57º SBPO
SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
Gramado • RS
DE 5 A 9 DE OUTUBRO DE 2025

PREVISÃO HÍBRIDA DE DEMANDA NA INDÚSTRIA DA MODA COM PROPHET, TABPFN E FEATURES TEMPORAIS

Lauro Alexandre Godinho Sampaio
Universidade Federal Fluminense (UFF) - Rio das Ostras
https://rgc.uff.br/
laurosampaio@id.uff.br

Dalton Borges
Universidade Federal Fluminense (UFF) - Rio das Ostras
https://rgc.uff.br/
daltonborges@id.uff.br

Leonardo Alves da Silva
Universidade Federal Fluminense (UFF) - Rio das Ostras
https://rgc.uff.br/
leonardo_as@id.uff.br

Ray Brandão de Miranda
Universidade Federal Fluminense (UFF) - Rio das Ostras
https://rgc.uff.br/
raybm@id.uff.br

Diogo Cabral Zaror
Universidade Federal Fluminense (UFF) - Rio das Ostras
https://rgc.uff.br/
diogozaror@id.uff.br

Getúlio Binote Júnior
Universidade Federal Fluminense (UFF) - Rio das Ostras
https://rgc.uff.br/
getuliobinote@id.uff.br

RESUMO

Este estudo investiga a previsão de demanda no setor de moda *fast fashion* por meio de uma abordagem híbrida que combina os modelos Prophet e TabPFN, explorando variáveis temporais como lags, sazonalidades e feriados. Utilizando a metodologia CRISP-DM, foram analisados dados semanais de vendas e estoque de uma empresa varejista, com foco na categoria de maior volume. A modelagem considerou múltiplas abordagens, desde métodos de referência até técnicas avançadas de aprendizado de máquina. A avaliação foi conduzida com métricas padronizadas (RMSE, MAE e MAPE), evidenciando o desempenho superior do TabPFN, com incorporação efetiva de *features* temporais geradas pelo treino do modelo Prophet, em relação aos demais modelos. Os resultados destacam o potencial de modelos híbridos, especialmente quando enriquecidos por atributos temporais, na captura da dinâmica complexa e volátil da demanda no varejo de moda.

PALAVRAS CHAVE. Previsão de Demanda, Indústria da Moda, Modelos Preditivos.

Tópicos (Previsão de Demanda no Setor de Moda, Abordagem Híbrida de Modelagem, Modelos Preditivos Tradicionais e Avançados, Engenharia de Features Temporais com Prophet, Avaliação de Desempenho dos Modelos)

https://proceedings.science/p/212266?lang=pt-br

ABSTRACT

This study investigates demand forecasting in the fast fashion sector through a hybrid approach that combines the Prophet and TabPFN models, exploring temporal variables such as lags, seasonality and holidays. Using the CRISP-DM methodology, weekly sales and inventory data from a retail company were analyzed, focusing on the highest volume category. The modeling considered multiple approaches, from reference methods to advanced machine learning techniques. The evaluation was conducted with standardized metrics (RMSE, MAE and MAPE), evidencing the superior performance of TabPFN, with effective incorporation of temporal features generated by training the Prophet model, in relation to the other models. The results highlight the potential of hybrid models, especially when enriched by temporal attributes, in capturing the complex and volatile dynamics of demand in fashion retail.

KEYWORDS. Demand Forecasting, Fashion Industry, Predictive Models.

Paper topics (Demand Forecasting in the Fashion Sector, Hybrid Modeling Approach, Traditional and Advanced Predictive Models, Temporal Feature Engineering with Prophet, Model Performance Evaluation)

1. Introdução

A previsão de demanda é uma atividade estratégica essencial, que influencia decisões de compras, produção e distribuição em diversos setores. Cada indústria apresenta dinâmicas próprias, moldadas por variáveis macroeconômicas, comportamentais e culturais [Hyndman e Athanasopoulos, 2018; Chopra e Meindl, 2007]. No setor de moda, essas particularidades são ainda mais intensas devido à volatilidade das tendências e à curta vida útil dos produtos, exigindo previsões em tempo real que considerem fenômenos digitais e sociais [Fildes et al., 2022; Nenni et al., 2013].

