PEDRO DE OLIVEIRA BRASCA PERES NOGUEIRA

PREVISÃO DE DEMANDA DE NOVOS PRODUTOS: REVISÃO DE PROCESSOS E APLICAÇÃO DE MÉTODOS PREDITIVOS EM UMA EMPRESA DO VAREJO DE MODA

São Paulo
2025

PEDRO DE OLIVEIRA BRASCA PERES NOGUEIRA

PREVISÃO DE DEMANDA DE NOVOS PRODUTOS: REVISÃO DE PROCESSOS E APLICAÇÃO DE MÉTODOS PREDITIVOS EM UMA EMPRESA DO VAREJO DE MODA

Trabalho apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do Diploma de Engenheiro de Produção.

São Paulo
2025

PEDRO DE OLIVEIRA BRASCA PERES NOGUEIRA

PREVISÃO DE DEMANDA DE NOVOS PRODUTOS: REVISÃO DE PROCESSOS E APLICAÇÃO DE MÉTODOS PREDITIVOS EM UMA EMPRESA DO VAREJO DE MODA

Trabalho apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do Diploma de Engenheiro de Produção.

Orientador:

Prof. Dr. Marco Aurélio de Mesquita

São Paulo
2025

Aos meus pais, por todas oportunidades que me proporcionaram ao longo da minha vida.

AGRADECIMENTOS

Aos meus pais, por todas as oportunidades, incentivos, amor, apoio e estrutura que sempre me ofereceram, e por me darem a possibilidade de sonhar alto.

À minha irmã, pelo amor e carinho que me dão forças para seguir em frente, especialmente nos momentos mais desafiadores.

À Escola Politécnica da USP, pela oportunidade de estar em um ambiente desafiador e de excelência acadêmica que me fez buscar ser melhor a cada dia, pela possibilidade de Aproveitamento de Estudos na *Sapienza Università di Roma* e por todas as outras oportunidades que dela vieram.

Ao Anglo Leonardo da Vinci, à Escola Granja Viana e a todos os professores que fizeram parte da minha formação, por construírem a base que me permitiu chegar até aqui e por contribuírem para minha formação como ser humano pensante e como cidadão. Agradeço, ainda, por me apresentarem grandes amigos de longa data, que estiveram comigo e foram essenciais durante toda essa jornada.

À Poli Júnior, por me possibilitar aprender muito, ser uma fuga em tempos de pandemia, desenvolver-me como pessoa, abrir portas profissionais e me apresentar muitos dos amigos que hoje são alguns dos meus mais próximos.

Aos meus amigos, que tornaram essa etapa da vida tão especial, por dividirem comigo tanto os melhores momentos quanto as noites viradas de estudo e trabalho.

À Gabriela, pelo seu amor, parceria, leveza e apoio constantes nesta jornada.

Ao professor doutor Marco Aurélio de Mesquita, por todo o suporte, atenção e paciência, bem como pelas valiosas orientações durante o desenvolvimento deste Trabalho de Formatura.

Aos meus companheiros de trabalho, por todo o aprendizado e pelo suporte para a realização deste trabalho.

À empresa *Fashion*, por possibilitar a realização deste trabalho, me permitir estudar um tema de grande interesse pessoal e proporcionar uma valiosa oportunidade de aprendizado e desenvolvimento.

Ao Poli *Rats*, por toda parceria, me apresentarem ao Futebol Americano, permitirem que eu jogasse e me desenvolver nesse esporte. Foi uma honra e muito divertido dividir os campos com vocês durante todos esses anos.

"É preciso imaginar Sísifo feliz."

Albert Camus

RESUMO

Este Trabalho de Formatura trata do desafio de dimensionar a primeira compra de novos produtos sem histórico de vendas em uma grande rede de *fast fashion*, buscando uma solução tecnicamente robusta e aderente ao processo real da empresa. Para isso, o estudo articula duas frentes complementares: mapeamento de processos e previsão de demanda. Na primeira frente, foi realizado o mapeamento do macroprocesso de lançamento de novas coleções e, em detalhe, o mapeamento do subprocesso de compras, utilizando a notação BPMN e entrevistas semiestruturadas com um gerente sênior e um gerente pleno. O mapeamento As Is tornou explícitas atividades, decisões, sistemas, papéis e dores do processo, escancarando um processo pouco padronizado e com forte heterogeneidade entre compradores. Os diagramas To Be elaborados no trabalho propõem um fluxo padronizado, com artefatos únicos de dados, responsabilidades claras e pontos específicos de inserção dos modelos preditivos, garantindo a aplicabilidade prática da solução proposta. A segunda frente desenvolve três modelos de previsão por atributos com Python: analogia em diferentes níveis da hierarquia mercadológica, regressão linear e um modelo baseado em *Quantile Regression Forests*. Os modelos são avaliados e comparados em *backtests* com métricas de desempenho de previsão utilizando dados reais da empresa. Os resultados indicam desempenho superior do modelo de *Quantile Regression Forests* em termos de erro, e também mostram que o modelo atende ao requisito dos gestores de poderem ajustar as previsões com base em seus *inputs* e julgamentos pessoais. Os resultados sustentam a recomendação para adoção de um novo método de previsão de demanda, tendo o modelo alcançado um erro percentual absoluto médio ponderado (wMAPE) de 45,1%, bem inferior ao valor de 86,2% da compra praticada nos lançamentos de 2025. Depois da remoção de *outliers*, reduziu-se o wMAPE para 35,5% em comparação aos 56,2% do praticado. Como próximos passos, recomenda-se a condução de um piloto em ambiente real, iniciando pela implementação do processo *To Be* proposto e pela adoção gradual do modelo de *Quantile Regression Forests*, de modo a avaliar seus impactos operacionais, refinar os modelos e orientar uma futura integração sistêmica mais ampla.

Palavras-chave: Previsão de Demanda. Novos Produtos. Mapeamento de Processos. Varejo da Moda. *Fast Fashion*. *Quantile Regression Forests*.

ABSTRACT

This Graduation Project addresses the challenge of sizing the initial purchase of new products, without sales history, in a large fast fashion retail chain, seeking a solution that is technically robust and aligned with the company's actual decision-making process. To this end, the study combines two complementary fronts: process mapping and demand forecasting. On the first front, the macroprocess of launching new collections was mapped and, in detail, the purchasing subprocess, using BPMN notation and semi-structured interviews with a senior manager and a mid-level manager. The As Is mapping makes explicit the activities, decisions, systems, roles, and pain points of the process, revealing a poorly standardized process with strong heterogeneity across buyers. The To Be diagrams developed in the study propose a standardized flow, with unique data artefacts, clear responsibilities, and specific touchpoints for inserting predictive models, thus ensuring the practical applicability of the proposed solution. The second front develops three attribute-based forecasting models in Python: analogy at different levels of the merchandising hierarchy, linear regression, and a model based on Quantile Regression Forests. The models are evaluated and compared in backtests using forecasting performance metrics and real company data. The results indicate superior performance of the Quantile Regression Forests model in terms of forecast error, while also meeting the managers' requirement to be able to adjust forecasts based on their own inputs and judgment. The results support the recommendation to adopt a new demand forecasting method, with the model achieving a weighted mean absolute percentage error (wMAPE) of 45.1%, substantially lower than the 86.2% observed in the current buying practice in the test with the 2025 launches. After removing outliers, the wMAPE is reduced to 35.5%, compared with 56.2% for the current practice. As next steps, a pilot in a real operational environment is recommended, starting with the implementation of the proposed To Be process and the gradual adoption of the Quantile Regression Forests model, in order to assess its operational impacts, refine the models, and guide a broader future system integration.

Keywords: Demand Forecasting. New Products. Process Mapping. Fashion Retail. *Fast Fashion. Quantile Regression Forests.*

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

ARIMA – *AutoRegressive Integrated Moving Average* (modelo autorregressivo integrado de médias móveis).
As Is – Situação atual do processo ("como é").
BPMN – *Business Process Model and Notation* (notação padrão para modelagem de processos de negócio).
MAPE – *Mean Absolute Percentage Error* (erro percentual absoluto médio).
MPE – *Mean Percentage Error* (erro percentual médio).
PLM – *Product Lifecycle Management* (gestão do ciclo de vida do produto).
QRF – *Quantile Regression Forests* (florestas aleatórias para regressão quantílica).
RMSE – *Root Mean Square Error* (raiz do erro quadrático médio).
SKU – *Stock Keeping Unit* (unidade de manutenção de estoque).
To Be – Situação futura desejada do processo ("como deve ser").
wMAPE – *Weighted Mean Absolute Percentage Error* (erro percentual absoluto médio ponderado).

SUMÁRIO

1 INTRODUÇÃO

Neste capítulo, apresenta-se a empresa objeto de estudo e caracteriza-se o problema de dimensionar as compras iniciais de novos produtos. Por motivos de confidencialidade, a companhia estudada será identificada como *Fashion*.

1.1 Contexto

1.1.1 A empresa

A *Fashion* é uma varejista de moda *fast fashion* de capital aberto, com atuação nacional, mais de 250 lojas distribuídas em todos os estados do país e operação de *e-commerce*. Seu posicionamento concentra-se majoritariamente nas classes B e C.

Suas lojas físicas possuem oferta de produtos adaptada ao perfil socioeconômico majoritário da localidade, às condições climáticas regionais e ao tamanho da loja. Essa diversidade de contextos traz desafios de planejamento de sortimento (variedade de produtos oferecida) e de abastecimento da rede, o que acarreta forte pressão por eficiência operacional, especialmente por ser uma companhia de capital aberto.

1.1.2 O setor

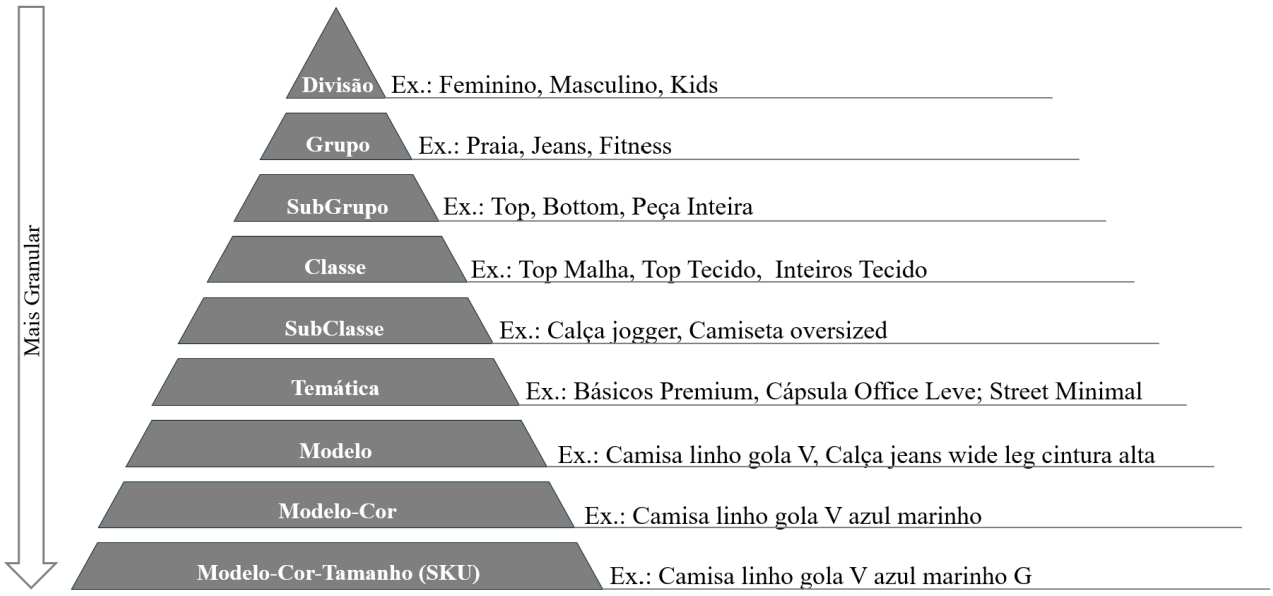
O modelo de negócios do nicho em que a *Fashion* está inserida baseia-se na ampla variedade de *Stock Keeping Unit*(SKU) (em português: Unidade de Manutenção de Estoque, código que identifica cada produto específico, no caso do varejo da moda, modelo-cor-tamanho), lançamentos frequentes, relançamento de produtos com bom desempenho e ciclos de vida curtos.Apenas uma fração reduzida do portfólio é composta por produtos constantes, enquanto a maior parte é composta por peças sazonais de coleções. Essa dinâmica demanda decisões ágeis de compra, alocação e reposição, sob risco de ruptura (indisponibilidade de estoque) ou excesso de estoque com necessidade de remarcações (reduções de preço planejadas para acelerar giro de estoque).

Além disso, a demanda no varejo de moda sofre forte influência de sazonalidade e de efeitos de calendário. Feriados e datas comemorativas como Dia das Mães, São João, Black Friday e Natal geram picos e vales de venda por categoria e por região. Como os ciclos de vida são curtos e as janelas de venda são estreitas, erros na construção de estoque nessas datas produzem impactos severos. Por isso, a previsão precisa capturar componentes sazonais múltiplos por família de produtos.

1.1.3 Macro estrutura interna

Para orientar o planejamento de sortimento e padronizar a comunicação entre áreas, a *Fashion* adota a hierarquia mercadológica (estrutura de agregação do produto) ilustrada na figura 1. Essa hierarquia organiza os produtos em níveis progressivos de detalhamento, dos agrupamentos mais amplos até o SKU, e serve como padrão de referência para análise, reporte e tomada de decisão entre áreas. À medida que se avança para os níveis inferiores, se aumenta o número de categorias e o grau de especificidade, refletindo a diversidade do portfólio da empresa, enquanto no topo concentram-se poucos agrupamentos amplos.

Figura 1: Hierarquia Mercadológica.



Fonte: Autoria própria.

O processo de lançamento de coleções é composto pelas macro etapas apresentadas na figura 2. Conforme ilustrado, o fluxo inicia-se pelo Planejamento Orçamentário, em que são definidos os tetos de gasto por Grupo (nível da hierarquia mercadológica). Na sequência, o planejamento de sortimento especifica a variedade (largura) e a profundidade (quantidade por item) de cada Classe, respeitando as restrições orçamentárias.

A etapa de Criação da Coleção é o momento em que os estilistas propriamente desenham as peças, criando um conjunto de *designs* maior do que o alvo final, de modo que as melhores combinações possam ser escolhidas posteriormente. Em seguida, a Prototipagem com Fornecedores gera peças-piloto para validar caimento, materiais e viabilidade produtiva, bem como para utilizá-las na próxima etapa.. Monta-se uma representação de exposição em loja, selecionam-se as peças que vão compor a coleção e define-se sua alocação por *clusters* de lojas, considerando clima, perfil dos clientes e tamanho das lojas.

Em seguida, a etapa de Sugestão de Compra é responsável pelos cálculos de previsão de demanda e recomendação de compra, levando em consideração os estoques em loja e todos *lead times* da cadeia. Por fim, há a etapa de Negociação e Emissão, em que se acordam condições comerciais, faseamento dos pedidos, tamanho e composição de *packs* (quando aplicável) e prazos, culminando na emissão dos pedidos.

Figura 2: Fluxo de Lançamento de Coleção.

Fonte: Autoria própria.

Concluída a emissão dos pedidos, inicia-se o acompanhamento da execução. Essa fase, denominada *In-Season*, possui seus próprios desafios e envolve monitorar a aderência do plano, aferir resultados e reagir a desvios. Como este Trabalho de Formatura foca na previsão de demanda para suporte à decisão de compra no pré-lançamento (*Pre-Season*), a etapa *In-Season* não é objeto de análise e, portanto, seus processos macros não foram mapeados.

Como a maior parte dos produtos da marca deriva de novas coleções, os lançamentos chegam ao mercado sem histórico. Nesses casos, as decisões de compra, alocação inicial e regras de reposição tendem a apoiar-se em lógicas de similaridade ao longo da hierarquia mercadológica ou em modelos que exploram de forma limitada os atributos disponíveis, por exemplo, todo tipo de característica do produto, como tecido, cor e preço. Diante desse contexto, o problema é enunciado na sequência.

1.2 Problema

Para itens com histórico, a *Fashion* já possui um método robusto de previsão de séries temporais. Já para lançamentos sem histórico, a prática corrente é realizar compras com

base em analises realizadas pelos gerentes do Grupo (da hierarquia mercadológica) que são bastante subjetivas a cada gestor e nada padronizadas. Assim, essa prática não explora de forma sistemática todos os atributos disponíveis na empresa e tampouco investiga as interações entre eles.

Na prática, as limitações descritas podem se converter em perdas na casa dos milhões de reais ao longo de uma coleção. Quando a previsão subdimensiona a demanda, há ruptura: perdem-se vendas e margem, clientes migram para concorrentes e o espaço de loja fica ocioso. Quando a previsão superdimensiona a demanda, forma-se excesso de estoque: capital imobilizado, necessidade de remarcações (gerando compressão de margem), maior custo logístico e risco de obsolescência. Em uma rede com centenas de lojas e milhares de SKUs, pequenos vieses sistemáticos, multiplicados por curtos ciclos de vida e janelas de sazonalidade fortes, degradam o giro, a cobertura e o retorno sobre o investimento, comprometendo todo o resultado da empresa.

Assim, apresenta-se a enunciação clara do problema da *Fashion* a ser estudado: "como prever melhor a quantidade da primeira compra de novos produtos com precisão proporcional aos modelos mais atuais da literatura, produzindo estimativas fiéis ao desempenho efetivo de vendas, com o objetivo de gerar retorno financeiro à empresa". O desempenho deve ser mensurado por métricas de acurácia adequadas, e a previsão deve ser estruturada em conformidade com o nível hierárquico e a frequência de decisão atualmente adotados na empresa.

1.3 Objetivos

O objetivo deste trabalho é melhorar a decisão de compra dos itens de cada nova coleção a partir da revisão dos processos de planejamento e dos modelos de previsão de demanda. A proposta busca produzir estimativas mais aderentes ao comportamento real de vendas de novos produtos, de modo a reduzir riscos de ruptura e de excesso de estoque e, consequentemente, melhorar giro, cobertura e margem.

Além da dimensão quantitativa, o trabalho também tem como objetivo mapear de forma estruturada o processo de lançamento e compra de novos produtos, explicitando atividades, decisões, sistemas e responsabilidades, de modo a propor um fluxo futuro que seja operacionalmente viável e útil na prática. Ao articular o diagnóstico do processo com a construção dos modelos preditivos, busca-se garantir que a solução recomendada possa ser efetivamente incorporada à rotina da empresa, com pontos claros de inserção das pre-

visões, entregas bem definidas entre áreas e requisitos mínimos para sua implementação.

Para isso, pretende-se realizar um estudo consistente sobre os modelos atualmente utilizados e discutidos na literatura recente, explicitando seus fundamentos, pressupostos, benefícios e limitações no contexto do varejo de moda. Em seguida, esses modelos serão implementados e comparados de forma aplicada sobre dados da *Fashion*, respeitando a mesma granularidade e frequência de decisão adotadas internamente. A avaliação será conduzida por meio de *backtests* (testes retrospectivos de uma estratégia) e métricas de acurácia e viés, de forma a permitir uma interpretação cuidadosa dos resultados, com ênfase na relevância operacional para dimensionar a primeira compra.

Em todo esse processo, também se busca analisar o que de fato explica a demanda, investigando sistematicamente variáveis de calendário, como sazonalidades, feriados nacionais e regionais e eventos promocionais, e todos os atributos de produto disponíveis, tais como tecido, cor, modelagem e faixa de preço. A análise contemplará a seleção de variáveis e a avaliação de sua importância preditiva, distinguindo fatores que agregam valor daqueles que não contribuem para a previsão. Como resultado esperado, pretende-se indicar a abordagem com melhor desempenho e oferecer recomendações objetivas para seu uso na decisão de compra inicial que seja aderente e implementável no processo de compras da empresa.

1.4 Relevância

Este trabalho endereça um problema central de Engenharia de Produção: gestão de estoque por meio da utilização de métodos de previsão de demanda. Embora não trate diretamente de um ambiente industrial tradicional, objeto de estudo principal do engenheiro de produção, esta é uma previsão aplicada a um contexto particularmente desafiador: o lançamento de novos produtos sem histórico em um tipo de varejo com produtos com ciclo de vida curto. Além disso, esse Trabalho de Formatura está ancorado em temas clássicos da área, planejamento de demanda, gestão de estoques, alocação de recursos e avaliação de desempenho, e traduz esses fundamentos para um cenário contemporâneo de alta incerteza. Ao propor, testar e comparar um método por atributos com métricas alinhadas ao negócio, o projeto produz conhecimento aplicável e replicável para a tomada de decisão do usuário final na *Fashion*.

A relevância é ainda maior devido ao modelo de negócios da empresa, no qual parte relevante do portfólio é composta por lançamentos sem histórico. Nesse contexto, a previsão

sem dados históricos torna-se recorrente e de alto impacto: decisões de compra, alocação inicial e regras de reposição precisam ser tomadas com pouca ou nenhuma informação. Erros nessa etapa propagam-se rapidamente pela rede, afetando indicadores-chave e, dado o posicionamento em classes de maior sensibilidade a preço, elevam o risco de perda de vendas e compressão de margem. Assim, aprimorar a capacidade preditiva pré-lançamento tem efeito direto sobre o resultado financeiro e a eficiência operacional.

