

# **João Luis Ract de Almeida Pereira**

Desenvolvimento de um sistema de previsão de demanda multi-item baseado em  
Machine Learning e análise multicritério no contexto do PCP 4.0: uma  
abordagem sob a ótica da OIPT

PROJETO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO  
APRESENTADO AO DEPARTAMENTO DE ENGENHARIA INDUSTRIAL  
DA PUC-RIO, COMO PARTE DOS REQUISITOS PARA OBTENÇÃO  
DO TÍTULO DE ENGENHEIRO DE PRODUÇÃO

Orientador: Rodrigo Caiado  
Coorientador: Renan Santos

Departamento de Engenharia Industrial  
Rio de Janeiro, 24 de Novembro de 2025.

## **Agradecimentos**

Aos meus pais, João Carlos e Cynthia, por terem me dado todo o suporte e incentivo para atingir todos os meus objetivos e por sempre acreditarem em mim. Nada disso seria possível sem vocês.

À minha irmã, Gabriella, pela amizade e estar sempre ao meu lado.

Ao Ainda e a todos os meus amigos, pelos momentos ao lado de vocês.

À minha namorada Mariana, por todo o carinho e apoio que você me dá.

Aos meus orientadores, Rodrigo e Renan, pelos ensinamentos e orientação durante todo o processo.

Ao Vasco da Gama, pelos amigos que fiz e momentos que vivi te acompanhando.

## Resumo

A previsão de demanda ocupa papel central na gestão da cadeia de suprimentos e no Planejamento e Controle da Produção (PCP), especialmente em ambientes marcados por alta incerteza, volatilidade e complexidade, como aqueles associados à Indústria 4.0 e ao PCP 4.0. Neste contexto, modelos estatísticos clássicos, embora amplamente utilizados, mostram limitações para lidar com múltiplos padrões sazonais, relações não lineares e grande volume de dados e variáveis exógenas. Amparado pela Teoria do Processamento de Informação Organizacional (OIPT), este trabalho parte da premissa de que a seleção de modelos de previsão deve estar alinhada ao nível de incerteza da demanda e à capacidade de processamento de informação da organização, buscando um fit entre complexidade da tarefa e sofisticação analítica da ferramenta. Por meio da metodologia de Design Science Research, é desenvolvido um artefato computacional sob a forma de um Sistema de Apoio à Decisão (SAD) que integra: (i) um pipeline de previsão de demanda multi–item, combinando modelos estatísticos e de Machine Learning; e (ii) um módulo de avaliação multicritério que considera simultaneamente métricas de acurácia, viés, estabilidade e desempenho computacional. O artefato é aplicado a dados reais extraídos do ERP de uma empresa industrial, conduzindo uma “competição de modelos” em múltiplas séries temporais no nível de SKU. Os resultados evidenciam que não existe um modelo dominante universal; ao contrário, diferentes algoritmos apresentam melhor desempenho em subconjuntos específicos de itens, reforçando a lógica contingencial da OIPT. A avaliação multicritério permite adaptar a recomendação de modelos a diferentes prioridades organizacionais (por exemplo, foco máximo em acurácia, equilíbrio entre acurácia e custo computacional ou busca por maior robustez e estabilidade), oferecendo ao gestor uma ferramenta prática para seleção de modelos alinhada à estratégia da empresa. O estudo contribui teoricamente ao operacionalizar a OIPT no contexto da previsão de demanda e, do ponto de vista prático, ao disponibilizar um protótipo de SAD que pode ser incorporado aos processos de S&OP e PCP 4.0, apoiando uma tomada de decisão mais informada, ágil e aderente às exigências de cadeias de suprimentos complexas e digitais.

**Palavras-chave:** Previsão de demanda; Machine Learning; PCP 4.0; Teoria do Processamento de Informação Organizacional; OIPT; Análise multicritério; Sistema de Apoio à Decisão; S&OP; IA

# Development of a Multi-Item Demand Forecasting System Based on Machine Learning and Multicriteria Analysis in the Context of PPC 4.0: An OIPT-Based Approach

## Abstract

Demand forecasting plays a central role in supply chain management and Production Planning and Control (PPC), especially in environments characterized by high uncertainty, volatility, and complexity, such as those associated with Industry 4.0 and PPC 4.0. In this context, traditional statistical models, although widely used, exhibit important limitations when dealing with multiple seasonal patterns, nonlinear relationships, and large volumes of data and exogenous variables. Grounded in Organizational Information Processing Theory (OIPT), this study assumes that the selection of forecasting models must be aligned with both the level of demand uncertainty and the organization's information processing capacity, seeking a fit between task complexity and analytical sophistication. Using the Design Science Research methodology, a computational artefact is developed in the form of a Decision Support System (DSS) that integrates: (i) a multi-item demand forecasting pipeline combining statistical and Machine Learning models; and (ii) a multicriteria evaluation module that simultaneously accounts for accuracy, bias, stability, and computational performance. The artefact is applied to real data extracted from the ERP system of an industrial company, enabling a "model competition" across multiple time series at the SKU level. The results show that there is no universally dominant model; instead, different algorithms perform best for specific subsets of items, reinforcing OIPT's contingency logic. The multicriteria assessment makes it possible to adapt model recommendations to distinct organizational priorities (e.g., maximum focus on accuracy, balance between accuracy and computational cost, or emphasis on robustness and stability), providing managers with a practical tool for model selection aligned with the firm's strategy. This study contributes to theory by operationalizing OIPT in the context of demand forecasting and, from a practical standpoint, by delivering a DSS prototype that can be embedded into S&OP and PPC 4.0 processes, supporting more informed, agile, and strategy-consistent decision-making in complex and digital supply chains.

**Keywords:** Demand forecasting; Machine Learning; PPC 4.0; Organizational Information Processing Theory; OIPT; Multicriteria analysis; Decision Support System; S&OP; AI

## Sumário

1. Introdução.....	1
2. Referencial Teórico .....	6
2.1 Indústria 4.0 .....	6
2.2 PCP 4.0 .....	8
2.3 O Panorama dos Modelos Preditivos na Previsão de Demanda.....	9
2.3.1 Modelos Estatísticos Clássicos.....	11
2.3.1.1 ETS (Error, Trend, Seasonality) .....	11
2.3.1.2 Método de Holt-Winters.....	12
2.3.1.3 ARIMA/SARIMA .....	12
2.3.2 Modelos de Machine Learning (ML) .....	13
2.3.2.1 Modelo Prophet .....	13
2.3.2.2 Random Forest.....	13
2.3.2.3 XGBoost (Extreme Gradient Boosting) .....	14
2.3.3 Modelos de Deep Learning (Redes Neurais).....	15
2.3.3.1 N-BEATS (Neural Basis Expansion Analysis for Interpretable Time Series) .....	15
2.4 OIPT .....	16
3. Metodologia.....	19
3.1 Design da pesquisa .....	19
3.1.1 Fase Qualitativa: Identificação do Problema e Motivação e Definição dos Objetivos da Solução .....	20
3.1.2 Fase Quantitativa: Modelagem Experimental .....	21
3.2 Métodos de Coleta de Dados .....	22
3.2.1 Entrevistas Não Estruturadas com Especialistas .....	22
3.2.2 Dados Secundários: Extração e Tratamento de Dados do Sistema ERP .....	23
3.3 Métodos de análise de dados .....	24
3.3.1 Pré-processamento.....	24
3.3.1.1 Diagnóstico de Distribuição e Testes de Normalidade.....	24
3.3.1.2 Verificação de Estacionariedade (Raiz Unitária) .....	26
3.3.1.3 Decomposição Estrutural e Sazonalidade.....	26
3.3.1.4 Tratamento de <i>Outliers</i> e Estabilização da Série.....	27
3.3.1.4.1 Detecção e Tratamento de <i>Outliers</i> .....	27
3.3.1.4.2 Transformações para Estabilização de Variância .....	28
3.3.1.5 Indução de Estacionariedade (Diferenciação) .....	29
3.3.1.6 Padronização de Escala para Machine Learning .....	29

3.3.2 Processamento .....	30
3.2.2.1 Estrutura de Autocorrelação: Variáveis de Defasagem ( <i>Lags</i> ) .....	30
3.2.2.2 Tendência Local: Estatísticas Móveis ( <i>Rolling Window</i> ).....	31
3.2.2.3 Sazonalidade Determinística: Componentes Calendários .....	31
3.2.2.4 Prevenção de Vazamento de Dados ( <i>Data Leakage</i> ) .....	31
3.2.2.5 Estratégia de Validação Temporal.....	32
3.2.2.6 Otimização de Hiperparâmetros (HPT).....	33
3.3.1. Pós processamento e Estruturação do Problema de Decisão Multicritério (MCDM) .....	34
3.3.5.2. Normalização e Agregação (Método SAW).....	35
4. Design e Desenvolvimento do Artefato.....	37
4.1 Página Inicial e Módulo de Upload de Dados .....	37
4.2 Módulo de Análise Exploratória .....	39
4.3. Módulo de Pré-Processamento .....	42
4.4. Módulo de Treinamento e Otimização de Modelos .....	44
4.5. Módulo de Testes e Validação com Dados Novos .....	48
4.6. Interface de Avaliação e Sistema de Apoio à Decisão (Dashboard) .....	50
4.7. Registro dos experimentos .....	56
5. Aplicação do artefato em um caso real.....	57
5.1. Análise de Desempenho Global e a Fronteira de Eficiência .....	58
5.2. Avaliação Multicritério: O Impacto da Estratégia na Seleção .....	59
5.2.1. Cenário 1: Foco Estratégico na Precisão .....	60
5.2.2. Cenário 2: Gestão de riscos .....	61
5.2.3. Cenário 3: Equilíbrio Operacional.....	63
5.3 Discussões .....	65
6. Conclusão .....	67
7. Referências .....	69

## **Lista de figuras**

Figura 1 – Pipeline metodológico. Fonte: Autoria própria, 2025.....	19
Figura 2- Tela de Upload de dados. Fonte: Autoria própria, 2025.....	39
Figura 3- Tela de Visualizações – Análise Exploratória. Fonte: Autoria própria, 2025. ....	39
Figura 4- Tela de Estatísticas descritivas – Análise Exploratória. Fonte: Autoria própria, 2025. ....	40
Figura 5- Tela de Testes de Estacionariedade – Análise Exploratória. Fonte: Autoria própria, 2025. ....	41
Figura 6- Tela de Decomposição e Sazonalidade – Análise Exploratória. Fonte: Autoria própria, 2025.....	42
Figura 7- Tela de Detecção de Outliers – Pré-Processamento. Fonte: Autoria própria, 2025. 43	
Figura 8- Tela de Transformações matemáticas – Pré-Processamento. Fonte: Autoria própria, 2025. ....	43
Figura 9- Tela de Normalização – Pré-Processamento. Fonte: Autoria própria, 2025. ....	44
Figura 10 - Tela de Seleção de modelos – Treinamento. Fonte: Autoria própria, 2025. ....	46
Figura 11- Tela de Sugestão automática de modelos – Treinamento. Fonte: Autoria própria, 2025. ....	47
Figura 12- Tela de Configuração – Treinamento. Fonte: Autoria própria, 2025. ....	48
Figura 13 - Tela de Visualizar Previsões – Testes. Fonte: Autoria própria, 2025.....	49
Figura 14- Tela de Comparação de modelos – Testes. Fonte: Autoria própria, 2025.....	50
Figura 15- Tela de Visão Geral – Dashboard. Fonte: Autoria própria, 2025.....	51
Figura 16- Tela de Análise Multicritério – Dashboard. Fonte: Autoria própria, 2025. ....	52
Figura 17- Tela de Comparação de erros – Dashboard. Fonte: Autoria própria, 2025. ....	53
Figura 18- Tela de Performance computacional – Dashboard. Fonte: Autoria própria, 2025.54	
Figura 19- Tela de Ranking final multicritério – Dashboard. Fonte: Autoria própria, 2025. ....	55
Figura 20- Tela de Previsão futura – Dashboard. Fonte: Autoria própria, 2025 .....	55
Figura 21- Tela de Registro de experimento. Fonte: Autoria própria, 2025 .....	56
Figura 22– Fronteira de eficiência dos modelos. Fonte: Autoria própria, 2025.....	59
Figura 23– Percentual de itens onde cada modelo obteve o melhor Score no cenário 1. Fonte: Autoria própria, 2025. ....	61
Figura 24– Percentual de itens onde cada modelo obteve o melhor Score no cenário 2. Fonte: Autoria própria, 2025. ....	63
Figura 25– Percentual de itens onde cada modelo obteve o melhor Score no cenário 3. Fonte: Autoria própria, 2025. ....	64

## **Lista de Tabelas**

Tabela 1– Comparação global dos modelos. Fonte: Autoria própria, 2025. ....	58
Tabela 2– Pesos de cada métrica no cenário 1. Fonte: Autoria própria, 2025 .....	60
Tabela 3– Pesos de cada métrica no cenário 2. Fonte: Autoria própria, 2025. ....	61
Tabela 4– Scores no cenário 2. Fonte: Autoria própria, 2025.....	62
Tabela 5– Scores no cenário 3. Fonte: Autoria própria, 2025.....	63
Tabela 6– Scores no cenário 3. Fonte: Autoria própria, 2025.....	64

## 1. Introdução

A gestão da cadeia de suprimentos (*Supply Chain Management* - SCM) se tornou um pilar estratégico para a competitividade das empresas em um mercado globalizado, dinâmico e cada vez mais centrado no cliente. Nesse cenário, a capacidade de antecipar as necessidades do mercado e alinhar os recursos produtivos e logísticos de forma eficiente não é apenas uma vantagem competitiva, mas uma condição essencial para a sobrevivência e prosperidade dos negócios. No centro dessa questão está a previsão de demanda, um processo crítico que serve como alicerce para virtualmente todas as decisões de planejamento, desde o nível estratégico até o operacional. (ATHANASOPOULOS et al., 2017)

A precisão na previsão de demanda impacta diretamente na eficiência operacional, satisfação do cliente e rentabilidade de uma empresa. (MAKRIDAKIS ET AL., 1998) No nível estratégico, projeções acuradas permitem identificar tendências de mercado, antecipar oportunidades de crescimento e se preparar contra possíveis ameaças. Taticamente, elas são indispensáveis para o planejamento de vendas e operações (S&OP), alocação de recursos, estratégias de marketing e gestão de portfólio de produtos. No âmbito operacional, as previsões guiam o planejamento e controle da produção (PCP), a gestão de estoques, a programação da produção, o planejamento de necessidades de materiais (MRP) e a logística de distribuição, buscando o equilíbrio ótimo entre o nível de serviço e os custos operacionais. (FILDES et al., 2022)

Contudo, prever a demanda com precisão tornou-se uma tarefa exponencialmente mais desafiadora. O ambiente de negócios contemporâneo é caracterizado por uma volatilidade crescente, ciclos de vida de produtos cada vez mais curtos, alta customização e a complexidade introduzida por novos modelos de negócio, como o varejo *omnichannel*. As necessidades dos clientes tornaram-se mais dinâmicas e complexas, aumentando as limitações dos modelos preditivos, que frequentemente enfrentam a escassez de séries históricas robustas e a dificuldade de capturar o comportamento errático das variáveis envolvidas. (HOFMANN e RUTSCHMANN, 2018)

Historicamente, as empresas têm se apoiado em métodos estatísticos consagrados para a análise de séries temporais, como os modelos de Suavização Exponencial (a exemplo da família ETS - Erro, Tendência e Sazonalidade) e os modelos Autorregressivos Integrados de Médias Móveis (ARIMA) (BOX et al., 2015). Embora eficazes em cenários com padrões de comportamento relativamente estáveis, esses modelos clássicos frequentemente pressupõem

linearidade e estacionariedade nos dados, premissas que são cada vez mais violadas pela complexidade dos mercados atuais. A dificuldade em modelar múltiplos padrões sazonais, tendências não lineares e o impacto de variáveis exógenas de forma integrada representa uma barreira significativa para a obtenção da acurácia necessária para uma gestão proativa e eficiente. (UZSOY, 2018)

É neste ponto que a ascensão da Inteligência Artificial (IA) e, mais especificamente, das áreas de Aprendizado de Máquina (*Machine Learning* - ML) e Aprendizado Profundo (*Deep Learning*), surge como uma mudança de paradigma para a previsão de demanda. Impulsionada pela Quarta Revolução Industrial (Indústria 4.0), a IA oferece um novo arsenal de ferramentas capazes de extrair padrões e insights de grandes volumes de dados (*Big Data*) com uma profundidade e precisão sem precedentes. (SHARMA et al., 2022). A Inteligência Artificial busca projetar sistemas computacionais capazes de imitar padrões de comportamento humano e gerar conhecimento relevante para a resolução de problemas. (GARTNER) Dentro da IA, o Aprendizado de Máquina permite que os algoritmos aprendam a partir de dados históricos sem serem explicitamente programados, identificando relações complexas e não lineares que passariam despercebidas pelos métodos tradicionais. Modelos como Prophet, XGBoost e Random Forest exemplificam o poder dessas técnicas, oferecendo flexibilidade e alta performance preditiva em diversos cenários.

A necessidade de adotar ferramentas mais sofisticadas é bem explicada pela Teoria do Processamento de Informações Organizacionais (Organizational Information Processing Theory - OIPT) (GALBRAITH, 1974). A teoria defende que a excelência organizacional é alcançada quando a capacidade de processamento de informações de uma empresa está alinhada com a incerteza e a complexidade do seu ambiente (GALBRAITH, 1974, 1977). Em outras palavras, quanto maior a incerteza, gerada pela volatilidade do mercado, complexidade de produtos e interdependência na cadeia de suprimentos, maior a quantidade de informação que a empresa precisa processar para tomar decisões eficazes (SRINIVASAN; SWINK; KIM, 2015). As ferramentas de previsão tradicionais representam uma capacidade de processamento limitada, adequada para ambientes mais estáveis, o que tem sido destacado em estudos recentes que aplicam a OIPT para explicar limitações de sistemas analíticos em cadeias de suprimentos complexas (ZHU et al., 2018). No cenário atual, com volume massivo de dados e dinâmica de mercado imprevisível, essa capacidade é excedida, gerando uma lacuna de desempenho que exige maior capacidade de processamento de informação (YANG et al., 2025). Modelos de IA e ML, com sua habilidade de processar grandes conjuntos de dados e modelar relações não

lineares complexas, surgem como uma evolução da capacidade de processamento de informações, permitindo que as empresas fechem essa lacuna e respondam de forma mais ágil e precisa aos desafios do mercado, em linha com estudos recentes que combinam OIPT com big data analytics e analytics em cadeias de suprimentos (YU et al., 2021).

A integração de modelos de previsão baseados em Inteligência Artificial tem o potencial de transformar radicalmente os processos de Planejamento de Vendas e Operações (S&OP) e de Planejamento e Controle da Produção (PCP). O S&OP, que funciona como um processo tático para equilibrar demanda e oferta, depende fundamentalmente de um plano de demanda confiável. Previsões mais acuradas e com menor viés, geradas por algoritmos de ML, resultam em um planejamento agregado da produção mais alinhado à realidade do mercado, otimizando os níveis de estoque, melhorando a alocação de recursos e aumentando a previsibilidade dos resultados do negócio. (BRAU, 2023) No nível operacional, o impacto no PCP é igualmente profundo. Previsões granulares e precisas no nível de SKU (*Stock Keeping Unit*) permitem um planejamento mais eficiente, reduzindo tanto o risco de falta de componentes (*stockout*) quanto o excesso de estoque de matéria-prima. A capacidade dos modelos de ML de prever picos e vales de demanda com maior antecedência possibilita um sequenciamento e programação da produção mais inteligentes, otimizando o uso da capacidade instalada e minimizando tempos ociosos. (SHARMA, 2022)

Essa evolução tecnológica, catalisada pela Indústria 4.0, está redefinindo o Planejamento e Controle da Produção (PCP), consolidando o que se entende por PCP 4.0. Diferentemente de abordagens anteriores, o PCP 4.0 foca na digitalização, automação e conectividade de ponta a ponta, por meio da integração de sistemas ciberfísicos (CPS) e da Internet das Coisas (IoT). O objetivo é transformar as funções gerenciais, tornando-as mais integradas e automatizadas para operar em redes de valor otimizadas em tempo real. (BUENO, 2020). Neste paradigma, a digitalização viabiliza a coleta massiva de dados em tempo real de toda a cadeia de suprimentos, promovendo "visibilidade e rastreabilidade. A integração vertical de ambientes de produção físicos e digitais, conectando sistemas de gestão (ERP) e de execução da manufatura (MES) com a inteligência máquina a máquina, permite uma visão holística e sincronizada das operações. A automação, por sua vez, é potencializada por algoritmos de Inteligência Artificial (IA) e Machine Learning (ML), que atuam como o motor analítico do PCP 4.0. Esses modelos são capazes de processar grandes volumes de dados (Big Data) para realizar previsões de demanda mais acuradas e identificar padrões complexos que seriam imperceptíveis aos métodos estatísticos tradicionais. (SHARMA, 2022)

Portanto, a transição para métodos de previsão baseados em IA não é meramente uma atualização tecnológica, mas uma redefinição dos processos de planejamento. A complexidade do mercado moderno exige uma capacidade de processamento de informações cada vez maior e a IA, com modelos como Prophet, XGBoost, Random Forest e redes neurais, oferece a capacidade computacional e analítica necessária para lidar com esse problema, permitindo que o PCP evolua de uma função puramente reativa para uma com capacidades preditivas e prescritivas, onde os modelos não apenas preveem, mas também apoiam a otimização das decisões.

Apesar de ser um tema com crescente popularidade, a grande maioria das pesquisas foca exclusivamente na acurácia preditiva como único critério para avaliar modelos, ignorando outros fatores críticos para a implementação prática, como custo computacional, tempo de execução e a necessidade de conhecimento especializado do gestor, que são cruciais para a viabilidade em um ambiente de negócios real. (KOLKOVÁ E NAVRÁTIL, 2021).

Diante desse contexto, formula-se a seguinte questão de pesquisa:

**RQ1:** Como alinhar a seleção de modelos de previsão à incerteza da demanda e à capacidade de processamento de informação sob diferentes prioridades organizacionais?

Para responder a essa questão, a proposta deste trabalho é desenvolver e aplicar um artefato (Sistema de Apoio à Decisão) fundamentado na metodologia de Design Science Research. O objetivo central é propor uma metodologia/protótipo que combine modelos de previsão baseados em Machine Learning com um método de análise multicritério, avaliando simultaneamente diferentes algoritmos segundo múltiplas métricas alinhadas à OIPT. A partir disso, o sistema recomenda, para cada item ou grupo de itens, o modelo com melhor ajuste entre desempenho preditivo, eficiência computacional e estabilidade, de modo a maximizar o fit organizacional entre a complexidade da demanda e a capacidade de processamento de informação disponível. Mais especificamente, os objetivos específicos do trabalho podem ser definidos da seguinte forma:

1. Desenvolver um Sistema de Apoio à Decisão que integre um pipeline de previsão de demanda multi–item baseado em Machine Learning e modelos estatísticos com um módulo de avaliação multicritério, operacionalizando os conceitos de capacidade de processamento de informação da OIPT.
2. Aplicar o artefato a dados reais da empresa, conduzindo uma “competição de modelos” em múltiplas séries, comparando desempenho em termos de acurácia, viés, tempo de processamento e medidas de incerteza.

Esses objetivos serão cumpridos através de uma abordagem multimétodos: (i) qualitativa, com entrevistas exploratórias que permitem enquadrar o problema sob a OIPT e definir requisitos da solução; e (ii) quantitativa, com desenvolvimento e teste do artefato computacional, que combina modelos de previsão e métodos multicritério em um ambiente experimental controlado com dados reais da empresa.

Este trabalho busca oferecer contribuições teóricas e práticas relevantes. Do ponto de vista teórico, a pesquisa contribui ao operacionalizar a OIPT como uma ferramenta analítica para a seleção de modelos de previsão, conectando um conceito clássico da administração à prática da engenharia de produção e da ciência de dados. Além disso, oferece um estudo empírico comparativo de modelos de IA de ponta em um contexto de demanda complexa e com dados reais, complementado por uma estrutura multicritério de decisão. No âmbito prático, a principal contribuição é o desenvolvimento de um artefato que pode ser diretamente utilizado pela empresa para aprimorar seu processo decisório. Por fim, este estudo serve como um guia metodológico para outras organizações que buscam avaliar e implementar soluções de IA em seus processos de planejamento de demanda, sob a perspectiva de *fit* organizacional proposta pela OIPT.

Este trabalho está estruturado da seguinte forma: na Seção 2, apresenta-se o referencial teórico, abordando Indústria 4.0, PCP 4.0, os principais modelos de previsão de demanda (estatísticos, de Machine Learning e Deep Learning) e a Teoria do Processamento de Informação Organizacional (OIPT). A Seção 3 descreve a metodologia adotada, baseada em Design Science Research, assim como os métodos de coleta e análise de dados. A Seção 4 apresenta o desenvolvimento do artefato proposto e os procedimentos da competição de modelos. Na Seção 5, são discutidos os resultados empíricos e as análises multicritério. Por fim, a Seção 6 traz as conclusões, contribuições, limitações e sugestões para pesquisas futuras.

## 2. Referencial Teórico

### 2.1 Indústria 4.0

A Indústria 4.0 (I4.0) representa o início da Quarta Revolução Industrial. Este conceito emergiu pela primeira vez em 2011 na Feira Industrial de Hannover, na Alemanha, sendo impulsionado como uma iniciativa estratégica nacional pelo governo. (ROŽANEC, 2022) A I4.0 é definida como uma tendência de digitalização e automação que transforma as redes de manufatura e a cadeia de suprimentos, evoluindo para fábricas inteligentes que permitem níveis sem precedentes de flexibilidade e transparência operacional. (BUENO, 2020)

Fundamentalmente, a I4.0 é caracterizada por três grandes tendências tecnológicas: conectividade, inteligência e automação flexível. (TORTORELLA, 2022) É um termo guarda-chuva que descreve a automação e a troca de dados em tecnologias e processos de fabricação, buscando a criação de redes de valor inteiramente digitalizadas, conectadas, inteligentes e descentralizadas. (IVANOV, 2021) O objetivo principal é tornar os processos de produção altamente flexíveis e reconfiguráveis para se adaptar às rápidas mudanças do mercado, viabilizando o ideal de customização em massa a um custo comparável ao da produção em massa. A I4.0, portanto, não se limita apenas à manufatura direta dentro da empresa, mas abrange toda a cadeia de valor. (CIANO, 2021)

O sucesso dessa transformação depende da sinergia de diversas tecnologias disruptivas. Entre os pilares tecnológicos habilitadores cruciais para a I4.0 estão os Sistemas Ciberfísicos (CPS), a Computação em Nuvem (Cloud Computing), a Realidade Aumentada, a Robótica Avançada e, centralmente, a Internet das Coisas (IoT), o Big Data Analytics (BDA), o Machine Learning (ML) e a Inteligência Artificial (IA). (ALEXOPOULOS, 2020)

A Internet das Coisas (IoT), especialmente a Industrial Internet of Things (IIoT), é a espinha dorsal da conectividade na I4.0. Ela é composta por uma rede de objetos físicos, como sensores embarcados, etiquetas RFID, eletrônicos e atuadores, que são interconectados e capazes de trocar informações através de infraestruturas de comunicação. A IoT/IIoT interconecta máquinas e ativos físicos para coletar dados em tempo real do chão de fábrica. Essa capacidade de coleta de dados em tempo real é o que permite a integração dos CPS, onde os ativos de fabricação se ligam ao ciberspaço. A conectividade fornecida pela IoT é fundamental para aumentar a transparência e a rastreabilidade ao longo da cadeia de valor.