Nesse contexto, observa-se um crescente uso de ferramentas avançadas — desde modelos estatísticos como o Prophet [Sean J. Taylor, 2017] até algoritmos de aprendizado de máquina como XGBoost e TabPFN [Hollmann et al., 2023] — capazes de capturar padrões complexos e não-lineares. Abordagens híbridas têm ganhado destaque na literatura por explorarem o potencial complementar dessas técnicas, especialmente com o uso de *features* temporais derivadas de séries históricas [Borges e Nascimento, 2022; Masini et al., 2023], contribuindo para melhorias significativas na acurácia preditiva.

O presente estudo tem como objetivo investigar uma abordagem híbrida que combina os modelos Prophet e TabPFN, enriquecida com *lags*, sazonalidades e feriados, para previsão de demanda no setor de *fast fashion*. A pesquisa compara o desempenho de modelos híbridos e individuais, quantifica os ganhos preditivos por meio de métricas como RMSE, MAE e MAPE, e avalia sua aplicabilidade prática com dados reais de uma varejista de moda, seguindo o framework CRISP-DM [Wirth e Hipp, 2000].

2. Previsão de Demanda na Indústria da Moda

A indústria da moda é marcada por alta oscilação, ciclos de vida curtos e forte sazonalidade, fatores que tornam a modelagem de séries temporais especialmente complexa [Fildes et al., 2022; Svetunkov e Boylan, 2020]. As preferências dos consumidores mudam rapidamente sob influência de redes sociais, influenciadores e eventos culturais, gerando picos de demanda que métodos tradicionais muitas vezes não conseguem antecipar.

Modelos como médias móveis, suavização exponencial e ARIMA oferecem uma base sólida para capturar tendências sazonais, mas apresentam limitações diante de mudanças abruptas — promoções relâmpago, viralizações ou lançamentos surpresa — comuns no *fast fashion* [Petropoulos et al., 2022; Huang et al., 2019].

Em resposta, cresce o uso de algoritmos de aprendizado de máquina e modelos híbridos, que combinam a robustez estatística com a flexibilidade das redes neurais. Essas abordagens permitem incorporar fontes não convencionais de dados — interações em mídias sociais, buscas online e métricas de engajamento — e ajustar-se em tempo real às variações do mercado [Masini et al., 2023; Tugay e Oguducu, 2020]. Ainda que demandem maior poder computacional, esses sistemas têm se mostrado mais eficazes na redução de estoques excedentes e na melhoria da experiência do cliente, assegurando às marcas uma vantagem competitiva sustentável.

3. Modelos de Previsão de Demanda

3.1. Modelos de linha de base

Os modelos *Naive*, média móvel, suavização exponencial, *Holt–Winters*, regressão linear e ARIMA são comumente adotados como linhas de base em estudos de *forecasting* com aprendizado de máquina, pois representam metodologias clássicas de previsão que capturam comportamentos elementares de séries temporais e estabelecem um patamar de comparação simples e interpretável [Zhang et al., 1998]. Assim, ao avaliar o desempenho de redes neurais ou outros algoritmos de Machine Learning, torna-se possível quantificar os ganhos obtidos e justificar a adoção de métodos mais sofisticados.

3.2. Prophet

O modelo Prophet foi desenvolvido pela equipe Core Data Science do Facebook (Meta) com o objetivo principal de facilitar a previsão acurada e rápida "em escala"para séries temporais de negócios que apresentam sazonalidade, sem exigir esforço manual extensivo, mas permitindo ajustes por analistas. Sua filosofia de design prioriza a usabilidade e a interpretabilidade, visando democratizar a capacidade de previsão em grandes organizações com numerosas e variadas métricas de negócio, implementando uma abordagem "analista-no-ciclo"(analyst-in-the-loop) [Sean J. Taylor, 2017].

A arquitetura fundamental do Prophet baseia-se em um modelo aditivo que decompõe a série temporal $y(t)$ em componentes distintos [Sean J. Taylor, 2017]:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

onde:

- $g(t)$: Representa a componente de tendência não periódica, modelada como uma função linear por partes ou logística, com detecção ou especificação manual de pontos de mudança (*changepoints*).

  $$g(t) = (k + a(t)^T\delta)t + (b + a(t)^T\gamma)$$

  onde, k é a taxa de crescimento, δ é a taxa de ajustes, b é parâmetro de offset e γ_j é definida como $-s_j\delta_j$ para que a função seja contínua.