1.5 Estrutura

Este trabalho está organizado em cinco capítulos. No Capítulo 1, como já desenvolvido, apresenta-se o contexto da empresa e do setor, o problema de pesquisa, os objetivos e a relevância do estudo. O Capítulo 2 reúne o referencial teórico, descrevendo o método de revisão bibliográfica e consolidando a literatura sobre previsão de demanda para novos produtos no varejo de moda, correção de vendas em ruptura, simulação, métricas de desempenho e mapeamento de processos. O Capítulo 3 detalha a metodologia adotada em duas frentes: a revisão do processo de compras e o estudo de simulação para comparação dos modelos de previsão de demanda. O Capítulo 4 apresenta os resultados do mapeamento As Is e To Be, bem como os resultados empíricos da aplicação dos métodos de previsão (analogia, regressão linear e *Quantile Regression Forests*) sobre os dados da *Fashion*. Por fim, o Capítulo 5 apresenta uma síntese do projeto, discute suas limitações e aponta desdobramentos e oportunidades futuras.

2 REVISÃO TEÓRICA

Este capítulo apresenta o referencial teórico que embasa o presente trabalho. Inicia-se pela descrição do método de revisão bibliográfica adotado, em seguida, consolida-se a literatura sobre previsão de demanda para novos produtos no varejo de moda, e, por fim, apresenta-se a fundamentação do mapeamento de processos.

2.1 Método de Revisão

A primeira etapa da revisão bibliográfica foi conduzida na base *Scopus*. Ela foi escolhida por ser um indexador de referência internacional, de grande abrangência e com rigoroso controle de qualidade editorial, condições que favorecem reprodutibilidade, transparência e atualidade do acervo. Dado seu escopo global e o alinhamento às tendências internacionais, adotou-se o inglês como idioma de busca e triagem, ampliando a cobertura de conteúdos relevantes e reduzindo vieses de seleção por idioma.

Assim, a estratégia de busca iniciou-se pela identificação e leitura de artigos de revisão da literatura. Essa opção foi adotada por dois motivos principais: oferecer uma visão abrangente do estado da arte e evolução ao longo do tempo, e servir de base para a seleção de estudos primários aderentes ao escopo deste trabalho. Etapa que é essencial ao projeto, pois constitui o ponto de partida para a aplicação da técnica de *snowball* (expansão da bibliografia a partir das referências das obras consultadas), permitindo ampliar de forma sistemática o conjunto de fontes relevantes e confiáveis.

Com relação à formulação da pesquisa na *Scopus*, estruturou-se em três núcleos principais: (i) uma referência à previsão de demanda; (ii) outra à novos produtos (ausência de histórico); e (iii) a última ao contexto do setor de moda. Testes preliminares mostraram que a remoção de qualquer um desses núcleos levava a resultados indesejados: sem o primeiro, predominavam discussões não quantitativas ou de planejamento genérico; sem o segundo, a busca retornava majoritariamente métodos para séries com histórico; sem o terceiro, emergiam estudos de outros setores, com baixa transferibilidade ao varejo de

moda. Além disso, foi adicionado um filtro para limitar trabalhos com mais de 10 anos da data de publicação para garantir atualidade e a consulta empregou operadores booleanos e de proximidade para não deixar de capturar estudos relevantes por detalhes de redação, como segue:

```
TITLE-ABS-KEY ( ( forecast* OR ‘‘demand forecast*’’ OR ‘‘sales forecast*’’
) AND ( ( new W/3 product* ) OR ( ( zero OR no OR without OR lack OR absence
) W/3 ( ‘‘past data’’ OR ‘‘sales data’’ OR ‘‘demand data’’ OR histor*
) ) ) AND ( ‘‘fashion*’’ OR ‘‘apparel’’ OR ‘‘clothing’’ OR ‘‘garment*’’
) ) AND PUBYEAR > 2014
```

Busca executada em 3 set. 2025.

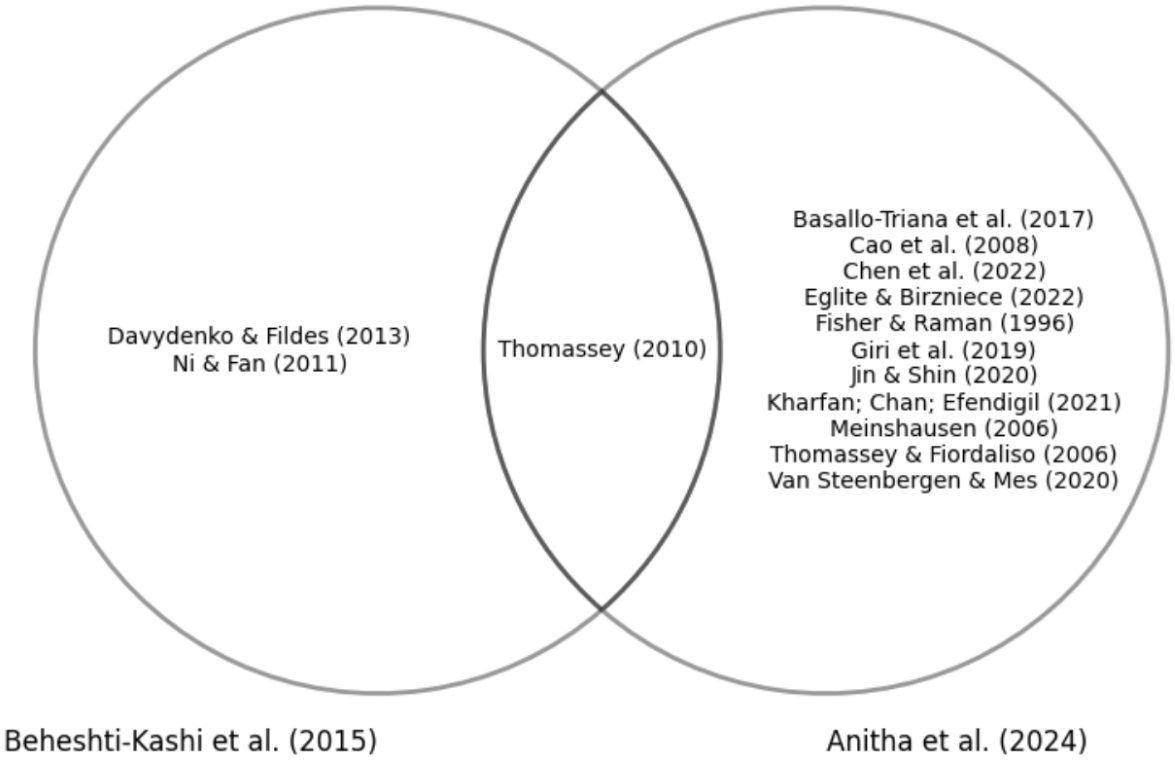
Nessa consulta, aplicou-se um filtro por tipo de documento de revisão para priorizar sínteses de alto nível e, com isso, obtiveram-se duas revisões diretamente pertinentes ao escopo deste trabalho: (i) *A survey on retail sales forecasting and prediction in fashion markets*, de Beheshti-Kashi et al. (2015), uma revisão de literatura, com 68 citações na própria base no momento da busca; e (ii) *Demand Forecasting New Fashion Products: A Review Paper*, de Anitha e Neelakandan (2025) no *Journal of Forecasting*, e já contabilizando 2 citações até o dia da coleta.

Tomadas em conjunto, essas duas fontes oferecem uma visão completa do campo de estudo: a de 2015 consolida fundamentos e linhas de pesquisa estruturantes na previsão de demanda em moda, enquanto a de 2025 atualiza o estado da arte especificamente para novos produtos de moda, discutindo frentes emergentes e soluções recentes. Esse pareamento equilibra robustez conceitual e atualidade temática, ampliando a cobertura sem perder casos relevantes por detalhes de redação, e fornece ponto de partida robusto para a técnica de *snowball*.

A partir dessas duas revisões, procedeu-se à expansão sistemática da bibliografia por *snowball*, rastreando referências-chave e trabalhos que as citaram. Esse processo é responsável por sustentar a parte teórica e metodológica e é ilustrado pelo Diagrama de Venn da Figura 3.

Além disso, dada a necessidade de fundamentar não apenas a aplicação dos métodos de previsão de demanda de novos produtos, mas também a sua formulação técnica, buscou-se recorrer a obras clássicas da literatura de estatística e aprendizado de máquina. Os temas foram sendo identificados e aprofundados no decorrer da pesquisa, o que reforçou a importância de consultar referências extremamente sólidas, capazes de sustentar con-

Figura 3: Diagrama de Venn Fonte dos Artigos.



Fonte: Autoria própria.

ceitualmente cada abordagem. A escolha dessas fontes teve como critério a relevância histórica, a recorrência em revisões bibliográficas e o reconhecimento internacional como obras consolidadas em seus respectivos campos. Dessa forma, além dos artigos relacionados à proposta de melhoria, incorporaram-se ao trabalho referências consideradas clássicas ou até mesmo fundadoras dos métodos estudados, de modo a assegurar uma base teórica robusta.

Nesse sentido, para a regressão linear múltipla, adotou-se o manual de Montgomery, Peck e Vining (2012), amplamente reconhecido como referência fundamental em estatística aplicada. Para a clusterização e modelos baseados em árvores de decisão, foram mobilizadas as contribuições de Breiman et al. (1984), autores da formulação original do método CART, e de Breiman (2001), responsável pela proposta inaugural do *Random Forest*, além da obra de Bishop (2006), referência em reconhecimento de padrões e aprendizado de máquina, que apresenta formalmente o algoritmo de *k-means*. Essas referências, por seu caráter seminal, fornecem sustentação teórica complementar às aplicações setoriais, fortalecendo o embasamento metodológico do presente trabalho.

Na mesma linha, para a construção do modelo de simulação utilizado na avaliação dos

impactos das alternativas de previsão e reposição, adotou-se como referência o *Handbook of Simulation*, de Banks (1998). Essa obra reúne de forma sistemática os princípios de concepção, modelagem, validação e análise de modelos de simulação discreta de eventos, oferecendo diretrizes práticas para etapas como definição dos objetivos do estudo, representação da dinâmica do sistema, especificação das entradas e delineamento dos experimentos computacionais. Com isso, o desenvolvimento do modelo de simulação seguiu boas práticas amplamente consolidadas na literatura, assegurando maior rigor na interpretação dos resultados obtidos a partir dos cenários testados,

Além disso, utilizou-se o livro BPMN Method and Style (SILVER, 2011) como base para a fundamentação teórica da modelagem de processos. A obra foi escolhida por oferecer uma visão clara sobre os fundamentos do BPMN, explicando não apenas como a notação funciona, mas também por que ela é útil e quais benefícios agrega em termos de padronização, clareza e comunicação entre áreas. Sua metodologia contribuiu para compreender de forma estruturada a lógica dos processos e orientar a elaboração dos diagramas apresentados.

2.2 Previsão de demanda para novos produtos

2.2.1 Previsão de Demanda no processo de Compras no Varejo

A previsão de demanda ocupa posição central no macroprocesso de compras do varejo: ela antecede e alimenta a decisão de compra, conectando o planejamento de demanda ao planejamento de suprimentos (EGLITE; BIRZNIECE, 2022). Dentro de um fluxo típico, desenvolvimento do produto, *sourcing* e compra, produção, distribuição e venda, a previsão é o gatilho que transforma expectativas de mercado em compromissos operacionais e financeiros ao longo da cadeia. Em termos práticos, isso significa emitir pedidos e planejar reposições com a antecedência necessária para absorver tempos de produção, transporte e controle de qualidade (THOMASSEY, 2010).

No eixo econômico-financeiro, a previsão baliza o uso do orçamento e do capital de giro. Estimativas mais confiáveis orientam quanto comprar sem imobilizar caixa além do necessário, reduzindo dependência de compras emergenciais e liberando recursos para outras frentes da operação. Como mostram Steenbergen e Mes (2020), previsões de demanda bem calibradas para novos produtos são essenciais para decisões de capacidade, compras e controle de estoques, pois evitam tanto rupturas quanto excesso de estoque e geram economias de inventário, com impacto direto na lucratividade e no nível de serviço

(STEENBERGEN; MES, 2020).

Esse efeito econômico emerge do *trade-off* ruptura e excesso. Subestimar a demanda eleva vendas perdidas e deteriora a margem; simulações indicam aumento de até 11% nas perdas e compressão da margem bruta entre 8% e 14% quando a previsão inicial erra (THOMASSEY, 2010). Por outro lado, superestimar resulta em sobras e remarcações que corroem o lucro (JIN; SHIN, 2020). Ao administrar esse equilíbrio, previsões mais precisas reduzem simultaneamente remarcações e indisponibilidades, com estudos de caso reportando ganhos lucro quando a compra é orientada por previsões acuradas (THOMASSEY, 2010). Em complemento, maior acurácia aumenta o giro, libera espaço e reduz custos operacionais (JIN; SHIN, 2020).

Do ponto de vista do nível de serviço e da disponibilidade, compras orientadas por previsão sustentam o *fill rate* nos canais de venda e reduzem vendas perdidas. O nível de serviço corresponde à proporção dos ciclos de reposição em que não houve ruptura de estoque, funcionando como métrica de confiabilidade do abastecimento. Já o *fill rate* mede a fração da demanda efetivamente atendida a partir do estoque disponível, isto é, quantos pedidos foram completados sem falta de produto. Métricas como essas tornam-se metas operacionais e, com modelos que quantificam incerteza, é possível dimensionar estoques de segurança e pontos de ressuprimento compatíveis com níveis-alvo (STEENBERGEN; MES, 2020).

Há ainda o cuidado metodológico com a demanda censurada. Em redes de moda com sortimento enxuto e níveis de estoque reduzidos por loja, as vendas observadas tendem a ficar muito próximas ao estoque disponível, de modo que esgotamentos fazem com que a quantidade vendida não reflita a demanda potencial. Chen et al. (2022), por exemplo, mostram que esse efeito de censura dificulta a estimação da procura real e propõem um modelo específico para aliviar o impacto da demanda censurada e das vendas perdidas sobre as previsões e decisões de estoque

Em cadeias com *lead times* relevantes, isto é, de grandes intervalos totais entre a emissão do pedido e a disponibilidade do produto para venda, a previsão de demanda assume papel ainda mais crítico. Quanto maior o *lead time*, maior é o intervalo entre a decisão de compra e o momento em que as vendas efetivamente ocorrem, ampliando a exposição a variações de mercado e a incertezas no comportamento do consumidor. Como destacam Fisher e Raman (1996), em contextos de produção com prazos longos, os pedidos precisam ser definidos antes do início da temporada de vendas, quando a demanda ainda é desconhecida, de modo que eventuais imprecisões na previsão têm impacto significativo

sobre a rentabilidade

A previsibilidade também é moeda de troca na negociação com fornecedores. Visibilidade de demanda favorece condições comerciais (volumes, prazos, janelas de entrega) e coordenação ao longo da cadeia, reduzindo a necessidade de compras de urgência (THOMASSEY, 2010). Exemplos de operadores globais mostram que compartilhar previsões permite reservar insumos genéricos e bloquear capacidade antes da definição final dos pedidos, ampliando a flexibilidade sem elevar o risco para as partes (CAO et al., 2008). Quando a previsão é mais precisa, o varejista consegue alinhar volumes aos pedidos mínimos com menor exposição e o fornecedor planeja capacidade com maior eficiência (THOMASSEY, 2010).

No sortimento, a previsão orienta simultaneamente a largura, isto é, quantas referências manter em linha, e a profundidade, quanto comprar de cada uma. Em níveis agregados de produtos, séries históricas ajudam a dimensionar o volume total. Já no nível de item, sobretudo quando se trata de lançamentos com histórico escasso ou inexistente, a literatura recente destaca o uso de modelos que exploram atributos de produto (por exemplo, preço, material e cor) e técnicas de regressão ou aprendizado de máquina para estimar a demanda (ANITHA; NEELAKANDAN, 2025).

Em operações multi-loja e omnicanal, previsões consistentes por local e canal evitam desequilíbrios e transferências internas. Essa segmentação por região permite que decisões de compra e de distribuição/reposição sejam alinhadas às diferenças de demanda entre mercados regionais, já que as previsões servem de base para planos de distribuição e reposição (CHEN et al., 2022).

Por fim, previsões que aproximam oferta da demanda real reduzem riscos, fortalecem *compliance* e sustentam a sustentabilidade da cadeia. Menos sobras significam menos liquidações agressivas e menor probabilidade de descarte, com ganhos econômicos e ambientais; além disso, cooperação baseada em previsões confiáveis melhora a eficiência global e a resiliência de longo prazo (ANITHA; NEELAKANDAN, 2025).

2.2.2 Conceitos e desafios da previsão de demanda

O horizonte de previsão é uma dimensão central no planejamento da demanda, pois define o período de tempo para o qual as estimativas são geradas e, consequentemente, o tipo de decisão que irão suportar (THOMASSEY, 2010). De modo geral, distinguem-se dois grandes grupos: as previsões de longo prazo, voltadas a decisões estratégicas e táticas,

e as de curto prazo, orientadas a ajustes operacionais (NI; FAN, 2011).

As previsões de longo prazo são elaboradas com meses ou até um ano de antecedência e têm como função principal sustentar decisões de grande impacto e longo *lead time*, como o planejamento de capacidade, a alocação inicial de recursos e a negociação com fornecedores. A necessidade desse horizonte é reforçada pelo fato de que, em cadeias de suprimento complexas e globalizadas, o tempo de fabricação e entrega de produtos costuma ser elevado e pouco flexível (KHARFAN; CHAN; EFENDIGIL, 2021).

Por sua vez, as previsões de curto prazo abrangem horizontes de dias ou poucas semanas e são utilizadas para apoiar decisões mais imediatas, como reposições, faseamento de pedidos e ajustes finos na distribuição. Seu diferencial é a incorporação de informações recentes de mercado, o que permite correções rápidas diante de mudanças na demanda (THOMASSEY, 2010).

Esse desdobramento leva a um trade-off fundamental: previsões elaboradas com grande antecedência garantem maior tempo de resposta, mas se baseiam em dados mais incertos. Já as previsões próximas ao consumo utilizam informações recentes e tendem a alcançar maior acurácia, porém deixam pouco espaço para replanejamento. Esse arranjo é explicitado por Ni e Fan (2011), que propõem um modelo em duas etapas, combinando uma previsão de longo prazo, voltada ao planejamento de produção e alocação inicial, com previsões de curto prazo, próximas ao consumo, usadas para ajustar quantidades e redistribuir estoques entre lojas. Assim, uma gestão eficaz da demanda exige a combinação entre diferentes horizontes, de modo a equilibrar robustez estratégica e flexibilidade operacional.

A granularidade da previsão refere-se ao grau de detalhe em que a demanda é estimada, podendo variar desde níveis altamente agregados, como o total de vendas da empresa ou de uma categoria, até o nível mais desagregado possível, como o SKU, representando um item específico em suas variações de cor e tamanho (THOMASSEY, 2010).

É um princípio estatístico bem estabelecido que previsões em níveis agregados tendem a ser mais estáveis e precisas do que previsões em níveis desagregados. Isso ocorre porque, ao agrupar diversos itens, flutuações individuais e ruídos aleatórios tendem a se compensar, suavizando a série histórica. Então, essa maior estabilidade torna as previsões agregadas particularmente adequadas para decisões de caráter estratégico ou tático, como o planejamento de capacidade, a definição de metas globais ou a negociação de volumes totais com fornecedores.

Por outro lado, previsões detalhadas em nível de SKU, ainda que mais incertas devido

à volatilidade intrínseca de séries muito específicas, são indispensáveis para decisões operacionais. Atividades como reposição de estoques e programação de pedidos dependem de estimativas detalhadas por item e ao longo do tempo, e não apenas de um valor agregado para todo o período (STEENBERGEN; MES, 2020).

Assim, a escolha da granularidade envolve um trade-off entre estabilidade estatística e aplicabilidade prática. Porém, a questão central não é apenas optar por previsões agregadas ou detalhadas, mas alinhar a granularidade da previsão à granularidade do problema enfrentado pela empresa. Em outras palavras, as previsões devem refletir o nível em que as decisões são tomadas, seja estratégico, tático ou operacional, buscando replicar essa estrutura no processo de previsão (THOMASSEY, 2010).

A discussão sobre granularidade da previsão leva naturalmente ao tema da hierarquia de produtos, já que os diferentes níveis estão organizados em uma estrutura progressiva. Essa hierarquia classifica os itens de um portfólio em níveis sucessivos, como categorias, subcategorias e unidades individuais de venda, o SKU (THOMASSEY, 2010). O principal papel dessa estrutura é alinhar decisões que ocorrem em diferentes níveis da organização. Enquanto decisões estratégicas e táticas se apoiam em previsões mais agregadas, a execução operacional exige estimativas em maior detalhe, no nível de item (NI; FAN, 2011).

Diante da necessidade de alinhar previsões mais agregadas, que apoiam decisões estratégicas e táticas, com estimativas detalhadas no nível de SKU, usadas na operação, muitas organizações deixam de depender de um único nível ou modelo de previsão. Em vez disso, utilizam previsões em diferentes níveis da hierarquia de produtos e combinam múltiplos modelos e abordagens, explorando melhor a informação disponível (ANITHA; NEELAKANDAN, 2025).

A discussão sobre granularidade e hierarquia da previsão evidencia que o nível em que a demanda é analisada influencia diretamente a tomada de decisão. Outro aspecto igualmente relevante são os padrões temporais, que moldam a evolução das vendas ao longo do tempo. Entre esses padrões, destaca-se a sazonalidade, uma das características mais marcantes e desafiadoras na previsão de demanda (THOMASSEY, 2010).

