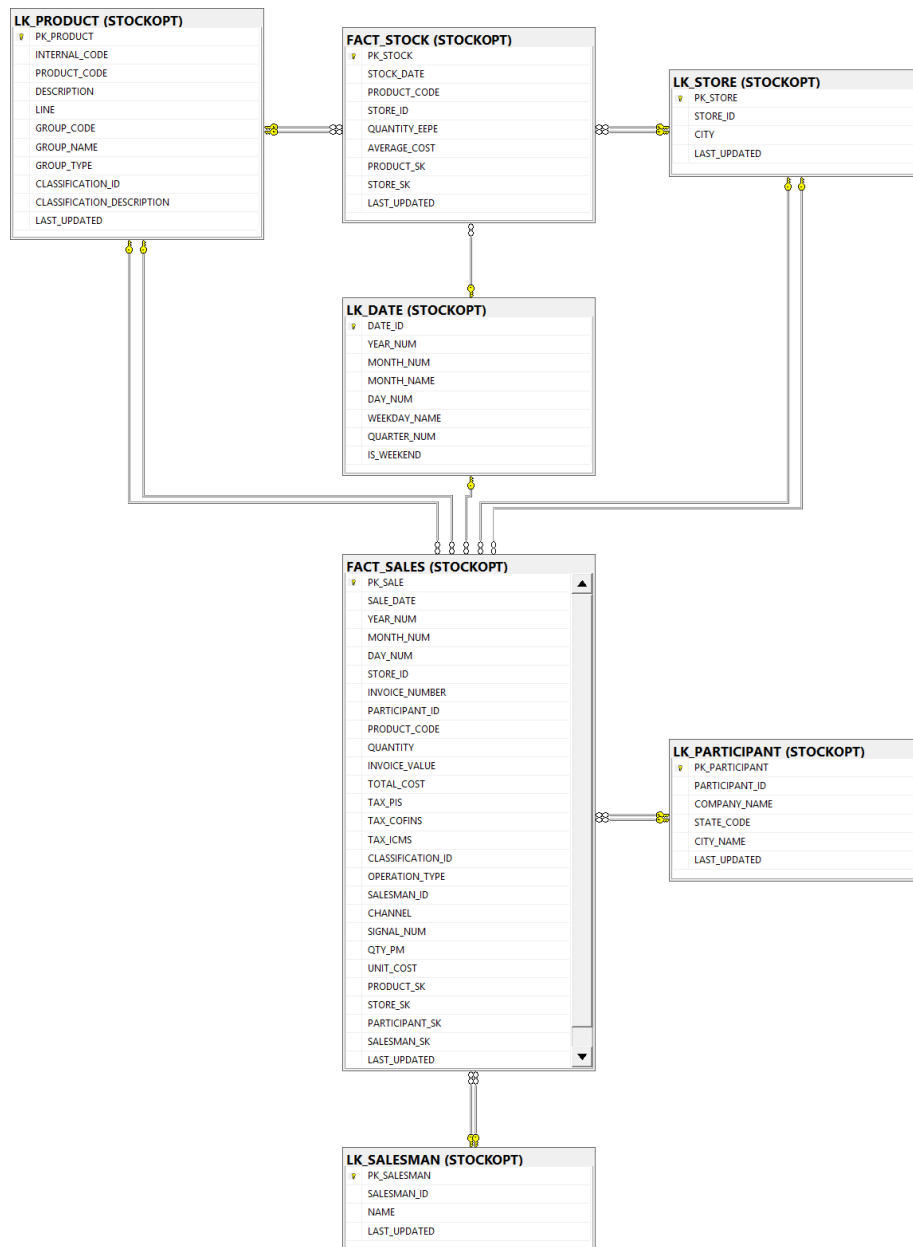


Figura 6.1 - Modelo relacional da base de dados operacional.

As tabelas de dimensão criadas foram:

- **LK_PRODUCT**: contém informação relativa aos produtos, incluindo código interno, descrição, grupo e subgrupo de produto;
- **LK_STORE**: contém informação relativa às lojas e localizações;
- **LK_DATE**: tabela calendário contém datas, meses, anos e outros atributos temporais;
- **LK_PARTICIPANT**: informação relativa aos clientes;

- **LK_SALESMAN:** informação relativa aos vendedores.