- $s(t)$: Modela a sazonalidade periódica (e.g., anual, semanal), utilizando séries de Fourier. A fórmula da série de Fourier para a sazonalidade é dada aproximadamente por:

  $$s(t) = \sum_{n=-N}^{N} c_n e^{i\frac{2\pi nt}{P}}$$

  onde P é o período da sazonalidade (e.g., 52 para semanal) e N é o número de termos utilizados.

- $h(t)$: Captura o impacto de feriados e eventos especiais com datas potencialmente irregulares. É modelado como a soma dos efeitos individuais κ_i para cada feriado i, ponderados por uma variável indicadora $I(t \in D_i)$ que é 1 se t pertence ao conjunto de datas D_i para o feriado i:

$$h(t) = \sum_{i=1}^{L} \kappa_i \cdot I(t \in D_i)$$

- ϵ_t: Representa o termo de erro residual, assumido como seguindo uma distribuição normal.

3.3. XGBoost

O XGBoost (eXtreme Gradient Boosting) é um algoritmo de aprendizado de máquina supervisionado, reconhecido por sua alta performance e precisão, que constrói modelos preditivos a partir de um conjunto de árvores de decisão. Sua metodologia central é o reforço de gradiente (*gradient boosting*), onde o modelo é aprimorado progressivamente pela adição sequencial de árvores. A predição final para uma determinada entrada x_i, denotada por $\hat{p}_i$, é obtida pela soma das contribuições de K árvores de decisão (f_k), cada uma pertencente a um espaço funcional $\mathcal{F}$ que engloba todas as estruturas de árvores possíveis, conforme expresso na Equação 1.

$$\hat{p}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \tag{1}$$

O aprendizado no XGBoost é orientado pela otimização de uma função objetivo, descrita na Equação 2. Esta função busca um equilíbrio criterioso: por um lado, minimiza o erro de predição, quantificado por uma função de perda L (como o erro quadrático médio, que mede a discrepância entre os valores reais p_i e os preditos $\hat{p}_i$); por outro, penaliza a complexidade excessiva do modelo através de um termo de regularização Ω. Cada função f_k representa uma árvore de regressão que, baseada em sua estrutura $q(x)$, atribui uma pontuação w à folha correspondente à entrada x_i.

$$\text{Obj}(\phi) = \sum_{i=1}^{n} L(p_i, \hat{p}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

Para evitar o sobreajuste (*overfitting*) e promover modelos mais generalizáveis, o termo de regularização $\Omega(f_k)$ de cada árvore é fundamental, conforme detalhado na Equação 3. Este termo é composto por duas partes: γT, onde T é o número total de folhas na árvore e γ penaliza a criação de muitas folhas (controlando a complexidade estrutural); e $\frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$, onde w_j é a pontuação da j-ésima folha e λ é o coeficiente de regularização ℓ_2, que suaviza as pontuações das folhas, evitando que algumas poucas tenham um peso desproporcional.

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{3}$$

3.4. TabPFN

O TabPFN (Tabular Prior-Data Fitted Network) é um modelo baseado em Transformer pré-treinado capaz de realizar classificação supervisionada em pequenos conjuntos de dados tabulares de forma extremamente rápida, geralmente em menos de um segundo, sem ajuste de hiperparâmetros [Hollmann et al., 2023]. Utiliza aprendizado no contexto (*In-Context Learning* - ICL), prevendo a partir de exemplos rotulados $(x, f(x))$ fornecidos diretamente na entrada, sem atualização de parâmetros, e gera previsões para todo o conjunto de teste em uma única passagem direta (*forward pass*).

Como instância de uma Rede Ajustada a Dados da Priori (*Prior-Data Fitted Network* - PFN), o TabPFN é treinado uma única vez, offline, para aproximar a inferência Bayesiana em dados sintéticos gerados a partir de uma distribuição a priori [Hollmann et al., 2023]. Essa priori cobre um amplo espaço de modelos causais estruturais, favorecendo estruturas mais simples para garantir boa capacidade de generalização [Hollmann et al., 2023].