A proliferação de entidades altamente conectadas e a onipresença de sensores levam à geração de uma quantidade massiva, heterogênea e dinâmica de dados. Esse volume crescente

de informações digitais, que pode chegar a vários gigabytes por dia do chão de fábrica, é o que constitui o Big Data (BD) e exige tecnologias eficientes para coleta (via IoT), armazenamento (via Cloud Computing) e processamento (usando técnicas de IA). O Big Data Analytics (BDA) é a ferramenta essencial que permite às organizações extraírem valor e percepções úteis dessa vasta gama de dados para a tomada de decisões. O BDA pode ser classificado em diferentes níveis de complexidade, como: Análise Descritiva (descreve a situação atual, auxiliando no diagnóstico de problemas), Análise Preditiva (prevê eventos e tendências futuras) e Análise Prescritiva (utiliza técnicas avançadas para prescrever e adaptar ações, oferecendo mecanismos de tomada de decisão). O BD, ao ser analisado, serve como um recurso crítico em estágios de processamento para obter insights funcionais.

O processamento e a análise desse Big Data são amplamente realizados pela Inteligência Artificial (IA) e seu subcampo, o Machine Learning (ML). A IA é um componente central para a I4.0 e para os sistemas baseados em IoT, sendo ativamente utilizada para processamento de dados, análise e tomada de decisões em ambientes industriais. O ML, em particular, foca em extrair conhecimento útil e permitir que o componente de ML tome decisões, aumentando a eficiência da fabricação. As técnicas de ML, incluindo as redes neurais profundas (*Deep Learning*), são ideais para desvendar padrões de produção complexos e fornecer suporte à decisão oportuno em diversas aplicações. A aprendizagem de máquina é categorizada, conforme o feedback disponível, em Aprendizagem Supervisionada (onde a resposta correta é fornecida), Aprendizagem por Reforço (onde o agente aprende a política ideal por interação com o ambiente) e Aprendizagem Não Supervisionada (onde se buscam padrões nos dados brutos sem avaliação). A IA e o ML, ao fornecerem inteligência acionável a partir dos dados coletados, transformam os modelos operacionais de reativos para proativos e prescritivos.

Os resultados da implementação da I4.0, habilitada por esta infraestrutura tecnológica integrada (IoT, Big Data, ML e IA), são percebidos de maneira abrangente nas operações e na gestão da cadeia de suprimentos. Em um nível genérico, as empresas que adotam essas estratégias baseadas em dados obtêm benefícios significativos no desempenho operacional e financeiro.

Um dos ganhos mais evidentes é a melhoria da eficiência e produtividade. A automação e a integração vertical dos sistemas (permitidas pela conectividade e pelos CPS) levam a processos mais simplificados e à redução de trabalho manual. Isso se manifesta em um aumento da produtividade, uma melhoria da Eficiência Geral do Equipamento (OEE) e uma redução nos

tempos de ciclo (lead times). A capacidade de monitoramento em tempo real fornecida pela IoT e a análise de BD possibilitam o controle e a regulação autônoma da produção.

Outro benefício crucial é a Manutenção Preditiva (PdM). Ao utilizar IA e ML para analisar grandes volumes de dados de sensores (Big Data), é possível prever falhas e estimar a vida útil remanescente de equipamentos. Essa capacidade preditiva evita paradas não planejadas (downtimes) e a substituição desnecessária de equipamentos, resultando em redução de custos de manutenção e aumento da segurança, disponibilidade e eficiência dos processos.

A qualidade do produto e do processo também é drasticamente melhorada. A I4.0 visa a produção de itens de maior qualidade ao menor custo possível. O uso de ML para inspeção e controle de qualidade permite a detecção precoce e precisa de defeitos, o que contribui para reduzir erros e correções durante o desenvolvimento e fabricação.

Além disso, a I4.0 fortalece a flexibilidade e a resiliência da cadeia de suprimentos. A maior flexibilidade na produção é alcançada através da reconfiguração dinâmica de ativos, e a capacidade de processamento de dados suporta a tomada de decisões em ambientes voláteis, ajudando as empresas a navegarem por riscos e disruptões.

O impacto da I4.0 na excelência operacional pode ser visto como uma transição que reforça a necessidade de uma perspectiva sistêmica e integrativa, onde a fusão eficaz entre o mundo físico e cibernético, possibilitada pela conectividade e inteligência, exige que as empresas integrem sistematicamente as novas tecnologias digitais como suporte às suas iniciativas contínuas de melhoria, em vez de focar apenas na adoção isolada de tecnologias.

## 2.2 PCP 4.0

O Planejamento e Controle da Produção (PCP) é a função central da gestão de operações, responsável por traduzir os planos estratégicos e a demanda de mercado em ações tangíveis no chão de fábrica. O PCP tradicional, estruturado em torno de conceitos como MRP (Material Requirements Planning) e MRP II (Manufacturing Resource Planning), opera de forma hierárquica e muitas vezes rígida. Ele depende de planos mestres de produção (MPS) que são definidos em lotes, com base em previsões de demanda de médio e longo prazo, e que são atualizados em ciclos periódicos (semanais ou mensais). A eficácia desse sistema é totalmente dependente da acurácia da previsão de demanda inicial; um erro na previsão gera o notório "efeito chicote" (bullwhip effect), amplificando distorções de estoque e cronograma ao longo de toda a cadeia de suprimentos (LEE; PADMANABHAN; WHANG, 1997).

O PCP 4.0 é a evolução direta desta função, infundida com os princípios e tecnologias da Indústria 4.0. Ele não é apenas um PCP digitalizado, mas um sistema de planejamento e controle fundamentalmente rearquitetado. Ao invés de operar em lotes rígidos, o PCP 4.0 busca gerenciar a "customização em massa" e o "Lote 1", onde cada produto pode ser único (ZHONG; XU; WANG, 2017). Para alcançar essa flexibilidade, o PCP 4.0 deve ser dinâmico, autônomo e orientado a dados em tempo real.

Neste novo paradigma, a importância da previsão de demanda não diminui; pelo contrário, ela se torna ainda mais crítica, mas sua natureza e seu papel mudam drasticamente. O PCP 4.0 não pode mais depender de uma única previsão estática de longo prazo para alimentar um MPS rígido. A volatilidade da demanda e a necessidade de resposta rápida exigem um modelo de previsão diferente: um que seja granular, de altíssimo prazo (muitas vezes em horizontes de dias ou horas) e contínuo, tornando tecnologias como a IA e ML os motores do PCP 4.0.

### **2.3 O Panorama dos Modelos Preditivos na Previsão de Demanda**

A previsão de demanda é, em sua essência, um problema de análise de séries temporais. Uma série temporal é uma sequência de pontos de dados coletados em intervalos de tempo sucessivos. No contexto da demanda, o objetivo é analisar o comportamento histórico desses dados, identificando padrões como tendência (crescimento ou declínio a longo prazo), sazonalidade (flutuações periódicas e previsíveis) e ciclicidade (padrões de médio prazo), para prever o comportamento futuro (HYNDMAN; ATHANASOPOULOS, 2018). O panorama industrial de metodologias para essa tarefa evoluiu significativamente, transitando de modelos puramente estatísticos para uma adoção generalizada de técnicas de Machine Learning (ML), culminando na atual preferência por abordagens híbridas e de alta performance.

Historicamente, o domínio da previsão de demanda pertencia aos modelos estatísticos clássicos, reverenciados por sua robustez teórica, interpretabilidade e eficácia em dados com padrões claros e relativamente estáveis. A família de métodos de Suavização Exponencial (Exponential Smoothing - ES), incluindo o método de Holt-Winters, tornou-se um padrão da indústria por sua capacidade de decompor a série em nível, tendência e sazonalidade. Paralelamente, a metodologia Box-Jenkins popularizou os modelos ARIMA (Autoregressive Integrated Moving Average), que modelam a demanda futura com base em sua própria inércia, ou seja, na dependência linear de seus valores passados e de erros de previsão passados (Box, Jenkins, Reinsel, & Ljung, 2015). Esses modelos univariados são eficazes, mas compartilham

uma limitação fundamental: eles lutam para incorporar de forma nativa e eficaz fatores externos (variáveis exógenas), como o impacto de promoções, ações de concorrentes, feriados móveis ou mudanças abruptas no sentimento do consumidor.

O panorama industrial contemporâneo, no entanto, é definido pela volatilidade e pela riqueza de dados. A ascensão do Big Data, impulsionada pela Indústria 4.0, disponibilizou um volume sem precedentes de informações contextuais que afetam a demanda. Isso expôs as limitações dos modelos estatísticos tradicionais e abriu caminho para a adoção em larga escala do Machine Learning (ML). Diferente dos modelos estatísticos, que partem de pressuposições sobre a estrutura dos dados, os modelos de ML são projetados para "aprender" padrões complexos, não-lineares e interações sutis diretamente dos dados (JAMES et al., 2013).

No cenário industrial atual, um grupo de modelos de ML se destaca: os algoritmos baseados em árvores de decisão. Modelos como Random Forest e, mais notavelmente, as Gradient Boosting Machines (GBMs), com implementações populares como XGBoost (CHEN; GUESTRIN, 2016) e LightGBM, tornaram-se a ferramenta de escolha para muitas organizações. A razão de seu sucesso é a capacidade de lidar com dados tabulares de forma muito eficaz. Eles podem processar nativamente centenas de variáveis exógenas (preço, investimento em marketing, dados de IoT, clima), lidar com dados ausentes, capturar relações não-lineares (como o efeito decrescente de um desconto) e fornecer métricas de "importância de característica", oferecendo um grau de interpretabilidade sobre *o que* está impulsionando a previsão.

Em paralelo, modelos de Deep Learning (DL), um subcampo do ML, ganharam tração para problemas de séries temporais mais complexos. Redes Neurais Recorrentes (RNNs) e suas variantes mais avançadas, como LSTMs (Long Short-Term Memory) e GRUs (Gated Recurrent Units), foram projetadas especificamente para dados sequenciais, possuindo uma "memória" interna que lhes permite capturar dependências de longo prazo na série temporal. Mais recentemente, arquiteturas baseadas em Transformers, que revolucionaram o processamento de linguagem natural, foram adaptadas para séries temporais (como o *Temporal Fusion Transformer*), demonstrando uma capacidade superior de modelar padrões temporais complexos em múltiplos horizontes de previsão (LIM et al. 2021).

Contudo, o panorama atual da indústria raramente é uma escolha binária entre estatística e ML. As famosas "M-Competitions", uma série de competições de previsão em larga escala, têm moldado as melhores práticas. A M4 Competition (2018) revelou que os métodos mais precisos eram frequentemente modelos híbridos, que combinavam a capacidade de

decomposição de modelos estatísticos (como a Suavização Exponencial) com a capacidade de aprendizado de padrões complexos de redes neurais (MAKRIDAKIS; SPILIOOTIS; ASSIMAKOPOULOS, 2018). A subsequente M5 Competition (2020), focada em dados reais de varejo e incluindo variáveis exógenas, foi um marco: os métodos vencedores foram quase exclusivamente baseados em variações do LightGBM, solidificando o domínio dos modelos de *gradient boosting* para problemas de previsão de demanda em larga escala e ricos em *features*.

### 2.3.1 Modelos Estatísticos Clássicos

Os modelos estatísticos tradicionais são fundamentais para a previsão de séries temporais, servindo como benchmarks robustos. Eles geralmente se baseiam na decomposição de padrões históricos, como tendência e sazonalidade.

#### 2.3.1.1 ETS (Error, Trend, Seasonality)

Os modelos ETS (Erro, Tendência e Sazonalidade) formam uma família de métodos de previsão de séries temporais também conhecida como Suavização Exponencial (Exponential Smoothing). A premissa central desta abordagem é o cálculo de médias ponderadas dos dados históricos, onde os pesos decrescem exponencialmente à medida que as observações se tornam mais antigas. Essencialmente, a técnica separa o padrão sistemático do ruído aleatório presente nos dados, permitindo que o suavizador atue como um filtro para obter uma estimativa mais precisa desse padrão.

O framework ETS decompõe a série temporal em seus três componentes fundamentais: Erro (E), Tendência (T) e Sazonalidade (S). Cada um desses componentes pode ser especificado de forma aditiva (A), multiplicativa (M) ou nula (N). A combinação dessas variações gera até 30 modelos distintos. É importante notar que modelos com erro aditivo e erro multiplicativo podem gerar previsões pontuais idênticas, diferindo, no entanto, nos intervalos de confiança calculados. A configuração mais simples é o ETS(A,N,N), conhecida como Suavização Exponencial Simples (SES), adequada para séries sem tendência ou sazonalidade, onde a previsão se baseia em uma média ponderada do valor observado e da previsão anterior. Outras configurações incluem o método de Holt (ETS(A,A,N)), que incorpora uma tendência linear, e modelos que utilizam uma tendência "amortecida" (Ad), útil para cenários onde se espera que o crescimento ou declínio diminua em horizontes futuros. A seleção do modelo ideal é frequentemente automatizada por funções que buscam minimizar os resíduos ou maximizar a máxima verossimilhança, utilizando critérios de informação como o Akaike Information Criteria (AIC) para selecionar a configuração mais competente.

### 2.3.1.2 Método de Holt-Winters

Embora tecnicamente seja um subconjunto da família ETS, o método de Holt-Winters é frequentemente destacado por sua aplicação histórica e específica no tratamento de dados sazonais. Este modelo é uma extensão direta do método de Holt (que já incorpora a tendência linear). O Holt-Winters adiciona um terceiro componente para modelar explicitamente a sazonalidade. Esta abordagem é ideal para séries temporais que apresentam padrões sazonais bem definidos e repetitivos, como vendas no varejo.

A arquitetura do modelo utiliza três equações de suavização distintas para atualizar, a cada período, o Nível ( $N_t$ ), a Tendência( $T_t$ ) e a Sazonalidade ( $S_t$ ). O comportamento dessas equações é controlado por três parâmetros de suavização:  $\alpha$  (alfa), associado ao nível;  $\beta$  (beta), associado à tendência; e  $\gamma$  (gama), associado à sazonalidade. Na prática, esses parâmetros são ajustados automaticamente por meio da minimização da soma dos erros quadrados entre os valores previstos e os observados.

### 2.3.1.3 ARIMA/SARIMA

Uma abordagem alternativa aos modelos de suavização exponencial é a família de modelos Autorregressivos Integrados de Médias Móveis (ARIMA), também conhecidos como modelos de Box & Jenkins. Enquanto os modelos ETS assumem que não há correlação entre os resíduos, o ARIMA é projetado especificamente para explorar a autocorrelação, ou seja, a conexão entre as observações passadas e os valores futuros. A família ARIMA tem demonstrado desempenho satisfatório na modelagem de séries temporais com diferentes níveis de complexidade.

O modelo ARIMA( $p,d,q$ ) é definido por três componentes:  $p$  (Ordem Autorregressiva - AR), que utiliza uma regressão linear dos valores passados da própria série para prever o futuro;  $q$  (Ordem de Médias Móveis - MA), que utiliza uma regressão linear dos erros de previsão passados; e  $d$  (Ordem de Integração - I), que representa o número de diferenciações necessárias para tornar a série estacionária (isto é, com média e variância constantes ao longo do tempo).

O modelo ARIMA básico não lida nativamente com sazonalidade. Para isso, utiliza-se a extensão SARIMA (Seasonal ARIMA). O SARIMA adiciona componentes sazonais ( $P,D,Q$ )\_m aos parâmetros não sazonais ( $p,d,q$ ), onde m representa o período da sazonalidade (ex: 12 para dados mensais). A extensão mais poderosa desta família é o SARIMAX, que permite a inclusão de variáveis explicativas (exógenas) ao modelo SARIMA. Esta abordagem, também chamada de Regressão Dinâmica, é extremamente relevante para a engenharia de

produção, pois permite que a previsão de demanda seja influenciada não apenas pelo seu comportamento passado, mas também por drivers externos, como promoções, feriados, investimentos em marketing.

## 2..32 Modelos de Machine Learning (ML)

Diferente dos modelos estatísticos que extrapolam padrões temporais, os modelos de ML aprendem uma função de mapeamento  $f(X) \rightarrow y$ , onde  $y$  é a demanda a ser prevista e  $X$  é um conjunto de "features" (características). Na previsão de demanda, essas features são tipicamente criadas por engenharia de features (ex: lags da demanda, médias móveis, dia da semana, mês, etc.).

### 2.3.2.1 Modelo Prophet

O modelo Prophet é uma biblioteca de código aberto desenvolvida pela equipe do Facebook (Meta), projetada especificamente para a previsão de dados de negócios. Ele é fundamentado em um modelo de decomposição aditivo (ou multiplicativo) expresso pela relação  $y_t = g_t + s_t + h_t + \epsilon_t$ , onde  $g_t$  representa a função de tendência,  $s_t$  as mudanças periódicas (sazonalidade),  $h_t$  o efeito de feriados ou eventos,  $\epsilon_t$  o termo de erro. Uma de suas principais vantagens para a prática empresarial é a facilidade em incorporar um calendário de eventos customizados. Por exemplo, o modelo pode ser alimentado com datas de feriados ou eventos específicos do negócio, como promoções ou eventos esportivos, capturando o impacto na demanda antes e depois da ocorrência.

### 2.3.2.2 Random Forest

O Random Forest (Floresta Aleatória) é um método de ensemble (conjunto) baseado em bagging (Bootstrap Aggregating). Esta técnica de Machine Learning opera construindo uma vasta coleção de árvores de decisão independentes durante a fase de treinamento. Para problemas de regressão, como a previsão de demanda, o resultado do modelo é obtido através da média das previsões de todas as árvores individuais que compõem a "floresta". A robustez do modelo advém de duas fontes principais de aleatoriedade: primeiro, cada árvore é treinada sobre uma amostra aleatória dos dados de treinamento (amostragem com reposição); segundo, ao construir cada nó da árvore, apenas um subconjunto aleatório de features (variáveis) é considerado para definir a divisão. Este processo duplo garante que as árvores sejam descorrelacionadas, o que reduz drasticamente a variância do modelo final e mitiga o risco de overfitting.

Para ser aplicado a problemas de séries temporais, que não é sua função nativa, o Random Forest exige que o problema seja reestruturado como um problema de regressão supervisionada. Isso é feito através de um processo de "engenharia de features", onde o valor futuro da demanda (o alvo da previsão,  $y$ ) é previsto com base em um conjunto de características ( $X$ ) construídas a partir de dados históricos. Essas características tipicamente incluem valores defasados (lags) da própria demanda (ex: demanda de 7 dias atrás), estatísticas de janela móvel (ex: média ou desvio padrão dos últimos 30 dias) e variáveis exógenas, como dados de calendário (dia da semana, mês, feriado) ou indicadores de negócios (ex: dias de promoção). A principal vantagem do Random Forest é sua alta performance com pouca necessidade de sintonização de hiperparâmetros e sua robustez a outliers. Sua maior limitação na previsão de demanda é ser um modelo inherentemente não-extrapolativo: ele não consegue prever valores que estejam fora do intervalo de dados (mínimo e máximo) observado durante o treinamento, tornando-o inadequado para séries com forte tendência de crescimento ou queda.

### 2.3.2.3 XGBoost (Extreme Gradient Boosting)

O XGBoost (Extreme Gradient Boosting) é outro modelo de ensemble baseado em árvores de decisão, mas que utiliza a técnica de boosting (reforço) em vez de bagging. O boosting funciona de forma sequencial: o modelo treina uma primeira árvore (geralmente "fraca"); em seguida, uma segunda árvore é treinada especificamente para corrigir os erros (resíduos) cometidos pela primeira. Este processo é repetido centenas ou milhares de vezes, com cada nova árvore focando nos erros residuais do conjunto anterior, permitindo que o modelo aprenda padrões extremamente complexos e não-lineares. O XGBoost é uma implementação específica e altamente otimizada do gradient boosting, que domina competições de ciência de dados (como o Kaggle) para dados tabulares.

Sua superioridade se deve a otimizações de sistema, como processamento paralelo, e avanços algorítmicos, notavelmente a inclusão de regularização (L1 - Lasso e L2 - Ridge) diretamente na função de perda. Essa regularização controla a complexidade das árvores e previne o overfitting, um problema comum em algoritmos de gradient boosting tradicionais. Tal como o Random Forest, sua aplicação na previsão de demanda exige uma robusta engenharia de features, transformando a série temporal em um conjunto de dados supervisionado com lags, médias móveis e variáveis exógenas. O XGBoost frequentemente apresenta acurácia superior ao Random Forest, mas é consideravelmente mais sensível à

sintonização de hiperparâmetros (como taxa de aprendizado, profundidade máxima das árvores e parâmetros de regularização). Similarmente ao RF, o XGBoost também é um modelo não-extrapolativo, limitando sua capacidade de prever valores além do espectro de dados de treinamento.

### 2.3.3 Modelos de Deep Learning (Redes Neurais)

Os modelos de Deep Learning (DL) são uma subárea do ML que utiliza redes neurais artificiais com múltiplas camadas (profundas) para aprender representações complexas dos dados. Para séries temporais, os modelos de DL são capazes de aprender padrões temporais diretamente dos dados, muitas vezes eliminando a necessidade de engenharia de features manual.

#### 2.3.3.1 N-BEATS (Neural Basis Expansion Analysis for Interpretable Time Series)

O N-BEATS é uma arquitetura de Deep Learning pura, que alcançou notoriedade ao demonstrar desempenho estado-da-arte na competição de previsão M4, superando métodos estatísticos e híbridos tradicionais. Diferente das LSTMs, o N-BEATS não utiliza células recorrentes ou camadas convolucionais; sua arquitetura é baseada inteiramente em camadas densas (fully connected layers), organizadas em uma estrutura de "stacks" (pilhas) e "blocos". O princípio fundamental do N-BEATS é a decomposição da série temporal, de forma análoga aos métodos clássicos. A rede processa a série histórica (lookback period) e a decompõe sequencialmente. Um primeiro stack modela um componente da série (como a tendência); o resíduo dessa modelagem (a série original menos a tendência prevista) é então passado para o próximo stack, que modela outro componente (como a sazonalidade).

Cada bloco dentro de um stack utiliza "funções de base" (basis expansion) para modelar esses componentes. Por exemplo, um bloco de tendência aprende coeficientes para um conjunto de polinômios, enquanto um bloco de sazonalidade aprende coeficientes para um conjunto de séries de Fourier. O modelo final agrupa as previsões de todos os blocos. Uma das principais vantagens do N-BEATS sobre outras arquiteturas de Deep Learning é a sua interpretabilidade. Na sua versão "interpretável" (N-BEATS-I), os stacks são pré-configurados para modelar especificamente a tendência e a sazonalidade, permitindo que o analista visualize os componentes que o modelo aprendeu, superando o problema da "caixa-preta" comum em LSTMs. Embora nativamente univariado, seu desempenho é robusto e demonstra a capacidade

das redes neurais profundas de aprenderem e generalizarem padrões clássicos de séries temporais de forma eficaz.

## 2.4 OIPT

A Teoria do Processamento de Informação Organizacional (OIPT), como articulada fundamentalmente por Jay Galbraith (1974, 1977), oferece um referencial teórico robusto para entender como as empresas estruturam suas operações e porque selecionam determinadas tecnologias, incluindo os modelos de previsão de demanda. A premissa central da OIPT é que as organizações são sistemas abertos de processamento de informação que devem lidar com a incerteza para executar suas tarefas. O desempenho organizacional, portanto, é contingente à habilidade da empresa em projetar estruturas e mecanismos capazes de processar a quantidade de informação necessária para lidar com a incerteza imposta por suas tarefas e seu ambiente (Galbraith, 1977).

O postulado central de Galbraith (1977) é que a incerteza, definida como a diferença (o gap) entre a quantidade de informação que uma organização precisa ter para tomar decisões e a quantidade de informação que ela efetivamente possui, é o principal impulsionador das escolhas de design organizacional. Quanto maior a incerteza da tarefa, maior a quantidade de informação que deve ser processada durante a execução da tarefa para que se atinja um nível de desempenho aceitável. No contexto específico da previsão de demanda, a incerteza não é uma abstração; ela se manifesta de formas concretas: alta volatilidade do mercado, ciclos de vida curtos de produtos, grande variedade de SKUs, forte impacto de variáveis exógenas (como promoções, ações de concorrentes, clima ou eventos macroeconômicos) e a complexa interdependência da cadeia de suprimentos (Tushman & Nadler, 1978). Um ambiente de alta incerteza, como um varejista de fast-fashion ou um e-commerce, gera uma necessidade de processamento de informação exponencialmente maior do que um fabricante de um produto commodity estável.

Diante dessa incerteza, a OIPT propõe que as organizações tenham duas estratégias básicas: (1) reduzir a necessidade de processamento de informação (criando recursos de folga, como estoques de segurança, ou criando tarefas autocontidas) ou (2) aumentar sua capacidade de processamento de informação (investindo em sistemas de informação verticais ou criando relações laterais) (Galbraith, 1974). A escolha de um modelo de previsão de demanda é, na visão da OIPT, uma decisão estratégica fundamental sobre o nível de capacidade de processamento de informação (IPC) que a empresa deseja instalar.

Modelos estatísticos tradicionais, como médias móveis, suavização exponencial ou mesmo modelos ARIMA univariados, representam um nível de IPC relativamente baixo. Eles são projetados para processar um fluxo de informação limitado e estruturado: a própria série temporal histórica da demanda. Eles são eficazes precisamente em ambientes de baixa incerteza, onde o passado é um bom preditor do futuro e a informação necessária para a decisão é contida na própria série de dados. Tentar usar um modelo tão simples em um ambiente de alta incerteza falha porque o modelo não tem capacidade para processar a vasta quantidade de informação adicional (promoções, dados de mídias sociais, etc.) que a organização precisa para reduzir essa incerteza.