Relativamente às tabelas de factos:

- **FACT_SALES:** registos de vendas;
- **FACT_STOCK:** registos de stocks.

6.3. Estrutura da Base de Dados Analítica

Após a construção da base de dados operacional, foram adicionadas à mesma base de dados novas tabelas destinadas aos resultados produzidos pelo modelo selecionado pela empresa.

A Figura 6.2 apresenta a estrutura relacional correspondente à componente analítica.

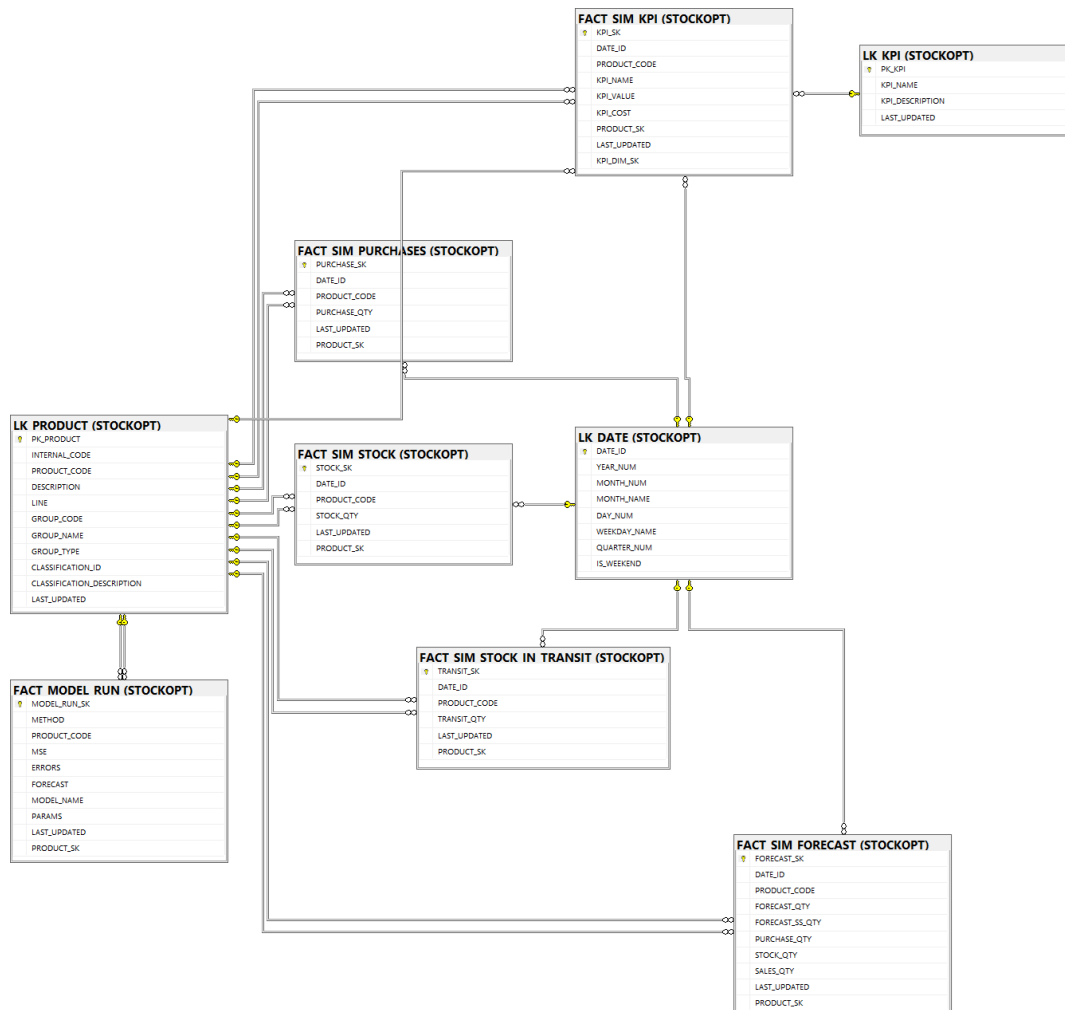


Figura 6.2 - Modelo relacional da base de dados analítica e dos modelos.