No arcabouço Bayesiano para aprendizado supervisionado, a Distribuição Preditiva Posterior (Posterior Predictive Distribution - PPD) para uma amostra de teste x_{test}, $p(\cdot|x_{test}, D_{train})$, é condicionada ao conjunto de amostras de treinamento $D_{train} := \{(x_1, y_1), ..., (x_n, y_n)\}$ [Hollmann et al., 2023]. A PPD pode ser obtida pela integração sobre o espaço de hipóteses Φ, conforme a Equação 4:

$$p(y|x, D) \propto \int_{\Phi} p(y|x, \phi)p(D|\phi)p(\phi)d\phi. \tag{4}$$

Nesta equação, $p(y|x, \phi)$ é a verossimilhança da saída y dado x e uma hipótese ϕ, $p(D|\phi)$ é a verossimilhança dos dados D dada a hipótese ϕ, e $p(\phi)$ é a probabilidade a priori da hipótese ϕ [Hollmann et al., 2023].

O treinamento de um PFN, denominado ajuste à priori sintética (synthetic prior-fitting), visa aproximar a PPD. O PFN é otimizado para fazer previsões para um subconjunto de dados $D_{test} \subset D$ condicionado ao restante do conjunto de dados $D_{train} = D \setminus D_{test}$ [Hollmann et al., 2023]. A função de perda para o treinamento do PFN, para um único ponto de teste $\{(x_{test}, y_{test})\} = D_{test}$, é dada pela Equação 5:

$$L_{PFN} = \mathbb{E}_{((x_{test},y_{test}),D_{train})\sim p(D)}[-\log q_\theta(y_{test}|x_{test}, D_{train})]. \tag{5}$$

Minimizar esta perda aproxima a verdadeira distribuição preditiva posterior Bayesiana [Hollmann et al., 2023].

4. Métricas de Desempenho

As métricas de desempenho são instrumentos para quantificar propriedades específicas de um algoritmo, abrangendo sua implementação e aplicação prática [Oblizanov et al., 2023]. Dessa forma, modelos de Inteligência Artificial são comumente avaliados por meio de métricas como o Erro Quadrático Médio (RMSE), o Erro Absoluto Médio (MAE) e o Erro Percentual Absoluto Médio (MAPE).

O RMSE é uma métrica comum para avaliação de modelos de previsão de demanda que calculam a raiz quadrada da média dos quadrados dos erros (diferenças entre os valores previstos e os reais). A métrica é descrita pela Equação 6

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \tag{6}$$

Em que X_i são os valores previstos, Y_i são os valores reais e m é o número total de observações.

O MAE, Equação 7, calcula a média das diferenças absolutas entre os valores previstos e reais, proporcionando uma métrica de erro mais robusta em relação a outliers.

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}|X_i - Y_i| \tag{7}$$

Onde:

- $|X_i - Y_i|$ é a diferença absoluta entre o valor previsto e o valor real.

O MAPE, Equação 8, calcula a média dos erros absolutos em termos percentuais, sendo útil quando é importante analisar variações relativas em vez de variações absolutas:

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{Y_i - X_i}{Y_i} \right| \tag{8}$$

Onde:

- $\left| \frac{Y_i - X_i}{Y_i} \right|$ é a diferença percentual entre o valor real e o valor previsto.

5. Materiais e métodos

Para direcionar o desenvolvimento dos modelos de previsão de demanda, adotou-se o *CRoss Industry Standard Process for Data Mining* (CRISP-DM). Conforme preconizado por Wirth e Hipp [2000], esta metodologia oferece uma estrutura iterativa e flexível, abrangendo fases desde a compreensão do negócio e dos dados, passando pela preparação, modelagem, avaliação, até a implantação, ajustando-se bem a diferentes contextos. O foco deste estudo comparou duas abordagens de treinamento para os modelos classicos a mais atuais, partindo para uma estratégia *stand-alone*, com os algoritmos ajustados diretamente sobre as variáveis originais, e outra utilizando *feature lags*.

A estratégia com *feature lags* foi implementada em duas etapas: primeiramente, um modelo Prophet foi ajustado para capturar tendências e sazonalidades da série temporal; em seguida, os resíduos desse treinamento foram extraídos e utilizados para gerar atrasos (*lags*), que enriqueceram o conjunto de variáveis explicativas dos modelos XGBoost e TabPFN. Esta abordagem inspira-se na proposta de Borges e Nascimento (2022), que evidenciaram melhorias de acurácia ao integrar resíduos do Prophet como *feature lags* em um modelo híbrido com LSTM. A lógica subjacente é que o Prophet modela componentes de longo prazo, enquanto seus resíduos contêm variações de curto prazo que podem ser exploradas por técnicas de aprendizado de máquina para refinar o desempenho preditivo [Borges e Nascimento, 2022].