Por outro lado, o advento de modelos de Machine Learning (ML), como XGBoost e LightGBM, e a infraestrutura de Big Data que os suporta, representam um investimento direto e maciço no aumento da capacidade de processamento de informação. Esses modelos são, por definição, mecanismos de alta IPC. Eles são projetados especificamente para processar simultaneamente centenas ou milhares de fluxos de informação (variáveis exógenas), identificar padrões não-lineares complexos e extrair significado de dados não estruturados (como sentimentos em redes sociais), informação que os modelos tradicionais são incapazes de processar. Na linguagem de Galbraith, a implementação de um sistema de previsão baseado em ML é um investimento em um "sistema de informação vertical" sofisticado, projetado para canalizar um grande volume de dados do ambiente para os tomadores de decisão de forma utilizável.

O sucesso, no entanto, não reside simplesmente em maximizar a capacidade de processamento. A OIPT é uma teoria de contingência, e seu conceito mais crucial é o *fit* (ajuste). O desempenho organizacional é alcançado quando há um alinhamento entre o nível de incerteza da tarefa e a capacidade de processamento de informação da organização (TUSHMAN; NADLER, 1978). O desalinhamento (*misfit*) é prejudicial. Se a incerteza da demanda é alta (mercado volátil, muitas promoções), mas a empresa utiliza um modelo de baixa capacidade de processamento de informação (como uma média móvel), existirá um *gap* de informação. A capacidade da empresa é menor que sua necessidade, resultando em previsões imprecisas, excesso de estoque e rupturas. Inversamente, se a incerteza é baixa (um produto estável, maduro), mas a empresa investe recursos excessivos em um complexo modelo de *Deep Learning* (alta capacidade de processamento de informação) que requer manutenção constante e vastos recursos computacionais, ela está sendo ineficiente. A capacidade excede a

necessidade, gerando custos desnecessários sem um ganho de performance correspondente (DAFT; LENGEL, 1986).

Portanto, a OIPT oferece um framework teórico poderoso para guiar a seleção de modelos de previsão. A escolha não deve ser puramente técnica, mas sim uma decisão de design organizacional. Uma empresa que busca o *fit* deve primeiro diagnosticar a incerteza de seu ambiente de demanda. Para SKUs de baixa incerteza, modelos estatísticos de baixa capacidade de processamento de informação são adequados e eficientes. Para SKUs de alta incerteza, a organização deve investir em modelos de alta capacidade de processamento de informação (como ML) para processar a informação necessária para reduzir essa incerteza e permitir um planejamento eficaz. A OIPT, em suma, justifica porque não existe um "melhor modelo" universal, mas sim um "modelo adequado" que equilibra a complexidade do ambiente com a capacidade informacional da ferramenta.

### 3. Metodologia

#### 3.1 Design da pesquisa

O trabalho adota como base metodológica o Design Science Research (DSR), ou Pesquisa em Ciência do Design, como mostrado na figura 1. Esta abordagem é fundamentalmente um paradigma de resolução de problemas, distinto das abordagens explicativas ou descritivas tradicionais. O objetivo central da DSR não é apenas descrever, explicar ou prever a realidade, mas sim alterá-la através da criação de "artefatos" novos e intencionais que endereçam problemas organizacionais práticos e relevantes (HEVNER, MARCH, PARK; RAM, 2004). Esses artefatos podem ser construtos, modelos, métodos ou instanciações de sistemas (MARCH; SMITH, 1995). A DSR busca, portanto, criar conhecimento útil, que é simultaneamente rigoroso em sua concepção e relevante em sua aplicação prática.

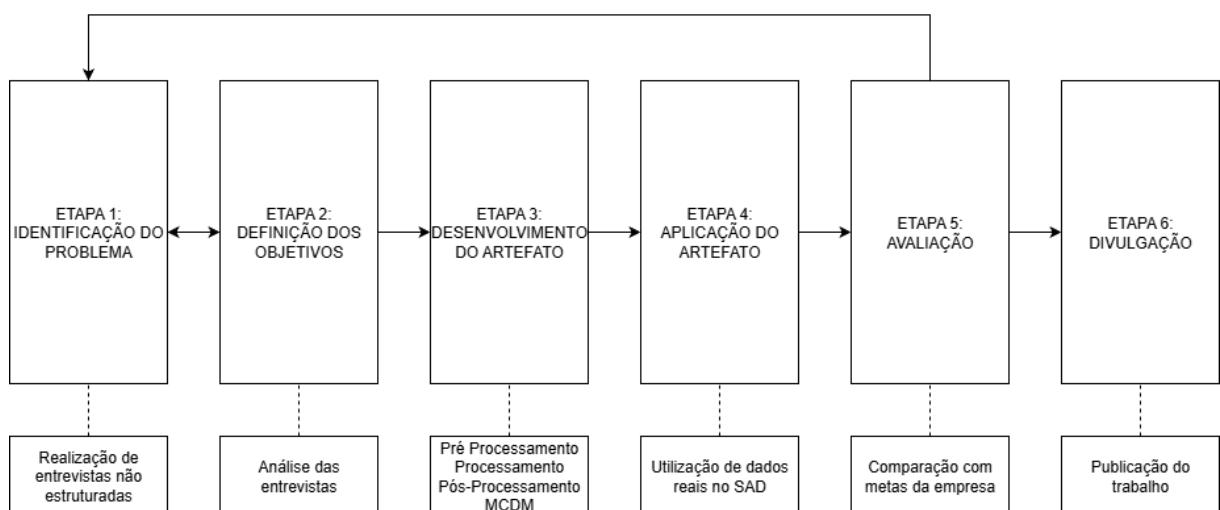


Figura 1 – Pipeline metodológico. Fonte: Autoria própria, 2025.

O processo da DSR é inherentemente iterativo, movendo-se entre a construção do artefato, sua avaliação e o refinamento do problema. Para estruturar esta pesquisa, foi adotado o modelo de processo de DSR proposto por Peffers, Tuunanen, Rothenberger e Chatterjee (2007), que consolida diversas abordagens em um processo nominal de seis etapas: Identificação do Problema e Motivação, Definição do objetivo, Desenvolvimento, Aplicação, Avaliação e Divulgação. Ao longo dessas seis etapas, foi utilizada uma abordagem multi-métodos, com etapas qualitativas e quantitativas. Essa abordagem multi-métodos advém da natureza do problema, onde o desafio não é apenas matemático. Enquanto a parte quantitativa é responsável por processar dados, rodar modelos e aplicar métodos de decisão multicritério

para comparar os resultados, a parte qualitativa é fundamental para o entendimento do contexto do problema e definir objetivos. Sob a ótica da OIPT, o desempenho organizacional depende do ajuste (*fit*) entre a incerteza da tarefa e a capacidade de processamento de informação. A fase qualitativa da pesquisa foca no lado da demanda da informação (diagnosticar a incerteza e os requisitos), enquanto a fase quantitativa foca no lado da oferta (desenvolver e testar a capacidade de processamento dos algoritmos).

### **3.1.1 Fase Qualitativa: Identificação do Problema e Motivação e Definição dos Objetivos da Solução**

A etapa inicial do Design Science Research (DSR) consiste na definição do problema de pesquisa específico e na justificativa da solução proposta, garantindo que o artefato a ser desenvolvido gere uma oportunidade de melhoria genuína e traga benefícios claros à organização (PEFFERS et al., 2007). Para evitar uma definição puramente teórica e basear o problema no contexto organizacional real, esta etapa foi conduzida através de uma abordagem qualitativa exploratória, utilizando entrevistas não estruturadas com especialistas em planejamento de demanda (*demand planners*) da empresa em estudo. A escolha por entrevistas não estruturadas permitiu uma exploração aberta e profunda das percepções dos especialistas sobre os desafios, limitações e pontos de dor do processo de previsão de demanda vigente (YIN, 2014).

A análise dessas entrevistas revelou que a organização opera em um mercado caracterizado por alta volatilidade, com um portfólio de produtos diversificado (SKUs) com diferentes ciclos de vida e os métodos de previsão atuais demonstram limitações significativas em lidar com essa complexidade, evidenciando um *gap* entre a necessidade de informação para um planejamento acurado e a capacidade dos sistemas atuais de processar e modelar essa informação, um desalinhamento sob a ótica da Teoria do Processamento de Informação Organizacional (GALBRAITH, 1977). O problema central não reside na ausência de uma previsão, mas na sub-otimização da sua qualidade, o que gera ineficiências operacionais como estoques de segurança excessivos e potenciais rupturas.

Uma vez identificado o problema, a metodologia DSR exige a definição de objetivos tangíveis e mensuráveis para a solução, permitindo que o artefato seja rigorosamente avaliado nas etapas posteriores (PEFFERS et al., 2007). Com base nas entrevistas, foram identificados os dois indicadores-chave de desempenho (KPIs) utilizados internamente para julgar a eficácia do planejamento: *Forecast Accuracy* (Acurácia da Previsão) e *Forecast Bias* (Viés da Previsão). O primeiro objetivo, aumentar a acurácia, é central para a eficiência operacional,

pois mede o quanto próximas as previsões estão dos valores de demanda reais (HYNDMAN; ATHANASOPOULOS, 2018). A baixa acurácia introduz ruído no planejamento, forçando a organização a proteger-se com estoques inflados ou resultando em perda de vendas e deterioração do nível de serviço (CHOPRA; MEINDL, 2016).

O segundo objetivo, diminuir o viés da previsão, é igualmente importante, pois a acurácia isolada é insuficiente se o erro não for aleatório. O viés mede a tendência sistemática de um modelo em superestimar ou subestimar consistentemente a demanda real (HYNDMAN; ATHANASOPOULOS, 2018). Um viés consistente de sub-previsão leva a rupturas de estoque e perda de receita, enquanto o viés de sobre-previsão resulta em acúmulo de estoque, custos de armazenagem e risco de obsolescência (GARDNER, 2006). Portanto, o artefato deve ter como meta gerar previsões com um viés estatisticamente próximo de zero, indicando que os erros são aleatórios e não sistemáticos.

### **3.1.2 Fase Quantitativa: Modelagem Experimental**

Após o diagnóstico qualitativo e a definição dos objetivos, a pesquisa avançou para as etapas de "Design e Desenvolvimento" e "Demonstração", conforme preconizado pelo ciclo metodológico do Design Science Research (PEFFERS et al., 2007). Esta fase adotou uma abordagem quantitativa e experimental, com o objetivo de materializar o artefato computacional proposto: um Sistema de Apoio à Decisão (SAD) multicritério capaz de operacionalizar e comparar diferentes modelos de previsão.

A estratégia central desta etapa consistiu na estruturação de um experimento comparativo robusto entre diferentes modelos. Inspirada nas metodologias aplicadas nas renomadas competições internacionais M4 e M5, esta abordagem rejeita a premissa de que existiria um algoritmo universalmente ótimo para todas as séries temporais (MAKRIDAKIS; SPILIOTIS; ASSIMAKOPOULOS, 2020). Em vez disso, o artefato foi construído para testar empiricamente o desempenho relativo de diferentes famílias de algoritmos, abrangendo desde métodos estatísticos clássicos, como a Suavização Exponencial, até técnicas avançadas de *Machine Learning* e *Deep Learning*, sob as mesmas condições de contorno e utilizando dados reais da organização.

O desenvolvimento do artefato seguiu um fluxo lógico estruturado em três estágios principais: Pré processamento, processamento e pós processamento, contando com uma avaliação multicritério, operacionalizando o conceito teórico de *fit* organizacional proposto pela Teoria do Processamento de Informação Organizacional (O IPT). Diferente de abordagens

tradicionais que observam exclusivamente a minimização do erro, o design desta pesquisa incorporou variáveis de eficiência computacional (tempo de processamento) e estabilidade (incerteza), reconhecendo que a complexidade do modelo deve ser proporcional à complexidade da tarefa (TUSHMAN; NADLER, 1978).

### **3.2 Métodos de Coleta de Dados**

A realização de uma pesquisa fundamentada no *Design Science Research* (DSR) exige uma estratégia de coleta de dados robusta, capaz de sustentar tanto a relevância prática do artefato quanto o rigor científico da sua construção. A validade de um Sistema de Apoio à Decisão (SAD) não reside apenas na sofisticação dos seus algoritmos, mas na qualidade e na representatividade das informações que o alimentam. Para compreender a complexidade do fenômeno estudado, esta pesquisa adotou uma abordagem de triangulação de dados (FLICK, 2009), combinando fontes primárias qualitativas (entrevistas) com dados secundários quantitativos (extração de ERP), além de uma revisão da literatura para compreender o estado da arte. Esta seção detalha os procedimentos, justificativas e protocolos adotados para cada modalidade de coleta.

#### **3.2.1 Entrevistas Não Estruturadas com Especialistas**

Na fase inicial de diagnóstico e identificação do problema (Etapa 1 do DSR), a coleta de dados teve como objetivo capturar o conhecimento tácito, as nuances processuais e as percepções subjetivas sobre os desafios do planejamento de demanda. O método selecionado para esta investigação foi a realização de entrevistas não estruturadas com especialistas da organização.

A escolha por entrevistas não estruturadas, em detrimento de questionários fechados ou entrevistas semiestruturadas rígidas, justifica-se pela natureza exploratória desta fase da pesquisa. Em estudos de caso em que as fronteiras entre o fenômeno e o contexto não são claramente evidentes, a flexibilidade é crucial (YIN, 2014). Questionários estruturados pressupõem que o pesquisador já conhece as variáveis relevantes a serem testadas, no entanto, no contexto complexo de *Supply Chain*, as causas raízes da ineficiência muitas vezes não são quantificáveis em um primeiro momento. A entrevista não estruturada permite que o entrevistado estabeleça a linha de raciocínio, oferecendo uma visão mais holística e menos enviesada pelas premissas do pesquisador (MARCONI; LAKATOS, 2003). Este método é ideal

para a obtenção de dados de natureza profunda, permitindo captar não apenas os fatos, mas o significado que os sujeitos atribuem a eles, essencial para diagnosticar o desalinhamento proposto pela Teoria do Processamento de Informação Organizacional.

Foram selecionados especialistas que atuam na linha de frente do processo: planejadores de demanda (*demand planners*), responsáveis pela execução técnica da previsão e pelas interfaces com as áreas comercial e financeira. O processo de coleta ocorreu em ambiente reservado, garantindo a confidencialidade necessária para que os participantes expusessem críticas ao processo vigente.

### **3.2.2 Dados Secundários: Extração e Tratamento de Dados do Sistema ERP**

Para as fases de "Design e Desenvolvimento" e "Aplicação" (Etapas 3 e 4 do DSR), a pesquisa exigiu dados quantitativos para o treinamento e validação dos modelos preditivos. A fonte utilizada foi a base de dados transacional do Sistema Integrado de Gestão (ERP - *Enterprise Resource Planning*) da organização parceira. Por questões de confidencialidade, as demandas foram multiplicadas por um número inteiro para a descaracterização de informação sensível.

A utilização de dados secundários extraídos diretamente de sistemas corporativos confere à pesquisa o que denominam de validade ecológica (HAIR et al., 2009). Ao contrário de dados simulados ou sintéticos, frequentemente usados em pesquisas teóricas de estatística, os dados de ERP contêm "ruídos" reais, como sazonalidades irregulares, tendências de mercado, efeitos de promoções e o impacto de intervenções humanas, que desafiam a robustez dos algoritmos. No contexto da Indústria 4.0 e do *Big Data Analytics*, a capacidade de extrair valor de grandes volumes de dados históricos é uma competência central (HOFMANN; RUTSCHMANN, 2018). Os dados coletados referem-se ao histórico de vendas *sell-in* (vendas da indústria para o varejista). A escolha pelo *sell-in* justifica-se pela posição da empresa na cadeia de suprimentos: como fabricante, seu planejamento de produção e estoque deve responder primariamente aos pedidos colocados pelos seus clientes diretos (distribuidores e varejistas), embora estes sejam derivados da demanda final (*sell-out*).

O *dataset* abrange um horizonte temporal de mais de 56 meses (dados a partir de 01/2021). Este período foi definido para garantir a captura de, no mínimo, dois a três ciclos sazonais completos, requisito fundamental para o treinamento eficaz de modelos como SARIMA e Holt-Winters, que dependem da identificação de padrões repetitivos (HYNDMAN;

ATHANASOPOULOS, 2018). Em termos de granularidade, os dados foram coletados no nível de SKU (*Stock Keeping Unit*) e mensalmente. A partir dele, foram coletadas as seguintes variáveis:

- Variável Alvo (Target): Volume de vendas unitárias faturadas.
- Variáveis Temporais: Data do faturamento, permitindo a extração de *features* derivadas como mês, trimestre e ano.
- Atributos Categóricos: Hierarquia do produto, incluindo Marca, Família e Subfamília. Metodologicamente, optou-se por realizar a modelagem preditiva no nível de agregação de "Subfamília". Esta decisão fundamenta-se na necessidade de mitigar os ruídos causados por rupturas de estoque pontuais que afetam SKUs individuais para previsões táticas de médio prazo (KOLASSA; SIEMSEN , 2016).

Durante a extração, foram aplicados filtros para excluir itens descontinuados ou lançamentos muito recentes com histórico inferior a 12 meses, visto que a escassez de dados históricos inviabilizaria o treinamento supervisionado de algoritmos de *Machine Learning* (BISHOP, 2006). A base de dados resultante da extração do ERP constituiu a matéria-prima empírica para a "competição de modelos".

### 3.3 Métodos de análise de dados

A análise de dados foi estruturada em três estágios fundamentais: (1) Pré-processamento e Engenharia de Atributos, visando adequar os dados brutos aos requisitos dos algoritmos; (2) Processamento e Modelagem, onde ocorre o treinamento e validação cruzada; e (3) Pós-processamento e Decisão Multicritério, focado na avaliação de desempenho e seleção do modelo ótimo.

#### 3.3.1 Pré-processamento

A etapa de pré-processamento não se limita à limpeza de dados, mas inclui um diagnóstico estatístico profundo para garantir que as séries temporais atendam aos pressupostos teóricos dos modelos a serem treinados.

##### 3.3.1.1 Diagnóstico de Distribuição e Testes de Normalidade

A análise de normalidade é uma etapa diagnóstica fundamental. O design do artefato não assume que o usuário saiba interpretar valores brutos de Assimetria e Curtose; em vez disso, executa um painel de testes estatísticos formais. Muitas técnicas de modelagem estatística e econômétrica, bem como alguns algoritmos de *machine learning*, assumem que os dados ou

seus resíduos seguem uma distribuição Gaussiana. A violação dessa premissa pode invalidar os resultados do modelo ou reduzir sua performance preditiva.

Além disso, a premissa de que os erros de previsão (ou a demanda) seguem uma distribuição normal é crucial para a construção de intervalos de confiança confiáveis. Para uma avaliação rigorosa, utiliza-se um nível de significância  $\alpha = 0,05$ . As hipóteses formais são:

- $H_0$ : A amostra provém de uma distribuição normal.
- $H_1$ : A amostra não provém de uma distribuição normal.

A rejeição de  $H_0$  ocorre se o p-valor for inferior a 0,05. Para mitigar a sensibilidade variada dos testes a desvios específicos (assimetria vs. caudas pesadas), adota-se um conjunto de três testes complementares (THADEWALD; BÜNING, 2007):

- **Teste de Shapiro-Wilk:** É o método primário devido à sua potência estatística superior para amostras de tamanho pequeno a moderado ( $N < 5000$ ), o que é comum em séries temporais de S&OP (SHAPIRO; WILK, 1965). Estudos confirmam sua eficácia sob essas restrições (RAZALI; WAH, 2011)5. A estatística  $W$  avalia a correlação entre dados ordenados e valores esperados:

$$W = \frac{\left( \sum_{i=1}^n a_i x_{(i)} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- **Teste de Jarque-Bera:** Teste assintótico que se baseia nos momentos da distribuição, verificando se a assimetria e a curtose da amostra se alinham aos de uma distribuição normal (JARQUE; BERA, 1987). Este teste é particularmente eficaz na identificação de "caudas pesadas, que frequentemente enviesam estimadores de Mínimos Quadrados Ordinários (BROOKS, 2019).
- **Teste de Kolmogorov-Smirnov (K-S):** Uma prova de aderência não paramétrica que mede a distância máxima ( $D$ ) entre a Função de Distribuição Acumulada (FDA) empírica e a FDA teórica (KOLMOGOROV, 1933; MASSEY, 1951). Sua inclusão justifica-se pela capacidade de avaliar a aderência global, sendo menos sensível a *outliers* extremos do que o Jarque-Bera (CHAKRAVARTI et al., 1967).

O artefato traduz a rejeição de  $H_0$  em ação prática, sugerindo transformações para a estabilização da variância. Esta etapa segue o trabalho de Box e Cox (1964), recomendando transformações como a logarítmica ou a família de transformações de potência de Box-Cox. Complementarmente, a descrição contextualiza o SKU analisado através do Coeficiente de Variação (CV%), métrica crucial para a classificação da previsibilidade de demanda (SILVER et al., 2016).

### 3.3.1.2 Verificação de Estacionariedade (Raiz Unitária)

Após a análise distributiva, prossegue-se para a verificação da estacionariedade. Uma série é classificada como estacionária se sua média, variância e autocovariância são constantes no tempo (HAMILTON, 1994). A modelagem de séries não estacionárias sem o devido tratamento pode resultar em "regressões espúrias", onde altos valores de  $R^2$  mascaram a falta de relação causal real entre as variáveis (GRANGER; NEWBOLD, 1974).

Reconhecendo o baixo poder estatístico dos testes de raiz unitária em amostras finitas, esta metodologia adota a estratégia de triangulação confirmatória. (Enders, 2014). O diagnóstico final emerge do consenso entre três testes com estruturas distintas:

1. **Teste Augmented Dickey-Fuller (ADF):** Testa a presença de uma raiz unitária, onde a hipótese nula ( $H_0$ ) é de que a série é não estacionária (DICKEY; FULLER, 1979). A formulação utiliza termos defasados para corrigir a autocorrelação serial.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t$$

2. **Teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS):** Para aumentar a confiabilidade, emprega-se o teste KPSS, que inverte a lógica das hipóteses, com a hipótese nula sendo que a série é estacionária (KWIATKOWSKI et al., 1992). É essencial para distinguir séries puramente não estacionárias.
3. **Teste Phillips-Perron (PP):** Aplicado como validador robusto (PHILLIPS; PERRON, 1988). Diferente do ADF, o teste PP aplica uma correção não paramétrica baseada no estimador de variância de longo prazo de Newey e West (1987), tornando-o resiliente a formas gerais de heterocedasticidade nos erros.

Se os testes indicarem não estacionariedade, a metodologia prescreve a aplicação de diferenciação sucessiva ( $\Delta^d$ ) até que a série se torne estacionária, um pré-requisito para a calibração de modelos ARIMA (BOX; JENKINS; REINSEL, 2015).

### 3.3.1.3 Decomposição Estrutural e Sazonalidade

A metodologia avança para a decomposição estrutural como ferramenta exploratória essencial (HYNDMAN; ATHANASOPOULOS, 2018). O objetivo é isolar três componentes latentes da série  $Y_t$ : Tendência ( $T_t$ ), Sazonalidade ( $S_t$ ) e Resíduo ( $R_t$ ). A metodologia clássica

define a relação entre os componentes (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998):

- **Modelo Aditivo:** Usado quando a amplitude da sazonalidade e do ruído é constante.

$$Y_t = T_t + S_t + R_t$$

- **Modelo Multiplicativo:** Comum em vendas com heterocedasticidade, onde a magnitude sazonal cresce com a tendência.

$$Y_t = T_t \times S_t \times R_t$$

A estimativa é feita pelo método de médias móveis (MACAULAY, 1930). Para quantificar a presença de padrões cílicos, utiliza-se a métrica "Força da Sazonalidade" (\$F\\_S\$), conforme proposto por Wang, Smith e Hyndman (2006). Ela mede a proporção da variância (após remoção da tendência) explicada pela sazonalidade:

$$F_S = \max \left( 0,1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)} \right)$$

O valor de \$F\_S\$ (entre 0 e 1) serve como um gatilho de decisão. Se \$F\_S\$ exceder um limiar (ex: 0.5), o sistema recomenda a transição para modelos preditivos com componente sazonal explícito, como SARIMA ou Holt-Winters.

### 3.3.1.4 Tratamento de *Outliers* e Estabilização da Série

A etapa de pré-processamento é um imperativo metodológico para assegurar a validade dos modelos preditivos. Dados brutos de demanda frequentemente apresentam anomalias, heterocedasticidade e não-estacionariedade que, se não tratadas, podem enviesar severamente a estimativa de parâmetros (HAIR et al., 2009). Este módulo estrutura-se em três etapas sequenciais: detecção e tratamento de valores atípicos, transformações para estabilização e padronização de escala.

#### 3.3.1.4.1 Detecção e Tratamento de *Outliers*

Outliers são observações que divergem significativamente do padrão global da série (HAWKINS, 1980). Em modelos baseados em Mínimos Quadrados Ordinários, tais pontos podem distorcer a regressão e violar pressupostos de normalidade. A metodologia emprega três abordagens complementares para sua detecção:

1. Z-Score: Método paramétrico que padroniza os dados assumindo uma distribuição Gaussiana. O escore \$z\_i\$ quantifica a distância da observação à média em unidades de desvio padrão:

$$z_i = \frac{y_i - \bar{y}}{s}$$

Adota-se o limiar  $|z_i| > 3$  para classificação de outliers extremos (HAIR et al., 2009). A limitação deste método reside na sua sensibilidade: outliers extremos podem inflar o desvio padrão amostral ( $s$ ), mascarando a própria detecção.

2. Intervalo Interquartil (IQR): Para contornar a falta de robustez do Z-Score, utiliza-se o método não-paramétrico do IQR, fundamentado na análise exploratória de Tukey (1977). Baseado em estatísticas de ordem (Mediana,  $Q_1$ ,  $Q_3$ ), este método é resiliente a valores extremos. Os limites de detecção são definidos como:

$$\text{Limites} = [Q_1 - 3 \cdot \text{IQR}, \quad Q_3 + 3 \cdot \text{IQR}]$$

Onde  $\text{IQR} = Q_3 - Q_1$ . Pontos fora destes limites são considerados anômalos.

3. **Isolation Forest (Floresta de Isolamento):** Como abordagem de *Machine Learning* não supervisionada, aplica-se o *Isolation Forest* (LIU; TING; ZHOU, 2008). O algoritmo isola anomalias baseando-se no princípio de que são "poucas e diferentes". Ao particionar os dados aleatoriamente em árvores de decisão, *outliers* tendem a ficar isolados mais próximos da raiz (menor caminho médio), dispensando pressupostos sobre a distribuição dos dados.