A sazonalidade manifesta-se em ciclos recorrentes, que podem ser anuais, mensais ou semanais, e deve ser necessariamente considerada em qualquer sistema de previsão robusto. Esses movimentos regulares costumam ser impulsionados por fatores de calendário, como feriados e datas comemorativas, bem como por condições climáticas que influenciam temporariamente o comportamento de consumo (THOMASSEY, 2010). No

varejo, a presença de sazonalidade é particularmente evidente, já que grandes volumes de vendas se concentram em períodos específicos, como campanhas promocionais ou datas festivas (KHARFAN; CHAN; EFENDIGIL, 2021).

É fundamental distinguir a sazonalidade previsível dos eventos imprevisíveis. A primeira inclui eventos fixos e conhecidos com antecedência, como o Natal em dezembro, que podem ser incorporados diretamente nos modelos de previsão e utilizados no planejamento de estoques e promoções (THOMASSEY, 2010). Já os choques imprevisíveis resultam de fatores exógenos e incertos, como variações climáticas fora do padrão ou crises econômicas súbitas, cuja ocorrência e intensidade são difíceis de antecipar, mas que podem alterar significativamente o padrão de consumo.

Incorporar a sazonalidade à previsão significa, em primeiro lugar, modelar explicitamente os padrões recorrentes de calendário. Isso pode ser feito por meio da decomposição das séries históricas ou da introdução de variáveis sazonais em modelos clássicos de séries temporais, como os baseados em suavização exponencial ou regressão (BEHESHTI-KASHI et al., 2015). Além disso, ciclos associados a fatores externos, como clima ou eventos macroeconômicos, podem ser tratados como variáveis explicativas adicionais em modelos híbridos, melhorando a acurácia das estimativas (NI; FAN, 2011). Por fim, para lidar com choques imprevisíveis, são recomendadas revisões frequentes da previsão e o monitoramento contínuo de variáveis externas, de modo a ajustar as estimativas quando ocorrerem desvios inesperados.

Além dos padrões sazonais e cíclicos, outra dimensão crítica da previsão de demanda está relacionada às estratégias comerciais de preço e promoção. Enquanto a sazonalidade decorre de fatores recorrentes e, em grande parte, exógenos, as promoções são geradas pelo próprio varejista e podem modificar substancialmente o comportamento de compra em curtos intervalos de tempo (THOMASSEY, 2010).

O efeito das promoções sobre a demanda é multifacetado. Em primeiro lugar, existe a dificuldade de separar a demanda "natural" daquela induzida pela ação promocional. Descontos e liquidações podem concentrar vendas em um período específico sem aumentar o volume total em longo prazo, tornando complexo isolar o impacto real dessas variáveis na série histórica (THOMASSEY, 2010).

Outro desafio é o problema da elasticidade-preço, ou seja, a sensibilidade da demanda às variações de preço. Determinar até que ponto um desconto gera incremento efetivo de consumo, em vez de apenas antecipar compras, exige medições consistentes que nem sempre estão disponíveis. Estimar elasticidades é particularmente difícil em situações de

dados limitados, mas continua sendo central para avaliar o efeito de diferentes políticas de preço (CARO; GALLIEN, 2012).

Todos esses elementos, granularidade, hierarquia, sazonalidade e preço, moldam o grau de complexidade da previsão no varejo em geral. Contudo, quando se trata de lançamentos, essa complexidade se intensifica, pois os novos produtos introduzem desafios adicionais que não podem ser capturados apenas por séries temporais, como ARIMA (*AutoRegressive Integrated Moving Average*, em português, modelo autorregressivo integrado de médias móveis) ou *Holt-Winters*, que dependem de séries de vendas suficientemente longas para identificar padrões e projetar tendências (ANITHA; NEELAKANDAN, 2025). Esses desafios são discutidos a seguir.

2.2.3 Contexto e métodos de previsão de demanda de novos produtos

A previsão de demanda para novos produtos apresenta especificidades que extrapolam os fundamentos discutidos anteriormente. Em coleções sazonais e lançamentos frequentes, a ausência de histórico de vendas e o ciclo de vida reduzido aumentam significativamente a incerteza da previsão e o risco de erros operacionais, pois tanto o excesso de estoque quanto a ruptura tornam-se mais prováveis nesse cenário (BASALLO-TRIANA; RODRÍGUEZ-SARASTY; BENITEZ-RESTREPO, 2017).

Diante dessa limitação, a literatura aponta para a necessidade de modelos alternativos às séries temporais. Nesse contexto, duas estratégias ganham destaque: utilizar atributos do produto como variáveis explicativas, como por exemplo cor, estilo, material, preço, categoria ou duração esperada do ciclo de vida, (KHARFAN; CHAN; EFENDIGIL, 2021), e recorrer a itens análogos, transferindo informações de produtos semelhantes para apoiar a previsão de novos SKUs (STEENBERGEN; MES, 2020).

De acordo com Thomassey (2010), do ponto de vista prático, isso significa que a previsão em ambientes de ciclo de vida curto deve começar pela definição de uma base de atributos descritivos confiáveis para cada produto. Em seguida, é necessário identificar grupos ou famílias de itens semelhantes, cujos perfis de vendas anteriores possam ser usados como referência. A previsão deixa, assim, de ser um exercício puramente de séries temporais para se tornar também um problema de classificação e regressão por atributos, em que o comportamento de novos SKUs é inferido a partir de padrões observados em produtos comparáveis (STEENBERGEN; MES, 2020). Dessa forma, o conceito de ciclo de vida curto não apenas descreve uma característica do varejo, mas orienta diretamente

a prática preditiva, exigindo que o foco da modelagem esteja tanto nos dados históricos disponíveis quanto nas informações descritivas dos produtos.

Enquanto o ciclo de vida curto aumenta a incerteza das previsões, a introdução de novos produtos sem histórico de vendas expõe a operação a dificuldades ainda maiores. Esse desafio, conhecido como problema do início sem dados ou *cold start*, é recorrente em diferentes setores.

Nessa linha, a literatura sobre previsão de novos produtos destaca abordagens que combinam informações de produtos análogos com atributos descritivos, partindo do princípio de que itens semelhantes em características físicas ou comerciais tendem a apresentar padrões de demanda próximos. Assim, o problema passa a ser formulado como um exercício de classificação ou regressão a partir desses atributos (ANITHA; NEELAKANDAN, 2025).

Os atributos mais recorrentes incluem aspectos físicos, como cor e material, e variáveis comerciais, como preço, categoria ou duração esperada do ciclo de vida (KHARFAN; CHAN; EFENDIGIL, 2021). Além disso, estudos destacam o papel de dados exógenos, como calendário de lançamentos, tendências de mercado ou indicadores econômicos, que podem enriquecer a previsão ao fornecer contexto adicional mesmo sem observações históricas do item em questão (ANITHA; NEELAKANDAN, 2025).

Dessa forma, este trabalho segue para a discussão de métodos específicos de previsão aplicados a novos produtos. São apresentados, em primeiro lugar, modelos que servirão como base de comparação, e, em seguida, métodos mais recentes que refletem as contribuições mais atuais da literatura sobre o tema.

2.2.3.1 Regressão Linear

A regressão linear múltipla constitui um dos métodos estatísticos mais tradicionais aplicados à previsão de demanda. Seu princípio básico consiste em estabelecer uma relação linear entre a variável dependente, no caso, a demanda de um produto, e um conjunto de variáveis independentes que representam seus atributos, como preço, categoria ou estação de lançamento. Trata-se de uma técnica amplamente utilizada em diferentes setores por sua simplicidade e transparência. Em muitos dos estudos analisados por Anitha e Neelakandan (2025), a regressão linear é empregada justamente como modelo de referência servindo de comparação para métodos mais sofisticados.

No contexto do varejo de moda, Chen et al. (2022) aplicaram a regressão linear como

baseline para comparação com modelos mais sofisticados, como o *Gradient Boosting Decision Tree*. A principal virtude desse método reside em sua interpretabilidade, uma vez que os coeficientes estimados permitem quantificar diretamente o efeito de cada atributo sobre as vendas projetadas. Por outro lado, a suposição de linearidade entre variáveis limita sua capacidade de capturar interações complexas e efeitos não lineares (ANITHA; NEELAKANDAN, 2025).

A formulação geral do modelo pode ser expressa pela seguinte equação de acordo com Montgomery, Peck e Vining (2012):

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \tag{2.1}$$

em que $\hat{y}_i$ representa a demanda prevista para o produto i; β_0 é o intercepto; β_j são os coeficientes que medem o impacto de cada atributo x_{ij}; e ε_i corresponde ao erro aleatório não explicado pelo modelo.

O procedimento de aplicação envolve a definição das variáveis explicativas e da variável resposta, a codificação de atributos categóricos por meio de variáveis *dummies*, variáveis que assumem valor 0 ou 1 para indicar a presença de uma categoria específica, a aplicação do modelo feita pelo método dos mínimos quadrados ordinários e a avaliação da acurácia por métricas de erro. Assim, a previsão de um produto novo é realizada pela inserção de seus atributos na equação estimada, gerando como saída um valor numérico de vendas esperadas (MONTGOMERY; PECK; VINING, 2012).

Entre suas principais vantagens, destacam-se a simplicidade de aplicação e a clareza interpretativa. No entanto, o modelo mostra desempenho limitado quando aplicado a contextos em que a relação entre variáveis não é estritamente linear. Nesse sentido, em Chen et al. (2022) a regressão linear é utilizada como método de referência para comparação com modelos mais sofisticados, os quais apresentam melhor desempenho em termos de acurácia ao explorar relações mais complexas entre atributos de produto, estoque e vendas.

2.2.3.2 Previsão por Analogia

A previsão por analogia é uma técnica amplamente utilizada para estimar a demanda de produtos novos, especialmente em setores nos quais o histórico de vendas é inexistente ou insuficiente (BASALLO-TRIANA; RODRÍGUEZ-SARASTY; BENITEZ-RESTREPO, 2017). A lógica central consiste em identificar produtos lançados anteriormente que apresentem atributos semelhantes ao item em análise, tais como categoria, faixa

de preço, composição, canal ou estação de lançamento, e utilizar seu desempenho como referência para projetar o comportamento do novo produto (STEENBERGEN; MES, 2020). Essa abordagem é particularmente relevante em contextos de ciclos de vida curtos, nos quais a substituição frequente de coleções torna inviável o acúmulo de séries temporais extensas.

No contexto do varejo de moda, uma implementação particularmente simples desse princípio consiste em utilizar um produto "representativo" da família como referência para todos os novos itens. Thomassey e Fiordaliso (2006) descreve esse procedimento sob a forma de um previsor baseado no perfil médio de vendas da família (*mean profile predictor*), que é empregado como referência para avaliar o desempenho de métodos mais sofisticados de previsão para novos produtos.

O ponto de partida é a construção de perfis de vendas normalizados para os produtos históricos de uma mesma família. Para cada artigo, as vendas semanais ao longo da estação são transformadas em uma curva de perfil por meio de dois ajustes sucessivos: normalização do volume, dividindo-se as vendas semanais pelo total vendido na estação, e normalização do tempo, reescalonando-se a duração da vida útil de cada produto para um horizonte comum de 52 semanas (THOMASSEY; FIORDALISO, 2006). O resultado é um conjunto de curvas comparáveis, que descrevem apenas a forma relativa do ciclo de vida de vendas, independentemente do nível absoluto de demanda e da duração real em loja.

A partir desses perfis normalizados, o previsor por perfil médio é obtido por meio da média, em cada semana normalizada, das vendas relativas de todos os produtos da família. Essa curva média passa a representar o comportamento "típico" da família e é utilizada como substituto analógico para qualquer novo artigo para o qual ainda não exista histórico de vendas (THOMASSEY; FIORDALISO, 2006). Assim, a previsão do novo produto corresponde ao perfil médio da família, eventualmente reescalonado para a duração específica planejada para aquele item.

Do ponto de vista operacional, o método exige apenas dados históricos básicos de vendas semanais e a classificação dos produtos por família, dispensando o uso explícito de atributos descritivos ou modelos paramétricos complexos. Por essa razão, Thomassey e Fiordaliso (2006) destaca que o perfil médio da família reflete a prática corrente em muitas empresas do setor e serve como referência natural para comparar abordagens mais avançadas. Nos experimentos apresentados pelo autor, o previsor por perfil médio é utilizado como baseline na avaliação de um sistema híbrido de previsão, permitindo

quantificar os ganhos de precisão obtidos quando, em vez de um único perfil médio, são considerados múltiplos perfis prototípicos de venda associados às características dos produtos.

2.2.3.3 *DemandForest*

O método desenvolvido por Steenbergen e Mes (2020) já acumula dezenas de citações em bases como a *Scopus* desde sua publicação no periódico *Decision Support Systems*, indicando rápida difusão da abordagem na literatura de previsão de novos produtos e, por isso, julgou-se um método interessante para este trabalho. No estudo, o *DemandForest* é aplicado a cinco empresas que atuam em diferentes segmentos: varejo de utilidades domésticas, *e-commerce* de iluminação, louças sanitárias, peças de máquinas agrícolas e ferramentas de jardim. O método é aplicado sempre em contextos em que há famílias de itens relativamente comparáveis dentro de cada empresa (combinações de categoria, faixa de preço,marca, canal, entre outros), de modo que novos SKUs possam ser relacionados a análogos históricos por meio de seus atributos de produto.

O método *DemandForest* combina, em um mesmo fluxo, clusterização *K-means*, *Random Forest* e *Quantile Regression Forest* para produzir previsões pré-lançamento de novos produtos e, ao mesmo tempo, quantificar a incerteza por meio de quantis preditivos. A proposta integra características de produto e históricos de itens análogos para prever o total do ciclo e o perfil temporal da demanda, apoiando diretamente decisões de estoque e níveis de serviço. Os autores validam o método com dados reais de múltiplas empresas e reportam ganhos potenciais relevantes em custos de inventário. O processo é dividido em duas fases, preparação/treino e teste/operação, destacando a passagem dos dados de produtos existentes para o modelo e, depois, a aplicação em itens novos.

Na fase de preparação, os perfis de demanda de itens existentes são agrupados via *K-means*. Em termos conceituais, o *K-means* segmenta as observações em k grupos de forma que, dentro de cada grupo, os perfis sejam o mais homogêneos possível (BISHOP, 2006). No contexto do *DemandForest*, os vetores de entrada representam perfis de demanda históricos ou atributos derivados, e a escolha de k é guiada por índices de validade de *clusters*.

Em seguida, aplica-se *Random Forest* para classificar o perfil provável de um novo produto a partir de seus atributos, enquanto *Quantile Regression Forest* estima a distribuição condicional do total demandado no horizonte de introdução. Com relação à formulação de modelos de aprendizado supervisionado, no caso das árvores de decisão, o

princípio consiste em dividir o espaço de atributos em regiões cada vez mais homogêneas por meio de partições sucessivas (BREIMAN et al., 1984). Cada nó da árvore escolhe um atributo x_j (por exemplo, preço, categoria ou características visuais) e um ponto de corte s, de modo a separar os dados em dois subconjuntos (adaptado de Breiman et al. (1984)):

$$R_1(j, s) = \{x \mid x_j \leq s\}, \quad R_2(j, s) = \{x \mid x_j > s\}. \tag{2.2}$$

Essa escolha é feita de forma a minimizar uma função de custo, geralmente o erro quadrático médio em problemas de regressão, garantindo que os valores de demanda dentro de cada subconjunto sejam mais consistentes entre si (BREIMAN et al., 1984). O processo é aplicado recursivamente até que o espaço seja dividido em várias regiões finais (folhas), e a previsão para uma nova observação é dada pelo valor médio associado à folha em que ela se enquadra.

O *Random Forest*, por sua vez, corresponde a um *ensemble* de B árvores de decisão. Cada árvore é treinada a partir de uma amostra diferente dos dados originais, obtida por meio de uma técnica de reamostragem que consiste em selecionar observações aleatórias com reposição (BREIMAN, 2001). Além disso, em cada divisão interna de uma árvore, apenas um subconjunto aleatório dos atributos disponíveis é considerado, o que aumenta a diversidade entre as árvores e reduz o risco de o modelo ficar ajustado apenas aos dados de treino.

No *DemandForest*, esse mecanismo é usado tanto para classificar o perfil de demanda provável (tarefa de classificação) quanto, em combinação com o QRF, para estimar a distribuição condicional do total demandado, explorando a robustez do *ensemble* na captura de relações não lineares entre atributos de produto e demanda.

Na fase de aplicação, as características do item novo alimentam ambos os modelos: o perfil previsto fornece as proporções semanais e os quantis do QRF definem o total e os limites inferior/superior, compondo assim a série semanal e seus intervalos de previsão. Além das previsões, o método possibilita também o estudo de insumos de interpretabilidade úteis ao planejamento pela análise da importância das variáveis por permutação.

Do ponto de vista teórico, o *Quantile Regression Forest* (QRF), proposto por Meinshausen (2006), estende o *Random Forest* de regressão para estimar não apenas a média, mas toda a distribuição condicional da variável objetivo dado os atributos disponíveis (MEINSHAUSEN, 2006).

Em vez de ajustar diretamente uma função de regressão para cada quantil, o QRF

reutiliza a estrutura das árvores da floresta: para um ponto x, ele identifica quais observações de treino caem nas mesmas folhas que x e atribui pesos a essas observações, construindo assim uma distribuição condicional empírica de Y (MEINSHAUSEN, 2006). Os quantis $\hat{Q}_\alpha(x)$ são então obtidos diretamente dessa distribuição estimada.

Sob condições regulares, Meinshausen (2006) mostra que essa distribuição estimada é consistente, de modo que, com dados suficientes, os quantis de demanda aproximam os quantis "reais"condicionados aos atributos do produto. Itens mais incertos geram intervalos de previsão mais amplos, enquanto produtos com análogos estáveis geram intervalos mais estreitos, enquanto itens com análogos estáveis geram intervalos mais estreitos. Essa adaptação local da largura dos intervalos é justamente o que permite ligar quantis preditivos a metas de *Cycle Service Level* sem impor uma forma paramétrica específica à distribuição da demanda.

2.2.4 Seleção de Variáveis

Um ponto metodológico central na construção de modelos preditivos é a definição das variáveis explicativas que irão compor cada técnica. Enquanto alguns métodos de aprendizado de máquina, como o *Random Forest*, realizam internamente um processo de escolha de variáveis ao selecionar atributos em cada divisão das árvores (BREIMAN, 2001), outros exigem um procedimento explícito de seleção. Já nos métodos mais simples, como a regressão linear múltipla e a previsão por analogia, a definição das variáveis explicativas é inteiramente dependente da decisão de quem está modelando, de modo que a ausência de um critério sistemático pode comprometer a validade do modelo.

Nesse contexto, uma das abordagens mais recorrentes na literatura é o método de *stepwise regression*, prática consolidada em estatística aplicada por oferecer um procedimento estruturado e replicável de seleção de variáveis. O método consiste em incluir ou remover variáveis do modelo de forma sequencial, de acordo com critérios estatísticos como, por exemplo, o p valor dos coeficientes (MONTGOMERY; PECK; VINING, 2012).

O *stepwise* atua sobre a regressão, avaliando a entrada ou saída de variáveis a cada iteração. O procedimento pode assumir três variações: *forward selection*, em que se parte de um modelo vazio e variáveis são incluídas uma a uma conforme seu ganho estatístico; *backward elimination*, em que se inicia com todas as variáveis e são removidas aquelas que não apresentam significância; e *stepwise* propriamente dito, que combina os dois procedimentos, permitindo tanto a inclusão quanto a exclusão de variáveis em etapas sucessivas (MONTGOMERY; PECK; VINING, 2012).

Um critério bastante utilizado é baseado no p-valor: uma variável candidata é incluída no modelo se seu coeficiente apresenta significância estatística abaixo de um limiar (por exemplo, $\alpha = 0{,}05$) e a variável é mantida apenas se sua inclusão reduzir o valor do critério (MONTGOMERY; PECK; VINING, 2012).

O resultado final do *stepwise regression* é um subconjunto de variáveis explicativas que compõem a equação de regressão, proporcionando um equilíbrio entre parcimônia e poder preditivo. Sua principal vantagem é a simplicidade de aplicação e a transparência metodológica, o que o torna adequado em contextos acadêmicos e de prática profissional, constituindo uma ferramenta robusta e amplamente aceita para situações em que se busca um critério objetivo e replicável para a seleção de variáveis.

2.2.5 Particularidades do varejo da moda

O varejo de moda apresenta especificidades que o tornam especialmente desafiador no campo da previsão de demanda. Este setor combina ciclos de vida curtos (THOMASSEY, 2010), em que produtos são substituídos rapidamente por novas coleções, uma alta complexidade de sortimento, resultante das múltiplas combinações de modelo, cor e tamanhos, longos prazos de fornecimento, decorrentes de cadeias produtivas globais (CAO et al., 2008), e o problema do *cold start*, em que lançamentos chegam ao mercado sem qualquer histórico prévio de vendas (ANITHA; NEELAKANDAN, 2025). Essas condições estruturais diferenciam a moda de outros segmentos varejistas mais estáveis e tornam a atividade preditiva sujeita a níveis mais elevados de incerteza.

Para além dessas características, a moda é fortemente marcada por fatores externos de influência, como a exposição em mídias sociais, o engajamento de celebridades e a disseminação de tendências em ambientes digitais. Esses elementos podem alterar bruscamente o ciclo de vida de um produto, reforçando a volatilidade já intrínseca ao setor (GIRI et al., 2019).