Ambas as estratégias de modelagem foram avaliadas com base nas métricas RMSE, MAE e MAPE, visando quantificar o impacto da inclusão dos *lags*. Embora a fase de implantação seja relevante em aplicações operacionais, ela extrapola o escopo deste trabalho, que se concentra na investigação e comparação de métodos preditivos para a demanda de produtos de uma empresa de *fast fashion*.

5.1. Compreensão do Negócio

A empresa analisada atua no dinâmico setor de *fast fashion*, oferecendo ampla variedade de vestuário para diferentes públicos e faixas etárias, com constante renovação de coleções e forte alinhamento às tendências, o que demanda planejamento ágil de estoques, compras e promoções diante da sazonalidade, volatilidade da demanda e comportamento do consumidor. Localizada em região de clima predominantemente quente, mantém elevada demanda por roupas leves e casuais durante todo o ano, enquanto eventos locais e feriados prolongados influenciam significativamente o fluxo de clientes e as vendas, especialmente na alta temporada, exigindo estratégias adaptadas a essas variações.

5.2. Compreensão dos Dados

Este estudo baseia-se em dados operacionais de uma empresa varejista especializada em moda unissex, com foco em roupas e acessórios. O conjunto contempla registros semanais de vendas e informações de estoque, organizados por categoria de produto, grupo de consumidores e data de início da semana comercial. Entre as variáveis disponíveis, destacam-se a categoria dos

itens, o público-alvo, o volume de peças vendidas, o faturamento obtido, além dos níveis de estoque — tanto disponíveis quanto em trânsito.

A estrutura detalhada desses registros proporciona uma visão granular do desempenho por categoria e grupo consumidor, possibilitando a identificação de padrões sazonais, tendências de consumo e variações na demanda ao longo do tempo. Essa riqueza informacional torna o conjunto uma base sólida para análises voltadas à melhoria dos processos logísticos e à tomada de decisão baseada em dados no contexto do varejo de moda.

5.3. Preparação dos Dados

A etapa de preparação teve início com a leitura de dois conjuntos principais: o primeiro, com dados semanais de vendas e faturamento; o segundo, composto por um calendário de eventos e datas comemorativas entre os anos de 2018 e 2023. As datas foram padronizadas e organizadas cronologicamente para garantir consistência nas análises temporais subsequentes.

Para garantir que a modelagem fosse condizente com cenários futuros, foram excluídas variáveis que não estariam disponíveis no momento da previsão, como os dados de estoque. Além disso, valores inconsistentes relacionados ao faturamento foram tratados: inicialmente marcados como ausentes, esses dados foram corrigidos, sempre que possível, com base em médias de preços observadas em semanas próximas e para produtos de características semelhantes.

Durante a engenharia de atributos, foram criadas variáveis que representassem o comportamento recente das vendas, permitindo ao modelo captar tendências e mudanças de padrão. Com o auxílio do calendário comemorativo, foram adicionados marcadores indicando semanas com feriados ou eventos locais, o que possibilita ao modelo considerar possíveis impulsos sazonais de demanda. Informações temporais como o número da semana e o mês também foram extraídas e incluídas como variáveis explicativas.

Como foco da análise, foi selecionada a categoria de produto com maior volume de vendas, garantindo representatividade nos resultados. Por fim, as quatro semanas mais recentes da série foram reservadas para a avaliação dos modelos preditivos, mantendo a integridade da sequência temporal dos dados.

5.4. Modelagem

A etapa de modelagem foi cuidadosamente estruturada para permitir uma comparação aprofundada entre três vertentes principais de previsão de demanda. O foco recaiu sobre os modelos Prophet, XGBoost e TabPFN, avaliados tanto em suas configurações isoladas (*standalone*) quanto enriquecidos com *feature lags* derivados do modelo Prophet. Inicialmente, o conjunto de dados passou por um processo de preparação, que incluiu a correção de valores problemáticos de vendas e a concentração da análise na categoria de maior volume: *Manga Curta Masculina*.