As principais tabelas criadas foram:

- **FACT_SIM_FORECAST:** previsões geradas pelo modelo;
- **FACT_SIM_STOCK:** tabela auxiliar para registar a evolução dos níveis de stock;
- **FACT_SIM_PURCHASES:** sugestões de compras dadas pelo modelo;
- **FACT_SIM_STOCK_IN_TRANSIT:** controlo do stock em trânsito;
- **FACT_SIM_KPI:** resultados dos indicadores de performance calculados durante as simulações;
- **FACT_MODEL_RUN:** registo dos modelos executados e respetivas configurações utilizadas.

Foi igualmente criada a dimensão:

- **LK_KPI:** descrição dos indicadores de performance.

Esta separação permitiu organizar melhor os dados e simplificar os processos de análise.

6.4. Integração com SSAS e Power BI

Após carregar os dados no SQL Server, foi desenvolvido um modelo tabular no SQL Server Analysis Services (SSAS). Neste modelo foram definidas as relações entre as tabelas, bem como as medidas DAX e restantes métricas utilizadas na análise dos dados.

A utilização do SSAS permitiu melhorar o desempenho no Power BI, uma vez que os cálculos e agregações passam a ser efetuados pelo próprio modelo tabular. Para além disso, esta abordagem permite reutilizar o mesmo modelo em diferentes relatórios e dashboards.

O Power BI foi ligado diretamente ao modelo tabular, que funciona como camada analítica responsável por disponibilizar a informação necessária para a visualização dos dados.

A Figura 6.3 ilustra o fluxo de informação da solução desenvolvida desde a obtenção dos dados no ERP empresarial até à construção do dashboard para análise e apoio à decisão no Power BI.

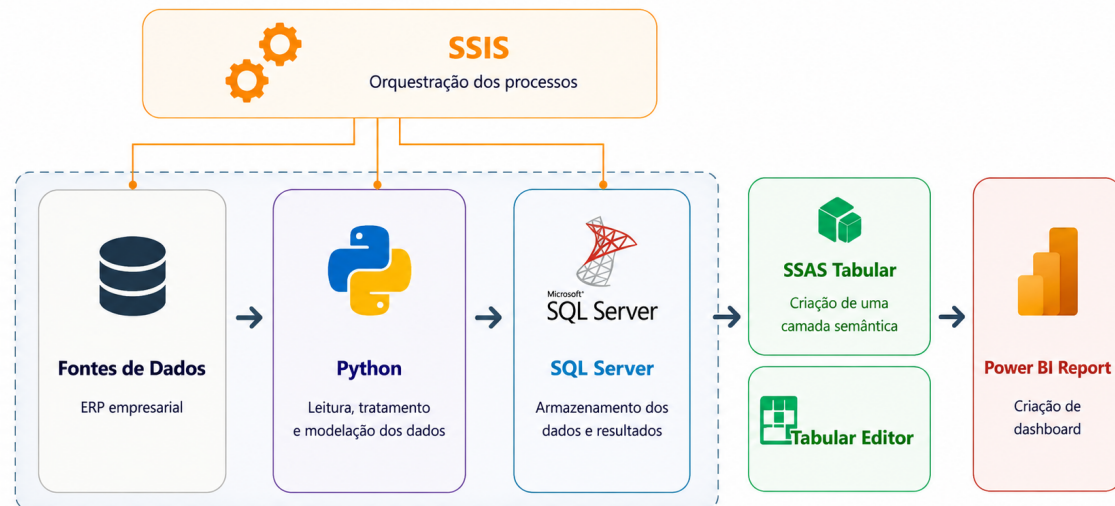


Figura 6.3 - Arquitetura da solução desenvolvida.

6.4.1. Dashboard de Suporte à Decisão

Com o objetivo de disponibilizar os resultados do trabalho desenvolvido, foi construída um dashboard em Power BI integrada na infraestrutura analítica da empresa. Este dashboard permite acompanhar a evolução das vendas e dos stocks, bem como consultar as sugestões de compra geradas pelo modelo proposto.

6.4.1.1. Análise de Vendas e Stocks

O primeiro separador do dashboard, apresentado na Figura 6.4, disponibiliza uma visão global da atividade comercial e dos níveis de stock da empresa.

No topo da página são apresentados indicadores agregados, que incluem o número total de clientes, o total de vendas, o valor monetário das vendas e o número de produtos vendidos. Adicionalmente, é disponibilizado um seletor que permite alterar entre a visualização em unidades físicas ou em valor monetário, que adapta automaticamente os gráficos e métricas.