Outro aspecto relevante é a crescente preocupação com sustentabilidade no consumo de moda, que incentiva modelos de negócio voltados à redução de sobras e remarcações por meio de melhor alinhamento entre oferta e demanda. Jin e Shin (2020) discutem como formatos de varejo apoiados em dados e em novos arranjos de consumo podem mitigar o problema de excesso de estoque e descarte de produtos, aproximando a gestão de demanda de objetivos ambientais e econômicos.

A heterogeneidade regional e de canais também adiciona complexidade: preferências

de consumo variam entre países, cidades e até mesmo bairros, além de se diferenciarem entre lojas físicas e comércio eletrônico (THOMASSEY, 2010).

Além dessas especificidades estruturais e comportamentais, a literatura também destaca manipulações desenvolvidas para lidar com problemas recorrentes do varejo de moda. Entre elas, destacam-se os ajustes voltados à correção das vendas perdidas por ruptura, decorrentes da censura da demanda, e os tratamentos relacionados à sazonalidade, que buscam incorporar ou remover efeitos de calendário. Esses pontos são detalhados a seguir.

2.2.5.1 Correção de vendas em ruptura

No varejo de moda, as vendas registradas não refletem necessariamente a demanda real, pois parte dos consumidores não consegue efetivar a compra devido à indisponibilidade de estoque. Esse fenômeno caracteriza o problema de censura da demanda, no qual a série histórica deixa de representar o volume potencial de procura e passa a refletir, em parte, a limitação de estoque.

Um exemplo de abordagem para tratar esse problema é o modelo em duas camadas proposto por Chen et al. (2022), no qual a primeira etapa consiste em estimar a demanda potencial total de cada produto, isto é, o volume que seria vendido caso não houvesse ruptura. Os autores ajustam uma regressão linear em que a demanda total D é explicada pelos atributos do produto e por variáveis de exposição comercial (categoria, faixa de preço, tecido, canal, entre outros). De forma simplificada, essa etapa pode ser representada por:

$$\hat{D}_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \tag{2.3}$$

em que $\hat{D}_i$ é a demanda total estimada para o produto i na ausência de ruptura, x_{ij} são os atributos observáveis do item e β_j são os parâmetros estimados da regressão (CHEN et al., 2022).

Na aplicação descrita por Chen et al. (2022), essa primeira camada de regressão substitui (ou complementa) as vendas observadas, reconstruindo o volume potencial de procura ao longo do ciclo de vida. A partir dessa demanda total estimada, os autores então aplicam o seu método específico para relacionar estoque inicial, vendas observadas e ruptura ao longo do tempo.

Assim, o procedimento de correção adotado pode ser interpretado como uma camada de pré-processamento dos dados: primeiro estima-se a demanda total esperada na ausência

de ruptura, e somente então as séries são usadas como entrada em modelos de previsão ou de simulação. Essa lógica é particularmente relevante em contextos de novos produtos de moda, em que a combinação de sortimento enxuto e alta rotatividade torna a ruptura frequente, amplificando o viés caso se usem diretamente as vendas observadas como referência para a demanda.

2.2.5.2 Sazonalidade

A sazonalidade é um dos elementos centrais que diferenciam a previsão de demanda no varejo de moda em relação a outros setores. Para além das coleções regulares de primavera/verão e outono/inverno, a procura é fortemente influenciada por datas comerciais (Natal, Dia das Mães, *Black Friday*), por períodos de liquidações e promoções, além de fatores climáticos que afetam diretamente categorias como roupas de inverno ou moda praia. Essa sobreposição de efeitos sazonais cria flutuações intensas e de curta duração, que dificultam a identificação de tendências mais estáveis e aumentam a incerteza preditiva (ANITHA; NEELAKANDAN, 2025).

Quando esses efeitos não são tratados de forma explícita, as previsões tendem a reproduzir oscilações superficiais, sem captar o verdadeiro padrão de procura. Isso resulta em decisões de compra desalinhadas, com risco elevado de excesso em determinados períodos e de rupturas em outros.

A literatura indica que esse desafio pode ser enfrentado por meio da incorporação de variáveis que representem os efeitos de calendário e uma forma comum é a criação de variáveis indicadoras (*dummies*), assim como já foi discutido com base em Montgomery, Peck e Vining (2012) anteriormente neste trabalho. Essas variáveis permitem representar estações, meses ou eventos específicos e ajustar regressões e previsões por analogia a padrões sazonais recorrentes

Assim, a incorporação dessas variáveis permite que os modelos capturem os efeitos periódicos característicos do calendário comercial e climático, ajustando as previsões conforme o comportamento sazonal observado em anos anteriores. Dessa forma, é possível reduzir parte da variabilidade aleatória das séries e tornar os modelos mais sensíveis aos ciclos reais de demanda do varejo de moda.

2.3 Simulação

O processo de condução de um estudo de simulação proposto por Banks (1998) pode ser organizado em uma sequência de etapas que vão da formulação do problema até a implementação das recomendações. Neste trabalho, esse fluxo é resumido em seis macroetapas: formulação do problema, definição dos objetivos e plano do projeto; modelo conceitual; coleta de dados; tradução do modelo; desenho experimental, produção de réplicas e análise; e documentação, relatório e implementação (BANKS, 1998).

2.3.1 Formulação do problema e definição dos objetivos e plano do projeto

O estudo começa com uma formulação clara do problema a ser tratado, alinhada entre analista e clientes internos/externos, seguida da definição dos objetivos específicos e dos indicadores de desempenho que a simulação deve estimar. Em paralelo, elabora-se um plano do projeto, que explicita escopo, cenários a serem avaliados, cronograma, papéis e responsabilidades, recursos computacionais e critérios de sucesso. Essa etapa inicial é crítica, pois condiciona todas as decisões subsequentes de modelagem e experimentação (BANKS, 1998).

2.3.2 Modelo conceitual

A partir da formulação do problema, o sistema real é representado por um modelo conceitual, composto por entidades, recursos, filas, eventos e regras de operação que capturam a lógica essencial do processo. Trata-se de uma abstração em nível lógico/matemático, ainda independente de software, em que se decide o grau de detalhe necessário para responder às perguntas de negócio. A recomendação de Banks (1998) é iniciar com um modelo relativamente simples e refinar a complexidade apenas quando a análise indicar necessidade, evitando modelos excessivamente detalhados que dificultem a verificação, a validação e a comunicação com os decisores.

2.3.3 Coleta de dados

Com o modelo conceitual definido, é necessário identificar e coletar os dados que alimentarão o modelo: padrões de chegada, tempos de processamento, capacidades, entre outros. Essa etapa envolve tanto o uso de dados históricos quanto a consulta a especialis-

tas e, frequentemente, ajustes estatísticos (como escolha de distribuições e parâmetros). Banks (1998) destaca que a qualidade do estudo de simulação é fortemente dependente da qualidade e da coerência dos dados de entrada, por isso, inspeções, limpeza, tratamento de *outliers* (observações muito distantes do comportamento típico dos dados e capazes de distorcer estimativas) e documentação das hipóteses de dados são partes integrantes dessa fase.

2.3.4 Tradução do modelo

A tradução corresponde à implementação do modelo conceitual em um ambiente de simulação discreta, por meio de um *software* ou linguagem apropriada. Nessa fase, o foco está em transformar a lógica conceitual em código executável e, em seguida, verificar se o modelo computacional faz exatamente o que se pretendeu (verificação) e se, como representação do sistema, é adequado para os objetivos do estudo (validação). Na prática, é comum um ciclo iterativo entre o modelo conceitual, os dados e a implementação, ajustando a estrutura ou o nível de detalhe sempre que surgem incongruências entre o comportamento simulado e o funcionamento observado do sistema (BANKS, 1998).

2.3.5 Desenho experimental, produção de réplicas e análises

Uma vez que o modelo computacional esteja verificado e validado, define-se o plano de experimentação: horizonte de simulação, condições de inicialização, número de repetições, política de geração de números aleatórios e cenários a comparar.Os códigos são então executados e seus resultados analisados com técnicas estatísticas apropriadas.

2.3.6 Documentação, relatório e implementação

Por fim, os resultados do estudo são consolidados em documentação técnica do modelo (assunções, estrutura, dados, testes de validação) e em relatórios gerenciais que traduzem as evidências em recomendações práticas. Banks (1998) ressalta que a utilidade da simulação depende da capacidade de comunicar resultados de forma compreensível aos decisores e de apoiar a implementação efetiva das mudanças sugeridas, seja na forma de novos parâmetros operacionais, novos layouts, novas políticas de controle ou investimentos em recursos. Assim, a última etapa fecha o ciclo ao transformar o conhecimento gerado pela simulação em ações concretas no sistema real (BANKS, 1998).

2.4 Métricas de Desempenho

A avaliação da qualidade das previsões é tão importante quanto o desenvolvimento dos próprios modelos. Diante da volatilidade da demanda, da curta vida útil dos produtos e da ausência frequente de histórico, torna-se essencial monitorar e comparar os métodos por meio de métricas de desempenho adequadas (THOMASSEY, 2010). Estudos comparativos indicam que a escolha da métrica influencia não apenas na interpretação da acurácia, mas também no ranking entre métodos concorrentes (DAVYDENKO; FILDES, 2013). Nesse sentido, autores acabam adotando a utilização de mais de uma medida, de forma complementar, para mitigar vieses e fornecer uma visão mais robusta (EGLITE; BIRZNIECE, 2022).

Entre as diversas medidas propostas pela literatura, este trabalho adota três grupos principais: *Bias* e *Mean Percentage Error* (MPE), para identificar tendências sistemáticas de super ou subestimação; o *Root Mean Square Error* (RMSE), que penaliza de forma mais intensa os grandes desvios; e o *Mean Absolute Percentage Error* (MAPE) e *Weighted MAPE* (wMAPE), que expressam erros em termos percentuais e ponderados por volume.

2.4.1 *Bias* e *Mean Percentage Error*

O *Bias* corresponde à média das diferenças entre valores previstos $\hat{y}_t$ e valores observados y_t, indicando se o modelo tende a superestimar ou subestimar a demanda. Já o *Mean Percentage Error* (MPE) expressa essa diferença relativa em termos percentuais, o que facilita a interpretação em contextos gerenciais.

$$\text{Bias} = \frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t) \tag{2.4}$$

$$\text{MPE} = \frac{100}{n}\sum_{t=1}^{n}\frac{(\hat{y}_t - y_t)}{y_t} \tag{2.5}$$

Na primeira fórmula, o erro de previsão é simplesmente a diferença entre previsto e realizado, e o *Bias* resulta da média dessas diferenças. Na segunda, o erro é normalizado pelo valor real, permitindo observar a magnitude relativa do desvio em porcentagem.

Essas métricas são úteis para diagnosticar se há viés sistemático nos modelos, mas apresentam limitações conhecidas. Em particular, medidas baseadas em erros percentuais podem produzir valores muito elevados quando a demanda observada é pequena, distor-

cendo a avaliação da acurácia (DAVYDENKO; FILDES, 2013). Além disso, o simples erro médio pode mascarar desvios relevantes, pois erros positivos e negativos tendem a se compensar, gerando valores próximos de zero mesmo quando os desvios absolutos são grandes.

2.4.2 Root Mean Square Error (RMSE)

O Root Mean Square Error (RMSE) mede a raiz quadrada da média dos erros ao quadrado. Essa formulação faz com que erros maiores tenham peso desproporcionalmente mais elevado, o que torna a métrica adequada para contextos onde grandes desvios têm custo econômico elevado, como é o caso da moda.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2} \tag{2.6}$$

O RMSE é obtido elevando-se cada erro ao quadrado, somando todos e tirando a média, seguido da raiz quadrada para retornar à escala original da variável de interesse. Assim, elimina-se o efeito de sinais positivos e negativos e aumenta-se a penalização dos maiores erros.

Na prática, o RMSE constitui uma alternativa eficaz para a avaliação de modelos de previsão em cenários com novos produtos, uma vez que atribui penalizações mais severas a erros que geram excessos expressivos de estoque ou rupturas de abastecimento. Por outro lado, é uma métrica dependente da escala da série, o que dificulta comparações entre categorias heterogêneas, como básicos de alto volume e itens de coleção cápsula.

2.4.3 Mean Absolute Percentage Error (MAPE) e Weighted MAPE (wMAPE)

O *Mean Absolute Percentage Error* (MAPE) corresponde à média dos erros absolutos em termos percentuais, enquanto o *Weighted MAPE* (wMAPE) pondera os erros absolutos pelo total de vendas, o que torna a métrica mais adequada em análises agregadas.

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \tag{2.7}$$

$$\text{wMAPE} = \frac{\sum_{t=1}^{n}|\hat{y}_t - y_t|}{\sum_{t=1}^{n} y_t} \times 100 \tag{2.8}$$

No MAPE, cada erro é normalizado pelo valor real e expresso em porcentagem, sendo depois calculada a média. No wMAPE, os erros absolutos são ponderados pelo total de vendas, de modo que produtos ou períodos de maior volume têm maior impacto no valor final.

Essas métricas são amplamente utilizadas na avaliação de modelos de previsão de demanda, em parte por sua interpretação relativamente simples (DAVYDENKO; FILDES, 2013), e aparecem com frequência em estudos aplicados ao varejo de moda, como Chen et al. (2022). Sua principal vantagem é a comunicação clara: gestores compreendem facilmente o significado de um erro de 10%. Contudo, o MAPE apresenta limitações, pois não pode ser calculado quando $y_t = 0$ e tende a inflar erros relativos em itens de baixa escala. Já o wMAPE resolve parcialmente essa limitação ao ponderar pelo volume, mas pode mascarar desvios críticos em produtos estratégicos de menor peso no faturamento. Assim, o uso do wMAPE é uma boa alternativa para análises de pré-venda e coleções inteiras, desde que acompanhado de métricas adicionais para garantir robustez.

2.5 Mapeamento de processos

2.5.1 Fundamentos e importância do mapeamento de processos

Na seção anterior foram discutidos aspectos quantitativos relacionados à previsão de demanda, porém é importante destacar que tanto a previsão quanto o processo de compras constituem processos de negócio que podem ser descritos, analisados e aperfeiçoados. Para representar graficamente esses fluxos de trabalho, uma das notações mais amplamente utilizadas é a BPMN (*Business Process Model and Notation*). Essa notação fornece uma linguagem visual padronizada para descrever, de forma clara e estruturada, as etapas, decisões e responsabilidades envolvidas em um processo. Com isso, torna-se possível compreender e comunicar com maior precisão como as atividades se encadeiam e onde há oportunidades de melhoria (SILVER, 2011).

A notação BPMN é hoje o padrão mais amplamente utilizado para modelagem de processos de negócio. Segundo Silver (2011), o principal propósito da notação é garantir clareza semântica e consistência, de modo que um mesmo diagrama possua uma única leitura possível – princípio que diferencia a BPMN de representações mais genéricas, como fluxogramas tradicionais (SILVER, 2011).

O mapeamento permite visualizar o encadeamento lógico das atividades, os pontos de decisão, os responsáveis e as entradas e saídas de cada etapa, tornando visíveis gargalos,

redundâncias e falhas de comunicação. Essa visão estruturada é o ponto de partida para qualquer iniciativa de melhoria contínua (SILVER, 2011).

Em termos metodológicos, o mapeamento de processos pode ser aplicado em dois momentos: o mapeamento "*As Is*", que representa o processo atual, e o mapeamento "*To Be*", que descreve o processo redesenhado após as melhorias propostas. Essa distinção permite comparar a situação real com a ideal e direcionar as mudanças de forma orientada por evidências (SILVER, 2011).

Segundo Silver (2011), um bom modelo de processo deve atender simultaneamente a três requisitos: correção, ou seja, representar de forma lógica e completa o comportamento do processo; clareza, tornando o fluxo facilmente compreensível mesmo para quem não participou de sua construção; e conformidade, garantindo aderência aos padrões da notação BPMN. Dessa forma, o modelo passa a ser não apenas um registro gráfico, mas uma ferramenta de comunicação e gestão.

2.5.2 Elementos e Aplicação Prática da BPMN

A aplicação prática da notação BPMN baseia-se em um conjunto de elementos gráficos que, em conjunto, descrevem de forma padronizada como um processo se inicia, se desenvolve e se encerra. A seguir, são apresentados os principais componentes que compõem um diagrama BPMN.

2.5.2.1 Eventos

Os eventos representam algo que ocorre durante o processo e que afeta o seu fluxo. Eles marcam o início, interrupções intermediárias ou o término de uma sequência de atividades. Segundo Silver (2011), um diagrama BPMN deve conter ao menos um Evento Inicial e um Evento Final, garantindo a integridade lógica do processo. Os eventos são representados por círculos e classificados conforme seu papel: o evento inicial (círculo simples) indica o gatilho que dá início ao processo; o evento intermediário (círculo duplo) sinaliza ocorrências que afetam o fluxo, como atrasos, exceções ou recebimento de mensagens; e o evento final (círculo com borda grossa) representa a conclusão do processo. Essa estrutura cíclica assegura que o modelo possua um ponto de partida e de encerramento claramente definidos, favorecendo a leitura e a compreensão do fluxo de atividades.

2.5.2.2 Atividades

As atividades correspondem às ações executadas dentro do processo, podendo envolver pessoas, sistemas ou ambos. São representadas graficamente por retângulos arredondados e indicam o trabalho que precisa ser realizado para transformar entradas em saídas. De acordo com Silver (2011), cada atividade deve possuir um nome no formato verbo no infinitivo e um substantivo, assegurando clareza e padronização terminológica. As atividades podem ser simples, quando realizadas em uma única etapa, ou compostas (subprocesses), quando agregam várias tarefas internas que podem ser detalhadas em diagramas complementares. Essa hierarquia é um dos pontos fortes da BPMN, pois permite representar processos em diferentes níveis de detalhe sem perder coerência entre as visões macro e micro.

2.5.2.3 *Gateways*

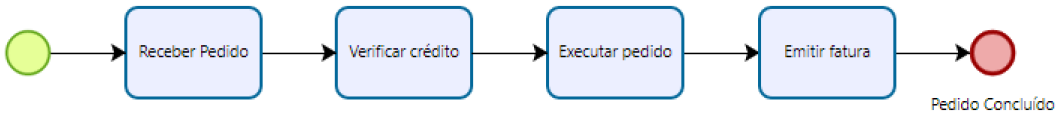
Os *Gateways* são elementos que controlam a divergência e a convergência do fluxo de processo, isto é, onde decisões são tomadas ou caminhos distintos se unem. Na notação BPMN, são representados por losangos e podem expressar diferentes tipos de lógica, como exclusão (representada por "X"), paralelismo (com símbolo "+") ou inclusão condicional (com círculo).

2.5.2.4 Fluxos de Sequência

Os fluxos de sequência são as setas que conectam os elementos de um processo, indicando a ordem lógica de execução. Cada fluxo representa uma transição direta entre eventos, atividades e *gateways*. Na BPMN, um fluxo de sequência nunca atravessa as fronteiras de uma piscina, pois ele descreve a lógica interna de um único processo (SILVER, 2011). Isso significa que interações entre diferentes participantes devem ser representadas não por fluxos de sequência, mas por fluxos de mensagem. Silver enfatiza que o traçado dos fluxos deve seguir o sentido de leitura natural e evitar cruzamentos desnecessários, de modo a maximizar a legibilidade do diagrama.

Após a introdução desses primeiros objetos, é possível visualizar sua aplicação integrada em um diagrama BPMN simples, conforme ilustrado na Figura 4. Esse tipo de representação demonstra o fluxo básico de um processo de negócio, do evento inicial ao final, passando pelas atividades e decisões que o compõem.

Figura 4: Exemplo de processo básico em BPMN com eventos, atividades e fluxos de sequência.



Fonte: Adaptado de Silver (2011, p. 34).

2.5.2.5 Piscinas e Raias

As Piscinas e Raias são elementos que estruturam o diagrama em faixas horizontais ou verticais, indicando os participantes ou as áreas responsáveis por cada atividade. Uma piscina representa um participante do processo, que pode ser uma organização, área ou sistema, enquanto as raias subdividem essa piscina em unidades menores, como departamentos ou funções (SILVER, 2011). Essa hierarquia facilita a identificação de quem executa cada etapa e como as responsabilidades se distribuem. Em contextos empresariais, a clareza dessa divisão é essencial para eliminar sobreposições e lacunas de responsabilidade.

A Figura 5 exemplifica a utilização de piscinas e raias em um processo, evidenciando como a BPMN permite associar atividades a diferentes áreas ou funções dentro de uma mesma organização.

Figura 5: Representação de raias dentro de uma piscina em BPMN.