Após a detecção, o sistema prioriza a Winsorização (GHOSH; VOGT, 2012). Ao contrário da remoção (que cria lacunas temporais), esta técnica limita os valores extremos aos quantis de 5% e 95%, preservando a estrutura temporal da série enquanto reduz a influência de eventos atípicos.

### 3.3.1.4.2 Transformações para Estabilização de Variância

Esta etapa visa adequar a estrutura estocástica da série temporal aos pressupostos teóricos dos modelos estatísticos, corrigindo fenômenos de heterocedasticidade onde a dispersão dos dados aumenta proporcionalmente ao nível da demanda média.

- Transformação Logarítmica: É a abordagem mais direta para linearizar tendências exponenciais e converter relações multiplicativas em aditivas. O artefato implementa a variante "Log-plus-one" para lidar com demandas nulas:

$$y'_t = \ln(y_t + 1)$$

- Transformação de Box-Cox: Para casos em que a transformação logarítmica é insuficiente, utiliza-se a transformação paramétrica de Box e Cox (1964). O método

estima, via Máxima Verossimilhança, um parâmetro  $\lambda$  ideal que maximiza a normalidade:

$$y_t^{(\lambda)} = \frac{y_t^\lambda - 1}{\lambda} \quad (\lambda \neq 0), \quad y_t^{(0)} = \ln(y_t)$$

### 3.3.1.5 Indução de Estacionariedade (Diferenciação)

A estacionariedade é um pré-requisito estrito para a família ARIMA. Baseando-se no diagnóstico dos testes de raiz unitária (ADF/KPSS), a metodologia aplica operadores de diferença (Delta) conforme preconizado por Box e Jenkins (2015):

- Diferenciação Simples (d): Remove tendências estocásticas e determinísticas lineares:

$$\Delta y_t = y_t - y_{t-1}$$

- Diferenciação Sazonal (D): Remove a correlação serial oriunda de ciclos repetitivos, subtraindo a observação atual daquela ocorrida no mesmo período do ciclo anterior (ex:  $s=12$ ):

$$\Delta_s y_t = y_t - y_{t-s}$$

### 3.3.1.6 Padronização de Escala para Machine Learning

Diferentemente da etapa anterior, esta subseção foca na eficiência computacional e na convergência numérica dos algoritmos de Aprendizado de Máquina. Modelos baseados em otimização por gradiente (como Redes Neurais) são sensíveis à magnitude das variáveis. O artefato oferece três técnicas:

1. Padronização (StandardScaler): Centraliza os dados na média zero e ajusta a variância para a unidade (BISHOP, 2006):

$$y' = \frac{y - \mu}{\sigma}$$

2. Normalização Min-Max (MinMaxScaler): Comprime os dados para o intervalo  $[0, 1]$ .

É mandatória para o treinamento de Deep Learning (LSTMs, N-BEATS) para evitar a saturação das funções de ativação e o problema do "desaparecimento do gradiente" (GOODFELLOW; BENGIO; COURVILLE, 2016):

$$y' = \frac{y - \min(y)}{\max(y) - \min(y)}$$

3. Escalonamento Robusto (RobustScaler): Utiliza a Mediana e o Intervalo Interquartil (IQR) para centralizar os dados, garantindo que a presença de anomalias não distorça a escala da maioria dos dados em séries ruidosas:

$$y' = \frac{y - \text{Mediana}}{\text{IQR}}$$

### 3.3.2 Processamento

Após o pré-processamento, a etapa de processamento de dados foca na transformação das séries temporais em estruturas adequadas para o treinamento de algoritmos supervisionados. Enquanto os modelos estatísticos clássicos (como SARIMA e ETS) possuem uma estrutura matemática desenhada explicitamente para lidar com a dependência temporal, os algoritmos de *Machine Learning* (ML), como *Random Forest* e *Gradient Boosting*, operam sob o pressuposto de independência entre as observações.

Para aplicar estes algoritmos eficazes à previsão de demanda, é necessário realizar uma transformação metodológica nos dados, convertendo a série temporal  $Y = \{y_1, y_2, \dots, y_T\}$  em uma matriz de regressão supervisionada. Este processo, denominado Engenharia de Atributos (*Feature Engineering*), tem como objetivo "desconstruir" a informação temporal (tendência, sazonalidade e autocorrelação) em um vetor de características explicativas  $X_t$ , permitindo que o modelo aprenda uma função de mapeamento  $f$  tal que  $\hat{y}_t = f(X_t)$ .

A metodologia adotada baseia-se na estratégia de Janela Deslizante (*Sliding Window*), conforme formalizado por Bontempi, Ben Taieb e Le Borgne (2013). O vetor de *features*  $X_t$  construído é composto por três categorias de variáveis: defasagens (*Lags*), estatísticas móveis e componentes calendários.

#### 3.2.2.1 Estrutura de Autocorrelação: Variáveis de Defasagem (*Lags*)

As variáveis mais críticas para a previsão de séries temporais com ML são os *lags*, que representam os valores passados da série alvo. Eles permitem que o modelo capture a estrutura de autocorrelação, ou a "memória" do processo. O método gera um conjunto de defasagens, com cada lag definido como  $x_{t,lag\_k} = y_{t-k}$ :

- **Lags Curtos** ( $t - 1, t - 2, t - 3$ ): Capturam a dependência imediata e a persistência de curto prazo (o nível recente da demanda).
- **Lag Semestral** ( $t - 6$ ): Captura ciclos de médio prazo.
- **Lag Sazonal** ( $t - 12$ ): Fundamental para capturar a sazonalidade anual. O valor de vendas do mesmo mês no ano anterior é frequentemente o preditor isolado mais forte no varejo e indústria.

### 3.2.2.2 Tendência Local: Estatísticas Móveis (*Rolling Window*)

Para mitigar o ruído aleatório de observações individuais e fornecer ao modelo uma visão da tendência local, calcula-se a Média Móvel simples dos últimos 3 períodos ( $ma_3$ ). É crucial notar a decisão metodológica de *design* para evitar o vazamento de dados (*data leakage*): a média móvel é calculada sobre os valores defasados (*shift*), garantindo que a média em  $t$  utilize apenas informações disponíveis até  $t-1$ .

A formulação matemática é dada por:

$$x_{t,ma_3} = \frac{1}{3} \sum_{i=1}^3 y_{t-i}$$

### 3.2.2.3 Sazonalidade Determinística: Componentes Calendários

Enquanto os *lags* capturam a sazonalidade estocástica (baseada nos valores passados), os modelos de ML beneficiam-se explicitamente de "dicas" determinísticas sobre a posição temporal da observação. O método extrai três componentes da data:

- **Mês:** Permite ao modelo aprender padrões sazonais fixos (ex: aumento de vendas em novembro).
- **Trimestre:** Captura padrões sazonais agregados.
- **Ano:** Serve como uma *proxy* para a tendência global de longo prazo (ex: crescimento anual do mercado).

### 3.2.2.4 Prevenção de Vazamento de Dados (*Data Leakage*)

Uma preocupação central na metodologia é garantir que nenhuma informação futura seja utilizada para prever o passado. Para assegurar a integridade do experimento, foram implementadas as seguintes salvaguardas:

1. **Concatenação Temporal:** Os dados de treino e teste são concatenados temporariamente para a geração de *features* (evitando perda de dados na fronteira da divisão), mas a ordem cronológica é estritamente mantida via ordenação por data.
2. **Divisão Determinística:** A separação final entre as matrizes de treino ( $X_{treino}, y_{treino}$ ) e teste ( $X_{teste}, y_{teste}$ ) é feita exclusivamente baseada na data de corte, correspondente à última observação do conjunto de treino original.
3. **Tratamento de NaNs na Origem:** A criação de *lags* gera valores nulos (*Nan*s) nas primeiras observações da série (ex: o  $lag_{12}$  não existe para os primeiros 12 meses).

Estas linhas são removidas antes do treinamento, pois modelos como *Random Forest* não lidam nativamente com valores faltantes na entrada.

Ao final deste processo, o problema de previsão de série temporal é transformado em um problema de aprendizado supervisionado padrão, onde o modelo aprende a função  $f$ :

$$\hat{y}_t = f(y_{t-1}, \dots, y_{t-12}, \overline{y_{t-1 \dots t-3}}, M_t, Q_t, Y_t)$$

Esta engenharia de atributos permite que algoritmos não-lineares e não-paramétricos, como árvores de decisão, detectem interações complexas (por exemplo, a correlação entre um mês específico e uma tendência de média móvel ascendente) que modelos lineares clássicos teriam dificuldade em capturar sem especificações manuais complexas.

### 3.2.2.5 Estratégia de Validação Temporal

A validação de modelos preditivos em séries temporais impõe desafios metodológicos distintos daqueles encontrados na aprendizagem supervisionada padrão. Em dados tabulares (*cross-section*), assume-se que as observações são independentes e identicamente distribuídas, o que permite o uso de embaralhamento aleatório (*shuffling*) para validação cruzada. No entanto, em séries temporais, a dependência temporal e a ordem cronológica são intrínsecas ao problema. O desrespeito a essa ordem resulta em vazamento de dados (*data leakage*), uma falha metodológica grave onde o modelo é treinado com informações futuras para prever o passado, gerando métricas de acurácia artificialmente infladas e inúteis para a prática (HYNDMAN; ATHANASOPOULOS, 2018).

Para garantir a robustez das previsões e a capacidade de generalização dos modelos, o artefato implementa uma Estratégia de Validação em Dois Níveis:

1. Avaliação Final via Hold-Out (Out-of-Sample):

Este nível simula o cenário real de uso do sistema de S&OP. O usuário define, através da interface, um percentual  $p$  (tipicamente 20%) dos dados mais recentes para compor o conjunto de teste. A série histórica  $Y = \{y_1, y_2, \dots, y_T\}$  é dividida deterministicamente em um ponto de corte  $T_{cut} = \lfloor T \times (1 - p) \rfloor$  em:

- Conjunto de Treinamento: utilizado exclusivamente para o ajuste dos parâmetros e treinamento dos modelos.
- Conjunto de Teste: utilizado exclusivamente para a avaliação final das métricas de desempenho (MAPE, RMSE).

Esta abordagem garante que as métricas reportadas no Dashboard reflitam a capacidade real do modelo de prever um futuro desconhecido ("Out-of-Sample"), sem nunca ter acessado esses dados durante a fase de ajuste.

## 2. Otimização via Validação Cruzada Temporal (Walk-Forward):

Para os modelos de Machine Learning (Random Forest, XGBoost, LightGBM) e Deep Learning (N-BEATS), a seleção dos hiperparâmetros (ex: profundidade da árvore, taxa de aprendizado) não pode ser realizada no conjunto de teste final, sob pena de overfitting específico a esse conjunto. O artefato implementa a técnica de Validação Cruzada em Janela Deslizante (Rolling Origin ou Walk-Forward Validation), utilizando a classe TimeSeriesSplit da biblioteca scikit-learn. Ao contrário do K-Fold tradicional, esta técnica cria múltiplas divisões (folds) de treino e validação que respeitam estritamente a ordem temporal:

- **Fold 1:** Treino  $[t_0, t_k]$ , Validação  $[t_{k+1}, t_{k+h}]$
- **Fold 2:** Treino  $[t_0, t_{k+h}]$ , Validação  $[t_{h+1}, t_{2h}]$
- **Fold n:** Treino  $[t_0, t_{T-h}]$ , Validação  $[t_{T-h+1}, t_T]$

Conforme defendido por Bergmeir e Benítez (2012), esta abordagem é estatisticamente superior para séries temporais, pois avalia a estabilidade do modelo em diferentes pontos do tempo e sob diferentes contextos de tendência e sazonalidade. O artefato calcula a métrica de erro média através desses *folds* para guiar a escolha definitiva dos hiperparâmetros

### 3.2.2.6 Otimização de Hiperparâmetros (HPT)

A busca pela combinação ótima de hiperparâmetros ( $\lambda^*$ ) que minimiza a função de perda ( $\hat{L}$ ) é automatizada pelo artefato. A estratégia de otimização varia conforme a natureza do algoritmo, garantindo eficiência computacional e precisão estatística:

#### A. Modelos Estatísticos (SARIMA, ETS): Critério de Informação

Para modelos clássicos, a otimização não utiliza validação cruzada computacionalmente intensiva, mas sim a minimização de Critérios de Informação, especificamente o AIC (Akaike Information Criterion):

$$AIC = 2k - 2 \ln(\hat{L})$$

Onde  $k$  é o número de parâmetros estimados e  $(\hat{L})$  é o valor máximo da função de verossimilhança. O artefato utiliza o algoritmo auto\_arima (HYNDMAN; KHANDAKAR, 2008) para buscar a combinação  $(p,d,q)(P,D,Q)_m$  que resulta no menor AIC, penalizando modelos excessivamente complexos para evitar overfitting.

#### B. Modelos de ML/DL: Busca Estocástica e Bayesiana

Para modelos de aprendizado de máquina, onde o espaço de busca é vasto e a função objetivo é não-convexa, o artefato oferece duas abordagens avançadas, configuráveis pelo usuário:

- **Random Search (RandomizedSearchCV):** Em vez de testar exaustivamente todas as combinações (*Grid Search*), o algoritmo amostra aleatoriamente N configurações do espaço de parâmetros. A busca aleatória é empiricamente mais eficiente que o *Grid Search* para encontrar ótimos globais, pois explora melhor as dimensões mais importantes dos hiperparâmetros (BERGSTRA; BENGIO ,2012).
- **Otimização Bayesiana (BayesSearchCV):** Esta é a abordagem estado-da-arte implementada no sistema. Ela trata a otimização como um problema de "caixa-preta", construindo um modelo probabilístico substituto da função objetivo. A cada iteração, o algoritmo utiliza os resultados passados para decidir "inteligentemente" qual conjunto de parâmetros testar a seguir, equilibrando a exploração de áreas incertas com a exploração de áreas promissoras (SNOEK; LAROCHELLE; ADAMS, 2012).

A função objetivo utilizada na otimização (a métrica a ser minimizada) é selecionada pelo usuário, permitindo alinhar o treinamento com o objetivo de negócio (ex: minimizar MAPE para penalizar erros relativos ou RMSE para penalizar grandes desvios).

### **3.3.1. Pós processamento e Estruturação do Problema de Decisão Multicritério (MCDM)**

Após o processamento de dados, a etapa de pós processamento consiste no cálculo de diferentes métricas relacionadas ao desempenho do modelo, como MAPE, MSE, Tempo computacional gasto e na análise de *trade-off* e seleção do modelo mais adequado para cada SKU. Tradicionalmente, a literatura de forecasting foca na minimização do erro (acurácia) como critério único de seleção. No entanto, sob a ótica da OIPT, a escolha de um sistema de previsão é uma decisão organizacional que deve equilibrar a necessidade de redução de incerteza com a capacidade de processamento de informação disponível.

A utilização de um único critério (como o MAPE) ignora restrições operacionais críticas, como o tempo de execução do algoritmo (*lead time* da informação) e a estabilidade da previsão. Para lidar com essa complexidade multidimensional, esta pesquisa adota uma abordagem de MCDM. A aplicação de métodos MCDM e *Machine Learning* de forma combinada permite integrar métricas de performance técnica com requisitos de negócio, oferecendo uma solução de compromisso mais robusta para o gerenciamento de estoques e planejamento (Vidal et al., 2022).

Neste trabalho, o problema de decisão é estruturado utilizando o método SAW (Simple Additive Weighting). A escolha pelo SAW justifica-se pela sua transparência, simplicidade computacional para implementação em tempo real no dashboard desenvolvido e pela capacidade de modelar preferências compensatórias, onde um desempenho ruim em um critério (ex: alto tempo de processamento) pode ser compensado por um desempenho excelente em outro (ex: acurácia superior).

A estrutura do problema de decisão é definida por uma matriz de decisão \$D\$, onde as linhas representam as **Alternativas** ( $A_i$ ) e as colunas representam os **Critérios** ( $C_j$ )

- **Alternativas (A):** O conjunto de modelos treinados e validados na etapa anterior.  
 $A$   
 $= \{\text{ETS, Holt-Winters, SARIMA, Random Forest, XGBoost, Prophet, N-BEATS, ...}\}$
- **Critérios (C):** Foram selecionadas três dimensões conflitantes que refletem o *trade-off* proposto pela OIPT entre qualidade da informação e custo de processamento:
  1. **Precisão (C<sub>1</sub>):** Mensurada pelo erro de previsão (MAPE ou RMSE). Representa a capacidade do modelo de reduzir a incerteza externa. O objetivo é minimizar.
  2. **Eficiência (C<sub>2</sub>):** Mensurada pelo Tempo de Treinamento e Inferência (em segundos). Representa o consumo de recursos da capacidade de processamento organizacional. O objetivo é minimizar.
  3. **Estabilidade (C<sub>3</sub>):** Mensurada pelo Desvio Padrão dos resíduos ou variância do erro nos *folds* de validação. Representa a confiabilidade do modelo e a consistência da informação gerada. O objetivo é minimizar.

### 3.3.5.2. Normalização e Agregação (Método SAW)

Como os critérios possuem unidades de medida distintas (percentual de erro vs. segundos de processamento), a aplicação direta da soma ponderada é inviável. A metodologia aplica, portanto, uma etapa de normalização linear.

Para critérios do tipo "custo" (onde menor é melhor, como Erro e Tempo), utiliza-se a seguinte função de transformação para normalizar o valor  $x_{ij}$  da alternativa  $i$  no critério  $j$  em um score normalizado  $r_{ij}$  que varia de 0 a 100:

$$r_{ij} = 100 \times \left( 1 - \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j) + \epsilon} \right)$$

Onde  $\epsilon$  é uma constante pequena ( $10^{-6}$ ) para evitar divisão por zero. Esta formulação garante que o modelo com o menor erro (ou menor tempo) receba score 100, enquanto o pior desempenho receba 0.

Após a normalização, o **Índice de Preferência Global ( $S_i$ )** ou "Score Geral" para cada modelo é calculado pela soma ponderada dos valores normalizados:

$$S_i = \frac{\sum_{j=1}^n w_j \cdot r_{ij}}{\sum_{j=1}^n w_j}$$

Onde  $w_j$  representa o peso de importância atribuído ao critério j.

A atribuição dos pesos (w) é o mecanismo pelo qual o artefato operacionaliza o fit organizacional. O dashboard desenvolvido permite que o gestor configure esses pesos dinamicamente para refletir o contexto estratégico, conforme ilustrado na implementação do sistema:

- **Cenário de Alta Criticidade (Curva A):** O gestor pode atribuir peso majoritário à Precisão (ex: $w_{erro} = 80, w_{tempo} = 0, w_{estabilidade} = 0$ ), priorizando a redução de incerteza independente do custo computacional.
- **Cenário de Alta Frequência/Restrição (Curva C):** Para itens de baixo valor ou limitações de *hardware*, o gestor pode aumentar o peso da Eficiência (ex: $w_{tempo} = 50$ ), penalizando modelos complexos (como Random Forest ou Prophet) que consomem recursos excessivos para ganhos marginais.

O modelo que apresentar o maior valor de  $S_i$  é recomendado pelo sistema como a solução ótima para o item analisado, garantindo que a escolha tecnológica esteja alinhada à necessidade de processamento de informação específica daquele SKU.

## 4. Design e Desenvolvimento do Artefato

Um artefato desenvolvido com base em Design Science Research só cumpre seu objetivo de relevância prática se for utilizável e agregar valor ao tomador de decisão (HEVNER et al., 2004). No contexto da previsão de demanda, onde a incerteza é alta, o julgamento humano especializado é insubstituível. Portanto, o artefato não foi concebido como uma "caixa preta" automatizada, mas como um sistema de suporte à decisão human-in-the-loop (centrado no ser humano).

Para materializar essa interface, foi escolhida a biblioteca Streamlit. A escolha desta ferramenta é estratégica: sendo um framework em Python, ela se integra com todo os pacotes de data science (Pandas, Scikit-learn, XGBoost, Prophet, etc.) utilizado no módulo de repositório de modelos. Isso permite a criação de um aplicativo web interativo de forma extremamente rápida, sem a necessidade de desenvolvimento front-end complexo, focando os esforços na funcionalidade analítica.

O dashboard em Streamlit atua como o principal painel de análise para os especialistas (demand planners), traduzindo os resultados complexos da competição de modelos em insights interpretáveis e personalizados de acordo com suas prioridades. Sua estrutura é projetada para facilitar a resposta às necessidades recorrentes do demand planner. O artefato foi concebido como um sistema multicritério de suporte à decisão para previsão de demanda. A implantação do multicritério no artefato é uma resposta direta à OIPT: reconhece-se que diferentes produtos (ou subfamílias de produtos) possuem diferentes níveis de incerteza. Um único modelo (representando uma única capacidade de processamento) seria inadequado. Portanto, o artefato é projetado como um portfólio de modelos, permitindo um fit entre o modelo e a natureza da série temporal analisada, que permite ao tomador de decisão escolher o modelo que melhor atende às suas necessidades.

Esta seção descreve o fluxo de uso completo e ideal do artefato, da perspectiva do usuário, mapeando cada tela e interação, desde o upload inicial dos dados até a geração da previsão final e exportação do plano de experimentos.

### 4.1 Página Inicial e Módulo de Upload de Dados

A jornada do usuário começa na Tela Inicial, que funciona como um "portal" para o artefato, apresentando uma breve descrição de suas funcionalidades e apresentando seus desenvolvedores. No final da tela inicial, há uma indicação levando o usuário para a página de upload de dados, onde o usuário vai começar a interagir de fato com o aplicativo.

O módulo Upload de Dados, mostrado na figura 2, é extremamente importante pois garante a qualidade de todo o pipeline que vem a seguir. O design deste módulo foca em três pilares: flexibilidade, robustez e feedback imediato.

Pensando na flexibilidade, o módulo oferece dois caminhos de entrada: o primeiro é o widget, que permite ao usuário carregar seus próprios dados em formato Excel e o segundo é Usar Dados de Exemplo, habilitando o aplicativo a ser testado com dados genéricos.

O processo da leitura destes dados começa com o Mapeamento Semântico, permitindo que o artefato funcione mesmo com falta de um padrão de nomenclatura nas bases de dados dos usuários de negócio. Nessa etapa, eles designam as colunas de Identificação do Item, Data/Período e Demanda. Isso permite que o artefato utilize qualquer base de dados que contenha essas informações. Além disso, os usuários podem, opcionalmente, selecionar "Variáveis Exógenas" para indicar eventos como promoções e feriados, que serão interpretados por alguns dos modelos de previsão.

Após o mapeamento, o usuário avança para o módulo de "Validação e Limpeza de Dados". Aqui, são oferecidas opções de pré-processamento básicas, exibidas como *checkboxes*, que incluem remover linhas com dados faltantes e substituir valores negativos por 0. Após decidir aplicar ou não essas transformações e clicar em "Aplicar Alterações e Validar", o módulo fornece um feedback visual imediato por meio de estatísticas, um gráfico de série temporal. Ao concluir, os dados processados são salvos no estado da sessão, garantindo sua persistência e disponibilidade para as etapas. O processo se encerra com um guia ativo, indicando a "Próximos passos: Vá para Análise Exploratória", concluindo assim o primeiro ciclo de interação do artefato

Figura 2- Tela de Upload de dados. Fonte: Autoria própria, 2025.

## 4.2 Módulo de Análise Exploratória

Após o upload e validação dos dados, o planejador avança para o Módulo de Análise Exploratória e Testes Estatísticos. Este componente é essencial para os objetivos de transparência do artefato e seu papel como sistema de apoio à decisão, partindo da premissa de que a seleção de modelos de previsão não deve ser um processo de "caixa-preta".

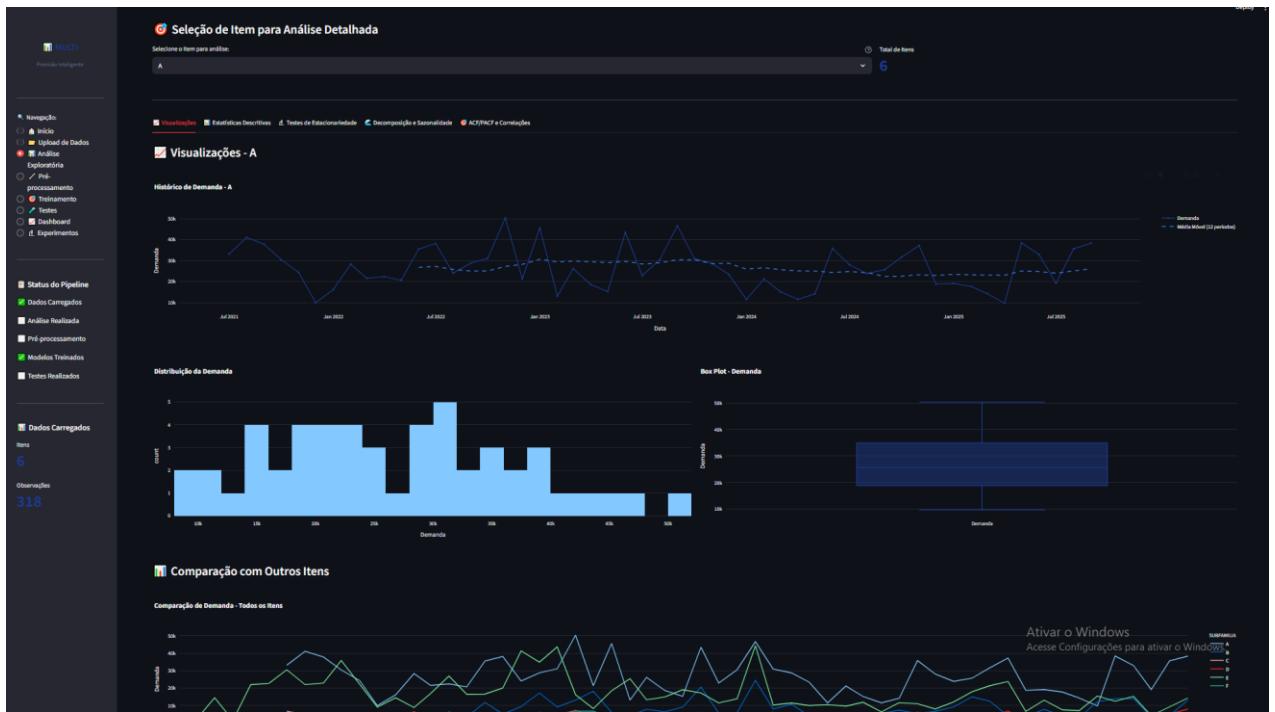


Figura 3- Tela de Visualizações – Análise Exploratória. Fonte: Autoria própria, 2025.

A interface é estruturada em quatro abas complementares. A primeira, "Visualizações", mostrada na Figura 3, foca na inspeção qualitativa, exibindo o gráfico da série temporal sobreposto por uma média móvel para identificação de tendências, além de histogramas e *boxplots* que revelam a distribuição dos dados e a presença visual de *outliers*.