A partir desta página é ainda possível aceder ao painel de filtros, ilustrado na Figura 6.5, onde o utilizador pode restringir a análise a anos específicos, lojas, clientes, vendedores, tipos de peça ou produtos individuais.

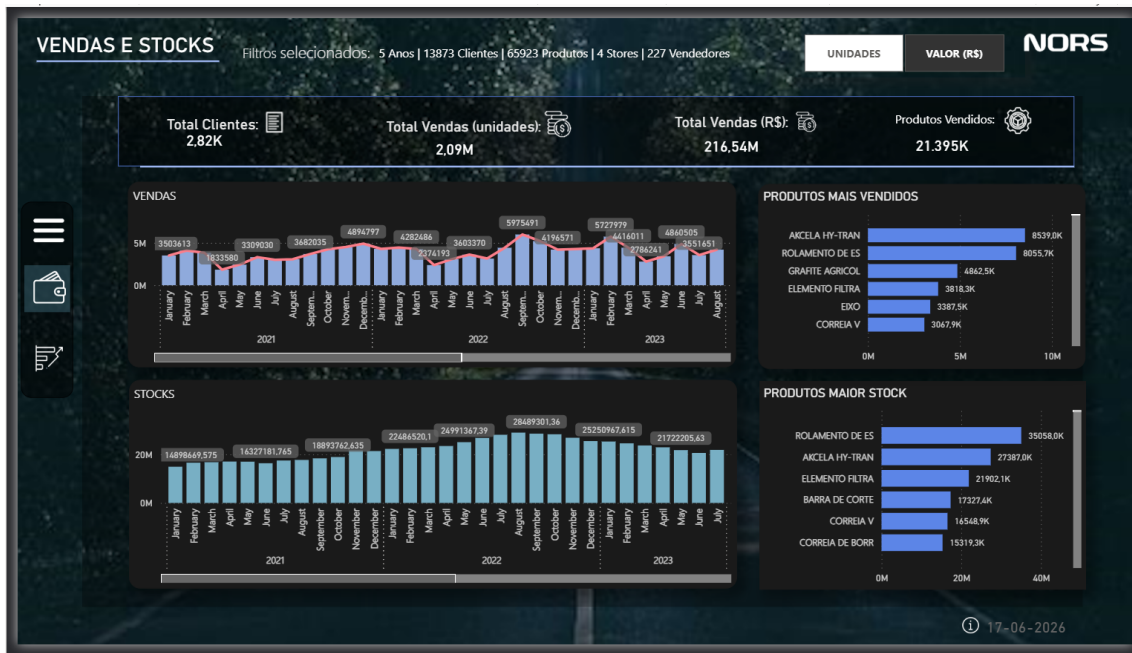


Figura 6.4 - Separador de análise de vendas e stocks.