O modelo Prophet foi configurado para operar com crescimento linear e sazonalidade semanal, além de incorporar os efeitos de feriados locais. Seu treinamento utilizou toda a série histórica, com exceção das últimas quatro semanas, que foram reservadas para a fase de teste. O XGBoost, por sua vez, foi avaliado sob duas configurações distintas: uma utilizando apenas variáveis temporais (sem a inclusão de *lags*) e outra incorporando *lags* de vendas recentes. Esta segunda abordagem permitiu capturar dinâmicas de curto prazo que não são modeladas explicitamente pelo Prophet. O TabPFN, um modelo de última geração baseado na arquitetura de transformadores para dados tabulares, também foi testado mediante duas abordagens: com e sem a inclusão de *feature lags* das vendas, sempre respeitando a limitação referente ao tamanho do conjunto de dados imposta pelo próprio modelo.

Modelos de referência clássicos, tais como Naive, Média Móvel, Suavização Exponencial, Holt-Winters, Regressão Linear e ARIMA, foram implementados com o propósito de estabelecer um parâmetro comparativo, embora não constituíssem o foco central desta análise. Em algumas

instâncias, tais modelos apresentaram métricas competitivas; contudo, o objetivo principal residia em avaliar o potencial dos métodos baseados em aprendizado de máquina e *deep learning* para capturar padrões complexos inerentes à demanda.

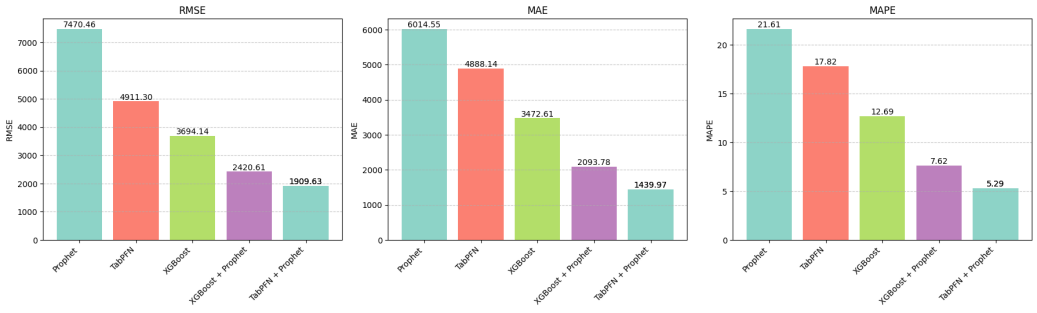
5.5. Avaliação

Um espectro de modelos foram desenvolvidos e submetidos à avaliação. Este conjunto abrangeu desde métodos de referência, como o Naive e a Média Móvel, até técnicas de aprendizado de máquina mais sofisticadas. Destacaram-se, neste último grupo, o Prophet, o XGBoost (em suas variantes *standalone* e com *lags* do Prophet) e o TabPFN (igualmente testado em versões *standalone* e com *lags* do Prophet). O desempenho de cada um dos modelos foi avaliado por meio de três métricas padrão no campo da previsão: RMSE (Raiz do Erro Quadrático Médio), MAE (Erro Médio Absoluto) e MAPE (Erro Percentual Médio Absoluto). Esta abordagem permitiu uma compreensão mais abrangente acerca das potencialidades e limitações de cada modelo no contexto específico do problema em estudo.

6. Resultados e Discussão

Ao analisar o desempenho dos diferentes modelos na tarefa de prever as vendas, com especial atenção ao conjunto de teste de quatro semanas para a categoria *Manga Curta Masculina*, observamos variações significativas em suas respectivas capacidades de antecipação. Os resultados, que são apresentados de forma visual na Figura 1 e detalhados numericamente na Tabela 1, indicam que a combinação do modelo TabPFN com as *features* de *lag* provenientes do Prophet (denominada "TabPFN + Prophet") alcançou o melhor desempenho geral. Em uma posição muito próxima, a versão do XGBoost enriquecida com essas mesmas *features* ("XGBoost + Prophet") também demonstrou um destaque considerável.

Figura 1: Comparativo de métricas por modelo em evidência



6.1. Análise Detalhada dos Resultados

Uma avaliação mais aprofundada dos números apresentados revela que o modelo TabPFN, quando potencializado com as *features* de *lag* derivadas do Prophet, exibiu um desempenho superior de forma consistente em todas as métricas de erro utilizadas. Especificamente, a configuração "TabPFN + Prophet"alcançou um RMSE de 1909.628, um MAE de 1439.969 e um MAPE de apenas 5.291%. Estes valores são significativamente inferiores aos obtidos pelos demais modelos testados. O modelo "XGBoost + Prophet", que se posicionou como o segundo melhor em termos de performance, apresentou um RMSE de 2420.606, MAE de 2093.785 e MAPE de 7.620%. Embora demonstre uma performance robusta e superior à maioria das alternativas, a vantagem do "TabPFN + Prophet"permanece evidente.