A análise avança para o campo quantitativo na aba de "Estatísticas Descritivas" (Figura 4). O usuário é apresentado a métricas de destaque: Média, Mediana, Desvio Padrão e o Coeficiente de Variação (CV%). O CV% é especialmente crítico no contexto de S&OP, pois quantifica a volatilidade relativa; valores elevados (ex: > 50%) sinalizam uma demanda errática, justificando a necessidade de modelos mais robustos.

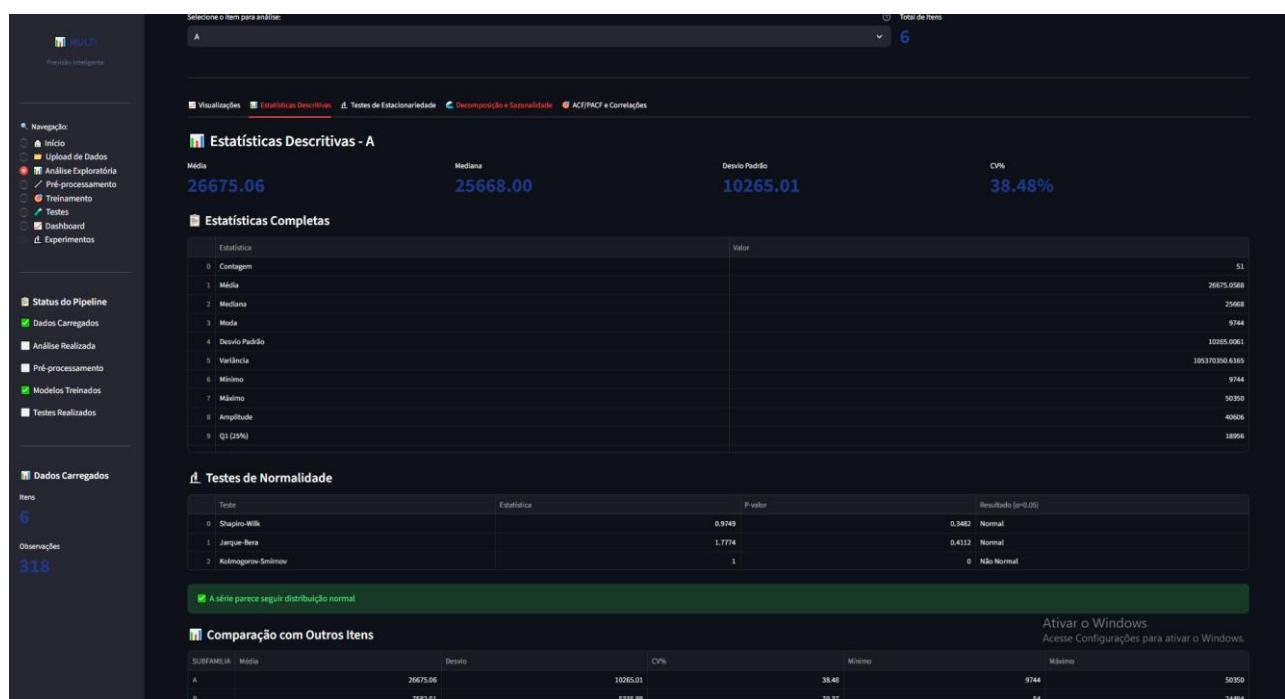


Figura 4- Tela de Estatísticas descritivas – Análise Exploratória. Fonte: Autoria própria, 2025.

Para um detalhamento técnico, o artefato exibe uma tabela de "Estatísticas Completas". Esta visão mais detalhada apresenta medidas como Media, Mediana, Moda, Variância, Amplitude, Quartis e Intervalo Interquartil, Assimetria (*Skewness*) e Curtose (*Kurtosis*).

A seguir, vem a seção " Testes de Normalidade". O design do artefato não assume que o usuário saiba interpretar os valores de Assimetria e Curtose. Em vez disso, ele executa um painel de testes estatísticos formais para verificar se os dados de demanda se assemelham a uma Curva de Gauss (distribuição normal). Nessa etapa, são realizados os testes de Shapiro Wilk, Jarque-Berra e Kolmogorov-Smirnov, e os resultados são exibidos em um quadro comparativo, com a Estatística, P-valor e o resultado.

O artefato desenvolvido traduz a rejeição de  $H_0$  ( $p < 0,05$ ) em ação prática, sugerindo transformações para a estabilização da variância. Esta etapa segue o trabalho seminal de Box e Cox (1964), recomendando transformações como a logarítmica ou a família de transformações de potência de Box-Cox para normalizar a distribuição antes da modelagem preditiva. Complementarmente, a descrição contextualiza o SKU analisado através do Coeficiente de Variação (CV%), métrica crucial para a classificação da previsibilidade de demanda (Silver et al., 2016).

Após a análise distributiva, prossegue-se para a verificação da estacionariedade (Figura 5). Uma série é classificada como estacionária se sua média, variância e autocovariância são constantes no tempo (HAMILTON, 1994). Reconhecendo o baixo poder estatístico dos testes de raiz unitária em amostras finitas, esta metodologia adota a estratégia de triangulação confirmatória (Enders, 2014). O diagnóstico final emerge do consenso entre três testes com estruturas distintas: ADF, KPSS e Phillips-Perron.

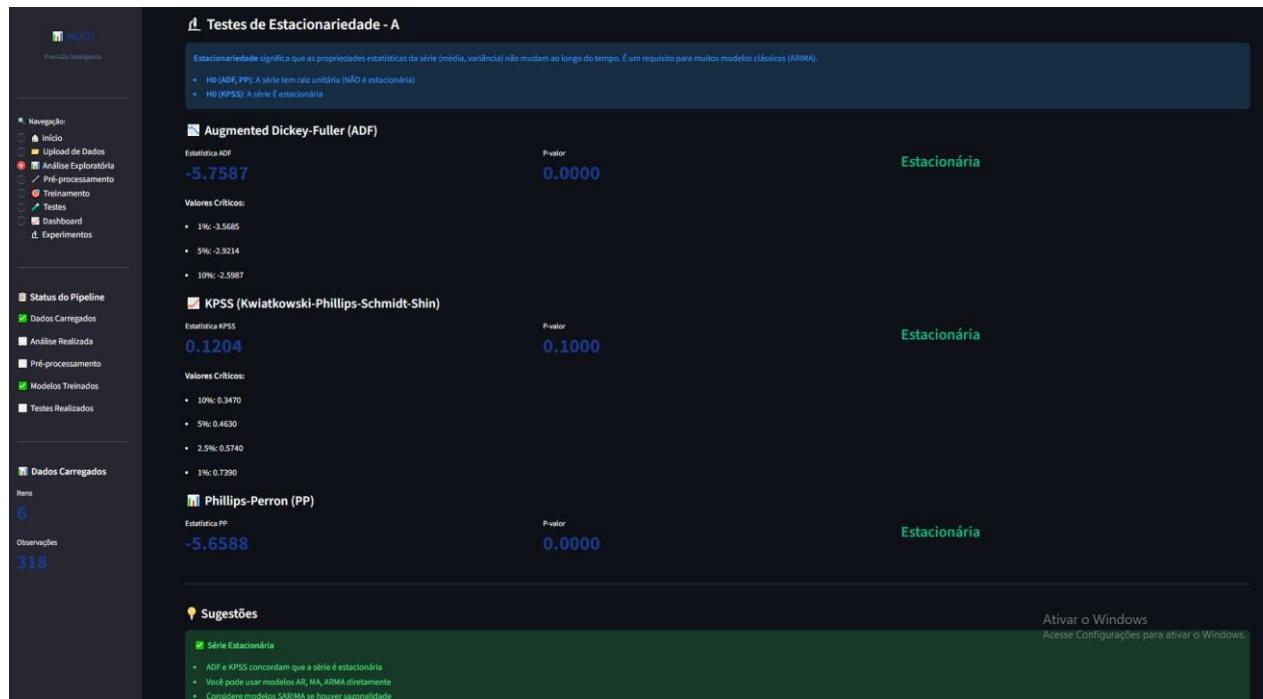


Figura 5- Tela de Testes de Estacionariedade – Análise Exploratória. Fonte: Autoria própria, 2025.

Avançando no fluxo do usuário, a metodologia segue para a decomposição estrutural como ferramenta exploratória essencial (HYNDMAN; ATHANASOPOULOS, 2018), com sua interface mostrada na Figura 6. O objetivo é isolar três componentes latentes da série  $Y_t$ : Tendência ( $T_t$ ), Sazonalidade ( $S_t$ ) e Resíduo ( $R_t$ ). Isso permite ao planejamento de S&OP distinguir se as variações na demanda são de longo prazo, cíclica ou aleatórias. Aqui, o usuário

pode escolher usar a decomposição aditiva ou multiplicativa e definir o período sazonal (valor padrão = 12 meses)

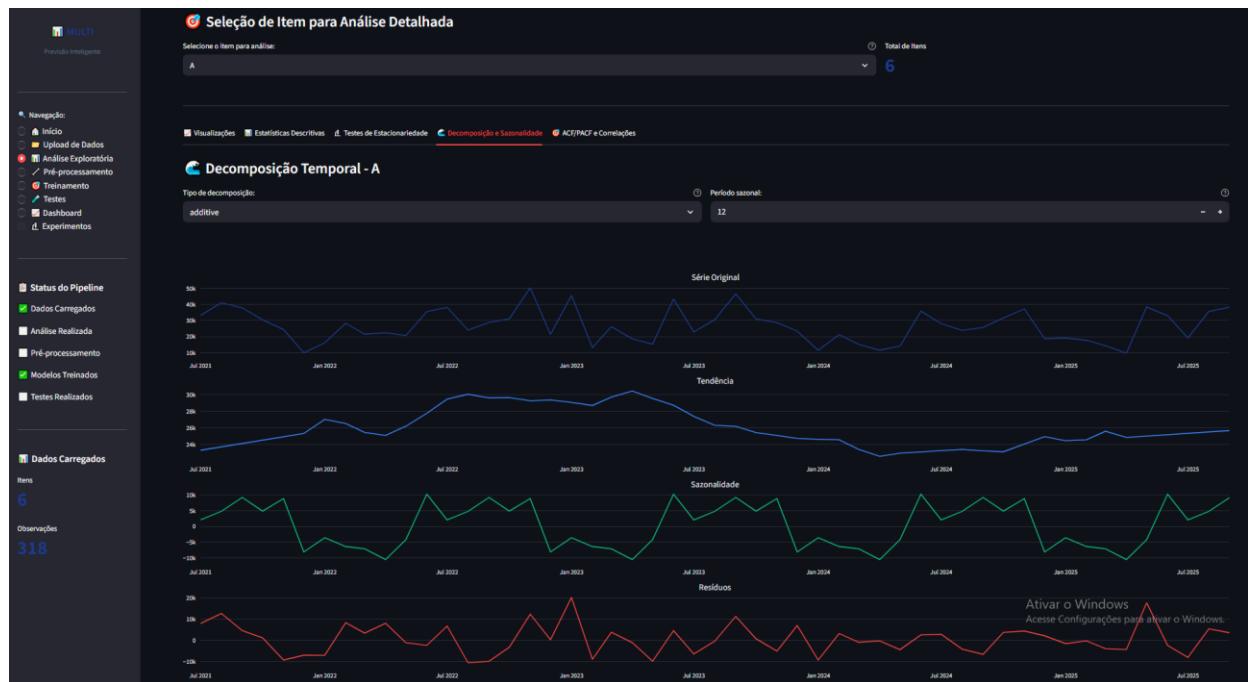


Figura 6- Tela de Decomposição e Sazonalidade – Análise Exploratória. Fonte: Autoria própria, 2025.

#### 4.3. Módulo de Pré-Processamento

A etapa de pré-processamento é um imperativo metodológico para assegurar a validade dos modelos preditivos. Dados brutos de demanda frequentemente apresentam anomalias, heterocedasticidade e não-estacionariedade que, se não tratadas, podem enviesar severamente a estimativa de parâmetros (HAIR et al., 2009).

Este módulo estrutura-se em três etapas sequenciais: (1) detecção e tratamento de *outliers*, (2) transformações para estabilização de variância e estacionariedade, e (3) normalização de escala para algoritmos de ML.



Figura 7- Tela de Detecção de Outliers – Pré-Processamento. Fonte: Autoria própria, 2025.

Na tela de detecção de outliers (Figura 7), o usuário pode escolher entre três métodos para o tratamento dos dados: Z-Score, IQR e Isolation Forest, sendo possível ajustar os thresholds desejados nas duas primeiras opções. Ao ser identificado um outlier na série, ele é indicado com um X vermelho no gráfico para facilitar a identificação.

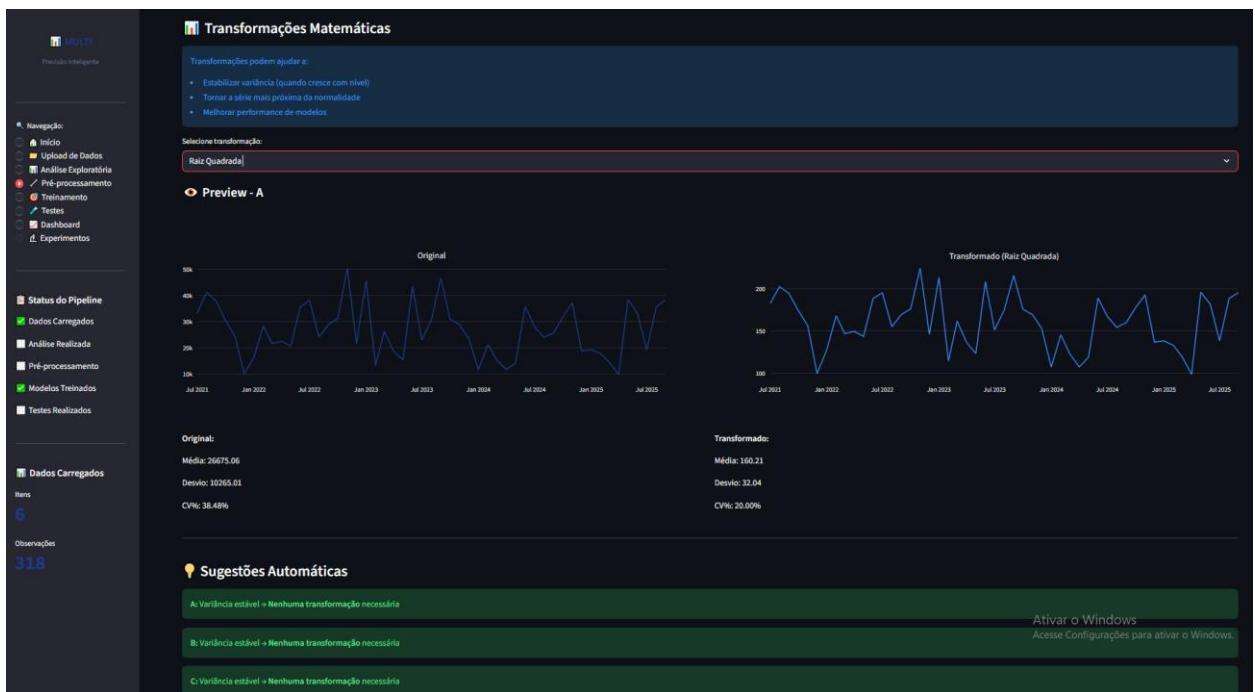


Figura 8- Tela de Transformações matemáticas – Pré-Processamento. Fonte: Autoria própria, 2025.

Na tela de transformações matemáticas (Figura 8), o usuário tem a oportunidade de aplicar diferentes transformações que podem ajudar a estabilizar a variância, tornar a serie mais próxima da normalidade e melhorar a performance dos modelos. Nele, há 5 transformações diferentes que podem ser aplicadas individualmente em cada uma das séries analisadas. São elas: Log, Raiz quadrada, Box-Cox, Diferenciação ordem 1 e Diferenciação Sazonal (lag12).

Ao selecionar uma das transformações, é feita uma comparação da série original com a transformada, exibindo informações de média, desvio e CV% de cada uma, além de mostrar a série de fato. No final da tela, há uma sugestão de transformação a ser aplicada baseada nas características da série.

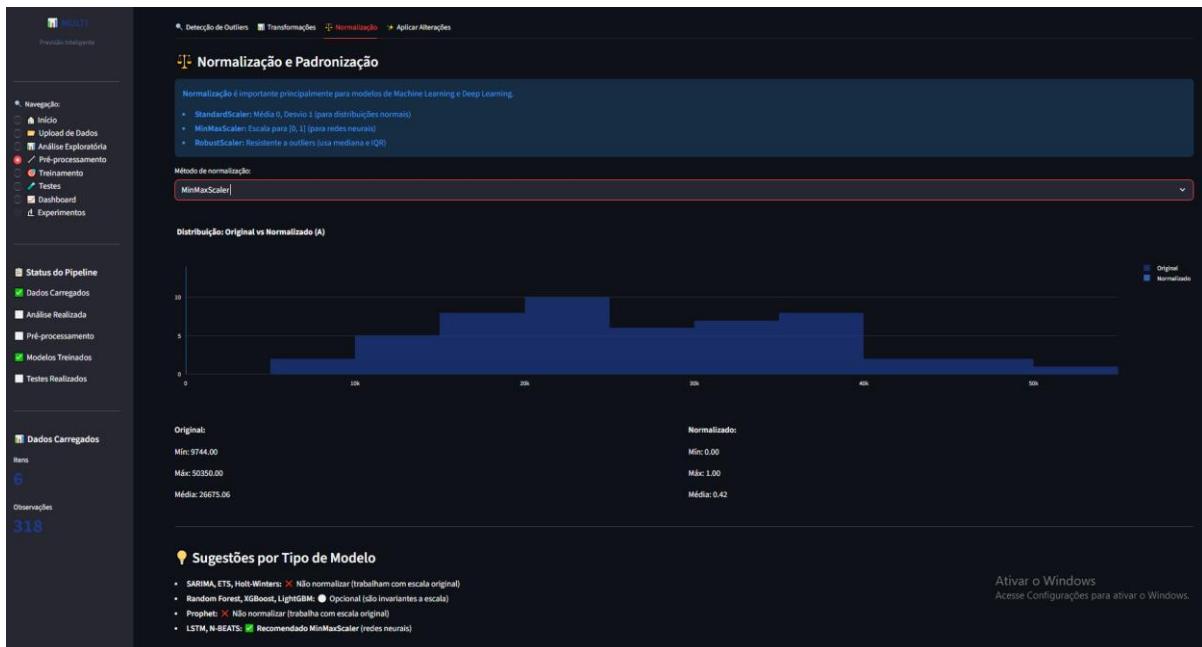


Figura 9- Tela de Normalização – Pré-Processamento. Fonte: Autoria própria, 2025.

Por fim, na tela de normalização (Figura 9), o usuário pode escolher aplicar uma das três transformações disponíveis no artefato de acordo com suas necessidades. No fim da tela, é sinalizado ao usuário qual transformação deve ser aplicada dependendo do modelo que será utilizado.

#### 4.4. Módulo de Treinamento e Otimização de Modelos

O Módulo de Treinamento constitui o motor analítico do artefato proposto. Esta etapa marca a transição do diagnóstico da análise estatística para a fase de experimentação. O objetivo deste módulo não é apenas ajustar um modelo isolado, mas orquestrar uma competição entre diferentes modelos para identificar a abordagem mais eficaz para cada série temporal individual.

A primeira fase do Módulo de Treinamento é a Configuração do Experimento. Mostrada na Figura 10, esta etapa representa a interface de controle entre o julgamento humano (o conhecimento de negócio do planejador) e o poder computacional dos algoritmos. O artefato impõe uma estrutura de decisão que obriga o usuário a definir explicitamente três dimensões fundamentais do problema de previsão: o escopo (quais itens), a metodologia (quais modelos) e a estratégia de validação.

A primeira decisão oferecida ao usuário é a definição da abrangência da análise. O componente de seleção permite dois modos de operação:

- **Execução em Lote (Batch):** "Todos os itens". Simula o processo rotineiro de S&OP, onde centenas de SKUs são previstos simultaneamente.
- **Execução Focada:** "Selecionar específicos". Permite uma análise detalhada.

O artefato apresenta ao usuário o portfólio de modelos disponíveis, segregados em duas categorias distintas: Modelos Clássicos/Estatísticos (SARIMA, ETS, Holt-Winters, TBATS) e Modelos de Machine/Deep Learning (Random Forest, XGBoost, LightGBM, Prophet, N-BEATS).

A decisão de design de permitir a múltipla seleção (checkbox) fundamenta-se no princípio da combinação de previsões e na evidência empírica de que não existe um modelo universalmente superior. Competições de previsão como a M4 (MAKRIDAKIS; SPILIOOTIS; ASSIMAKOPOULOS, 2018) demonstraram que diferentes famílias de algoritmos performam melhor em diferentes horizontes e frequências de dados. Ao permitir que o usuário selecione, por exemplo, "SARIMA" e "XGBoost" simultaneamente, o artefato configura um experimento competitivo para identificar empiricamente qual abordagem modela melhor a série de determinado item.



Figura 10 - Tela de Seleção de modelos – Treinamento. Fonte: Autoria própria, 2025.

Um dos principais diferenciais do artefato é a funcionalidade "Sugestão Automática" (Figura 11). Este componente atua como um sistema de recomendação que traduz as características estatísticas da série em uma seleção ótima de modelos.

Este sistema baseia-se em um conjunto de regras heurísticas derivadas das melhores práticas de *forecasting* (HYNDMAN; ATHANASOPOULOS, 2018):

- **Comprimento da Série (n\_obs):**

- Se  $n < 24$  meses: O sistema recomenda modelos parcimoniosos como ETS e Holt-Winters Aditivo, alertando que modelos de ML/DL (que exigem grandes amostras para treinamento) tendem a sofrer de *overfitting* ou falhar por falta de dados.
- Se  $n > 48$  meses: O sistema habilita e recomenda modelos de alta complexidade como N-BEATS e Random Forest.

- **Volatilidade (cv):**

- Se o Coeficiente de Variação é alto ( $> 0.7$ ), indicando demanda errática, o sistema prioriza modelos não-lineares baseados em árvores (XGBoost, Random Forest), que lidam melhor com variância alta do que os modelos lineares clássicos.

- **Sazonalidade (sazonal):**

- o Se detectada pelos testes de autocorrelação (ACF), o sistema recomenda explicitamente modelos com componentes sazonais nativos (SARIMA, Holt-Winters, Prophet).
- **Presença de Outliers (outliers):**
  - o Se detectados, o sistema sugere modelos robustos a ruído (Random Forest, Prophet) e desencoraja modelos sensíveis a valores extremos (como modelos baseados em médias móveis simples sem tratamento prévio).

Esta lógica de recomendação reduz a barreira cognitiva para o usuário, garantindo que modelos inadequados sejam filtrados antes mesmo do início do processamento computacional.

The screenshot shows the 'Seleção de Modelos' (Model Selection) screen. At the top, a message says 'Nesta etapa, você pode treinar múltiplos modelos para cada item.' (In this step, you can train multiple models for each item.) Below that, a note says 'Itens com poucos dados (< 24 meses) ou gips podem faltar ou ser pulados.' (Items with few data (< 24 months) or gaps may be missing or skipped.)

The main section is titled 'Seleção de Modelos' (Model Selection) and shows the following details:

- Modo de seleção:  Todos os itens  Selecionar específicos
- Itens Selecionados: 6
- Seleção Manual:  Sugestão Automática:  Configurações:
- Sugestões Automáticas por Item**: Esta aba analisa as características estatísticas de cada série (até 5) e sugere os modelos mais adequados da sua lista de seleção.
- Analise: E** (Analysis E):
 

Tamando (meses)	Estacionária?	Sazonal?	Contém Outlier?
57	Sim	Não	Não

 Resumo da Análise:
  - Série longa (>=48 meses). Boa para todos os modelos.]
 Modelos Recomendados (da sua seleção):
  - N-BEATS, Random Forest, SARIMA, XGBoost
- Analise: F**
- Analise: A**
- Analise: C**
- Analise: D**

Figura 11- Tela de Sugestão automática de modelos – Treinamento. Fonte: Autoria própria, 2025.

Na tela de configurações (Figura 12), o usuário consegue ter controle total sobre o plano operacionalização do modelo. O artefato dispõe de diversos dispositivos para que a experiência possa ser ideal para as necessidades do demand planner. A primeira escolha a ser feita é a divisão treino/teste através de um slider, com padrão definido para 20% de teste. A seguir, deve ser decidido se o artefato utilizará ou não a otimização de hiperparâmetros (HPT). Caso opte pela otimização, deve ser definido o método de otimização (RandomSearch ou Bayesian Search), o número de iterações que serão realizadas para achar o modelo ótimo e a métrica de otimização desejada (MSE, RMSE, MAPE ou MAE).

Caso o demand planner não opte pela otimização de hiperparâmetros, os parâmetros dos modelos podem ser definidos manualmente, da maneira que ele achar melhor baseado em sua experiência.

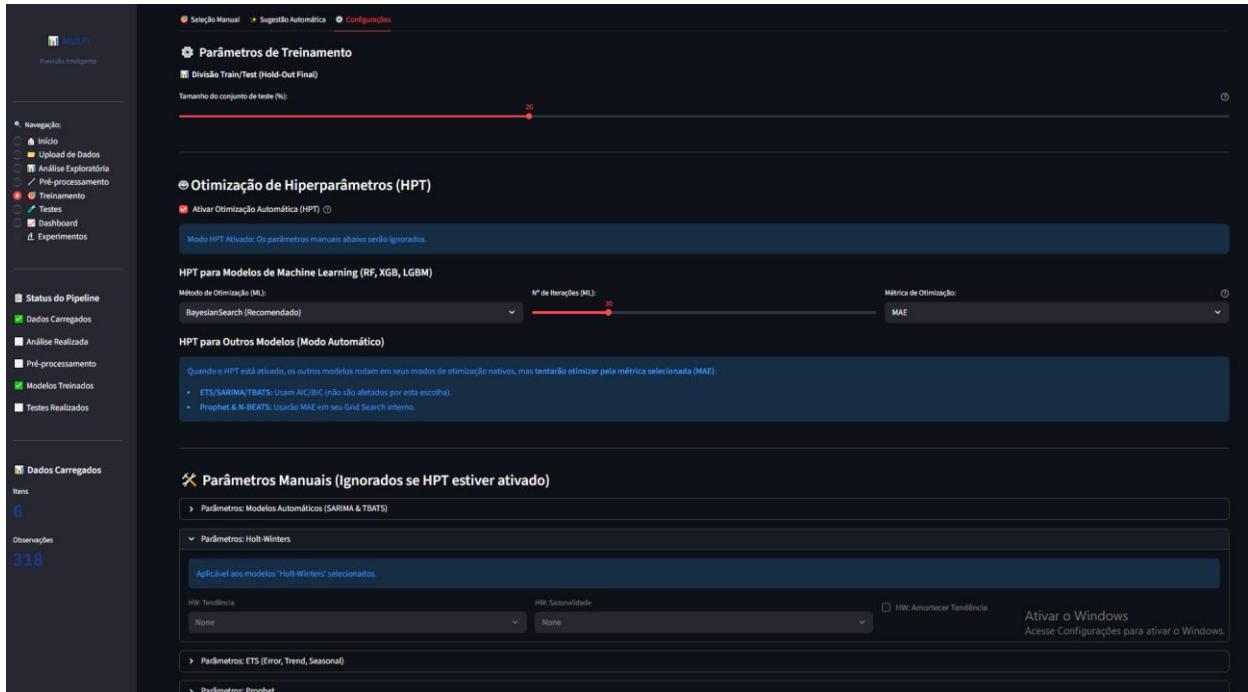


Figura 12- Tela de Configuração – Treinamento. Fonte: Autoria própria, 2025.