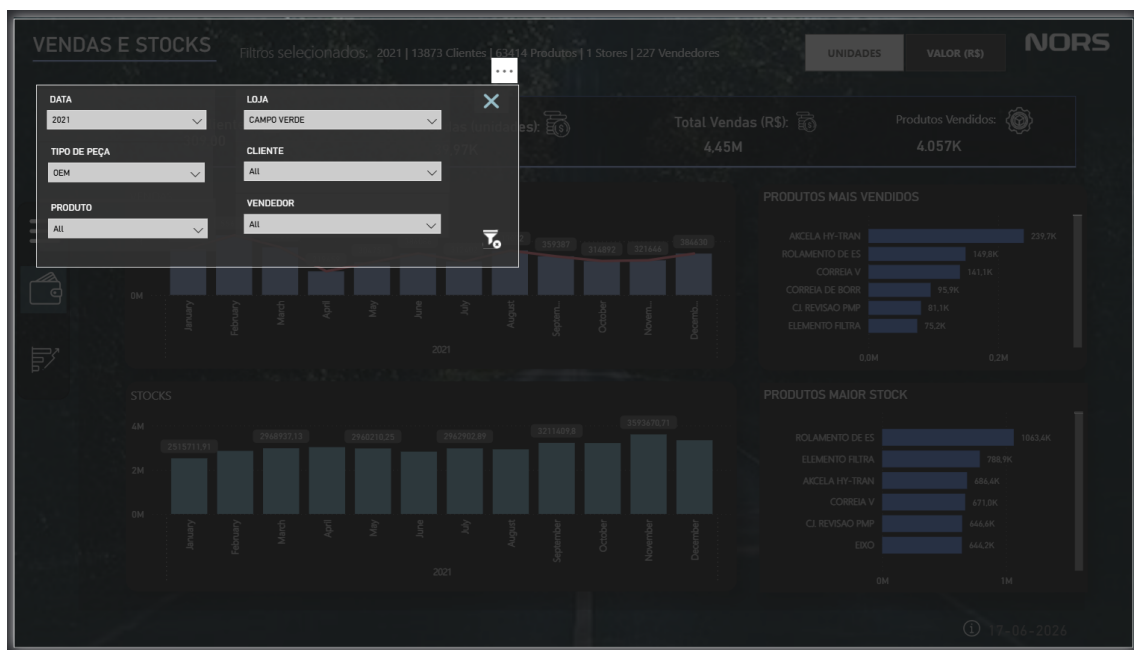


Figura 6.5 - Painel de filtros do dashboard.

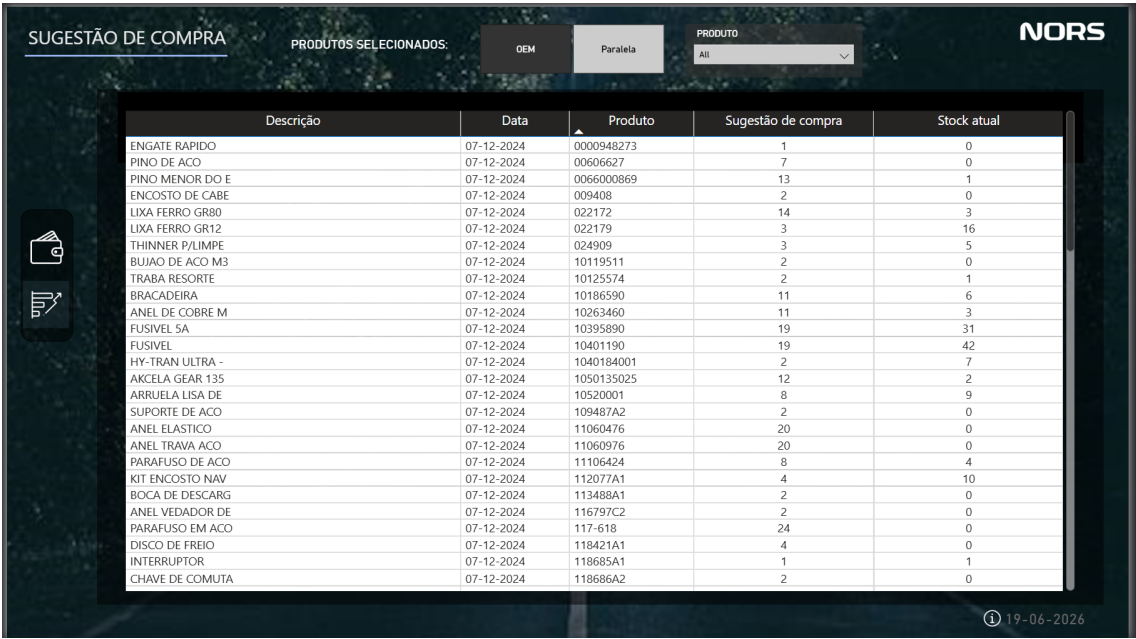
6.4.1.2. Sugestão de Compra

O segundo separador do dashboard, apresentado na Figura 6.6, apresenta as sugestões de compra geradas pelo modelo de gestão de stocks desenvolvido.

Para cada produto são disponibilizadas informações relevantes, estas incluem a descrição da referência, a data da recomendação, a quantidade sugerida para compra e o stock disponível.

Tal como no separador anterior, o utilizador pode aplicar filtros para selecionar produtos específicos ou restringir a análise a determinados tipos de peças.

O principal objetivo desta página consiste em disponibilizar a quem for utilizar, uma ferramenta operacional que permite exportar as recomendações geradas pelo modelo e utilizá-las diretamente no processo de encomendas.