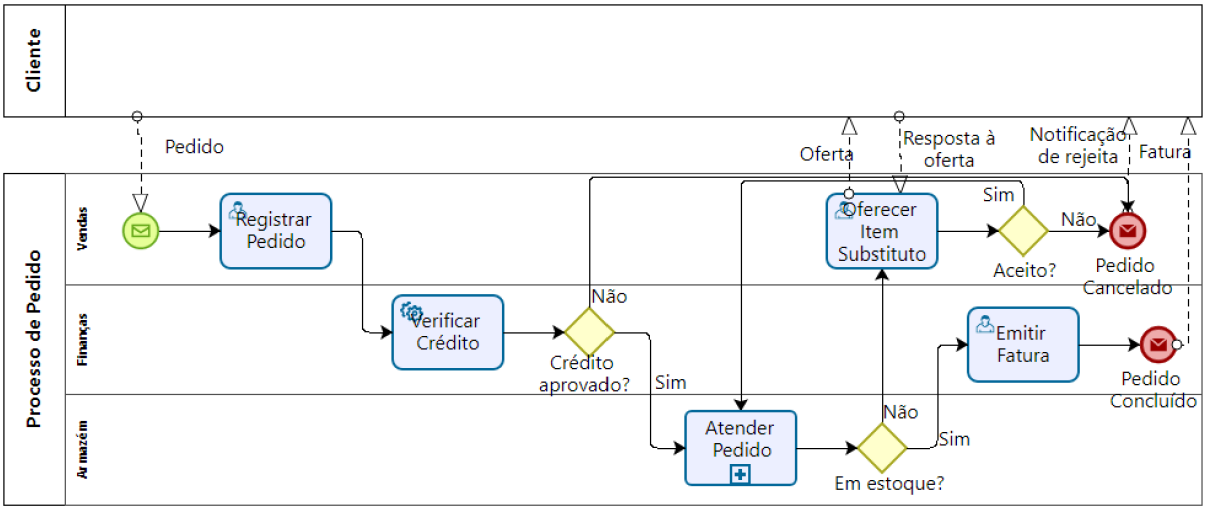
Fonte: Adaptado de Silver (2011, p. 48).

2.5.2.6 Fluxos de Mensagem

Os fluxos de mensagem representam a comunicação entre diferentes participantes, ou entre duas áreas distintas de uma organização. São desenhados como linhas tracejadas com um círculo no início e uma seta no fim, ligando duas piscinas separadas. Ao contrário dos fluxos de sequência, os fluxos de mensagem não descrevem a ordem de execução, mas a troca de informações ou solicitações entre processos independentes (SILVER, 2011). Esses fluxos são cruciais para representar colaborações e integrações externas, tornando o modelo mais realista e útil para o redesenho de processos interdepartamentais. Em termos práticos, permitem visualizar onde ocorrem trocas de dados, aprovações e dependências externas, aspectos fundamentais para diagnosticar gargalos de comunicação e oportunidades de automação.

Na Figura 6, observa-se um exemplo de diagrama de colaboração, no qual duas piscinas interagem por meio de fluxos de mensagem. Esse tipo de representação é especialmente útil para mapear processos que envolvem trocas entre diferentes áreas ou organizações, como o relacionamento entre a área de compras e os fornecedores.

Figura 6: Exemplo de diagrama de colaboração em BPMN com piscinas e fluxos de mensagem.



Fonte: Adaptado de Silver (2011, p. 40).

3 MÉTODO

Conforme apresentado no capítulo introdutório, este trabalho busca enfrentar o problema das compras não aderentes para novos produtos em uma grande rede de moda *fast fashion* atuante no mercado brasileiro. Neste capítulo, é estruturada a metodologia para abordar o problema e propor um método e um processo efetivo, incluindo um modelo de apoio à decisão, viável e com aderência organizacional.

Com esse intuito, são adotadas duas frentes, que serão detalhadas a seguir: revisão de processos e modelo de previsão de demanda

3.1 Revisão de processos

Para mapear e compreender o funcionamento atual (*As Is*) da previsão de demanda na empresa e, a partir desse diagnóstico, em uma etapa posterior deste trabalho, propor um método de melhoria alinhado às necessidades do negócio, adotou-se a metodologia BPMN. Esta é uma referência internacional, que explicita eventos, atividades e integrações tecnológicas, favorece a leitura estruturada do fluxo, a atribuição de responsabilidades e a identificação de atividades, além de estabelecer uma linguagem comum entre áreas e um padrão de mapeamento claro, consistente e replicável (SILVER, 2011).

O método de mapeamento de processos deste trabalho seguirá a abordagem *top-down*: com o problema definido, o primeiro passo é delimitar o escopo. Nessa etapa, acorda-se onde o processo começa e termina e o que cada instância representa. Assim, evitam-se ambiguidades antes de entrar no detalhamento do fluxo (SILVER, 2011). Com isso estabelecido, elabora-se o mapeamento macro, isto é, o mapeamento das grandes atividades envolvidas no processo.

A partir desse mapa, as atividades macro desdobram-se em subprocessos detalhados em diagramas próprios. Nessa fase, esclarecem-se frequências, granularidades, responsabilidades e requisitos. Além disso, o mapeamento mais minucioso evidencia problemas e

variações, sejam por exceções, retrabalhos ou desvios, tornando importante seu detalhamento e descrição para sustentar o diagnóstico e orientar as melhorias.

Com base nesse diagnóstico, elabora-se o *To Be* nos mesmos níveis de granularidade e frequência. A ideia é construir um mapa de processo ideal, no qual os principais problemas sejam corrigidos. Como não é o foco principal deste projeto, não serão utilizadas metodologias de melhoria mais profundas que buscam por causas raiz, ainda assim, essa base já permite uma excelente compreensão dos problemas do processo e suas causas como um todo.

Para consolidar esses entendimentos, serão conduzidas entrevistas semiestruturadas com dois profissionais com perfis complementares: um gerente sênior, com visão ponta a ponta do processo, e um gerente pleno, responsável pela execução das rotinas e pelo uso dos artefatos e sistemas associados.

Será utilizado um roteiro semiestruturado para esclarecer: objetivos do processo, eventos de início e término, atividades e decisões, exceções e retrabalhos, entregas entre áreas, frequências, responsabilidades, integrações sistêmicas e principais pontos de dor. Esses elementos, sempre que possível, serão explicitados por meio dos objetos do mapeamento (por exemplo, raias, eventos, *gateways* e fluxos de mensagem), apoiando a atribuição clara de papéis e a governança do processo (SILVER, 2011).

Para essas entrevistas, são definidas previamente perguntas que orientam a conversa e, ao mesmo tempo, preservam a flexibilidade para explorar pontos emergentes. As perguntas a seguir têm justamente o foco de delimitar escopo, explicitar responsabilidades, identificar decisões e tornar visíveis as integrações sistêmicas e os problemas no processo:

1. Qual é o objetivo do processo e como seu desempenho é aferido?
2. Quais são os eventos de início e de término do processo e com que frequência ocorrem?
3. Quais são as atividades macro, seus propósitos, e como elas se encadeiam entre planejamento, compras e calendário de lançamento?
4. Em quais pontos ocorrem decisões, quais critérios são aplicados, e qual é o racional esperado na escolha entre alternativas?
5. Que exceções, variações e retrabalhos ocorrem, com que frequência aparecem, quais são suas causas e como são tratados?

6. Quais são as entregas entre as etapas, quem são os responsáveis e quais são os formatos delas?

7. Quais são as frequências, as etapas de planejamento e os níveis de detalhe utilizados em cada atividade (Grupo, subgrupo, SKU)?

8. Quem é responsável pelo quê em cada etapa e como as atividades são distribuídas?

9. Que sistemas e artefatos são utilizados (planilhas, relatórios) e que objetos de dados são trocados?

10. Quais são as principais dores sentidas no processo (gargalos, riscos e impactos em operações) e quais são as mais impactantes no resultado do processo?

Com base nas informações coletadas a partir dessas questões, a consolidação dos mapas ocorrerá em ciclos: após cada rodada de entrevistas, os diagramas serão desenvolvidos no programa Bizagi e validados com os mesmos gerentes da *Fashion*, verificando se o fluxo *As Is* representa corretamente o trabalho real, se há variações relevantes ainda não capturadas e se o proposto de fato reflete a lógica praticada.

Adicionalmente, os diagramas *To Be* também serão validados com os mesmos gerentes, com foco na aderência operacional e na viabilidade das mudanças sugeridas. Esse procedimento reforça a coerência do modelo e sua utilidade como base para as propostas de melhoria a serem apresentadas no *To Be* (SILVER, 2011).

3.2 Modelos de previsão de demanda

Para desenvolver e propor um método de previsão de demanda, este trabalho desenvolve um processo de simulação para testar diferentes modelos de previsão, que segue passos similares aos propostos por Banks (1998), descritos a seguir:

3.2.1 Formulação do problema e dos objetivos

Como já exposto na Introdução, o problema em foco é o mau dimensionamento da primeira compra de SKUs sem histórico, que gera rupturas e excessos de estoque. Portanto, o objetivo da simulação é identificar, entre modelos de previsão de demanda para produtos novos, aquele que apresenta a melhor performance com base em indicadores de assertividade, a serem definidos, que mensuram a aderência entre o valor previsto e as

vendas no ciclo de vida esperado, bem como comparar essa performance com a do modelo atualmente utilizado.

3.2.2 Modelo Conceitual

No Modelo Conceitual, é delimitado o que o sistema preditivo deve entregar, sem detalhar procedimentos de implementação. Parte-se do entendimento aprofundado do processo (mapeamento de processos) para explicitar o comportamento esperado do modelo, as entradas que o alimentam, a granularidade de atuação, o nível de previsão, o horizonte temporal e a periodicidade de atualização.

Nesta etapa, descreve-se também a relação do modelo com as dinâmicas das coleções (ciclos de vida, fases de lançamento, remarcações e políticas de reposição), a granularidade de compra (lotes mínimos, grades e regras de arredondamento), os critérios de alocação às lojas (clusterização, potencial de demanda e canal) e a frequência de compra e revisão (cadência de geração de previsões e gatilhos de replanejamento).

3.2.3 Coleta de dados

Nesta etapa, os dados disponíveis são mapeados, identificando-se tabelas e colunas, limites de histórico de datas, e aplicam-se filtros para reter apenas o que é útil ao problema. Avaliam-se ainda lacunas, duplicidades e inconsistências, com procedimentos de limpeza e verificação de consistência.

Adicionalmente, são tratadas rupturas mediante identificação de dias com estoque zero e marcação de ruptura. Na literatura, Chen et al. (2022) estimam a demanda total via regressão linear, porém neste trabalho, adota-se uma simplificação dado que este não é o foco principal: propaga-se a venda média por loja para todas as lojas que recebem o determinado produto.

Por fim, ainda nessa etapa, podem ser realizados cortes de escopo por limitações computacionais preservando a representatividade do conjunto. Sendo assim, o resultado é uma base de dados pronta para a etapa de modelagem.

3.2.4 Tradução do modelo

Nesta etapa, são desenvolvidos os modelos que traduzem o modelo conceitual em implementações computacionais. São implementados três métodos diferentes com confi-

gurações comparáveis: uma previsão por analogia mais simples para servir como *baseline* (referência de comparação), uma regressão linear com seleção de variáveis por *stepwise* guiado por *wMAPE*, e uma adaptação do método de Steenbergen e Mes (2020) utilizando uma variação do método de *Random Forest*.

O *baseline* por analogia segue a lógica de tomar uma referência média de vendas de produtos semelhantes, conceito próximo ao *mean profile* usado como previsor para itens novos por (THOMASSEY; FIORDALISO, 2006). Neste trabalho, essa média é testada em diferentes níveis da hierarquia mercadológica (mais agregados versus mais granulares) para identificar o nível que maximiza o desempenho nos indicadores.

A regressão linear é utilizada para estimar a demanda a partir de atributos do produto, em linha com um dos métodos testados por Chen et al. (2022). Aqui, a seleção de variáveis é feita por *stepwise* orientado por *wMAPE* para alinhar a escolha de variáveis ao objetivo operacional de redução de erro na primeira compra.

Por fim, toma-se como referência o método de Steenbergen e Mes (2020), que combina *K-means* para agrupar perfis de demanda, *Random Forest* para classificar o perfil de novos produtos e *Quantile Regression Forests* para prever demanda total e quantis. No artigo original, essa arquitetura é aplicada a portfólios de tipos de produtos mais heterogêneos, com maior variedade de curvas de venda, o que justifica a etapa de *clustering*. No presente projeto, dado o escopo restrito a novos produtos de moda, optou-se por não reproduzir integralmente esse fluxo e aplicar diretamente o *Quantile Regression Forest* sobre os atributos dos produtos, preservando a lógica probabilística da abordagem original com uma arquitetura mais simples.

3.2.5 Planejamento, execução e análise de experimentos

Nesta seção, define-se como os modelos serão comparados: estabelece-se a divisão entre treino, validação (para ajuste e seleção de variáveis) e teste (para avaliação final), assegurando que o conjunto de teste permaneça intocado até o fim para garantir imparcialidade. Adota-se uma partição próxima de 75%/25% (treino/validação vs. teste), em linha com a prática utilizada por Steenbergen e Mes (2020). São feitas variações de treino e validação com diferentes sementes aleatórias (variável que fixa o comportamento aleatório e a divisão entre conjuntos de dados) e consolidação dos resultados no teste.

A avaliação prioriza o wMAPE, métrica adotada em estudos recentes de previsão de novos produtos de moda por combinar leitura intuitiva (erro percentual médio ponderado

pelas vendas) e maior adequação a dados com diferentes magnitudes e níveis de agregação (KHARFAN; CHAN; EFENDIGIL, 2021), e MPE, para diagnosticar viés médio, indicar se as previsões tendem a superestimar ou subestimar. Reportam-se os resultados com e sem *outliers* (dados extremos positivos ou negativos) para aferir robustez frente a desempenhos excepcionalmente altos/baixos não modeláveis.

3.2.6 Documentação e implementação

Por fim, com relação à documentação, todo o desenvolvimento é realizado em um *script Python* e é disponibilizado à *Fashion*. A documentação técnica e funcional é consolidada neste trabalho, servindo como referência para manutenção e evolução do código, bem como para entendimento do histórico do processo.

Com esse conjunto de etapas metodológicas, estabelece-se a base necessária para avaliar, de forma aplicada, a proposta deste trabalho. No capítulo seguinte, são apresentados os resultados obtidos com a revisão de processos, a aplicação dos modelos de previsão, a comparação com a prática atual de compras e suas implicações para a operação da *Fashion*.

4 RESULTADOS

Neste capítulo são apresentados os resultados da aplicação dos dois métodos propostos na seção anterior.

4.1 Mapeamento de Processos

Esta frente é dividida em duas fases principais: o mapeamento *As Is*, tratando do estado atual do processo, e o mapeamento *To Be*, novas propostas de fluxo para o processo.

4.1.1 Mapeamento As Is

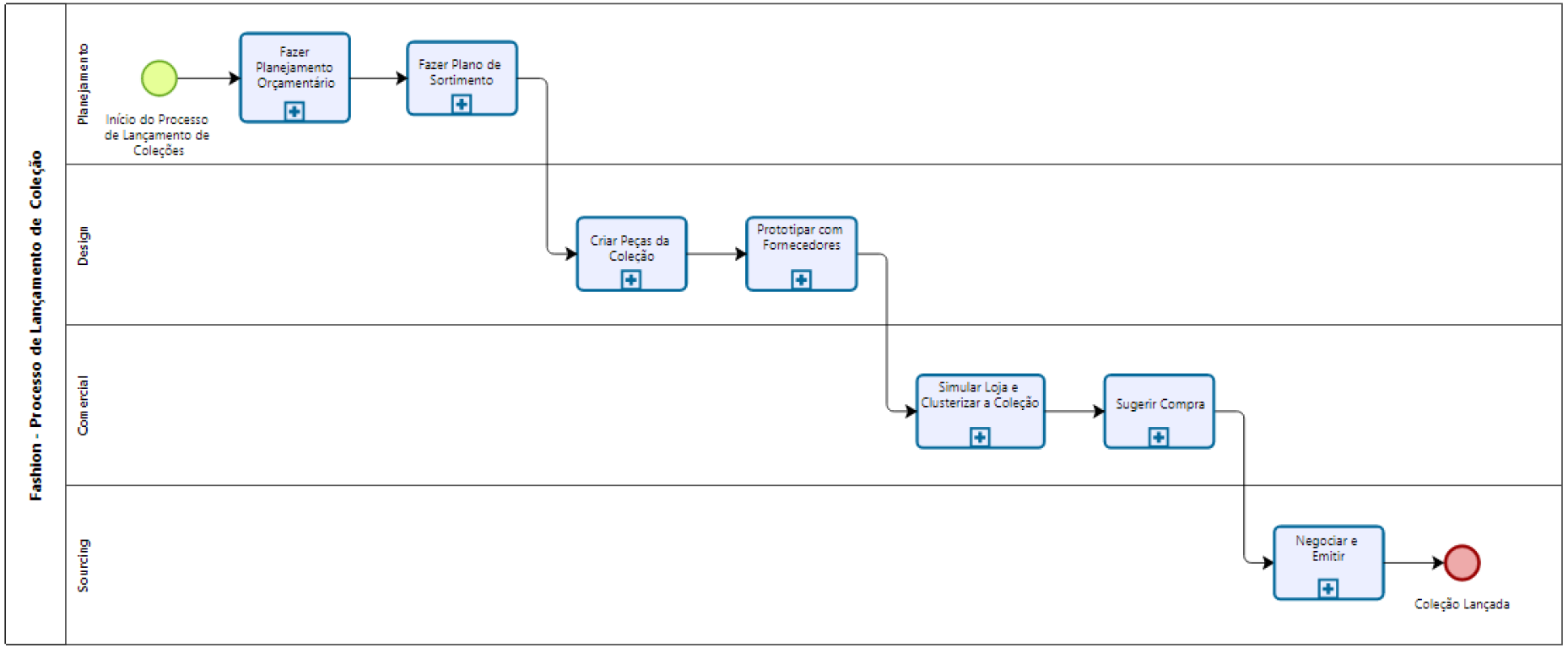
Na etapa de mapeamento de processos, foram realizadas entrevistas com dois perfis complementares na *Fashion*: primeiramente, um gerente sênior de Planejamento, com visão ampla e responsabilidade pela supervisão do fluxo e, em segundo lugar, um gerente pleno, responsável por alguns Grupos (nível da hierarquia de produtos), diretamente envolvido na execução das rotinas e no uso dos sistemas e artefatos.

A combinação dessas duas perspectivas assegurou equilíbrio entre visão estratégica e prática operacional, oferecendo a base necessária para a aplicação do BPMN.

Com base na entrevista com o gerente sênior, foram desenhados e validados com a equipe dois mapeamentos: a Figura 7, que apresenta a visão macro do processo de lançamento de coleções, definindo escopo e principais interações entre áreas, e a Figura 8, que descreve, em termos gerais, o processo de compras, destacando etapas, decisões e integrações sistêmicas relevantes.

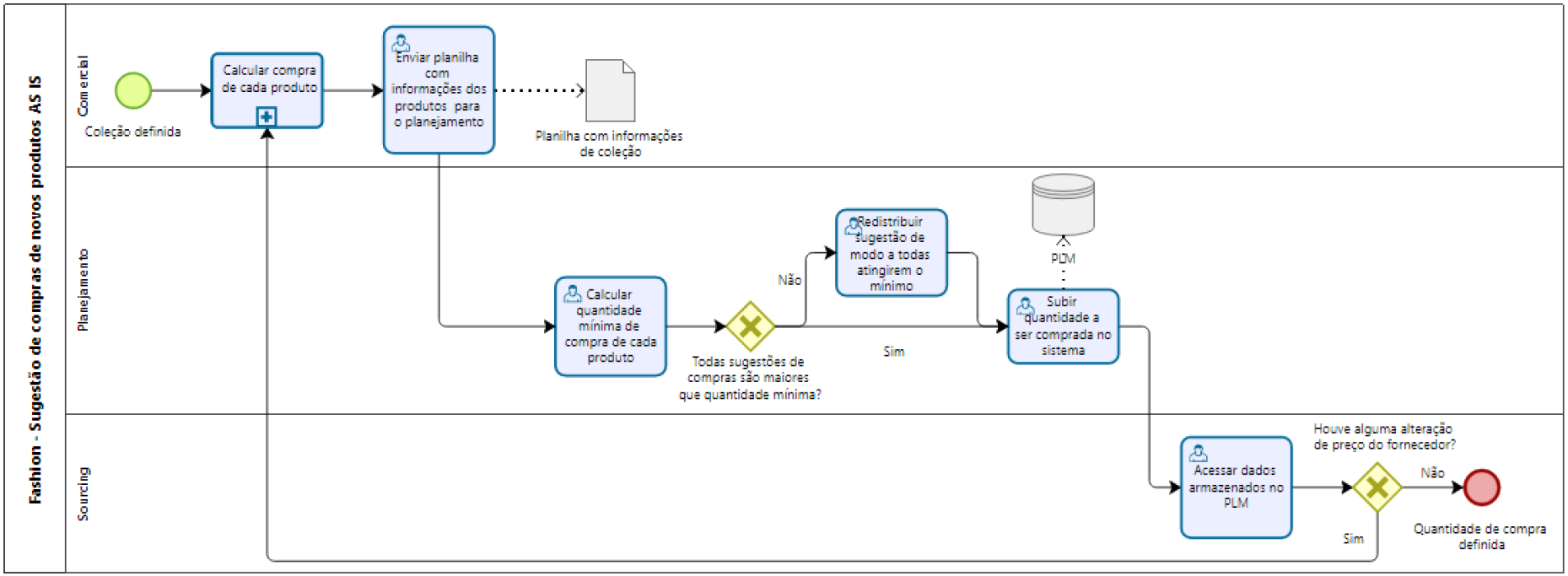
Na sequência, a entrevista com o gerente pleno permitiu avançar para o nível operacional. A partir dessa conversa, a Figura 9 e a Figura 10 foram construídas, exemplificando como a sugestão de compra é calculada na prática, explicitando regras aplicadas, parâmetros utilizados, fontes de dados consultadas, validações e variações recorrentes.