## 4.5. Módulo de Testes e Validação com Dados Novos

Após a fase de treinamento e otimização, a jornada entra na fase de testes e validação. Na literatura de *forecasting*, é amplamente reconhecido que um modelo pode apresentar um ajuste excelente aos dados de treino (*in-sample*), mas falhar ao prever novos dados (*out-of-sample*), fenômeno conhecido como *overfitting* (MAKRIDAKIS et al., 2018).

O módulo de testes, foi desenhado para mitigar esse risco e fornecer ao planejador de S&OP uma ferramenta de auditoria. O objetivo deste módulo é duplo: (1) Permitir o diagnóstico dos erros através da análise de resíduos e (2) Simular o ambiente de produção, testando os modelos treinados contra arquivos de dados totalmente novos, simulando a chegada de novos períodos de venda.

A interface é estruturada em três abas funcionais que operacionalizam a avaliação metodológica: Visualização de Previsões (Figura 13), Teste com Novos Dados e Comparaçao de Modelos (Figura 14).

A primeira aba foca na análise do comportamento do modelo selecionado no conjunto de teste (*hold-out*) definido durante o treinamento.

Metodologicamente, a análise dos resíduos ( $e_t = y_t - \hat{y}_t$ ) é o teste definitivo da adequação de um modelo. Se um algoritmo extraiu com sucesso toda a informação estrutural (sinal) disponível nos dados, os resíduos devem se comportar como Ruído Branco. O artefato fornece três ferramentas para esta verificação:

1. **Gráfico de Resíduos no Tempo:** Permite verificar a premissa de média zero ( $E[e_t] = 0$ ) e a homoscedasticidade (variância constante). Se os resíduos apresentam padrões ondulatórios ou crescentes, indica que o modelo falhou em capturar a sazonalidade ou a volatilidade da série, respectivamente (HYNDMAN; ATHANASOPOULOS, 2018).
2. **Histograma de Resíduos:** Permite verificar a premissa de normalidade ( $(e_t \sim N(0, \sigma^2))$ ). Uma distribuição assimétrica ou bimodal nos resíduos sugere que o modelo está enviesado, subestimando ou superestimando sistematicamente a demanda.
3. **Métricas de Erro:** O cálculo do MAE (Erro Médio Absoluto) e RMSE (Raiz do Erro Quadrático Médio) oferece uma quantificação da precisão. Enquanto o MAE é mais interpretável, o RMSE penaliza grandes desvios, sendo uma métrica importante em situações onde não pode haver rupturas ou sobre estoque, por exemplo.

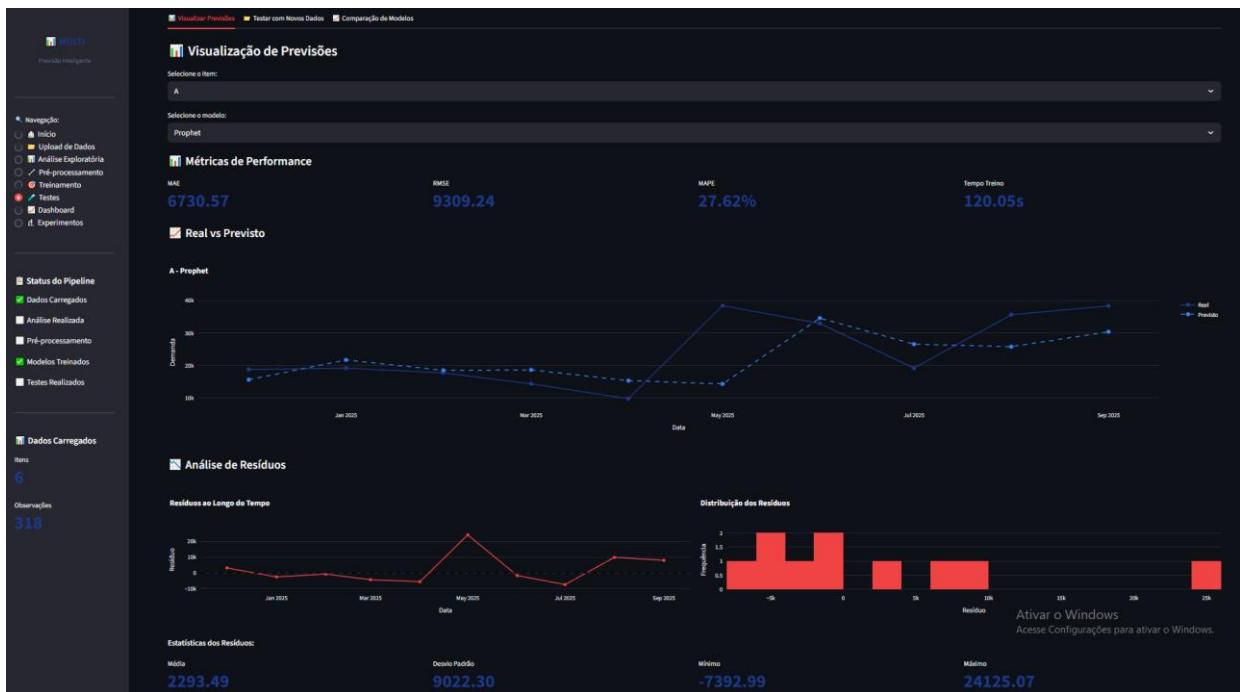


Figura 13 - Tela de Visualizar Previsões – Testes. Fonte: Autoria própria, 2025.

Dentro do modulo de testes, há também a funcionalidade de testes com novos dados. Diferente da validação *hold-out* (que separa uma parte do histórico existente), este teste permite que o usuário faça o *upload* de um arquivo contendo dados que o sistema nunca viu. Este processo simula a realidade operacional do S&OP: o modelo é treinado no mês M, e no mês M+1 sua performance é auditada com os dados reais de vendas recém-fechados.

A última etapa consolida os resultados, permitindo uma análise comparativa preliminar. O artefato gera um relatório visualizado através de dois gráficos. Primeiro, um Boxplot de Distribuição de Erros (MAPE) revela a estabilidade dos modelos. Em S&OP, a consistência é

frequentemente preferível a uma precisão média alta com falhas ocasionais graves (Armstrong, 2001). Segundo um Gráfico de Barras de Tempo Computacional permite avaliar se o ganho de acurácia de um modelo complexo justifica o custo adicional de processamento frente a modelos mais simples. Finalmente, o módulo gera uma tabela de Ranking onde você pode ordenar os modelos pela métrica desejada.

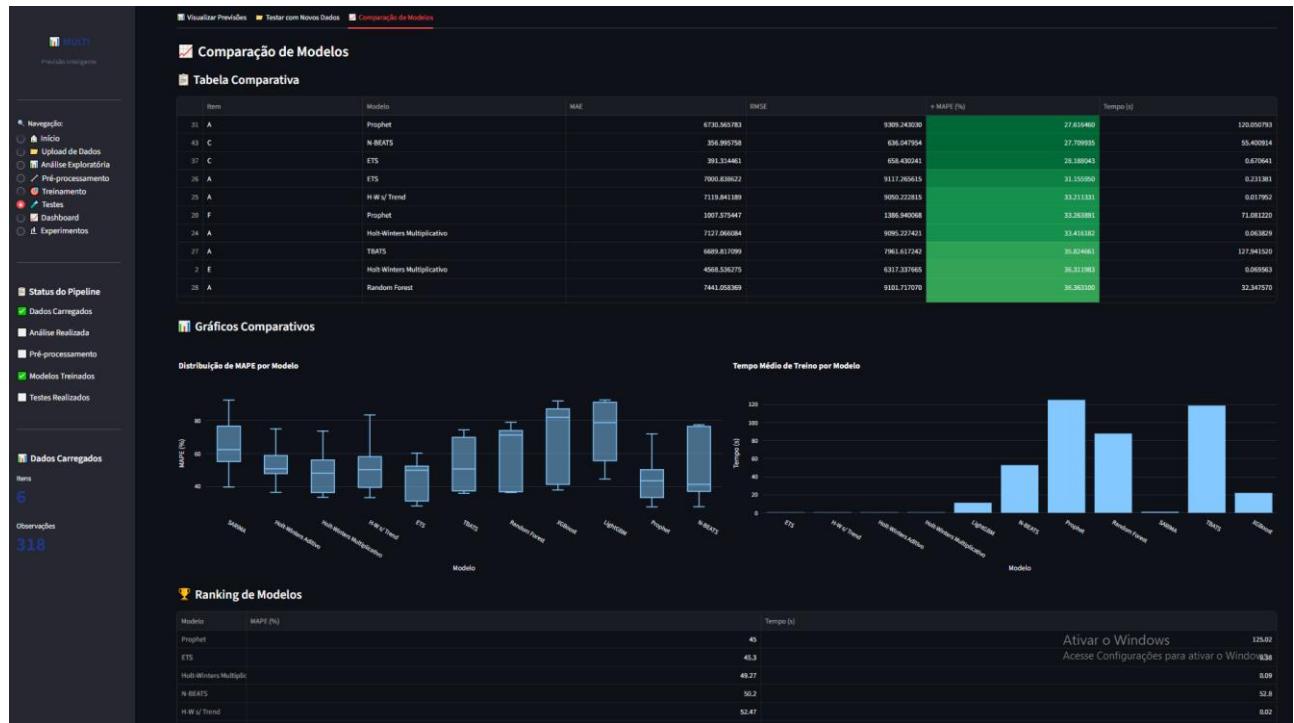


Figura 14- Tela de Comparação de modelos – Testes. Fonte: Autoria própria, 2025.

#### 4.6. Interface de Avaliação e Sistema de Apoio à Decisão (Dashboard)

A última etapa do artefato é a construção de um Dashboard Analítico. Esta interface tem como objetivo ir além da visualização de dados, funcionando como um Sistema de Apoio à Decisão (SAD) interativo. Nela, são processados os resultados brutos da etapa de treinamento e é fornecido ao planejador de demanda uma plataforma para a seleção racional, baseada em diferentes variáveis, dos modelos de previsão. A tela de dashboard está estruturada em 6 abas, cada uma desenhada para responder a uma pergunta específica do processo de planejamento.

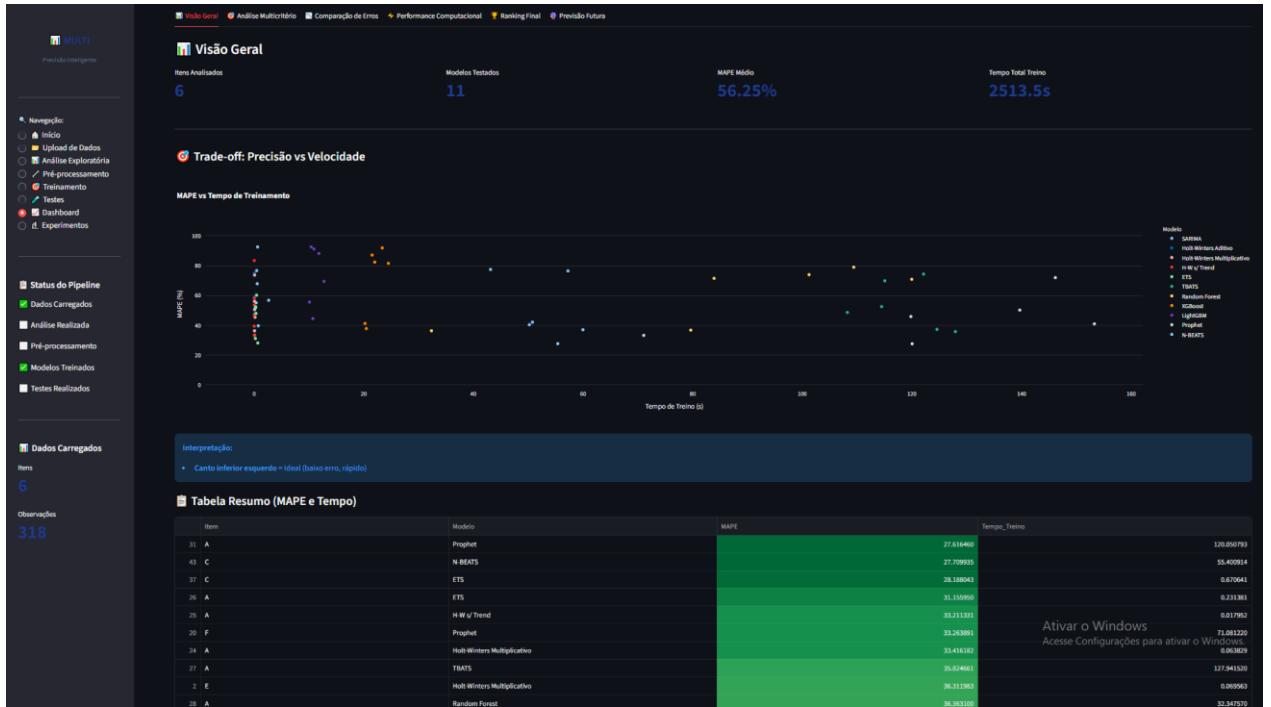


Figura 15- Tela de Visão Geral – Dashboard. Fonte: Autoria própria, 2025.

O primeiro módulo, denominado "Visão Geral", mostrado na Figura 15, atua como um painel de controle executivo. Sua função metodológica é sintetizar o desempenho global dos experimentos, permitindo uma avaliação rápida do desempenho dos modelos. O sistema agrupa métricas de todos os modelos treinados para calcular KPIs agregados, como o MAPE médio global e o tempo total de processamento computacional.

O destaque desta seção reside na implementação visual da "Fronteira de Eficiência". Foi desenvolvido um gráfico de dispersão que correlaciona duas variáveis conflitantes: o Tempo de Treinamento (eixo X) e o Erro MAPE (eixo Y). Cada ponto no gráfico representa um par Modelo-Item. Esta visualização é crítica para identificar outliers positivos (modelos que se situam no quadrante inferior esquerdo, com alta precisão e baixo tempo) e descartar modelos ineficientes. Além disso, o artefato gera uma tabela resumo interativa, que permite a ordenação dos dados por qualquer coluna, facilitando a identificação rápida dos melhores desempenhos em cada uma das métricas.

Reconhecendo que a minimização do erro médio não é o único critério para a escolha de um modelo em ambiente industrial, como defendido na OIPT, o segundo módulo, mostrado na Figura 16, implementa um método Decisão Multicritério (MCDM). A metodologia adotada baseia-se no Método da Soma Ponderada (SAW - Simple Additive Weighting). Através de controles deslizantes (sliders), o usuário define o peso de três dimensões: Precisão (inverso do Erro), Eficiência (inverso do Tempo) e Estabilidade (inverso da Incerteza/Desvio Padrão dos

resíduos). Como as métricas possuem unidades e escalas distintas (segundos, porcentagem e unidades de produto), o algoritmo aplica uma normalização linear para transformar todos os valores em uma escala adimensional de 0 a 100. Para métricas onde "menor é melhor" (erro, tempo e incerteza), é aplicada uma inversão. O Score Geral é calculado pela soma ponderada dos scores normalizados das três dimensões.

Visualmente, esta análise é suportada por Gráficos de Radar, que permitem comparar a "cobertura" de desempenho de cada modelo. Tabelas detalhadas exibem os scores parciais de cada dimensão, garantindo transparência sobre o motivo pelo qual um modelo obteve determinada pontuação final.

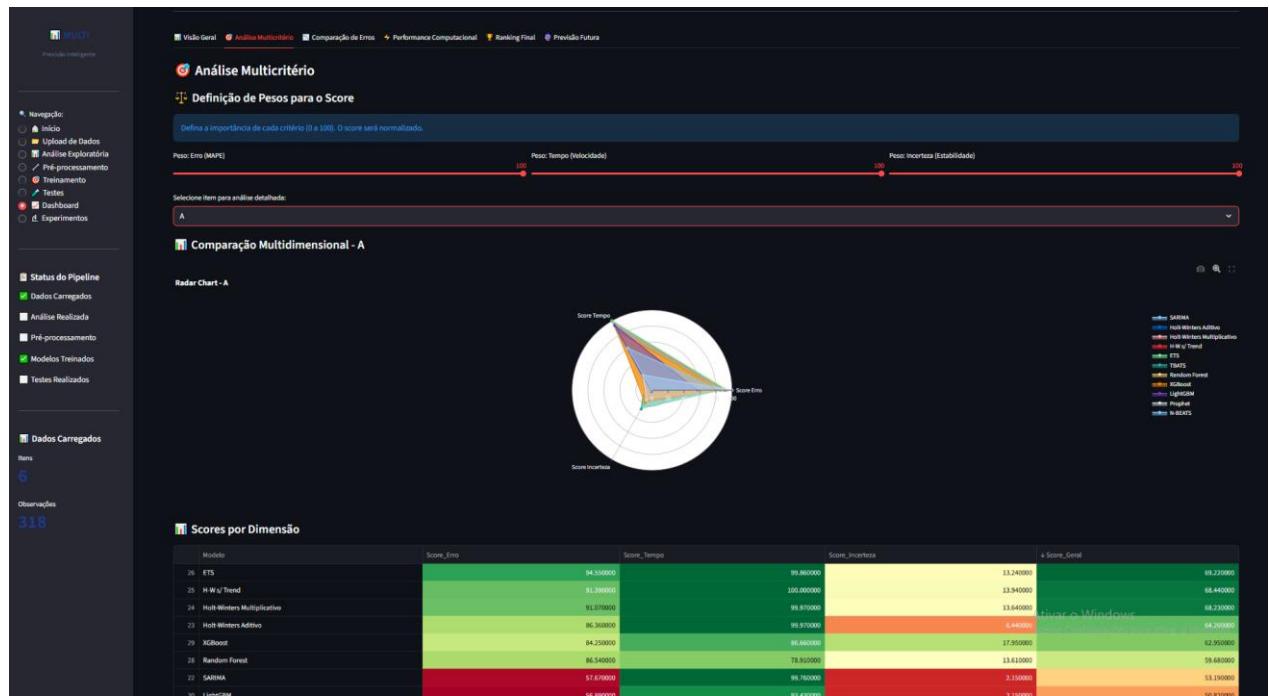


Figura 16- Tela de Análise Multicritério – Dashboard. Fonte: Autoria própria, 2025.

O módulo de "Comparação de Erros" (Figura 17) aprofunda a validação estatística, movendo-se da análise de médias para a análise de distribuições. O código implementa Box Plots para as métricas MAPE e MAE. Um modelo pode apresentar um MAPE médio baixo, mas possuir uma alta variância ou diversos outliers, indicando instabilidade em períodos específicos, problema esse revelado pelo Box Plot.

Adicionalmente, foi implementada uma matriz de calor cruzando a dimensão "Itens" com a dimensão "Modelos". A intensidade da cor representa a magnitude do erro. Esta ferramenta visual permite detectar padrões sistêmicos de falha ou sucesso, identificando se determinados algoritmos funcionam melhor com determinados tipos de item. Tabelas de

estatística descritiva (média, desvio padrão, mínimo e máximo do erro) complementam a análise visual.

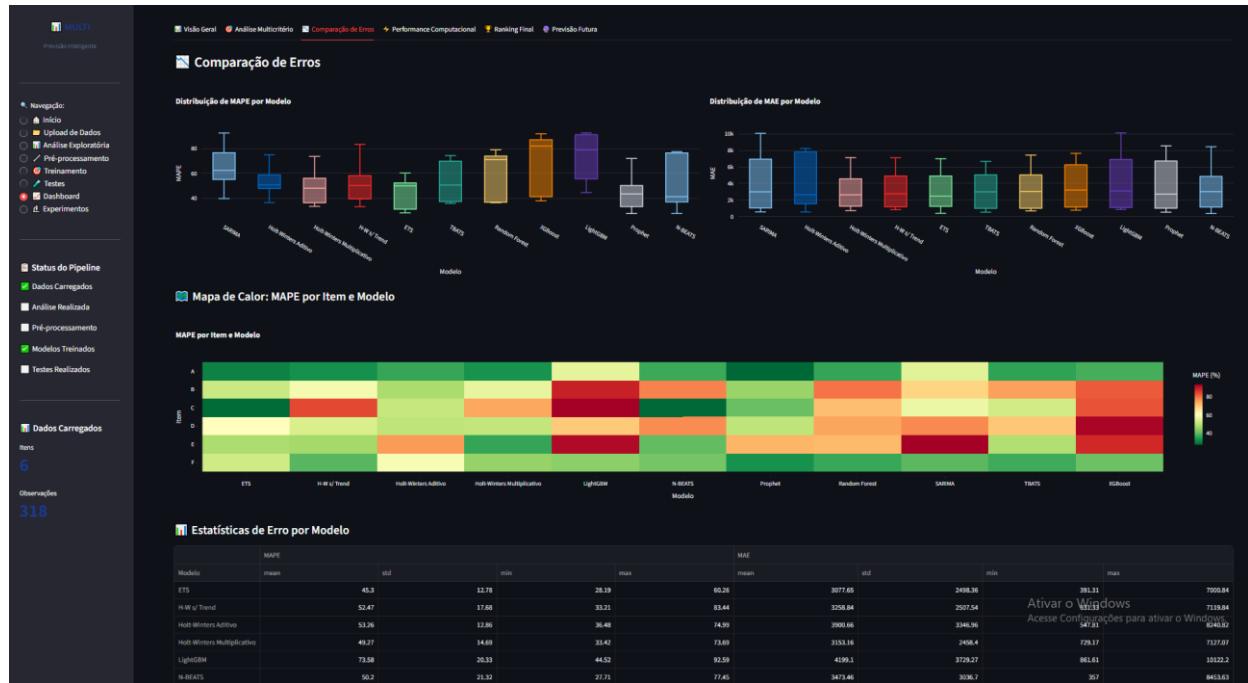


Figura 17- Tela de Comparação de erros – Dashboard. Fonte: Autoria própria, 2025.

Dada a aplicação do artefato no contexto do PCP 4.0, cada vez mais impactado por um grande volume de dados (Big Data) gerados a partir de suas tecnologias, como o IoT, a eficiência computacional é tratada como um requisito crítico. A quarta aba, mostrada na Figura 18, dedica-se a analisar o custo de processamento dos algoritmos. A visualização dos dados ocorre através de gráficos de barras. Esta análise é vital para dimensionar a infraestrutura tecnológica necessária para utilizar modelos mais complexos ou adaptar os modelos que serão utilizados para que sejam compatíveis com a capacidade computacional existente.

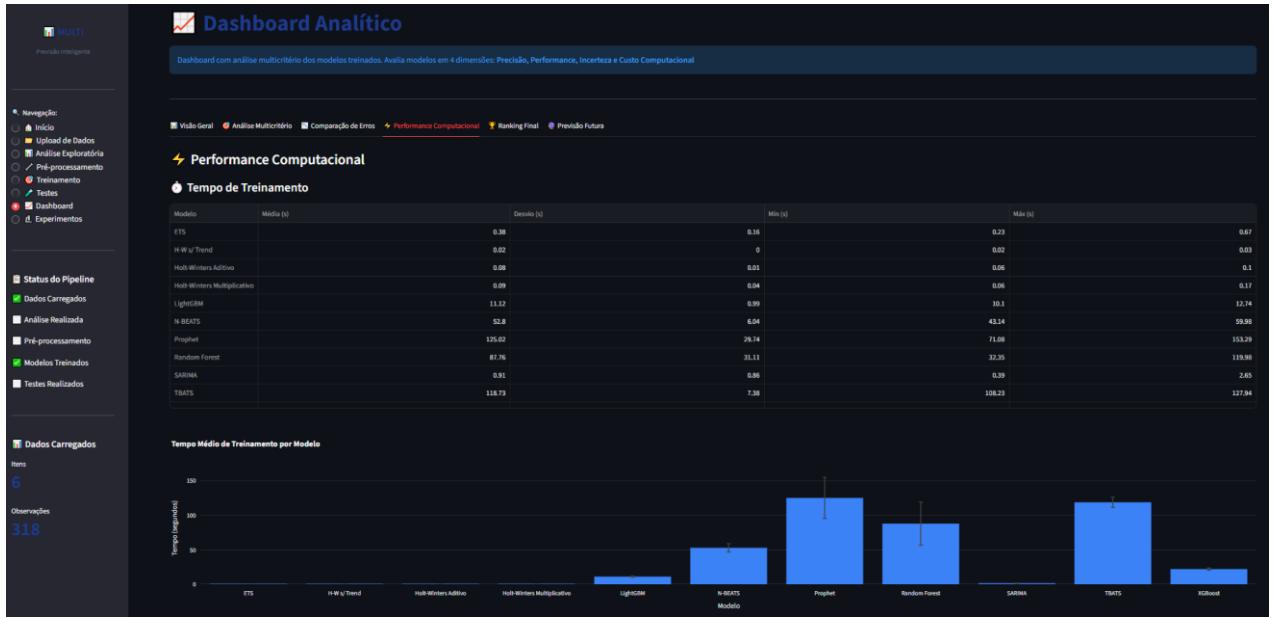


Figura 18- Tela de Performance computacional – Dashboard. Fonte: Autoria própria, 2025.

O quinto módulo (Figura 19) operacionaliza a escolha do modelo. Utilizando os pesos definidos na etapa multicritério, o sistema gera um ranking consolidado. O algoritmo identifica o "Campeão Geral" (modelo com maior média de Score Geral) e, crucialmente, executa uma lógica de seleção granular, determinando o Melhor Modelo por Item.

Esta funcionalidade oferece uma solução personalizada e alinhada com a OPIT, rejeitando a premissa de que um único algoritmo deve ser universalmente aplicado, e sim um algoritmo que se adapte às particularidades de cada série. O resultado é apresentado em tabelas dinâmicas que listam, para cada SKU, qual foi o modelo vencedor, seu score final e o erro associado. Um gráfico de distribuição (Pizza) ilustra o share de cada algoritmo no portfólio, oferecendo uma visão do desempenho dos diferentes modelos.