Descrição	Data	Produto	Sugestão de compra	Stock atual
ENGATE RAPIDO	07-12-2024	0000948273	1	0
PINO DE ACO	07-12-2024	00606627	7	0
PINO MENOR DO E	07-12-2024	0066000869	13	1
ENCOSTO DE CABE	07-12-2024	009408	2	0
LIXA FERRO GR80	07-12-2024	022172	14	3
LIXA FERRO GR12	07-12-2024	022179	3	16
THINNER P/LIMPE	07-12-2024	024909	3	5
BUJAO DE ACO M3	07-12-2024	10119511	2	0
TRABA RESORTE	07-12-2024	10125574	2	1
BRACADEIRA	07-12-2024	10186590	11	6
ANEL DE COBRE M	07-12-2024	10263460	11	3
FUSIVEL 5A	07-12-2024	10395890	19	31
FUSIVEL	07-12-2024	10401190	19	42
HY-TRAN ULTRA -	07-12-2024	1040184001	2	7
AKCELA GEAR 135	07-12-2024	1050135025	12	2
ARRUELA LISA DE	07-12-2024	10520001	8	9
SUPORTE DE ACO	07-12-2024	109487A2	2	0
ANEL ELASTICO	07-12-2024	11060476	20	0
ANEL TRAVA ACO	07-12-2024	11060976	20	0
PARAFUSO DE ACO	07-12-2024	11106424	8	4
KIT ENCOSTO NAV	07-12-2024	112077A1	4	10
BOCA DE DESCARG	07-12-2024	113488A1	2	0
ANEL VEDADOR DE	07-12-2024	116797C2	2	0
PARAFUSO EM ACO	07-12-2024	117-618	24	0
DISCO DE FREIO	07-12-2024	118421A1	4	0
INTERRUPTOR	07-12-2024	118685A1	1	1
CHAVE DE COMUTA	07-12-2024	118686A2	2	0

Figura 6.6 - Separador de sugestões de compra.

7. Conclusões

O presente trabalho teve como principal objetivo desenvolver um modelo de gestão de stocks para peças OEM numa empresa que opera no setor agrícola brasileiro, utilizando métodos de previsão e diferentes políticas de stock de segurança.

A análise e tratamento dos dados permitiram compreender melhor a estrutura das bases de dados disponibilizadas pela empresa. A construção das séries temporais e a segmentação dos produtos revelaram diferenças significativas entre os padrões de procura das várias referências, que evidenciaram a presença de procura intermitente e irregular neste contexto.

Ao longo do trabalho foram desenvolvidos e avaliados diferentes modelos de previsão, tais como métodos estatísticos tradicionais, abordagens específicas para procura intermitente, modelos probabilísticos e técnicas de *Machine Learning*. Os resultados obtidos indicaram que o desempenho dos modelos depende fortemente das características da procura, uma vez que foram observados comportamentos muito distintos entre os diferentes segmentos dos artigos comercializados.

Relativamente às políticas de stock de segurança, os resultados mostraram que abordagens mais ajustadas à variabilidade da procura permitiram reduzir o valor global de inventário e manter simultaneamente níveis de serviço elevados. As simulações realizadas evidenciaram ainda a importância da correta modelação dos erros de previsão no cálculo dos níveis de stock de segurança e dos pontos de encomenda.

Para além da componente analítica, todo o trabalho desenvolvido foi integrado na infraestrutura tecnológica da empresa através da criação de processos automáticos de ETL, persistência e modelação dos dados e um dashboard de suporte à decisão em Power BI. Esta integração permitiu automatizar grande parte do processo de atualização da informação e criar uma solução facilmente aplicável em contexto empresarial real.

Como trabalho futuro, poderá ser interessante explorar metodologias mais avançadas de otimização de stocks, nomeadamente abordagens multi-echelon, bem como a incorporação de variáveis externas relacionadas com sazonalidade agrícola, fatores económicos ou condições climáticas. Poderá igualmente ser relevante analisar modelos de *Deep Learning* e arquiteturas híbridas aplicadas à previsão da procura intermitente.