Figura 7: Lançamento de coleções (visão macro)



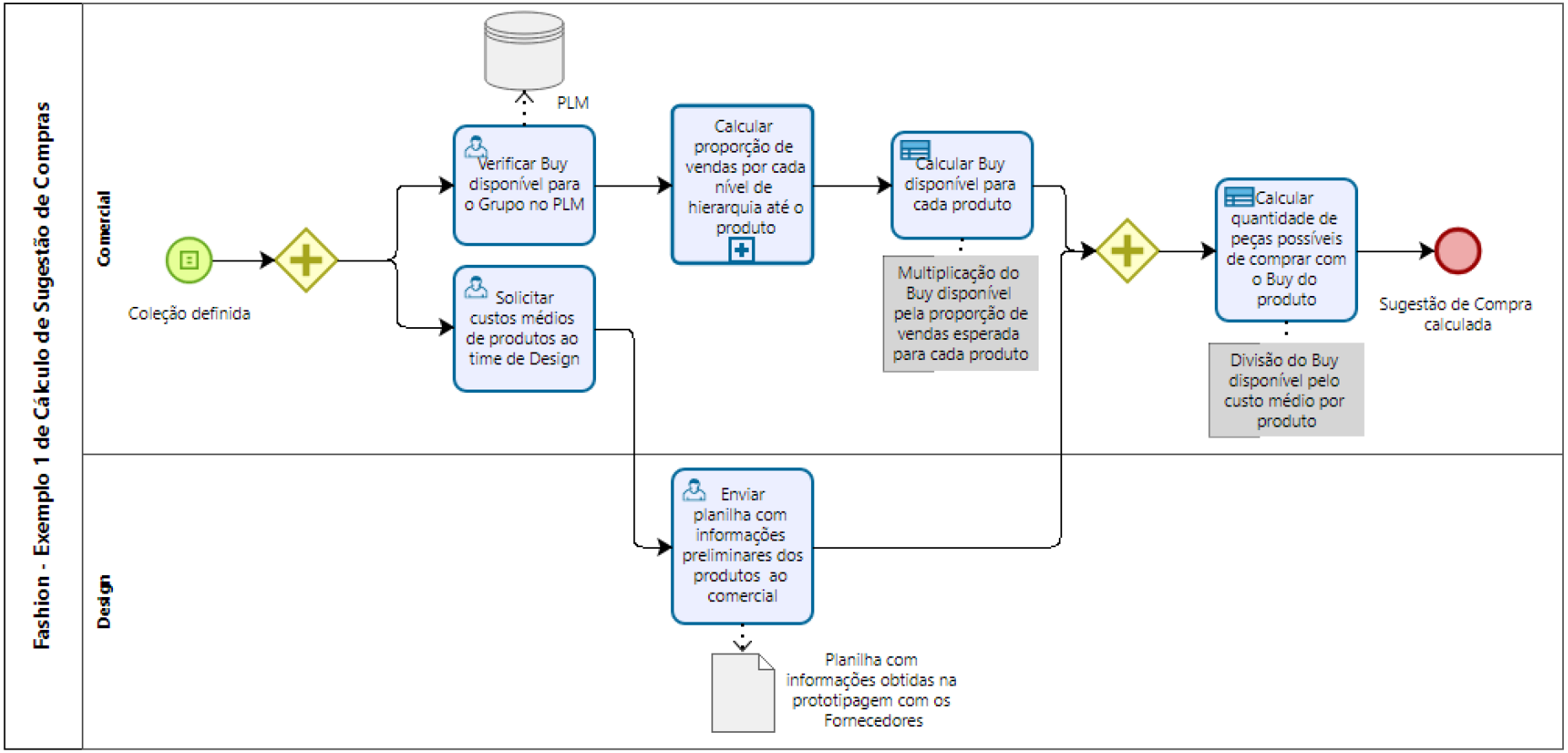
Fonte: Autoria própria.

Figura 8: Sugestão de Compras As Is



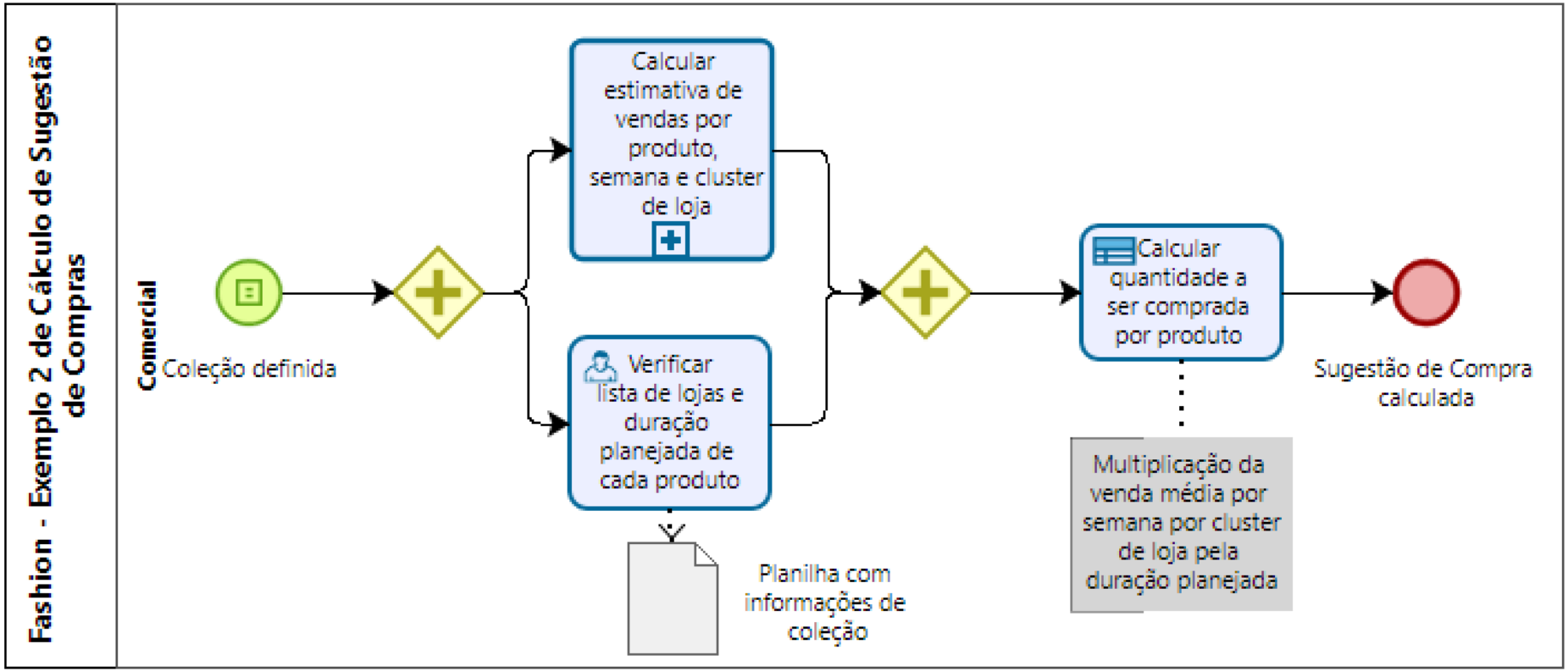
Fonte: Autoria própria.

Figura 9: Exemplo 1 de cálculo da sugestão de compra



Fonte: Autoria própria.

Figura 10: Exemplo 2 de cálculo da sugestão de compra



Fonte: Autoria própria.

A Figura 7, de nível macro, reúne as principais etapas do lançamento de coleções e indica quem influencia cada uma delas. Fica claro que o fluxo atravessa várias áreas e envolve sucessivos repasses entre times. O diagrama mostra o caminho principal, mas, por trás de cada etapa, existem subprocessos maiores e mais detalhados, com regras, dados e exceções que exigem mapeamentos próprios.

Vale destacar alguns aprendizados que ficam fora dos mapeamentos formais. Primeiro, a compra não é definida por loja específica, e sim por grupos de lojas (com mais de mil combinações possíveis). Inicialmente a lista de lojas é definida e, depois, como em qualquer outro produto, aplica-se a lógica de distribuição da empresa para direcionar o envio por unidade. Essa lógica é ampla e complexa, combinando critérios de demanda, capacidade e diretrizes comerciais para ajustar a alocação.

Outra compreensão importante diz respeito à categorização dos produtos. Foi explicado que a maior parte dos novos produtos lançados entra como "C", categoria que não conta com informações adicionais ou testes prévios. Já os produtos novos "B" derivam de algum modelo similar já trabalhado na rede, e os "A", são contínuos e não têm produtos novos. Em geral, os 'C' são produtos mais diferenciados e planejados para um ciclo de vida curto, de aproximadamente oito semanas.

Quanto às descobertas de calendário, dois momentos concentram os inícios de coleção: novembro (preparação, principalmente, para Natal, e também para *Black Friday*) e fevereiro/março (troca de coleção ao fim do verão). Há ainda janelas relevantes em maio e agosto. Mesmo assim, ocorrem entradas menores de novos produtos ao longo do ano, para ajustar a oferta entre as viradas principais.

A Figura 8 é o aprofundamento do subprocesso de "Sugerir Compras"da Figura 7. Nele também podemos ver os responsáveis por cada atividade e o encadeamento delas, além de registrar também o conjunto de ferramentas de dados utilizado, com destaque para o uso de planilhas pouco estruturadas, o que revela uma base de dados menos organizada, mais dependente de controles manuais e trocas pontuais de arquivos.

Também se evidencia um subprocesso importante que não está detalhado no fluxograma: o cálculo da compra de cada produto. De acordo com o que foi aprendido nas entrevistas, cada gerente pleno, no nível de Grupo, define a própria forma de chegar ao número necessário, não há diretriz padronizada, exige-se apenas o resultado. Na prática, essa estrutura gera métodos distintos entre Grupos, variando conforme as análises que cada gerente pleno decide realizar.

Ainda na Figura 8, há outro ponto de atenção. Quando há mudança no preço do

fornecedor por questões comerciais, o processo precisa ser reiniciado e, na maioria das vezes, essa divergência só é identificada no final, causando grande retrabalho.

Por fim, a Figura 9 e a Figura 10 mostram, na prática, como o subprocesso de cálculo de compra por produto (apontado na Figura 8) é efetivamente executado. Registraram-se duas abordagens usualmente adotadas por gerentes plenos em seus respectivos Grupos, evidenciando variações de método e nível de estruturação.

Na Figura 9, a lógica é *top-down*: parte-se do *buy* (orçamento total de compras do Grupo) e realiza-se uma quebra sucessiva do orçamento até chegar ao nível de produto. A partir da parcela atribuída a cada item, calcula-se quanto comprar. Nessa abordagem, não há diretriz para estimar, de forma estruturada, o potencial de vendas do produto, além disso, o próprio cálculo de compra não é padronizado, ficando sujeito a critérios locais e ajustes subjetivos.

A Figura 10 descreve um segundo exemplo de prática, também sem estrutura fixa, em que se busca estimar o "ritmo de vendas", venda semanal do produto por loja, por meio de diferentes procedimentos. Com base nessa estimativa, deriva-se a compra pretendida. Embora traga um olhar mais próximo da demanda, a abordagem permanece dependente de julgamento e de escolhas metodológicas do gerente pleno.

Após a discussão, identificaram-se três problemas centrais do fluxo atual. A seguir, cada ponto é apresentado junto de sua causa principal:

1. Processos heterogêneos entre áreas e Grupos, gerando resultados distintos. Causa: ausência de padronização dos modelos e dos critérios de cálculo.

2. Retrabalho recorrente quando mudanças de preço de fornecedores são percebidas tardiamente. Causa: comunicação lenta das alterações e ausência de validações automáticas.

3. Desorganização operacional e baixa rastreabilidade ao longo das etapas. Causa: uso muito comum de planilhas sem padrão, com controles manuais, versões paralelas e trocas de arquivos.

4.1.2 Mapeamento To Be

A partir daquelas análises, iniciou-se a conversa sobre o mapeamento *To Be* e destacaram-se premissas que o novo processo deve preservar. Em primeiro lugar, considerou-se essencial manter espaço para o julgamento dos gerentes plenos sobre o desempenho esperado

dos produtos, com a possibilidade de sinalizar explicitamente se um item é percebido como mais arriscado, mais certo ou neutro. Em segundo lugar, reconheceu-se a necessidade de assegurar a decisão sobre como redistribuir o *buy* quando a compra sugerida ultrapassar esse limite, garantindo que cada gerente pleno conduza esse ajuste de acordo com o método que julgar mais correto.

A partir dos *inputs* e das dores mapeadas, um novo fluxo foi desenhado e validado e está exposto na Figura 12.

Neste, propõe-se um mapeamento *To Be* que redefine papéis e integrações. Elimina-se a checagem de regras de negócio pelo Planejamento e introduz-se a atuação de TI como quem executa o processamento técnico. Após a definição da coleção, o Comercial preenche uma planilha padronizada, no formato da Figura 11 e a encaminha a TI.

Esse desenho aproveita um caminho já comum na *Fashion*, em que equipes sem perfil técnico acionam TI para operacionalizar rotinas de maior complexidade. TI roda o *script* (o modelo de previsão de demanda de novos produtos a ser desenvolvido) e devolve ao Comercial um *output* com a sugestão de compra por SKU, além de checagens automáticas de regras de negócio de mínimo de quantidade por loja.

Figura 11: Excel Padronizado com Atributos dos SKUs.

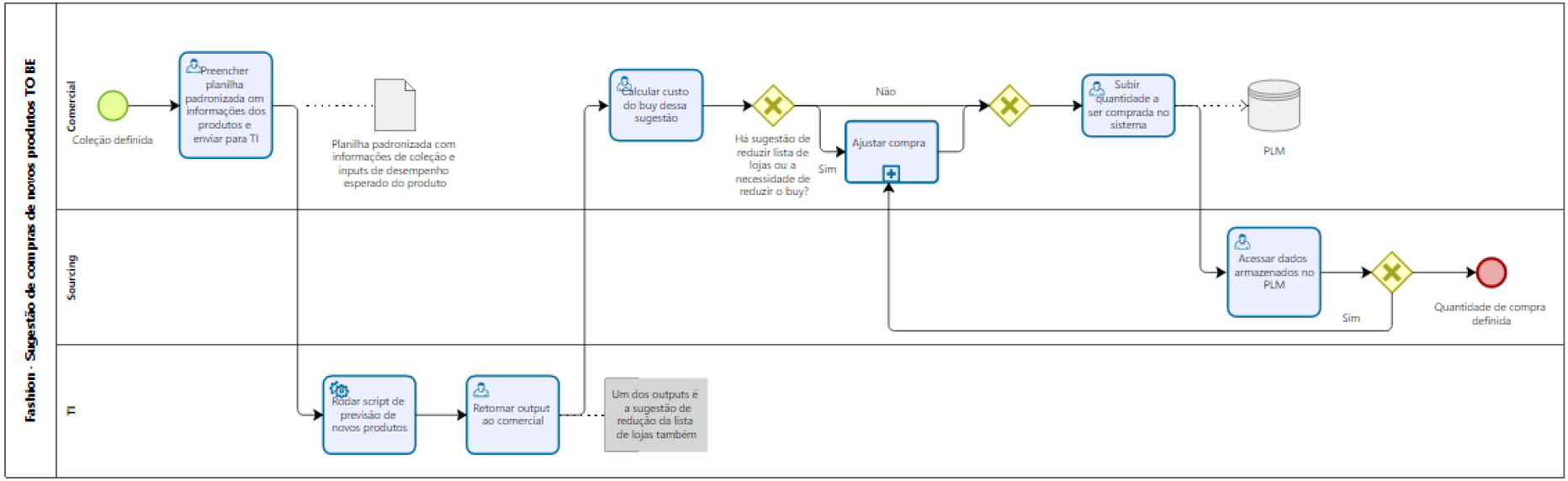
Tamanho	Divisão	Grupo	Subgrupo	Classe	Subclasse	Temática	Modelo	Cor	Faixa de Preço	Tipo de Coleção	Estampa	Origem	Mês de laçamento

Fonte: Autoria própria.

Após receber o *output* (que passa a indicar se há necessidade de reduzir a lista de lojas), o Comercial calcula o custo da compra frente ao *buy* disponível. Em seguida, verifica se é preciso ajustar a lista de lojas e/ou reduzir o *buy* de alguns produtos. Assim, nesses casos, preserva-se também o espaço de julgamento do gerente pleno para conduzir os ajustes. Concluídas as correções, o Comercial sobe a base comprada no PLM (*Product Lifecycle Management*, em português, gestão do ciclo de vida do produto), sem nova passagem pelo Planejamento.

Quanto às mudanças de preço de fornecedor, reconhece-se que continuarão a ocorrer, porém, no novo fluxo, o retorno se dá apenas ao passo de ajuste de compra, uma vez que já há um *output* do modelo padronizado e registro do histórico da sugestão. Isso reduz retrabalho e evita reiniciar o processo inteiro, mantendo rastreabilidade sobre o que foi recomendado e o que foi alterado.

Figura 12: Sugestão de Compras To Be



Fonte: Autoria própria.

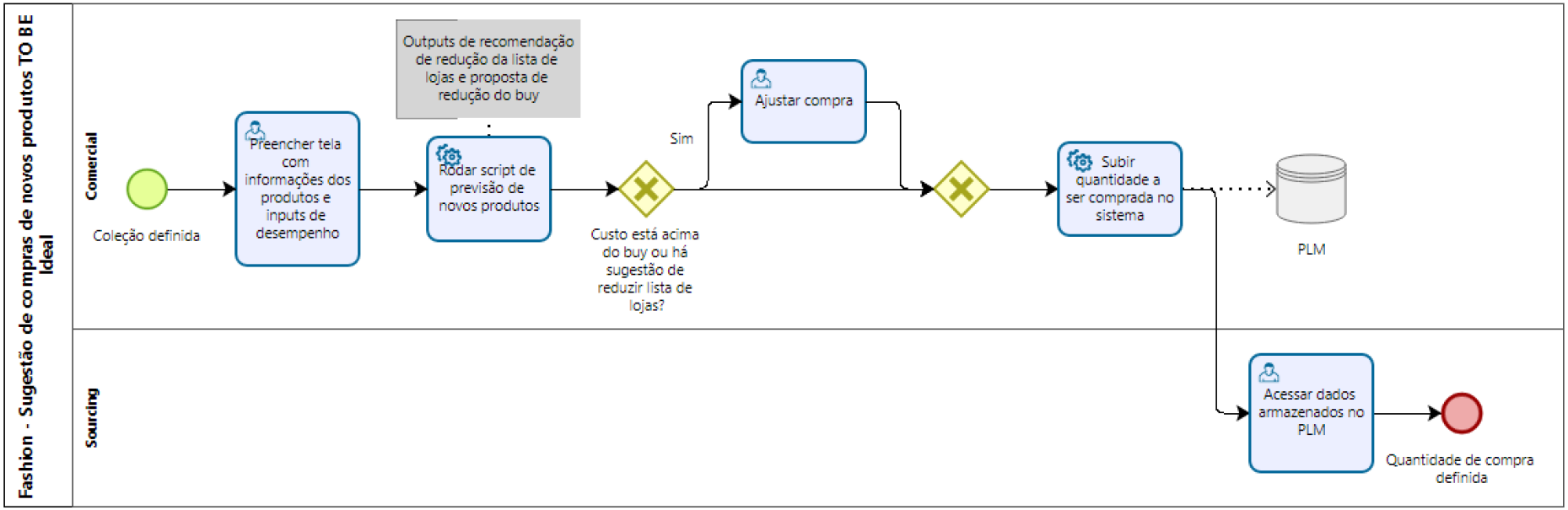
Em relação aos três problemas principais levantados no *As Is*, o processo *To Be* ataca o essencial: cria padronização e rastreabilidade no processo de compras por meio de uma planilha de entrada única e de um *output* padronizado, reorganiza etapas com *checkpoints* claros e reduz a desorganização antes causada por controles paralelos. Soma-se a isso o uso de um modelo de previsão de demanda para novos produtos, que gera uma estimativa consistente, servindo de *baseline* comum e diminuindo a dependência de julgamentos subjetivos. Quanto às mudanças de preço, o desenho encurta o retrabalho: a revisão do *buy* retorna apenas ao passo de ajuste, porém agora com histórico de recomendação registrado.

Pensando no longo prazo, delineou-se também a Figura 13, um mapeamento *To Be* ideal construído com o time da *Fashion*, pensando no melhor cenário para esse fluxo. Nesse desenho, o processo passa exclusivamente pelo Comercial até a etapa final, quando ocorre a transferência para o *Sourcing* apenas para a efetivação da compra com base nos *outputs* gerados, como em todos outros cenários. Ou seja, o Comercial torna-se o único responsável pelo fluxo decisório e operacional, e todo o trabalho é realizado em telas desenvolvidas no ambiente interno da *Fashion*, iniciativa que já vem sendo adotada recentemente, embora com prazos de implementação mais longos.

Na Figura 13, as telas concentram a coleta de insumos, a execução do *script* e a aplicação automática das regras de negócio. O usuário do Comercial consegue registrar percepções de risco (arriscado, certo ou neutro), ajustar a lista de lojas e redistribuir o *buy* quando necessário, diretamente na interface. As mesmas telas realizam a integração com o PLM, eliminando trocas de planilhas, e incorporam verificações imediatas de preço do fornecedor para evitar retrabalhos e recalculações fora do sistema. Ao final, o *Sourcing* recebe os *outputs* consolidados e dá sequência à execução da compra.

A adoção desse fluxo integralmente em sistema reduz transferências internas desnecessárias, padroniza critérios, melhora a rastreabilidade e encurta ciclos de validação. Embora demande esforço tecnológico e tempo de construção das telas, a Figura 13 estabelece a direção desejada para consolidar regras, dados e decisões em um único ambiente, com governança clara e menor dependência de intervenções manuais.