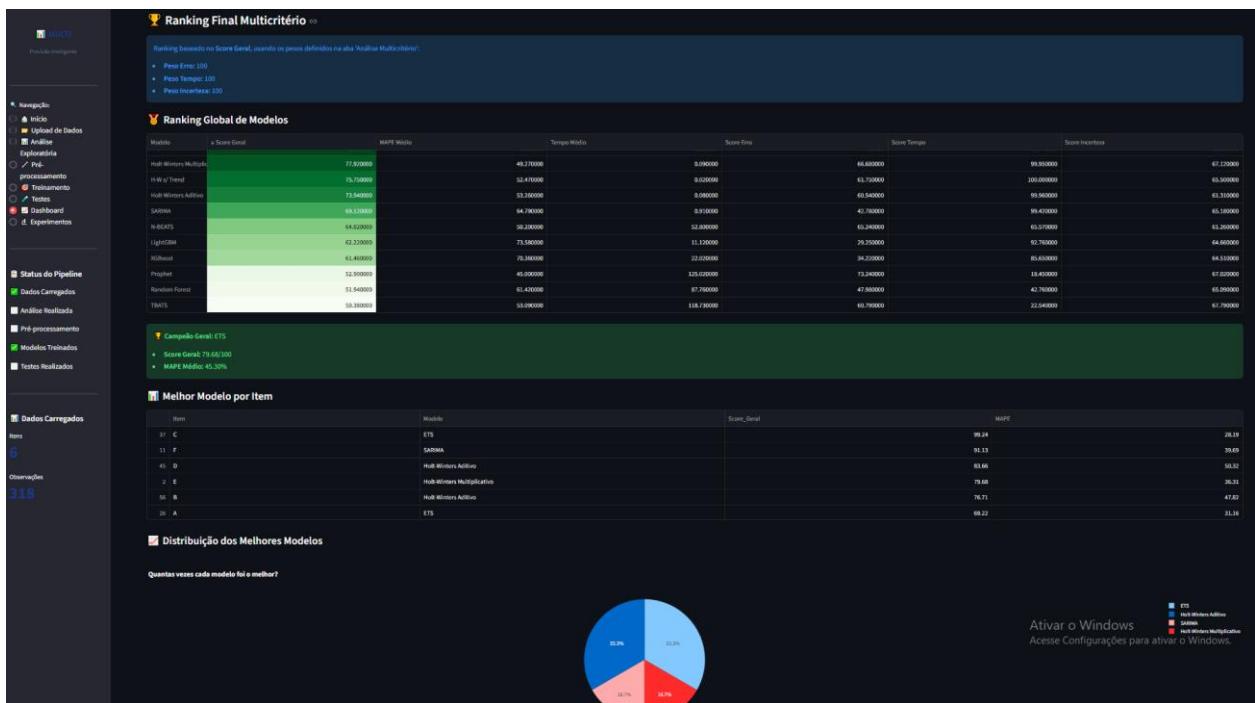


Figura 19- Tela de Ranking final multicritério – Dashboard. Fonte: Autoria própria, 2025.

A última etapa do fluxo metodológico é a "Previsão Futura" (Figura 20). Diferente da etapa de validação (que separa dados em treino e teste), esta função executa um retreinamento utilizando 100% do histórico disponível. O código recupera os hiperparâmetros ótimos armazenados na sessão e reconstrói o modelo.

Para projetar um horizonte de 12 meses, os resultados (histórico + previsão) são exportáveis para Excel, garantindo interoperabilidade com outros sistemas corporativos.

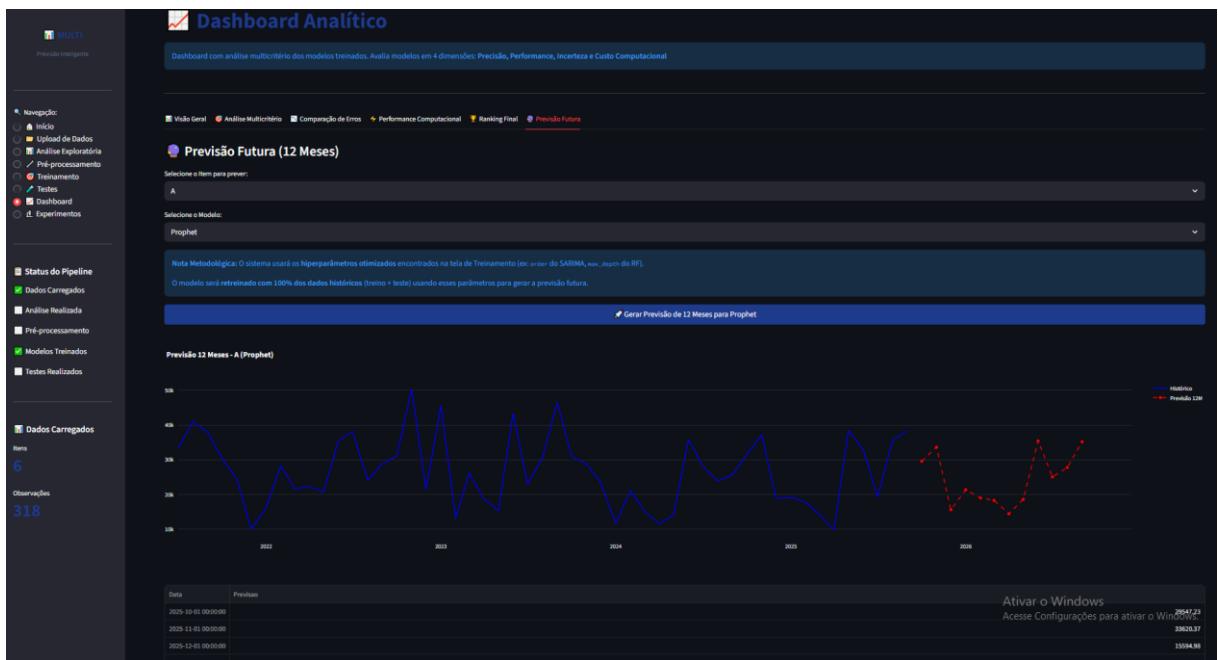


Figura 20- Tela de Previsão futura – Dashboard. Fonte: Autoria própria, 2025

## 4.7. Registro dos experimentos

Por fim, na tela de Registro de Experimentos (Figura 21), após o fim do ciclo de previsão de demanda e escolha de modelo se encerrarem, o usuário pode exportar em formato Excel um plano de experimentos detalhado, com todas as transformações nas séries realizadas nas etapas de pré-processamento, previsões realizadas por cada modelo nas etapas de treino e teste, um resumo dos parâmetros utilizados no treinamento e os hiperparâmetros ótimos encontrados, caso a opção tenha sido habilitada.

Figura 21- Tela de Registro de experimento. Fonte: Autoria própria, 2025

## 5. Aplicação do artefato em um caso real

A seção anterior detalhou a concepção do artefato de previsão multicritério, fundamentado nos princípios do Design Science Research (DSR). A metodologia culminou no design de um sistema que incorpora uma competição de modelos (estatísticos, machine learning e deep learning) e um pipeline robusto de pré-processamento, com o objetivo de endereçar as lacunas identificadas na Etapa 1 da DSR. Esta seção avança no ciclo do DSR, focando nas etapas de Demonstração (Etapa 4) e Avaliação (Etapa 5) (PEFFERS et al., 2007). O artefato será agora aplicado em um ambiente operacional real para demonstrar sua utilidade e medir rigorosamente seu desempenho e aplicabilidade.

Para a realização do experimento, foram selecionados todos os produtos de duas das marcas presentes no portfólio da empresa. A avaliação empírica foi conduzida utilizando séries temporais reais de vendas B2B (sell-in) desses produtos, agregados por subfamília, do portfólio da organização parceira. Essa agregação por subfamília já é realizada hoje no processo de previsão de demanda da empresa e se dá para que as séries sejam menos afetadas pelas rupturas, decorrentes de uma demanda volátil e uma cadeia de suprimentos complexa na qual a empresa está inserida, com lead times longos e disponibilidade de matéria prima limitada em determinados momentos. Atualmente, na empresa, são utilizados apenas os modelos estatísticos clássicos para o suporte ao processo de previsão de demanda.

Toda a discussão a seguir é fundamentada na Teoria do Processamento de Informação Organizacional (O IPT). Nesta seção, é testada a hipótese de que o desempenho superior no planejamento de demanda não advém da escolha de um modelo universalmente ótimo, mas sim do fit entre a incerteza inerente à tarefa de previsão, a capacidade de processamento computacional alocada e as prioridades estratégicas definidas pelo planejador de demanda. Para demonstrar a flexibilidade e a robustez do artefato frente a diferentes realidades de negócio, os resultados foram segmentados e analisados sob dois cenários de decisão distintos: a priorização exclusiva da acurácia, e a busca pelo equilíbrio e eficiência operacional, ambas feitas na granularidade de subfamília. Em ambos os cenários de avaliação multicritério, está sendo avaliado o desempenho dos modelos com a otimização de parâmetros ligada com o número de instâncias da otimização bayesiana igual a 30 (valor padrão do SAD) e com MSE definida como métrica de otimização, como é utilizado na empresa hoje.

## 5.1. Análise de Desempenho Global e a Fronteira de Eficiência

A avaliação empírica do artefato iniciou-se com uma análise de desempenho, comparando todas as famílias de algoritmos (estatísticos, *machine learning* e *deep learning*) aplicadas às séries temporais de *sell-in* das subfamílias selecionadas. Esta etapa tem como objetivo estabelecer uma linha de base de performance e verificar a existência de um modelo universalmente superior, hipótese frequentemente debatida na literatura de *forecasting*.

Tabela 1 – Comparação global dos modelos. Fonte: Autoria própria, 2025.

<b>Modelo</b>	<b>MAPE</b>	<b>Tempo</b>	<b>Score</b>	<b>Score</b>	<b>Score</b>
	Médio (%)	Médio (s)	Erro	Tempo	Incerteza
Prophet	45.00	125.02	73.24	18.45	67.02
ETS	45.30	0.38	72.78	99.77	66.50
Holt-Winters Multiplicativo	49.27	0.09	66.68	99.95	67.12
N-BEATS	50.20	52.80	65.24	65.57	61.26
H-W s/ Trend	52.47	0.02	61.75	100.00	65.50
TBATS	53.09	118.73	60.79	22.54	67.79
Holt-Winters Aditivo	53.26	0.08	60.54	99.96	61.31
Random Forest	61.42	87.76	47.98	42.76	65.09
SARIMA	64.79	0.91	42.78	99.42	65.18
XGBoost	70.36	22.02	34.22	85.65	64.51
LightGBM	73.58	11.12	29.25	92.76	64.66

Os resultados consolidados revelam uma disparidade significativa entre as classes de modelos, tanto na precisão das previsões quanto, e principalmente, no consumo de recursos computacionais. A Tabela 1 apresenta um resumo das métricas médias obtidas no experimento.

Ao analisar a dimensão da eficiência, observa-se que os modelos estatísticos clássicos, notadamente a família de Suavização Exponencial (ETS e Holt-Winters), mantiveram sua reputação de robustez e velocidade. O modelo ETS, por exemplo, registrou um tempo médio de treinamento de apenas 0,38 segundos por série. Em contraste, os modelos de inteligência artificial exigiram uma capacidade de processamento exponencialmente maior. O algoritmo Prophet demandou, em média, 125,02 segundos para processar as mesmas séries, um aumento de aproximadamente 328 vezes no custo computacional em comparação ao ETS. O modelo de *Deep Learning* N-BEATS situou-se em um patamar intermediário de custo, com média de 52,8 segundos.

Na dimensão da precisão (mensurada pelo MAPE - Erro Percentual Absoluto Médio), a relação não é linear. Embora o modelo Prophet tenha apresentado o menor erro médio global em alguns cenários de teste, a diferença de acurácia em relação aos métodos estatísticos foi, em

muitos casos, marginal. Já os algoritmos baseados em árvores de decisão, como XGBoost e Random Forest, apresentaram uma variância de desempenho grande, com erros médios (MAPE) superiores a 60% e 70% em determinadas situações. Este comportamento sugere a ocorrência de *overfitting* em séries com históricos mais curtos ou com alta relação ruído-sinal, onde a complexidade do modelo excedeu a informação disponível nos dados.

Estes achados corroboram a Teoria do Processamento de Informação Organizacional (O IPT) aplicada ao PCP 4.0. Observa-se claramente uma Fronteira de Eficiência. Para a maioria dos itens de comportamento estável, os modelos estatísticos oferecem o melhor *fit* organizacional, entregando precisão competente com custo computacional desprezível. A aplicação de modelos pesados de IA (como Prophet ou N-BEATS) só se justifica estatisticamente em itens de alta volatilidade ou "Curva A", onde uma redução marginal no erro de previsão traduz-se em ganhos financeiros que compensam o alto *lead time* de processamento da informação.

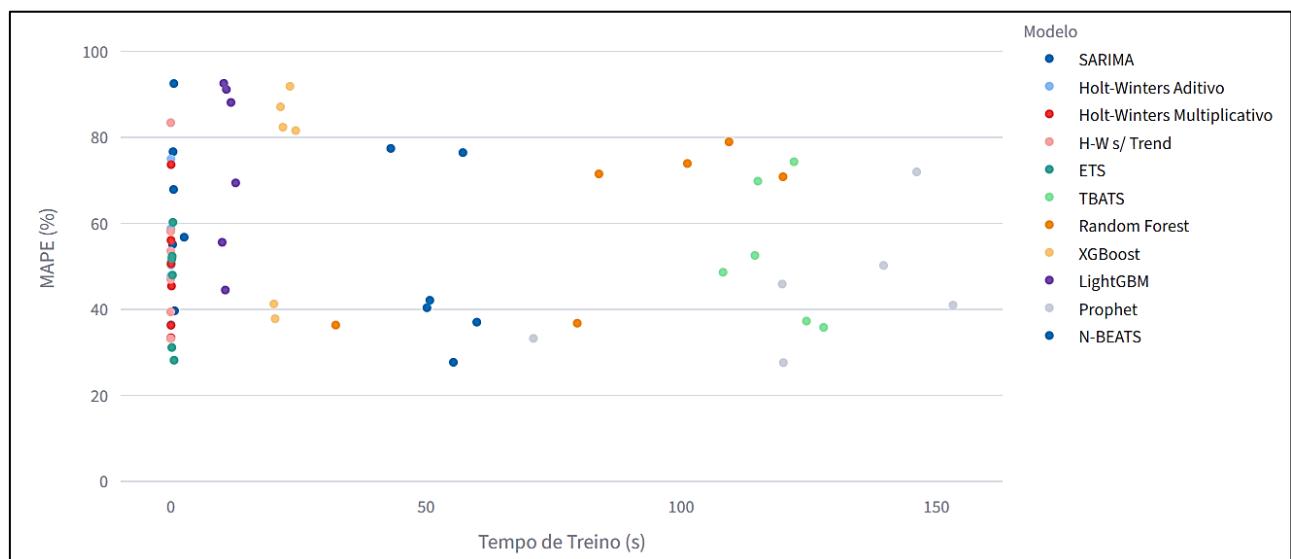


Figura 22– Fronteira de eficiência dos modelos. Fonte: Autoria própria, 2025.

Portanto, a superioridade de um modelo não é absoluta, mas depende da situação. Enquanto a IA demonstrou capacidade de capturar padrões complexos que escaparam à estatística clássica em itens específicos, sua aplicação indiscriminada em todo o portfólio geraria uma ineficiência sistêmica, consumindo recursos de processamento desproporcionais ao ganho de acurácia obtido.

## 5.2. Avaliação Multicritério: O Impacto da Estratégia na Seleção

Reconhecendo que a minimização do erro médio não é o único critério para a escolha de um modelo em ambiente industrial, conforme defendido pela Teoria do Processamento de

Informação Organizacional (O IPT), esta etapa aplicou o método de decisão multicritério SAW (*Simple Additive Weighting*). O objetivo foi demonstrar como a definição de "melhor modelo" é dinâmica e dependente das prioridades estratégicas do planejador de demanda.

Para validar a robustez do artefato, foram simulados três cenários distintos de pesos para os critérios de Precisão (Erro), Eficiência (Tempo) e Estabilidade (Incerteza).

### 5.2.1. Cenário 1: Foco Estratégico na Precisão

No primeiro cenário de avaliação, o sistema foi configurado para simular um ambiente de alta criticidade, onde a redução da incerteza é a prioridade máxima. Neste contexto, assume-se que o custo da falta de produto ou o custo de oportunidade de vendas perdidas superam largamente quaisquer preocupações com o custo computacional ou tempo de execução dos modelos. Os pesos do algoritmo multicritério do artefato foram ajustados para maximizar a acurácia, não penalizando o consumo de recursos de hardware. Sob a ótica da O IPT, este cenário reflete a busca por uma capacidade de processamento de informação máxima, onde a organização decide investir pesadamente em análise de dados para mitigar a volatilidade do mercado. A configuração de pesos priorizou exclusivamente a acurácia (*Erro = 100%, Tempo = 0%, Estabilidade = 0%*).

Tabela 2– Pesos de cada métrica no cenário 1. Fonte: Autoria própria, 2025

	Erro (MAPE)	Tempo (s)	Incerteza (estabilidade)
Peso	100	0	0

Nesse cenário, os scores globais correspondem ao Score de erro, mostrado na tabela 1. Os resultados globais consolidados neste cenário demonstram uma vantagem dos métodos de Machine Learning, especificamente, o modelo Prophet. Ele alcançou a primeira posição no ranking global, com um Score Geral de 73,24, seguido de perto pelo modelo ETS com 72,78. A liderança do Prophet justifica-se pela sua capacidade superior em modelar sazonalidades complexas e mudanças de tendência, resultando no menor MAPE médio do experimento (45,00%). Contudo, a margem estreita para o ETS (apenas 0,46 pontos de diferença no score) indica que, para o conjunto de dados analisado, o ganho de precisão da abordagem de *Machine Learning* foi marginal em relação ao método estatístico clássico.

Nesse cenário considerando exclusivamente o lado operacional, com 100% da pontuação vinda do score de erro, o modelo Prophet se destacou, sendo o melhor modelo em 4 das 6 subfamílias analisadas (A, B, D, F), enquanto o Holt-Winters multiplicativo obteve o

melhor resultado na subfamília E e o N-BEATS na C. Na figura 23, podemos ver o gráfico de pizza representando a porcentagem de subfamílias onde cada modelo obteve um melhor desempenho.

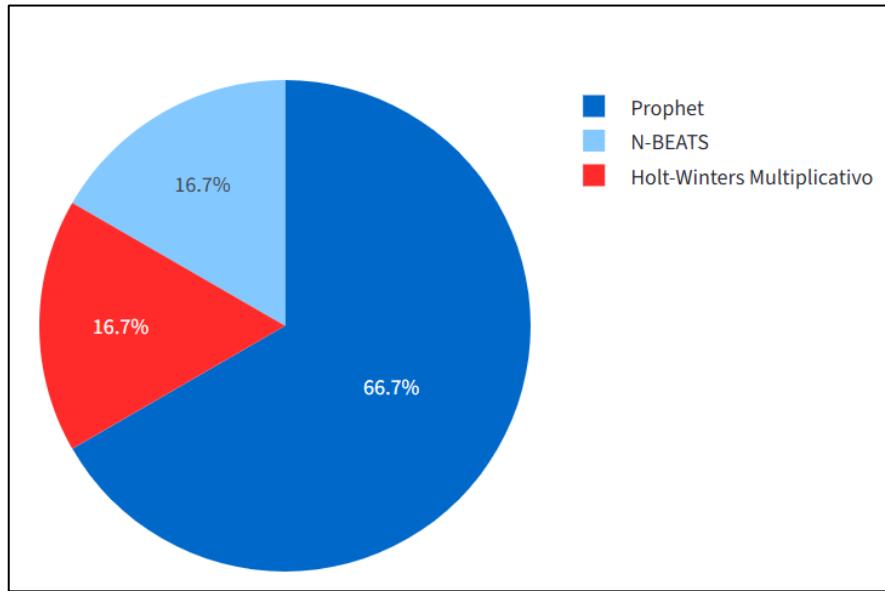


Figura 23– Percentual de itens onde cada modelo obteve o melhor Score no cenário 1. Fonte: Autoria própria, 2025.

### 5.2.2. Cenário 2: Gestão de riscos

Neste cenário, a análise transita de uma perspectiva puramente focada na magnitude do erro para uma abordagem orientada à gestão de riscos e confiabilidade da informação. Na prática do S&OP, um modelo de previsão que apresenta uma acurácia média excelente, mas que falha drasticamente em períodos aleatórios (alta variância), é considerado perigoso. A instabilidade nos resíduos de previsão impacta diretamente o cálculo dos estoques de segurança, que são dimensionados justamente para cobrir a variabilidade do erro durante o lead time. Esta configuração penaliza severamente modelos que apresentam outliers em seus resíduos ou comportamento errático entre diferentes janelas de validação, privilegiando a robustez. Os pesos foram distribuídos igualmente entre as dimensões Erro e Estabilidade (*Erro = 100, Tempo = 0, Estabilidade = 100*).

Tabela 3– Pesos de cada métrica no cenário 2. Fonte: Autoria própria, 2025.

	Erro (MAPE)	Tempo (s)	Incerteza (estabilidade)
Peso	100	0	100

Ao aplicar esta ponderação aos resultados experimentais, a configuração do ranking se altera. O Prophet segue na liderança (Score 70,13), demonstrando que sua vantagem não está apenas na média do erro, mas na consistência das previsões (Score de Incerteza de 67,02, o

segundo maior do grupo). O modelo TBATS, que havia tido desempenho mediano nos outros cenários, subiu para a 4<sup>a</sup> posição, destacando-se pela estabilidade de seus componentes sazonais. Já modelos como LightGBM e XGBoost permaneceram nas últimas posições, evidenciando a alta variância e o risco de overfitting associados a algoritmos de árvore de decisão em séries temporais curtas.

*Tabela 4– Scores no cenário 2. Fonte: Autoria própria, 2025.*

Modelo	Score Geral	MAPE Médio (%)	Tempo Médio (s)	Score Erro	Score Tempo	Score Incerteza
Prophet	70.13	45.00	125.02	73.24	18.45	67.02
ETS	69.64	45.3	0.38	72.78	99.77	66.50
Holt-Winters Multiplicativo	66.90	49.27	0.09	66.68	99.95	67.12
TBATS	64.29	53.09	118.73	60.79	22.54	67.79
H-W s/ Trend	63.63	52.47	0.02	61.75	100.00	65.50
N-BEATS	63.25	50.20	52.80	65.24	65.57	61.26
Holt-Winters Aditivo	60.92	53.26	0.08	60.54	99.96	61.31
Random Forest	56.54	61.42	87.76	47.98	42.76	65.09
SARIMA	53.98	64.79	0.91	42.78	99.42	65.18
XGBoost	49.37	70.36	22.02	34.22	85.65	64.51
LightGBM	46.96	73.58	11.12	29.25	92.76	64.66

Nesse cenário, o modelo Prophet mais uma vez liderou, sendo o melhor modelo em 3 das 6 subfamílias analisadas (B, D, F). Além dele, podemos ver a entrada do N-BEATS, modelo de deep learning como o melhor modelo para a subfamília C, obtendo um MAPE de 27,71% para essa subfamília, o segundo menor entre todos os modelos testados em todos as subfamílias, e um Score total de 99,29 para a subfamília C. Do lado dos modelos estatísticos, o Holt-Winters Multiplicativo segui com a subfamília E enquanto o TBATS aparece como o melhor modelo da subfamília A. Na Figura 24, podemos observar a porcentagem de subfamílias onde cada modelo obteve um melhor desempenho considerando esse novo cenário.

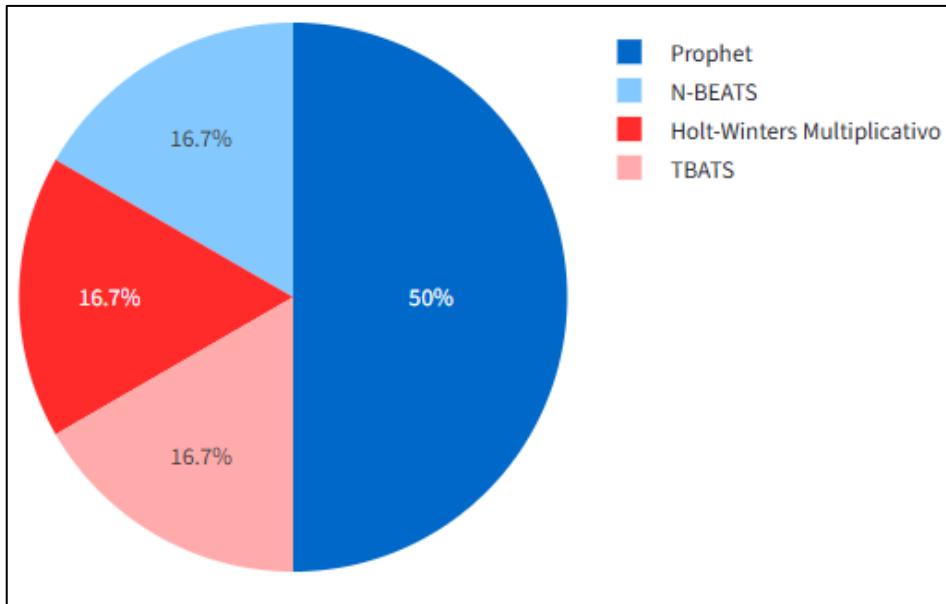


Figura 24– Percentual de itens onde cada modelo obteve o melhor Score no cenário 2. Fonte: Autoria própria, 2025.

### 5.2.3. Cenário 3: Equilíbrio Operacional

O terceiro cenário de teste inverte a lógica de decisão, priorizando a agilidade e a eficiência do processo de S&OP. Este cenário simula uma realidade operacional comum em grandes organizações de varejo e bens de consumo: a necessidade de re-planejamento ágil, muitas vezes em períodos mais curtos, ou a existência de restrições de infraestrutura de TI ao lidar com bases de dados massivas. Neste arranjo, os pesos do sistema multicritério foram reconfigurados para considerar de maneira igual a qualidade de previsão e o esforço computacional.

A alteração dos critérios de seleção provocou uma reordenação drástica no ranking de modelos, ilustrando a sensibilidade do artefato às preferências gerenciais. Os pesos foram distribuídos igualmente entre as três dimensões (*Erro* = 100, *Tempo* = 100, *Estabilidade* = 100).

Tabela 5– Scores no cenário 3. Fonte: Autoria própria, 2025.