Referências

- Axsäter, S. (2006). *Inventory Control*. 2ª ed. Springer (Citado nas páginas 9, 27, 29).
- Bacchetti, A. e Sacconi, N. (2012). “Spare Parts Classification and Demand Forecasting for Stock Control: Investigating the Gap Between Research and Practice”. Em: *Omega* 40.6, pp. 722–737 (Citado na página 3).
- Barrow, D. K. e Kourentzes, N. (2016). “Distributions of Forecasting Errors of Forecast Combinations: Implications for Inventory Management”. Em: *International Journal of Production Economics* 177, pp. 24–33 (Citado nas páginas 9, 27, 29).
- Barrow, D. K., Kourentzes, N. e Sandberg, R. (2017). “On the Calculation of Safety Stocks When Demand Is Forecasted”. Em: *European Journal of Operational Research* 256.2, pp. 454–461 (Citado nas páginas 9, 29).
- Capkun, V., Hameri, A.-P. e Weiss, L. A. (2009). “On the relationship between inventory and financial performance in manufacturing companies”. Em: *International Journal of Operations & Production Management* 29.8, pp. 789–806 (Citado na página 3).
- Chen, T. e Guestrin, C. (2016). “XGBoost: A Scalable Tree Boosting System”. Em: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 785–794. DOI: 10.1145/2939672.2939785 (Citado na página 8).
- Christ, G. D., Strauch, A. G. N., Santos, L. P. d. e Shikida, P. F. A. (2022). “The Competitiveness of Brazilian Agribusiness in International Trade”. Em: *Revista de Economia e Sociologia Rural* 60.1 (Citado na página 4).
- Costantino, F., Di Gravio, G. e Shaban, A. (2018). “Demand Forecasting for High-Turnover Spare Parts in Agricultural and Construction Machines: A Case Study”. Em: *International Journal of Engineering Business Management* 10 (Citado na página 4).
- Croston, J. D. (1972). “Forecasting and Stock Control for Intermittent Demands”. Em: *Operational Research Quarterly* 23.3, pp. 289–303 (Citado nas páginas 7, 20).
- Cruz, M. B. et al. (2020). “Order and Stock Costs Optimization in an Automotive Spare Parts Wholesaler”. Em: *Computational Science and Its Applications – ICCSA 2020*. Springer. DOI: 10.1007/978-3-030-61844-5_9 (Citado nas páginas 7, 9, 17, 28, 36).
- Freund, Y. e Schapire, R. E. (1997). “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”. Em: *Journal of Computer and System Sciences* 55.1, pp. 119–139. DOI: 10.1006/jcss.1997.1504 (Citado na página 8).
- Friedman, J. H. (2001). “Greedy function approximation: a gradient boosting machine”. Em: *Annals of statistics*, pp. 1189–1232 (Citado na página 26).

- Gallagher, T., Mitchke, M. D. e Rogers, M. C. (2005). “Profiting from Spare Parts”. Em: *The McKinsey Quarterly* (Citado na página 3).
- Ghobbar, A. A. e Friend, C. H. (2003). “Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model”. Em: *Computers & operations research* 30.14, pp. 2097–2114 (Citado na página 3).
- Golfarelli, M. e Rizzi, S. (2009). *Data Warehouse Design: Modern Principles and Methodologies*. 1st. New York: McGraw-Hill Osborne Media (Citado nas páginas 10, 63).
- Goltsos, T. E., Syntetos, A. A. e Boylan, J. E. (2022). “On the Link Between Forecasting and Inventory Control: A Review and Research Agenda”. Em: *European Journal of Operational Research* 299.1, pp. 397–419 (Citado na página 10).
- Greene, W. H. (1994). “Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models”. Em: *NYU Working Paper* (Citado nas páginas 23, 25).
- Hyndman, R. J. e Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. 2ª ed. OTexts (Citado nas páginas 17, 19, 21, 34, 35).
- Hyndman, R. J. e Koehler, A. B. (2006). “Another Look at Measures of Forecast Accuracy”. Em: *International Journal of Forecasting* 22.4, pp. 679–688. DOI: 10.1016/j.ijforecast.2006.03.001 (Citado na página 9).
- IBGE (2023). *Levantamento Sistemático da Produção Agrícola*. Acesso em: 4 jun. 2026. URL: <https://www.ibge.gov.br/> (Citado na página 4).
- IHL Group (2023). *Inventory Distortion: The Good, the Bad and the Ugly*. Acesso em: 5 Jun. 2026. URL: <https://www.ihlservices.com> (Citado na página 3).
- Ke, G. et al. (2017). “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”. Em: *Advances in Neural Information Processing Systems* 30 (Citado nas páginas 8, 26, 27).
- Lambert, D. (1992). “Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing”. Em: *Technometrics* 34.1, pp. 1–14 (Citado na página 24).
- Microsoft (2025). *Semantic models in the Power BI service*. <https://learn.microsoft.com/power-bi/connect-data/service-datasets-understand>. Acesso em: 4 Jun. 2026 (Citado na página 10).
- Microsoft SQL Server 2022 (2022). *Microsoft SQL Server 2022*. Acesso em: 4 jun. 2026. URL: <https://www.microsoft.com/sql-server/> (Citado na página 5).
- Python 3.11 (2022). *Python Language Reference, version 3.11*. Acesso em: 4 jun. 2026. URL: <https://www.python.org/> (Citado nas páginas 5, 35).