Figura 13: Sugestão de Compras To Be Ideal



Fonte: Autoria própria.

Para que os mapeamentos *To Be* e *To Be Ideal* deixem de ser apenas propostas conceituais e se convertam em mudanças concretas, é necessário estruturar um plano de ação específico para revisão do processo de compras e adoção do novo modelo de previsão.

Um primeiro movimento consiste em indicar, dentro da *Fashion*, um dono formal do projeto, com mandato explícito para conduzir a implementação. Esse responsável deve ter interlocução direta com Comercial, Planejamento, TI e Operações, além de ter parte de sua agenda protegida para o tema, evitando que a iniciativa dependa apenas de esforço residual. Cabe a essa função coordenar o cronograma, consolidar decisões, registrar definições de regras de negócio e organizar a governança mínima do projeto (reuniões periódicas, acompanhamento de indicadores e registro de pendências).

Em seguida, o mapa *To Be* apresentado nesta seção pode ser desdobrado em um plano tático, que traduza cada etapa do fluxo em atividades concretas. Isso inclui, por exemplo, padronizar e publicar o modelo de planilha de entrada, definir quem preenche e em que momento, formalizar o procedimento de entrega dos arquivos à TI, versionar o *script* de previsão e estabelecer critérios para o registro das percepções de risco pelos gerentes plenos.

Nessa fase, também são definidos o escopo do piloto (por exemplo, um conjunto de produtos distribuídos para todas as lojas), o período de teste, os indicadores a serem monitorados (wMAPE, MPE, ruptura, excesso de estoque, prazos de aprovação) e o plano de treinamento dos usuários, garantindo que todos compreendam o novo papel do modelo na rotina de decisão.

Por fim, recomenda-se conduzir o piloto em ciclos curtos, nos quais o Comercial passa a utilizar a planilha padronizada, a TI executa o *script* do modelo de previsão e os gerentes plenos ajustam o *buy* com base no *output* gerado, sempre comparando os resultados com a prática anterior.

Os problemas observados nesses ciclos, como gargalos operacionais, dificuldades de uso, lacunas de dados ou conflitos com regras de negócio, devem retroalimentar tanto a parametrização do modelo quanto o desenho do processo, permitindo ajustes graduais antes de qualquer automação mais profunda.

Uma vez estabilizado o fluxo *To Be* e comprovados ganhos mínimos, o plano de ação pode evoluir para a ampliação do escopo e para o início da jornada tecnológica em direção ao *To Be* Ideal, com desenvolvimento de telas internas, maior integração com o PLM e redução progressiva da dependência de planilhas.

4.2 Previsão de Demanda

Após a apresentação dos resultados da revisão de processos, passa-se aos resultados dos modelos de previsão de demanda, de forma que estes sejam incorporados e passem a compor os processos desenhados anteriormente. Nesta etapa, adota-se a mesma estrutura do capítulo de Método, preservando apenas os tópicos pertinentes a esta seção e em alinhamento com o fluxo *To Be* proposto.

4.2.1 Modelo Conceitual

Tendo em vista tudo que foi mapeado na frente anterior, este trabalho adota uma modelagem de previsão de demanda apenas para os produtos do tipo C, em que se concentram os lançamentos que não contam com informações extras (por exemplo, tendências de moda em alta), com o objetivo de apoiar a primeira decisão de compra em nível de rede, sem envolver políticas de reposição nem o motor de alocação por loja. A escolha de um escopo mais estreito permite maior profundidade analítica no problema de itens sem histórico e separa a previsão de demanda das regras operacionais de distribuição.

4.2.1.1 Unidade de análise e nível de agregação.

A unidade de análise do modelo é o SKU, e a previsão é realizada no nível de rede (agregado), isto é, demanda total do SKU em toda a rede, e não por ponto de venda. Isso está alinhado com a decisão de compra centralizada: primeiro estima-se o volume global a adquirir e decisões de alocação/grade por loja são tratadas por sistemas específicos fora do escopo.

4.2.1.2 Horizonte e ciclo de vida

Considera-se que o comportamento típico e esperado de um lançamento segue um ciclo de aproximadamente 8 semanas. Portanto, o horizonte-alvo de previsão corresponde às oito primeiras semanas do ciclo a partir do início operacional do produto.

4.2.1.3 Início da coleção

Para mitigar distorções de *lead time* e diferenças logísticas em uma rede nacional, considera-se como início do ciclo de vida de cada SKU a primeira data em que ele está disponível em pelo menos 80% da lista planejada de lojas ou, alternativamente, quando as

vendas acumuladas alcançam 5% do total observado na janela analisada, o que acontecer primeiro.

Essa regra evita iniciar a contagem com exposição muito parcial e, ao mesmo tempo, impede atrasos quando o produto já demonstra tração de vendas.

4.2.1.4 Critérios de elegibilidade

Entram no escopo apenas SKUs enviados para pelo menos 80% da rede. O objetivo é reduzir vieses provenientes de decisões táticas de alocação (por exemplo, pilotos regionais, testes de sortimento) que poderiam contaminar a estimativa da demanda agregada do lançamento.

4.2.1.5 Variáveis de saída (alvos)

O alvo é a série semanal de vendas por loja com estoque no horizonte definido e a demanda potencial correspondente no nível rede. Os procedimentos de cálculo estão detalhados em Coleta de dados.

4.2.1.6 Variáveis de entrada

Utilizam-se apenas informações disponíveis no momento da decisão (atributos de produto, calendário/coleção e resumos históricos agregados). E a descrição detalhada das variáveis e chaves também está em Coleta de Dados.

4.2.1.7 Formulação do problema

A previsão de lançamentos é tratada como um problema de previsão com *cold start*: treina-se com coleções passadas e aplica-se em coleções subsequentes, avaliando-se sempre fora do período de treino. O objetivo prático é estimar, desde o início do ciclo, a trajetória semanal e o total das oito semanas no nível rede-lista de lojas, fornecendo um insumo direto para a decisão de compra.

4.2.1.8 Remarcações ao final do ciclo

As remarcações praticadas no encerramento do ciclo não são modeladas, pois entende-se que os descontos de fim de ciclo integram a dinâmica natural do ciclo de vida e, portanto, permanecem fora do escopo desta análise.

4.2.1.9 Objetivo de desempenho

Por fim, o objetivo é reduzir erro absoluto ponderado (wMAPE) e viés (MPE) em relação a *baselines* operacionais.

4.2.2 Coleta de dados

Nesta etapa, os dados disponíveis são mapeados, com identificação de tabelas, colunas, limites de histórico e aplicação de filtros para reter apenas o que é útil ao problema.

Em seguida, realizam-se checagens de qualidade (duplicidades e inconsistências), bem como procedimentos de limpeza e verificação de consistência. Ao final, obtém-se uma base pronta para modelagem, representativa do escopo definido na seção Modelo Conceitual.

4.2.2.1 Fontes, estrutura e histórico

Existem três principais bases: tabela de vendas e estoque diários por produto-loja (registra apenas dias com movimento), tabela de produto e tabela de local. Por ser uma parte extremamente relevante do processo de previsão de novos produtos, é importante explicitar os dados disponíveis na tabela de produto: toda a hierarquia mercadológica (Divisão, Grupo, Subgrupo, Classe, Subclasse, Temática, Modelo, Cor, Tamanho, ID) além de faixa de preço, tipo de coleção (por exemplo, inverno, verão, ano todo), tipo de moda (por exemplo *street*), tipo de estampa (liso, xadrez, *silk*) e origem (nacional e importado).

As bases analisadas foram extraídas e consolidadas em 18/10/2025 e o histórico disponível cobre até 2022, por isso, as análises de novos produtos consideram apenas lançamentos a partir de 2023, quando passa a existir janela histórica para validação de modelos novos.

4.2.2.2 Recortes de escopo (capacidade computacional e comparabilidade)

Para garantir viabilidade computacional e manter comparabilidade entre itens, foram adotados os Grupos, todos eles dentro do Feminino: Básico Feminino, Moda e Ajeitada.

4.2.2.3 Construção das séries e janelas

Para cada SKU elegível, retém-se a janela dos primeiros quatro meses de vida do produto. Essa margem cobre o horizonte definido e sua vizinhança, permitindo consolidar medidas semanais consistentes para previsão e avaliação. Em seguida, é feito o agrupamento e a delimitação das datas de início e, depois, aplica-se um filtro de oito semanas.

4.2.2.4 Filtro de rede e cobertura

Calcula-se o percentual de cobertura como a fração de lojas da lista que apresentam estoque positivo do produto. Esse indicador é utilizado para verificar o atendimento ao critério de elegibilidade definido na seção Modelo Conceitual.

4.2.2.5 Tratamento de rupturas e estimativa potencial

Para lidar com lojas sem estoque, transforma-se a série de vendas em "vendas por loja com estoque" e esse indicador torna-se o objeto de previsão. E para se ter essa visão agregada da rede, agregam-se os dados por todo o período e assim estima-se a trajetória das oito semanas iniciais do ciclo.

Já para obter a demanda potencial total do período, multiplica-se a previsão agregada de vendas por loja pelo número total de lojas da lista, simulando um cenário hipotético em que todas as lojas têm estoque exatamente até o último dia do ciclo de vida, para fins de simplificação.

4.2.2.6 Tratamento de variáveis categóricas e de calendário

As variáveis categóricas, por sua vez, são convertidas em colunas contendo 1 ou 0 apenas para níveis que apresentem, no mínimo, 20 SKUs na base de treino. Esse limiar, equivalente a aproximadamente quatro combinações modelo-cor, foi adotado como critério de frequência mínima para evitar categorias muito raras, reduzindo a dimensionalidade e o risco de sobreajuste.

Para variáveis de calendário associadas a eventos específicos (por exemplo, datas comemorativas), cria-se um indicador binário que assume valor 1 quando a primeira entrada do produto ocorre até 45 dias antes do evento, e 0 caso contrário. A hipótese é que produtos lançados nessa janela podem ter sido planejados para aquele evento. Cabe observar que já existem variáveis de mês na base, que capturam parte dos efeitos sazonais e de

eventos pontuais, de modo que os indicadores de eventos funcionam como refinamento adicional dessa informação.

4.2.2.7 Limpeza e verificações de consistência

Realizam-se rotinas de limpeza (produto-loja-data), tratamento de nulos e padronização de chaves. Períodos afetados por inconsistências ou rupturas prolongadas são excluídos.

4.2.2.8 Resultado da etapa

Como resultado, obtém-se uma base de dados no nível rede-dia, por lista de lojas, contendo: identificadores do SKU e de sua lista, métricas de cobertura semanal, a série de vendas por loja com estoque e as variáveis agregadas necessárias para compor a trajetória e o total do horizonte. Essa base é a entrada para o desenho experimental e para a etapa de modelagem, na qual os cortes de avaliação e os procedimentos de previsão são aplicados.

4.2.3 Tradução do modelo

Nesta etapa o modelo conceitual é traduzido em dados, implementando três abordagens comparáveis: uma previsão por analogia por vários níveis hierárquicos, uma regressão linear com seleção de variáveis por *stepwise* orientado por *wMAPE*, e uma aplicação de *Quantile Regression Forests (QRF)* inspirada em Van Steenbergen & Mes (2020), utilizada aqui sem a etapa de *clustering* por já haver um pré-filtro de perfis.

4.2.3.1 Previsão por Analogia

Constrói-se uma previsão por analogia que estima a demanda a partir de médias históricas calculadas em níveis de uma hierarquia (do mais detalhado ao mais agregado). Os dados de treino são divididos em 80% para treino de fato e 20% para validação, mantendo modelos do mesmo grupo juntos para evitar mistura entre grupos.

Exploram-se duas alavancas, profundidade da hierarquia utilizada e tamanho mínimo de grupo, e para cada combinação, produzem-se previsões e calculam-se *wMAPE*. A melhor configuração (menor *wMAPE*) é então treinada com os dados de treino completos e testada com os dados de teste, e reportam-se *wMAPE* e *MPE* finais.

```
Algoritmo Previsão por Analogia:
1) Definir ano_teste, separar treino (< ano_teste) e dividir treino/validação (80%/20%)
2) Para cada (profundidade, quantidade_min):
     - Calcular médias até ‘‘profundidade’’
     - Fazer a previsão no nível quantidade_min (se quantidade > quantidade_min),
       - Se não, prever em níveis mais agregados
     - Calcular wMAPE
3) Escolher menor wMAPE e aplicar ao teste, reportar wMAPE e MPE
```

4.2.3.2 Regressão Linear

Aplica-se uma regressão linear que inicia vazia e adiciona conjuntos de variáveis aos poucos, com a utilização do *stepwise*, buscando a redução do *wMAPE* médio na validação acima de um limiar para poder adicionar a variável.

Com os blocos aceitos, ajusta-se o modelo nos anos de treino e, no ano de teste, e reportam-se o *wMAPE* e o *MPE*. Para evitar valores negativos, aplica-se um piso igual ao menor valor positivo observado no treino.

```
Algoritmo Regressão Linear:
1) Definir ano_teste, separar treino (< ano_teste) e teste (= ano_teste)
2) Fazer grupos de treino com: treino e validações 80%/20%
3) Iniciar modelo sem preditores
4) Para cada bloco na ordem definida:
     - Avaliar inclusão do bloco nas validações (wMAPE médio)
     - Se reduzir wMAPE acima do limiar, manter o bloco
5) Ajustar modelo final no treino com os blocos aceitos
6) Prever no teste; aplicar piso; calcular e apresentar wMAPE e MPE
```

4.2.3.3 Quantile Regression Forests

Emprega-se o método de *Quantile Regression Forests* (QRF) para estimar a mediana condicional ($q = 0{,}5$) da demanda no ano de teste. Em cada árvore, reúnem-se os valores observados nas folhas alcançadas e, ao agregar esse conjunto ao longo das árvores, toma-se o quantil desejado para obter a previsão. O procedimento divide os dados em 80%/20% por grupos em várias repetições e realiza uma seleção por blocos de variáveis, em ordem predefinida: um bloco é incorporado quando reduz o *wMAPE* médio de validação acima de um limiar, similar à regressão. Com os blocos aceitos, o QRF é ajustado no período de treino e, no ano de teste, calculam-se *wMAPE* e *MPE*. Para evitar valores negativos, aplica-se um piso igual ao menor valor positivo observado no treino.

```
Algoritmo Quantile Regression Forests:
1) Definir ano_teste, separar treino e teste, fixar quantil q=0,5 e piso (>0)
2) Fazer grupos de treino com treino e validações 80%/20%
3) Iniciar sem preditores
4) Para cada bloco na ordem definida:
     - Treinar QRF nos grupos e prever o quantil 50%
     - Calcular wMAPE médio, se reduzir acima do limiar, aceitar o bloco
5) Ajustar QRF final no treino com os blocos aceitos
6) Prever no teste, aplicar piso, reportar wMAPE e MPE
```

4.2.4 Planejamento, execução e análise de experimentos

Os algoritmos foram executados com divisão temporal fixa: treino e validação em 2023–2024 e teste em 2025. Nas validações, utilizaram-se cinco réplicas 80%/20% preservando os grupos de itens. Após os filtros de elegibilidade, o conjunto de teste ficou com aproximadamente 200 SKUs (cerca de 40 modelos–cor) e o conjunto de treino/validação com cerca de 500 SKUs (cerca de 100 modelos-cor).

Com relação ao método de análogos por hierarquia, o resultado da calibração do modelo com os dados de treino, dividindo-o em treino e validação é expresso na Tabela 1. O modelo foi calibrado variando a profundidade da hierarquia na ordem Tamanho do SKU (P,M,G e segue), Divisão e Grupo, isto é, em qual desses níveis da hierarquia de produto é calculada a média de vendas, e o tamanho mínimo de grupo para esse cálculo (e quando não atinge esse número, sobe um nível na hierarquia de produto).

As cinco melhores combinações ficaram praticamente empatadas, com valores de wMAPE muito próximos e diferença inferior a 0,7 de pontos percentuais. Entre elas, para utilizar no teste final para prever os dados de 2025, optou-se pelo primeiro que calcula a média de vendas no nível de Tamaho do SKU quando existem pelo menos 19 SKUs do tamanho na base de treino.

Tabela 1: Top 5 combinações por wMAPE na validação 80/20 por modelo (previsão da venda por loja total).

Nível de Hierarquia	*SKUs mínimos*	wMAPE (%)
1	19	49.78
2	20	49.78
3	2	50.24
4	18	50.39
5	3	50.39

Autoria Própria.

Na regressão linear adotou-se seleção progressiva de blocos, aceitando a inclusão apenas quando a média do wMAPE em validação melhorava pelo menos 0,5 pontos percentuais.

Como observado na Tabela 2, o processo começou com desempenho em torno de 55,3% e apresentou o maior ganho quando se adicionaram os meses do ano, refletindo sazonalidade explícita (queda de 6,6 pontos percentuais). Em seguida, o bloco de tamanhos trouxe nova redução (–1,7 pontos percentuais) e, por fim, a informação de origem do produto (nacional ou importado) ainda contribuiu com pequena melhora (–0,6 pontos percentuais).

Eventos pontuais de calendário e boa parte dos blocos da taxonomia de produto não foram incorporados por não atingirem o limiar de ganho médio, e alguns chegaram a piorar o erro por provável adequação demasiada aos dados de treino. Assim, o modelo final de validação manteve 16 variáveis e wMAPE médio de aproximadamente 46,47% na validação.

Tabela 2: Sequência do stepwise na regressão linear com limiar de 0,5 p.p. no wMAPE médio de validação (previsão da venda por loja total).

Passo	Variáveis	Bloco	Aceito?	wMAPE (%)	Δ p.p.
1	3	Faixa de preço	não	55.29	0.08
2	11	Meses do ano	sim	48.76	6.61
3	1	Calendário: Natal	não	50.27	-1.50
4	1	Calendário: Black Friday	não	48.65	0.11
5	1	Calendário: Dia das Mães	não	48.83	-0.07
6	1	Calendário: Dia dos Pais	não	48.70	0.06
7	1	Calendário: Dia dos Namorados	não	49.68	-0.92
8	1	Calendário: Sábado de Carnaval	não	48.95	-0.19
9	1	Calendário: Ano Novo	não	49.57	-0.81
10	1	Calendário: Dia das Crianças	não	49.13	-0.37
11	4	Produto: Tamanho	sim	47.09	1.67
12	2	Produto: Grupo	não	47.95	-0.86
13	2	Produto: Subgrupo	não	48.40	-1.31
14	9	Produto: Classe	não	49.78	-2.69
15	9	Produto: Subclasse	não	52.95	-5.85
16	1	Produto: Cor	não	47.65	-0.55
17	4	Produto: Tipo de estampa	não	47.15	-0.06
18	1	Produto: Origem (nacional/importado)	sim	46.47	0.62
19	1	Produto: Sazonalidade (verão/ano todo)	não	46.41	0.06
20	4	Produto: Estilo (street, easy etc.)	não	48.69	-2.22

Autoria Própria.

Na tabela 3 estão os coeficientes estimados e respectivos p valores do resultado. De modo geral observa-se: efeitos sazonais marcantes no fim do ano (novembro e dezembro positivos), reduções em meses do outono/inverno para este portfólio (março a maio e agosto negativos), gradiente de tamanho com maior intensidade em tamanhos M e P (positivos) e penalização em PP e GG (negativos), além de um efeito adicional para itens importados.

Como os blocos mais específicos de produto não foram retidos, o modelo preserva caráter mais geral, guiado pelo calendário de lançamento, pela curva de tamanhos e pela condição de importação.

Tabela 3: Coeficientes e p valores do modelo de regressão linear final. Efeitos relativos à categoria de referência de cada grupo de dummies.

variável	coef.	valor-p
Mes_02	-0.377	0.042
Mes_03	-0.444	0.010
Mes_04	-0.473	0.022
Mes_05	-0.718	0.0002
Mes_06	-0.009	0.967
Mes_07	-0.188	0.352
Mes_08	-0.788	0.002
Mes_09	0.084	0.673
Mes_10	-0.412	0.046
Mes_11	0.868	2.48×10^{-6}
Mes_12	1.118	4.46×10^{-10}
Tamanho__GG	-0.247	0.026
Tamanho__M	0.742	4.72×10^{-11}
Tamanho__P	0.432	0.00013
Tamanho__PP	-0.220	0.044
Origem (importado)	0.205	0.015

Autoria Própria.

Por fim, o método *Quantile Regression Forests* foi construído nas mesmas duas etapas: primeiro, definiram-se as bases de treino/validação (2023–2024) e teste (2025) e organizaram-se validações 80%/20% por grupos, depois, aplicou-se uma seleção em blocos de variáveis, aceitando cada bloco apenas quando a média do wMAPE em validação melhorava pelo menos 0,5 pontos percentuais.

O ponto de partida foi a mediana global do treino em cada grupo de validação. Na sequência, a inclusão dos meses do ano trouxe a maior queda de erro, a informação de tamanhos acrescentou novo ganho e a origem do produto (importado versus nacional) consolidou a melhor configuração igual à regressão. Esse encadeamento reduziu o wMAPE de cerca de 55,37% para aproximadamente 46,47%, como observado ena Tabela 4.

Tabela 4: Evolução do wMAPE em validação (média dos 5 splits) ao longo da seleção por blocos no QRF (previsão da venda por loja total).