	Erro (MAPE)	Tempo (s)	Incerteza (estabilidade)
Peso	100	100	100

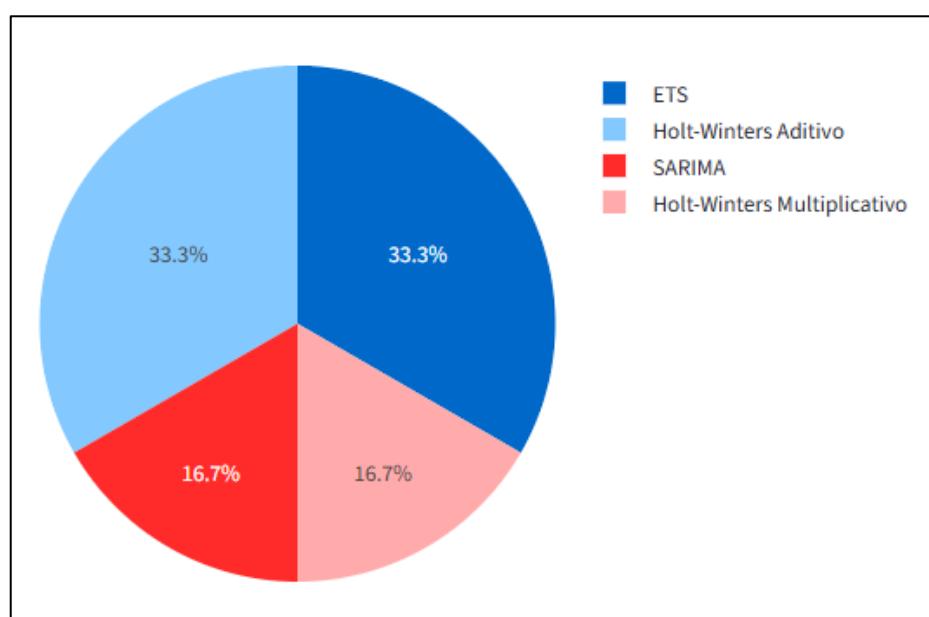
A introdução da variável "Tempo" alterou drasticamente o ranking. O modelo ETS ascendeu à liderança absoluta com um Score Geral de 79,68, seguido pelo Holt-Winters Multiplicativo (Score 77,92). Em contraste, os modelos de Machine Learning enfrentaram uma queda drástica. Entre eles, vale destacar o Prophet, líder no primeiro cenário e que caiu para a 9<sup>a</sup> posição (Score 52,90), penalizado severamente pelo seu alto tempo de processamento

(125,02 segundos em média, contra 0,38s do ETS). Este resultado valida a premissa de que o excesso de capacidade de processamento, modelos lentos e complexos geram ineficiência organizacional quando não acompanhados de um ganho proporcional em redução de incerteza ou aumento de sua capacidade computacional.

*Tabela 6– Scores no cenário 3. Fonte: Autoria própria, 2025.*

Modelo	Score Geral	MAPE Médio (%)	Tempo Médio (s)	Score Erro	Score Tempo	Score Incerteza
ETS	79.68	45.30	0.38	72.78	99.77	66.50
Holt-Winters Multiplicativo	77.92	49.27	0.09	66.68	99.95	67.12
H-W s/ Trend	75.75	52.47	0.02	61.75	100.00	65.50
Holt-Winters Aditivo	73.94	53.26	0.08	60.54	99.96	61.31
SARIMA	69.12	64.79	0.91	42.78	99.42	65.18
N-BEATS	64.02	50.20	52.80	65.24	65.57	61.26
LightGBM	62.22	73.58	11.12	29.25	92.76	64.66
XGBoost	61.46	70.36	22.02	34.22	85.65	64.51
Prophet	52.90	45.00	125.02	73.24	18.45	67.02
Random Forest	51.94	61.42	87.76	47.98	42.76	65.09
TBATS	50.38	53.09	118.73	60.79	22.54	67.79

Nesse novo cenário considerando um equilíbrio entre os 3 parâmetros, os modelos estatísticos clássicos obtiveram larga vantagem, sendo escolhidos o melhor modelo em todas 6 subfamílias analisadas, como podemos observar na Figura 25. Para as famílias A e C, foi escolhido o ETS, para B e D, Holt-Winters aditivo, para a subfamília E, Holt-Winters Multiplicativo e para a F, SARIMA.



*Figura 25– Percentual de itens onde cada modelo obteve o melhor Score no cenário 3. Fonte: Autoria própria, 2025.*

### 5.3 Discussões

Os resultados obtidos na aplicação empírica do artefato confirmam a premissa central deste estudo de que não existe um algoritmo universalmente superior para a previsão de demanda no contexto do PCP 4.0. A variação de desempenho observada entre as subfamílias e a mudança de ranking conforme a reconfiguração dos pesos do sistema multicritério evidenciam que a escolha do “melhor” modelo depende do contexto de decisão e das prioridades gerenciais, e não apenas da dimensão operacional da acurácia.

Sob a lente da Teoria do Processamento de Informação Organizacional (O IPT), esses achados reforçam empiricamente o conceito de fit entre a incerteza da tarefa e a capacidade de processamento de informação da organização (GALBRAITH, 1974). Em termos de capacidade de processamento de informação, as séries com demanda mais errática e alta volatilidade geram maior necessidade de processamento de informação. Nesses casos, os modelos de *Machine Learning* e *Deep Learning*, como Prophet e N-BEATS, representam uma elevação da capacidade de processamento de informação, oferecendo melhor capacidade de captura de padrões complexos, ainda que com custo computacional significativamente maior.

Por outro lado, para subfamílias com padrões mais estáveis ou com menor criticidade em termos de risco de ruptura, modelos estatísticos clássicos, como ETS, Holt-Winters e SARIMA, mostraram-se suficientes para atender às necessidades de processamento, oferecendo previsões aceitáveis com consumo de recursos muito inferior. Nesses casos, investir em modelos mais complexos implicaria em um aumento de custo computacional sem contrapartida proporcional em redução de incerteza, o que, segundo a O IPT, caracteriza desperdício organizacional e misfit entre tarefa e estrutura de informação.

É nesse ponto que o componente multicritério do artefato passa a operar como um “agente” da O IPT no processo decisório. Ao integrar simultaneamente as dimensões de erro (MAPE), tempo computacional e estabilidade, o sistema de apoio à decisão transforma os resultados do treinamento e dos testes em um mecanismo explícito de balanceamento entre necessidade de processamento de informação e capacidade de processamento de informação. A possibilidade de configurar pesos diferentes para cada dimensão permite que o planejador ajuste, de forma declarada, o nível de capacidade de processamento que a organização está disposta a mobilizar para responder à incerteza observada em cada série, tornando o *fit* uma escolha gerencial parametrizável e auditável, em vez de uma consequência implícita da escolha de um único modelo.

Nos cenários em que a acurácia é priorizada (Cenário 1), o sistema multicritério aproxima a organização de uma estratégia de “maximização” da capacidade de processamento de informação, aceitando maiores tempos de processamento e maior complexidade algorítmica em troca de menor incerteza nas previsões. Na lógica da OIPT, o artefato atua como um mecanismo de aumento deliberado da capacidade de processamento de informação para lidar com um ambiente considerado como altamente incerto, típico de contextos em que o custo da falta de produto ou de perda de vendas é elevado.

Já nos cenários que incorporam tempo e estabilidade (Cenários 2 e 3), o mesmo artefato passa a funcionar como um dispositivo de racionalização da capacidade de processamento. Ao penalizar modelos lentos ou instáveis, o sistema multicritério ajuda a evitar situações de “excesso” de capacidade de processamento, como o uso de modelos complexos em séries de baixa incerteza, que, de acordo com a OIPT, não contribuem para reduzir a incerteza residual e apenas adicionam custo e complexidade ao processo organizacional (GALBRAITH, 1974). Na prática, o sistema orienta a organização a operar mais próxima de uma fronteira eficiente entre esforço computacional e qualidade da informação, ajustando a estrutura de processamento às características informacionais de cada série.

Dessa forma, o artefato proposto contribui para a literatura ao demonstrar empiricamente um mecanismo concreto pelo qual a OIPT pode ser operacionalizada em ambientes de PCP 4.0 e S&OP. O módulo multicritério não apenas seleciona modelos “melhores” do ponto de vista isolado da acurácia, mas atua como um mediador entre os requisitos de informação (incerteza, criticidade, volatilidade) e a capacidade de processamento (tipo de modelo, tempo de execução, complexidade computacional), promovendo um fit dinâmico e contextualizado entre a demanda e as prioridades da empresa.

## 6. Conclusão

A gestão da cadeia de suprimentos contemporânea, impulsionada pela I4.0, exige que as organizações transitem de modelos reativos para sistemas preditivos e adaptáveis. A volatilidade dos mercados atuais e o advento do *Big Data* expuseram as limitações dos métodos tradicionais de previsão, exigindo a adoção de novas abordagens analíticas (HOFMANN; RUTSCHMANN, 2018). Este trabalho de conclusão de curso propôs-se a investigar essa lacuna, desenvolvendo e avaliando um artefato computacional alinhado aos princípios do PCP 4.0, fundamentado metodologicamente no Design Science Research (DSR) (PEFFERS et al., 2007) e teoricamente na Teoria do Processamento de Informação Organizacional (OIPT).

O objetivo central desta pesquisa foi responder como a seleção de modelos de previsão de demanda pode ser otimizada considerando o trade-off entre precisão, tempo de processamento e incerteza. Para tanto, foi desenvolvido um sistema multicritério de suporte à decisão que orquestrou uma competição entre modelos estatísticos clássicos e algoritmos de *Machine Learning* e *Deep Learning*. A aplicação deste artefato em dados reais permitiu validar empiricamente os conceitos de ajuste (fit) organizacional. Os resultados corroboram a visão de que o desempenho não é absoluto, mas contingencial ao alinhamento entre a incerteza da tarefa e a capacidade de processamento de informação da organização (TUSHMAN; NADLER, 1978).

Nos testes realizados, observou-se que modelos de *Machine Learning*, especificamente o Prophet e N-Beats, demonstraram uma capacidade superior de reduzir a incerteza em itens de alta complexidade (Curva A). Para a subfamília C, a aplicação dessas técnicas resultou em uma redução do erro em comparação aos métodos estatísticos. Isso valida a premissa de que tarefas de alta incerteza demandam alta capacidade de processamento de informação, superando as limitações inerentes aos modelos lineares e estacionários.

Contudo, a pesquisa também iluminou o custo oculto dessa sofisticação. A análise de eficiência demonstrou que modelos de *Machine Learning* podem ser ordens de magnitude mais lentos que métodos clássicos como o ETS e o Holt-Winters. Esse achado é crucial, pois a complexidade computacional e o tempo de execução são fatores críticos para a viabilidade prática da implementação de modelos preditivos em ambientes de negócios reais (KOLKOVÁ; NAVRÁTIL, 2021). O artefato provou seu valor ao identificar que, para itens de menor volatilidade, modelos estatísticos simples entregam resultados mais eficientes, evitando o desperdício de recursos e o excesso de processamento desnecessário (DAFT; LENGEL, 1986).

Uma das principais contribuições teóricas deste trabalho foi a operacionalização da OIPT no contexto do PCP. Ao demonstrar que a "melhor previsão" varia dependendo se o objetivo estratégico é a acurácia ou a eficiência, o estudo refuta a visão de que o modelo mais complexo é sempre o melhor. A fragmentação das vitórias reforça o conceito de equifinalidade e confirma achados recentes de que métodos híbridos e a combinação de previsões tendem a ser superiores a abordagens monolíticas em competições de larga escala (MAKRIDAKIS et al., 2018; 2020).

No entanto, é fundamental delimitar o escopo da aplicação da ferramenta desenvolvida. Deve-se reconhecer que o artefato atua estritamente como um Sistema de Apoio à Decisão (SAD) e não substitui a necessidade de um processo robusto de Planejamento de Vendas e Operações (S&OP). A previsão estatística ou algorítmica é apenas o ponto de partida; o resultado ótimo organizacional depende da integração colaborativa entre as áreas de Vendas, Marketing, Finanças e Operações, que aportam inteligência de mercado qualitativa não capturada pelos modelos matemáticos (KAHN, 2003; BRAU, 2023). O artefato serve, portanto, para reduzir a incerteza analítica, mas o consenso do plano de demanda permanece uma responsabilidade humana e processual.

Como oportunidades para pesquisas futuras, sugere-se a expansão do repositório de modelos do artefato para incluir arquiteturas de Redes Neurais Recorrentes mais sofisticadas, como as LSTMs (Long Short-Term Memory), que são projetadas especificamente para capturar dependências de longo prazo em sequências temporais complexas (KOLKOVÁ; NAVRÁTIL, 2021). Adicionalmente, recomenda-se a evolução do artefato para além da previsão de demanda pura, buscando sua integração horizontal com outras áreas do S&OP. Isso incluiria a conexão com módulos de gerenciamento de estoques, utilizando a métrica de incerteza do artefato para calibrar estoques de segurança dinâmicos, e com o planejamento de transportes, otimizando a malha logística com base na previsão granular de volume.

Em suma, este trabalho conclui que a transição para o PCP 4.0 não é apenas uma atualização tecnológica, mas uma evolução na capacidade de digitalização e inteligência organizacional (BUENO, 2020). O artefato validado serve como um protótipo dessa nova era, onde a precisão analítica e a eficiência operacional caminham juntas, guiadas pela inteligência humana apoiada por dados.

## 7. Referências

- ALEXOPOULOS, K.; NIKOLAKIS, N.; CHRYSSOLOURIS, G. Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing. *International Journal of Computer Integrated Manufacturing*, v. 33, n. 5, p. 429-439, 2020.
- ARMSTRONG, J. S. (Ed.). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Boston: Kluwer Academic, 2001.
- ATHANASOPOULOS, G.; HYNDMAN, R. J.; KOURENTZES, N.; PETROPOULOS, F. Forecasting with temporal hierarchies. *European Journal of Operational Research*, v. 262, n. 1, p. 60-74, 2017.
- BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, v. 191, p. 192-213, 2012.
- BERGSTRA, J.; BENGIO, Y. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, v. 13, p. 281-305, 2012.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- BONTEMPI, G.; BEN TAIEB, S.; LE BORGNE, Y.-A. Machine learning strategies for time series forecasting. In: AUFAURE, M.-A.; ZIMÁNYI, E. (Eds.). *Business Intelligence. eBISS 2012. Lecture Notes in Business Information Processing*, v. 138. Berlin: Springer, 2013.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 26, n. 2, p. 211-243, 1964.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time Series Analysis: Forecasting and Control*. 5. ed. Hoboken: John Wiley & Sons, 2015.
- BRAU, R. I.; ALOYSIUS, J. A.; SIEMSEN, E. Demand planning for the digital supply chain: How to integrate human judgment and predictive analytics. *Journal of Operations Management*, v. 69, n. 6, p. 946-972, 2023.
- BREIMAN, L. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5-32, 2001.
- BROOKS, C. *Introductory Econometrics for Finance*. 4. ed. Cambridge: Cambridge University Press, 2019.
- BUENO, A.; GODINHO FILHO, M.; FRANK, A. G. Smart production planning and control in the Industry 4.0 context: A systematic literature review. *Computers & Industrial Engineering*, v. 149, p. 106774, 2020.
- CHAKRAVARTI, I. M.; LAHA, R. G.; ROY, J. *Handbook of Methods of Applied Statistics*. New York: John Wiley & Sons, 1967. v. 1.

- CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22., 2016, New York. Proceedings... New York: ACM, 2016. p. 785-794.
- CHOPRA, S.; MEINDL, P. Supply Chain Management: Strategy, Planning, and Operation. 6. ed. Boston: Pearson, 2016.
- CIANO, M. P. et al. Lean Production and Industry 4.0: The state of the art and future perspectives. *Total Quality Management & Business Excellence*, v. 32, n. 13-14, p. 1381-1400, 2021.
- DAFT, R. L.; LENGEL, R. H. Organizational Information Requirements, Media Richness and Structural Design. *Management Science*, v. 32, n. 5, p. 554-571, 1986.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, v. 74, n. 366a, p. 427-431, 1979.
- ENDERS, W. *Applied Econometric Time Series*. 4. ed. New York: John Wiley & Sons, 2014.
- FILDES, R.; MA, S.; KOLASSA, S. Retail forecasting: Research and practice. *International Journal of Forecasting*, v. 38, n. 4, p. 1283-1318, 2022.
- GALBRAITH, J. R. Organization Design: An Information Processing View. *Interfaces*, v. 4, n. 3, p. 28-36, 1974.
- GALBRAITH, J. R. Organization Design. Reading: Addison-Wesley, 1977.
- GARDNER, E. S. Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, v. 22, n. 4, p. 637-666, 2006.
- GARTNER. What Is Artificial Intelligence (AI). Disponível em: <<https://www.gartner.com/en/topics/artificial-intelligence>>. Acesso em: set. 2025.
- GHOSH, D.; VOGT, A. Outliers: An evaluation of methodologies. In: JOINT STATISTICAL MEETINGS, 2012, San Diego. Proceedings... Alexandria: American Statistical Association, 2012. p. 3455-3460.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. Cambridge: MIT Press, 2016.
- GRANGER, C. W. J.; NEWBOLD, P. Spurious regressions in econometrics. *Journal of Econometrics*, v. 2, n. 2, p. 111-120, 1974.
- HAIR, J. F. et al. Multivariate Data Analysis. 7. ed. Upper Saddle River: Prentice Hall, 2009.
- HAMILTON, J. D. Time Series Analysis. Princeton: Princeton University Press, 1994.

- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. New York: Springer, 2009.
- HAWKINS, D. M. *Identification of Outliers*. London: Chapman and Hall, 1980.
- HEVNER, A. R.; MARCH, S. T.; PARK, J.; RAM, S. Design Science in Information Systems Research. *MIS Quarterly*, v. 28, n. 1, p. 75-105, 2004.
- HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. *Neural Computation*, v. 9, n. 8, p. 1735-1780, 1997.
- HOFMANN, E.; RUTSCHMANN, E. Big data analytics and demand forecasting in supply chains: a conceptual analysis. *The International Journal of Logistics Management*, v. 29, n. 2, p. 739-766, 2018.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 2. ed. Melbourne: OTexts, 2018.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, v. 27, n. 3, p. 1-22, 2008.
- HYNDMAN, R. J.; KOEHLER, A. B.; ORD, J. K.; SNYDER, R. D. *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin: Springer, 2008.
- IVANOV, D.; DOLGUI, A. A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Production Planning & Control*, v. 32, n. 9, p. 775-788, 2021.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. *An Introduction to Statistical Learning: with Applications in R*. New York: Springer, 2013.
- JARQUE, C. M.; BERA, A. K. A Test for Normality of Observations and Regression Residuals. *International Statistical Review*, v. 55, n. 2, p. 163-172, 1987.
- KAHN, K. B. The impact of demand management on supply chain performance. *Journal of Business Logistics*, v. 24, n. 1, p. 1-20, 2003.
- KOLASSA, S.; ROSTAMI-TABAR, B.; SIEMSEN, E. *Demand Forecasting for Executives and Professionals*. Boca Raton: CRC Press, 2024.
- KOLKOVÁ, A.; NAVRÁTIL, M. Demand Forecasting in Python: Deep Learning Model Based on LSTM Architecture versus Statistical Models. *Acta Polytechnica Hungarica*, v. 18, n. 8, p. 123-141, 2021.
- KOLMOGOROV, A. N. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, v. 4, p. 83-91, 1933.

- KWIATKOWSKI, D.; PHILLIPS, P. C. B.; SCHMIDT, P.; SHIN, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, v. 54, n. 1-3, p. 159-178, 1992.
- LEE, H. L.; PADMANABHAN, V.; WHANG, S. Information distortion in a supply chain: The bullwhip effect. *Management Science*, v. 43, n. 4, p. 546-558, 1997.
- LIM, B.; ARIK, S. O.; LOEFF, N.; PFISTER, T. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. *International Journal of Forecasting*, v. 37, n. 4, p. 1748-1764, 2021.
- LIU, F. T.; TING, K. M.; ZHOU, Z. H. Isolation Forest. In: *IEEE INTERNATIONAL CONFERENCE ON DATA MINING*, 8., 2008, Pisa. Proceedings... Pisa: IEEE, 2008. p. 413-422.
- LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, v. 65, n. 2, p. 297-303, 1978.
- MACAULAY, F. R. *The Smoothing of Time Series*. New York: National Bureau of Economic Research, 1930.
- MAKRIDAKIS, S.; SPILLOTIS, E.; ASSIMAKOPOULOS, V. The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, v. 34, n. 4, p. 802-808, 2018.
- MAKRIDAKIS, S.; SPILLOTIS, E.; ASSIMAKOPOULOS, V. The M5 Accuracy competition: Results, findings and conclusions. *International Journal of Forecasting*, v. 36, n. 4, p. 1349-1364, 2020.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. *Forecasting: Methods and Applications*. 3. ed. New York: Wiley, 1998.
- MARCH, S. T.; SMITH, G. F. Design and natural science research on information technology. *Decision Support Systems*, v. 15, n. 4, p. 251-266, 1995.
- MASSEY, F. J. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, v. 46, n. 253, p. 68-78, 1951.
- MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. *Introduction to Time Series Analysis and Forecasting*. 2. ed. Hoboken: John Wiley & Sons, 2015.
- NEWHEY, W. K.; WEST, K. D. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, v. 55, n. 3, p. 703-708, 1987.

- ORESHKIN, B. N.; CARPOV, D.; CHAPADOS, N.; BENGIO, Y. N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR), 2020. Proceedings... 2020.
- PEFFERS, K.; TUUNANEN, T.; ROTHENBERGER, M. A.; CHATTERJEE, S. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, v. 24, n. 3, p. 45-77, 2007.
- PHILLIPS, P. C. B.; PERRON, P. Testing for a unit root in time series regression. *Biometrika*, v. 75, n. 2, p. 335-346, 1988.
- RAZALI, N. M.; WAH, Y. B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, v. 2, n. 1, p. 21-33, 2011.
- ROŽANEC, J. et al. Actionable cognitive twins for decision making in manufacturing. *International Journal of Production Economics*, v. 241, p. 108252, 2021.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3/4, p. 591-611, 1965.
- SHARMA, R.; SHISHODIA, A.; GUNASEKARAN, A. The role of artificial intelligence in supply chain management: mapping the territory. *International Journal of Production Research*, v. 60, n. 24, p. 7564-7591, 2022.
- SILVER, E. A.; PYKE, D. F.; PETERSON, R. *Inventory Management and Production Planning and Scheduling*. 4. ed. Hoboken: John Wiley & Sons, 2016.
- SRINIVASAN, R.; SWINK, M. Leveraging supply chain integration through planning comprehensiveness: An organizational information processing theory perspective. *Decision Sciences*, v. 46, n. 5, p. 823-861, 2015.
- UZSOY, R. et al. Rolling horizon, multi-product production planning with chance constraints and forecast evolution for wafer fabs. *International Journal of Production Research*, v. 56, n. 18, p. 6112-6134, 2018.
- YANG, Y.; ABD. RAHMAN, A.; ABDAN, K.; AZIZ ABDUL, Y.; LI, Y. The application of organizational information processing theory in supply chain management strategy research: A bibliometric review. *Corporate & Business Strategy Review*, v. 6, n. 1 (Special Issue), p. 378-391, 2025. DOI: 10.22495/cbsrv6i1siart14.
- YIN, R. K. *Case Study Research: Design and Methods*. 5. ed. Thousand Oaks: Sage, 2014.

## Apendice

### Resultados de todos os testes por item

Item	Modelo	MAPE	Tempo_Treino
A	Prophet	27.61646	120.0508
A	ETS	31.15595	0.231381
A	H-W s/ Trend	33.21133	0.017952
A	Holt-Winters Multiplicativo	33.41618	0.063829
A	TBATS	35.82466	127.9415
A	Random Forest	36.3631	32.34757
A	Holt-Winters Aditivo	36.48137	0.064827
A	N-BEATS	37.04203	59.98471
A	XGBoost	37.84699	20.46552
A	SARIMA	55.11998	0.385639
A	LightGBM	55.62847	10.095
B	Prophet	45.91068	119.8236
B	Holt-Winters Aditivo	47.81512	0.07679
B	ETS	51.84404	0.258404
B	Holt-Winters Multiplicativo	56.09227	0.067878
B	H-W s/ Trend	58.13908	0.021943
B	SARIMA	67.90665	0.566774
B	TBATS	74.35426	122.1303
B	N-BEATS	77.45307	43.13559
B	Random Forest	78.97987	109.3954
B	XGBoost	81.59653	24.49828
B	LightGBM	88.14408	11.8167
C	N-BEATS	27.70994	55.40091
C	ETS	28.18804	0.670641
C	Prophet	40.98833	153.2894
C	Holt-Winters Aditivo	51.13248	0.095464
C	TBATS	52.56336	114.4748
C	SARIMA	56.80234	2.653694
C	Random Forest	70.86907	119.9752
C	Holt-Winters Multiplicativo	73.68868	0.085553
C	XGBoost	82.39245	22.00214
C	H-W s/ Trend	83.44303	0.024962
C	LightGBM	92.58829	10.41121
D	Prophet	50.23982	139.6851
D	Holt-Winters Aditivo	50.32415	0.085742
D	Holt-Winters Multiplicativo	50.63842	0.093748
D	H-W s/ Trend	53.58031	0.018949
D	ETS	60.25541	0.448811
D	LightGBM	69.43826	12.74357
D	TBATS	69.85304	115.0513
D	Random Forest	73.95462	101.2436

D	N-BEATS	76.48307	57.25959
D	SARIMA	76.69556	0.46985
D	XGBoost	91.90481	23.37494
E	Holt-Winters Multiplicativo	36.31198	0.069563
E	N-BEATS	40.39935	50.21722
E	H-W s/ Trend	46.99323	0.025078
E	ETS	47.99617	0.352618
E	TBATS	48.65051	108.2335
E	Random Forest	71.52436	83.90885
E	Prophet	71.99372	146.1705
E	Holt-Winters Aditivo	74.98879	0.075003
E	XGBoost	87.12414	21.5316
E	LightGBM	91.18439	10.91911
E	SARIMA	92.54983	0.633729
F	Prophet	33.26389	71.08122
F	Random Forest	36.80285	79.66143
F	TBATS	37.29781	124.5756
F	H-W s/ Trend	39.43416	0.01895
F	SARIMA	39.69241	0.759727
F	XGBoost	41.27108	20.2304
F	N-BEATS	42.12697	50.7786
F	LightGBM	44.51757	10.71429
F	Holt-Winters Multiplicativo	45.44737	0.172538
F	ETS	52.37288	0.297482
F	Holt-Winters Aditivo	58.79722	0.060837