- Romeijnnders, W., Teunter, R. e Jaarsveld, W. van (2012). “A two-step method for forecasting spare parts demand using information on component repairs”. Em: *European Journal of Operational Research* 220.2, pp. 386–393 (Citado nas páginas 17, 18).
- Russo, M., Ferrari, A. e Webb, C. (2012). *Microsoft SQL Server 2012 Analysis Services: The BISM Tabular Model*. Microsoft Press (Citado na página 10).
- Shehadeh, A., Alshboul, O., Al Mamlook, R. E. e Hamedat, O. (2021). “Machine learning-based models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression”. Em: *Automation in Construction* 129, p. 103827. DOI: 10.1016/j.autcon.2021.103827 (Citado na página 8).
- Silver, E. A. (1981). “Operations Research in Inventory Management: A Review and Critique”. Em: *Operations Research* 29.4, pp. 628–645 (Citado na página 4).
- Silver, E. A., Pyke, D. F. e Thomas, D. J. (2017). *Inventory and Production Management in Supply Chains*. 4ª ed. Boca Raton, FL: CRC Press (Citado nas páginas 10, 19, 29).
- Strijbosch, L. W. G., Heuts, R. J. M. e Schoot, E. H. M. van der (2000). “A Compound-Poisson Bayesian Approach for Spare Parts Inventory Forecasting”. Em: *International Journal of Production Economics* 63.3, pp. 217–229. DOI: 10.1016/S0925-5273(99)00031-6 (Citado nas páginas 9, 27).
- Syntetos, A. A. e Boylan, J. E. (2001). “On the bias of intermittent demand estimates”. Em: *International journal of production economics* 71.1-3, pp. 457–466 (Citado na página 2).
- Syntetos, A. A., Babai, M. Z., Davies, J. e Stephenson, D. (2009). “Forecasting and stock control: A study in a wholesaling context”. Em: *International Journal of Production Economics* 118.1, pp. 103–111 (Citado na página 7).
- Syntetos, A. A. e Boylan, J. E. (2005). “The Accuracy of Intermittent Demand Estimates”. Em: *International Journal of Forecasting* 21.2, pp. 303–314 (Citado nas páginas 8, 20).
- Syntetos, A. A., Boylan, J. E. e Croston, J. D. (2005). “On the categorization of demand patterns”. Em: *Journal of the Operational Research Society* 56.5, pp. 495–503 (Citado nas páginas 7, 14, 15, 30).
- Teunter, R. H., Syntetos, A. A. e Babai, M. Z. (2011). “Intermittent demand: Linking forecasting to inventory obsolescence”. Em: *European Journal of Operational Research* 214.3, pp. 606–615 (Citado nas páginas 21, 30).
- Willemain, T. R., Smart, C. N. e Schwarz, H. F. (2004). “A New Approach to Forecasting Intermittent Demand for Service Parts Inventories”. Em: *International Journal of Forecasting* 20.3, pp. 375–387 (Citado nas páginas 4, 8, 31, 32).

- Zhang, W. et al. (2020). "State-of-the-art review of soft computing applications in underground excavations". Em: *Geoscience Frontiers* 11.4, pp. 1095–1106 (Citado na página 8).
- Zierz, O. et al. (2026). "Business Intelligence Tools in Organizations with a Focus on Power BI Applications in Civil Construction: A Systematic Literature Review". Em: *Buildings* 16.4, p. 869 (Citado na página 11).