Marco	wMape (%)	ganho vs. anterior (p.p.)	#Variáveis
baseline (mediana)	55.37	0.00	0
+ Meses do ano	48.76	6.61	11
+ Produto: Tamanho	47.09	1.67	15
+ Produto: Origem (importado)	46.47	0.62	16

Autoria Própria.

Após fixar a configuração, o modelo foi treinado no período de 2023–2024 e avaliado em 2025. A interpretação das importâncias (por redução de impureza na floresta subjacente), exposta na Tabela 5, indica que o tamanho tem um papel extremamente importante, seguido da sazonalidade de fim de ano e de alguns meses específicos, além do efeito de itens importados.

Tabela 5: Importâncias (redução de impureza) do QRF final - principais variáveis.

Variável	Importância
Tamanho__GG	0.253
Tamanho__M	0.206
Mes_12	0.109
Tamanho__PP	0.089
Mes_11	0.070
Origens: importado	0.032
Mes_05	0.026
Tamanho__P	0.022
Mes_03	0.018
Mes_02	0.010

Autoria Própria.

Em termos práticos, o QRF é útil por estimar diretamente um quantil da distribuição condicional (aqui, a mediana), combinando muitas árvores para capturar relações não lineares e interações sem exigir especificação funcional. O resultado final mantém um caráter geral: depende sobretudo do calendário de lançamento (meses, com destaque para novembro e dezembro), da curva de tamanhos (M e P com maior contribuição, PP e GG em sentido oposto) e da condição de importação. Não emergem sinais fortes de atributos

muito específicos de produto, o que favorece a aplicação ampla do modelo ao portfólio e às janelas sazonais analisadas.

Com os modelos calibrados e avaliados, procede-se à comparação entre os três métodos propostos (previsão por analogia da hierarquia mercadológica, regressão linear e *Quantile Regression Forests*) e a referência de compra realizada, que espelha o procedimento atualmente adotado nos Grupos.

Para a compra realizada, considera-se, por SKU, a soma do estoque remanescente no último dia do ciclo de vida com as vendas do ciclo e seu erro a comparação com a venda ajustada ao número de lojas. E, como definido em Método, a avaliação utiliza wMAPE e MPE para sintetizar erro e viés ao longo do portfólio, obtendo o resultado exposto na Tabela 6.

Tabela 6: Resultados no teste de 2025 (previsão da venda por loja total).

Método	MPE	wMAPE
Compra realizada	-73.2	86.2
Análogos hierárquicos	-32.6	56.3
Regressão linear	8.7	47.9
Quantile Regression Forests	10.2	45.1

Autoria Própria.

Os números mostram uma clara progressão de desempenho, a compra realizada apresenta subestimação sistemática (MPE negativo) e o maior erro absoluto, o método por análogos reduz bem o erro, embora ainda com tendência a falta, a regressão linear traz novo ganho e muda o sinal do viés, enquanto a *Quantile Regression Forests* alcança o menor wMAPE do grupo, com viés positivo moderado.

Para reduzir a influência de observações extremas e focar a análise em casos operacionalmente mais informativos, restringiu-se a amostra aos produtos cuja razão entre venda ajustada e compra permaneceu no intervalo $\left[\frac{1}{3}, 3\right]$. Assim, foram excluídos itens com desvios muito acentuados, comuns de situações atípicas de ruptura ou excesso raros, que poderiam distorcer as comparações entre métodos, e o resultado exposto na Tabela 7 foi alcançado.

Tabela 7: Resultados no teste de 2025 após filtro na razão venda/compra $[\frac{1}{3}, 3]$ (previsão da venda por loja total).

método	MPE	wMAPE
Compra realizada	-41.8	56.2
Análogos hierárquicos	-10.0	36.2
Regressão linear	25.4	36.8
Quantile Regression Forests	25.5	35.5

Autoria Própria.

Com a retirada dos casos extremos, o erro absoluto cai para todos os métodos e as diferenças entre eles ficam menores. A *Quantile Regression Forests* mantém a menor taxa de erro, seguida muito próxima pela regressão linear e pelos análogos. O viés da previsão pela hierarquia de produtos permanece negativo, enquanto os modelos preditivos tendem ligeiramente a superestimar.

Apesar dos ganhos obtidos, os métodos propostos não atingem níveis muito baixos de wMAPE. Porém, cabe relembrar que a primeira compra tem essa complexidade por natureza: trata-se de prever a demanda de um item totalmente novo, sem histórico do próprio produto, em um portfólio com grande diversidade de categorias e formas.

Em boa parte da literatura, os cenários analisados reúnem itens mais semelhantes entre si, o que tende a facilitar a aprendizagem. Ainda assim, observou-se redução expressiva do erro quando se compara com a *Quantile Regression Forests*: a queda de wMAPE ficou em torno de 45% frente à compra realizada, e ao filtrar casos extremos, a redução permaneceu próxima de 40%. Esses resultados sugerem uma melhora consistente na qualidade da primeira compra, mesmo em um ambiente de alta heterogeneidade.

Além disso, a *Quantile Regression Forests* acrescenta uma flexibilidade operacional relevante. Por estimar diretamente um quantil da distribuição condicional, é possível alinhar a previsão às expectativas do gerente pleno, requisito descrito no processo de mapeamento.

Uma regra prática que pode ser adotada é: para itens percebidos como mais arriscados, usar um quantil mais baixo (por exemplo, 35%–40%), para itens neutros, manter a mediana (50%), para itens com expectativa positiva, adotar um quantil superior (60%–65%), e, em apostas mais fortes, avançar para 70%–75%. Essa lógica pode também ser replicada no método por análogos, em que, em vez de tomar a média das vendas no nível hierárquico selecionado, calculam-se os mesmos quantis sobre o conjunto de análogos,

mantendo coerência entre os modelos e abrindo espaço para incorporação explícita do julgamento gerencial.

Outro aspecto observado é que, para todos os modelos, as variáveis retidas convergem para atributos gerais: calendário de lançamento, distribuição de tamanhos e condição de importação. Não emergiram efeitos robustos associados a especificações muito granulares de produto. Esse caráter mais geral é compatível com o estágio da decisão e com as restrições de dados desta aplicação, mas levanta o questionamento sobre se não seria possível extrair mais ganho com uma maior representatividade dos dados.

Por fim, do ponto de vista de uso, o modelo de previsão devolve, para cada SKU, uma quantidade sugerida de compra em nível de rede. A partir desse valor, realiza-se a comparação com a lista de lojas planejadas, gerando, na própria base de dados, um indicador que sinaliza se o volume proposto é suficiente para atender à cobertura mínima por loja, conforme previsto no desenho *To Be*.

Essa lógica operacional está diretamente articulada à revisão de processos: a planilha de entrada padronizada, os pontos formais de checagem de regras de negócio e a etapa de ajuste do *buy* pelos gerentes plenos foram concebidos justamente para receber o *output* do modelo e incorporá-lo ao fluxo decisório. Assim, a previsão deixa de ser um cálculo isolado e passa a funcionar como insumo estruturado e rastreável dentro do processo de compras.

O fluxo de utilização torna-se, assim, coerente com o processo redesenhado: preenche-se o cadastro do SKU com atributos de produto e variáveis de calendário (já organizados na tabela de entrada proposta na revisão de processos), TI executa o *script* de previsão, o modelo gera a quantidade sugerida e os indicadores de cobertura por loja, e o Comercial utiliza essas informações como *baseline* comum para ajustar a compra, respeitando o *buy* disponível e suas percepções de risco. A interação entre modelo e processo garante rastreabilidade, padronização dos insumos e clareza sobre em que etapa e de que forma a previsão passa a influenciar a decisão.

Em seguida, o próximo capítulo retoma os objetivos propostos, sintetiza as principais contribuições do estudo, discute suas limitações e aponta desdobramentos e oportunidades para futuras aplicações.

5 CONCLUSÃO

Este capítulo encerra o trabalho, sintetizando os principais resultados alcançados, registrando as limitações do escopo desenvolvido e discutindo os desdobramentos práticos e as oportunidades de evolução para a *Fashion*.

5.1 Síntese

Este Trabalho de Formatura partiu do desafio de dimensionar a primeira compra de produtos sem histórico em uma grande rede *fast fashion*, voltada às classes B e C, com alta variedade de SKUs e ciclos de vida curtos. A pesquisa combinou duas frentes complementares: o mapeamento do processo de compras e de lançamento de produtos em BPMN, nas visões *As Is*, *To Be* e *To Be Ideal*, e a modelagem para previsão de demanda de novos produtos, com comparação de métodos e avaliação por métricas operacionais aderentes ao uso gerencial.

No eixo de processos, as entrevistas com gestores sênior e pleno permitiram descrever o fluxo ponta a ponta, tornando explícitas decisões, responsabilidades, integrações sistêmicas e pontos de dor. O mapeamento *As Is* revelou problemas críticos para a operação: baixa padronização, forte heterogeneidade entre compradores, uso inconsistente de métodos de previsão e decisões pouco rastreáveis, o que se traduzia em compras com erros elevados.

A partir desse diagnóstico, o trabalho trouxe implicações práticas diretas para a empresa: maior clareza e alinhamento sobre como o processo realmente funciona, a identificação estruturada dos principais gargalos (especialmente a inconsistência de métodos e o nível de erro observado nas compras) e a formulação de uma proposta de evolução em etapas.

Nas visões *To Be* e *To Be Ideal*, essa evolução é organizada em um roteiro de mudanças: em um primeiro momento, ações de menor complexidade (padronização de ar-

tefatos e indicadores, explicitação do papel da previsão no fluxo e regras mínimas de governança) e, em um horizonte futuro, um processo mais robusto, sistematizado e suportado por sistemas e modelos preditivos. Com isso, o trabalho não apenas descreve o processo, mas entrega um caminho concreto de melhoria contínua.

No eixo quantitativo, o trabalho propõe um modelo de previsão de demanda sem histórico, baseado em atributos dos produtos, comparando três abordagens: previsão por analogia, regressão linear com seleção de variáveis orientada por wMAPE e uma adaptação de *Quantile Regression Forests* (arquitetura inspirada em Steenbergen e Mes (2020)). O desenho experimental contemplou separação entre treino, validação e teste, foco em wMAPE (interpretação gerencial) e MPE (viés médio da compra), além de análises com e sem *outliers*.

Os resultados mostram que o modelo de *Quantile Regression Forests* supera tanto os métodos alternativos quanto a prática atual de compra, reduzindo o erro e o viés e oferecendo previsões em quantis, que podem ser ajustadas conforme o apetite de risco.

Em conjunto, o redesenho de processo e o estudo empírico não apenas respondem ao problema acadêmico de previsão de novos produtos, mas entregam à empresa um caminho pragmático: um processo mais claro e consistente hoje, um diagnóstico objetivo dos erros e incoerências atuais e um plano de melhoria por etapas rumo a um processo decisório de compras mais padronizado, analítico e robusto no futuro.

5.2 Limitações

No que se refere às limitações deste Trabalho de Formatura, a principal diz respeito ao fato de o projeto ainda não ter sido efetivamente implementado na empresa. As análises realizadas se baseiam em modelagem, simulações e discussões de processo, o que permite comparar cenários e quantificar ganhos potenciais, mas não observar o comportamento real dos usuários, dos sistemas e dos indicadores de negócio ao longo do tempo.

Sem um piloto em ambiente produtivo, não é possível mensurar, por exemplo, o impacto efetivo sobre o erro de previsão, ruptura, excesso de estoque ou giro na prática, nem capturar eventuais efeitos colaterais (como sobrecarga operacional, necessidade de retrabalho ou ajustes finos nas regras de negócio).

Além disso, os mapas *To Be* e *To Be Ideal* indicam ganhos claros (padronização, redução de retrabalho e maior integração de dados), mas sua execução demanda uma jornada de implementação complexa. Como discutido ao longo do trabalho, seria necessário

desenvolver ou evoluir ferramentas, telas e integrações sistêmicas, além de redefinir papéis, explicitar responsáveis e formalizar decisões hoje dispersas em planilhas e trocas informais. Isso implica alinhar áreas (Comercial, Planejamento, TI, Operações), priorizar demandas em meio a outros projetos concorrentes, treinar usuários para novos fluxos e estabelecer uma governança contínua de dados e processos. Essa trilha de implementação (priorização, execução e governança) não esteve no escopo do estudo e condiciona a captura integral dos benefícios aqui estimados.

Ademais, é importante relembrar as simplificações que este trabalho precisou fazer para viabilizar um estudo aprofundado. Para garantir comparabilidade e capacidade computacional, adotaram-se recortes de escopo na etapa de dados, o que reduziu a abrangência e pode ter afetado a acurácia máxima alcançável. Em particular:

- Utilização de uma única lista de lojas para a rede inteira (em vez de múltiplas listas ou segmentações).
- Foco em apenas três Grupos da hierarquia mercadológica, dentre cerca de 20 existentes na empresa.

Essas opções, embora justificadas metodologicamente, limitam a generalização dos resultados e podem subestimar ganhos adicionais que surgiriam ao modelar especificidades por *cluster* de loja e por um número maior de Grupos.

Adicionalmente, mesmo que a empresa esteja posicionada no universo *fast fashion*, o foco da *Fashion* nas classes B e C pressiona por maior variedade de sortimento. Como discutido na literatura, nessas condições o desempenho de um produto depende mais do conjunto da proposta de valor (combinação de atributos, tendência, preço relativo e exposição) do que de atributos isolados. Essa característica eleva a variância não explicada por modelos baseados apenas em atributos de produto e torna mais difícil alcançar previsões de alta precisão.

Nesse contexto, mesmo com os ganhos obtidos pelos métodos testados, a experiência e o julgamento dos gestores seguem sendo fundamentais. Sendo a experiência do gestor fundamental para o ajuste fino das decisões de compra, reforçando o caráter complementar da abordagem preditiva proposta.

5.3 Desdobramentos e oportunidades

Com relação aos desdobramentos e oportunidades evidenciados ao longo deste trabalho, o processo de mapeamento e discussão foi conduzido com gerentes com os quais o autor tem elevada abertura, o que favoreceu tanto a coleta de informações quanto a validação dos resultados. Os ganhos já obtidos foram apresentados a esses gestores, juntamente com a perspectiva de ganhos adicionais decorrentes de um uso mais intensivo da base de dados de atributos já disponível na *Fashion*.

Nesse contexto, os gerentes se mostraram receptivos à proposta e passaram a avaliar a realização de um piloto aplicando o modelo de *Quantile Regression Forests* a um conjunto de produtos distribuídos para todas as lojas, como forma de testar, em ambiente real, o potencial da abordagem preditiva sugerida.

Dado esse alinhamento inicial, um primeiro passo concreto de implementação seria colocar em prática o processo descrito no mapa *To Be*, tomando-o como ponto de partida antes de perseguir o cenário *To Be Ideal*. Na prática, isso significaria organizar as rotinas de previsão de demanda em torno do fluxo proposto, padronizando entradas e saídas de informação e incorporando o uso sistemático dos modelos desenvolvidos em um ciclo piloto. Esse piloto poderia ser conduzido justamente sobre o conjunto de produtos já distribuídos para todas as lojas, permitindo comparar, em condições controladas, o desempenho do novo processo em relação às práticas atuais.

Em termos práticos, para que esse movimento saia do papel, seria necessário designar um responsável claro pelo projeto por parte da empresa, preferencialmente alguém com interface transversal com Planejamento, Comercial, TI e Operações. Essa pessoa precisaria ter parte de sua carga de trabalho formalmente alocada ao projeto, coordenando um plano de atividades e priorizando, junto às áreas envolvidas, o que entra no escopo inicial do piloto.

A execução de alguns ciclos nesse arranjo simplificado permitiria identificar problemas esperados (gargalos operacionais, resistências, lacunas de dados) e ajustar tanto os parâmetros dos modelos quanto o desenho do processo, antes de investir em integrações sistêmicas mais profundas.

Em um horizonte de médio prazo, os mapas *To Be* e *To Be Ideal* podem servir como guia para a evolução da solução após a validação do piloto. Além do interesse já manifestado pelos gestores em ampliar o escopo para um número maior de listas de lojas e de Grupos da hierarquia mercadológica, seria possível aproveitar de forma mais

intensa os recursos tecnológicos da empresa para explorar os dados de todos os Grupos, automatizar fluxos, integrar previsões aos sistemas transacionais e consolidar painéis e rotinas de governança.

E esse aprofundamento permitiria refinar em maior grau os modelos propostos, capturar melhor a heterogeneidade da rede e dos portfólios e aproximar ainda mais a solução desenvolvida das necessidades reais da operação.

REFERÊNCIAS

ANITHA, S.; NEELAKANDAN, R. Demand forecasting new fashion products: A review paper. *Journal of Forecasting*, v. 44, n. 2, p. 270–280, mar. 2025. Disponível em: ⟨https://doi.org/10.1002/for.3192⟩.

BANKS, J. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. [S.l.]: John Wiley & Sons, 1998.

BASALLO-TRIANA, M. J.; RODRÍGUEZ-SARASTY, J. A.; BENITEZ-RESTREPO, H. D. Analogue-based demand forecasting of short life-cycle products: a regression approach and a comprehensive assessment. *International Journal of Production Research*, v. 55, n. 8, p. 2336–2350, 2017. Disponível em: ⟨https://doi.org/10.1080/00207543.2016.1241443⟩.

BEHESHTI-KASHI, S. et al. A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering*, v. 3, n. 1, p. 154–161, 2015. Disponível em: ⟨https://doi.org/10.1080/21642583.2014.999389⟩.

BISHOP, C. M. *Pattern Recognition and Machine Learning*. [S.l.]: Springer, 2006.

BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. Disponível em: ⟨https://doi.org/10.1023/A:1010933404324⟩.

BREIMAN, L. et al. *Classification and Regression Trees*. [S.l.]: Wadsworth International Group, 1984.

CAO, N. et al. How are supply chains coordinated? an empirical observation in textile-apparel businesses. *Journal of Fashion Marketing and Management*, v. 12, n. 3, p. 384–397, 2008. Disponível em: ⟨https://doi.org/10.1108/13612020810889326⟩.

CARO, F.; GALLIEN, J. Clearance pricing optimization for a fast-fashion retailer. *Operations Research*, v. 60, n. 6, p. 1404–1422, 2012. Disponível em: ⟨https://doi.org/10.1287/opre.1120.1102⟩.

CHEN, D. et al. Sales forecasting for fashion products considering lost sales. *Applied Sciences*, v. 12, p. 7081, 2022. Disponível em: ⟨https://doi.org/10.3390/app12147081⟩.

DAVYDENKO, A.; FILDES, R. Measuring forecasting accuracy: The case of judgmental adjustments to sku-level demand forecasts. *International Journal of Forecasting*, v. 29, p. 510–522, 2013. Disponível em: ⟨https://doi.org/10.1016/j.ijforecast.2012.09.002⟩.

EGLITE, L.; BIRZNIECE, I. Retail sales forecasting using deep learning: Systematic literature review. *Complex Systems Informatics and Modeling Quarterly*, n. 30, p. 53–62, 2022. Disponível em: ⟨https://doi.org/10.7250/csimq.2022-30.03⟩.

FISHER, M.; RAMAN, A. Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research*, v. 44, n. 1, p. 87–99, 1996. Disponível em: ⟨https://www.jstor.org/stable/171907⟩.

GIRI, C. et al. Forecasting new apparel sales using deep learning and nonlinear neural network regression. In: *2019 International Conference on Engineering, Science, and Industrial Applications (ICESI)*. IEEE, 2019. p. 1–6. Disponível em: ⟨https://doi.org/10.1109/ICESI.2019.8863024⟩.

JIN, B. E.; SHIN, D. C. Changing the game to compete: Innovations in the fashion retail industry from disruptive business models. *Business Horizons*, v. 63, p. 301–311, 2020. Disponível em: ⟨https://doi.org/10.1016/j.bushor.2020.01.004⟩.

KHARFAN, M.; CHAN, V. W. K.; EFENDIGIL, T. F. A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. *Annals of Operations Research*, v. 303, p. 159–174, 2021. Disponível em: ⟨https://doi.org/10.1007/s10479-020-03666-w⟩.

MEINSHAUSEN, N. Quantile regression forests. *Journal of Machine Learning Research*, v. 7, p. 983–999, 2006. Disponível em: ⟨http://www.jmlr.org/papers/v7/meinshausen06a.html⟩.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to Linear Regression Analysis*. 5. ed. [S.l.]: John Wiley & Sons, Inc., 2012.

NI, Y.; FAN, F. A two-stage dynamic sales forecasting model for the fashion retail. *Expert Systems with Applications*, v. 38, p. 1529–1536, 2011. Disponível em: ⟨https://doi.org/10.1016/j.eswa.2010.07.065⟩.

SILVER, B. *BPMN Method and Style, Second Edition, with BPMN Implementer's Guide*. [S.l.]: Cody-Cassidy Press, 2011.

STEENBERGEN, R. M. van; MES, M. R. K. Forecasting demand profiles of new products. *Decision Support Systems*, v. 139, p. 113401, 2020. Disponível em: ⟨https://doi.org/10.1016/j.dss.2020.113401⟩.

THOMASSEY, S. Sales forecasts in clothing industry: The key success factor of the supply chain management. *International Journal of Production Economics*, v. 128, p. 470–483, 2010. Disponível em: ⟨https://doi.org/10.1016/j.ijpe.2010.07.018⟩.

THOMASSEY, S.; FIORDALISO, A. A hybrid sales forecasting system based on clustering and decision trees. *Decision Support Systems*, v. 42, p. 408–421, 2006. Disponível em: ⟨https://doi.org/10.1016/j.dss.2005.01.008⟩.