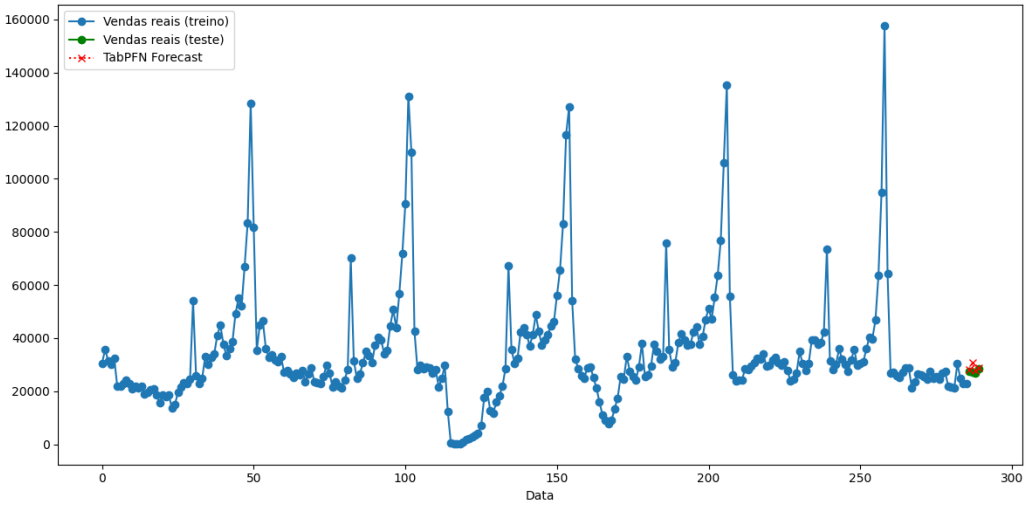
Em contrapartida, as versões *standalone* destes mesmos modelos apresentaram resultados inferiores. O XGBoost isolado registrou um RMSE de 3694.144 e um MAPE de 12.693%,

Tabela 1: Tabela Comparativa de Métricas

Modelo	RMSE	MAE	MAPE
ARIMA	4786.219	4698.154	17.170%
Holt-Winters (Dupla)	5401.044	4835.801	17.494%
Holt-Winters (Tripla)	3754.158	3355.133	12.195%
Média Móvel	2940.561	2869.188	10.427%
Naive	6964.868	6937.530	25.327%
Prophet	7470.464	6014.549	21.610%
Reg. Linear	9892.503	9875.457	36.030%
Suavização Exp.	4503.129	4154.538	15.044%
TabPFN	4911.297	4888.136	17.823%
TabPFN + Prophet	**1909.628**	**1439.969**	**5.291%**
XGBoost	3694.144	3472.607	12.693%
XGBoost + Prophet	2420.606	2093.785	7.620%

enquanto o TabPFN isolado obteve um RMSE de 4911.297 e um MAPE de 17.823%. Esta comparação sublinha o benefício da incorporação das *features* de *lag* do Prophet. Para contextualizar esses resultados, modelos de base como a Média Móvel (RMSE de 2940.561, MAPE de 10.427%) e o Holt-Winters (Tripla) (RMSE de 3754.158, MAPE de 12.195%) superaram as versões *standalone* do XGBoost e TabPFN em algumas métricas, o que evidencia a complexidade da tarefa de previsão e a importância crucial da engenharia de *features*. O Prophet, quando utilizado isoladamente, alcançou um RMSE de 7470.464 e um MAPE de 21.610%.

Figura 2: Análise modelo TabPFN com *features de lag*



A Figura 2 oferece uma representação visual que permite comparar as previsões geradas pelo modelo "TabPFN + Prophet"com os valores reais observados durante o período de teste. Nota-se que o modelo "TabPFN + Prophet"demonstrou capacidade de acompanhar a tendência geral dos dados, resultando em erros relativamente pequenos quando comparados aos valores verdadeiros. Um aspecto que merece destaque é a habilidade do modelo em capturar a direção das variações na demanda, ainda que possam existir algumas discrepâncias nos valores exatos previstos. A fi-

gura ilustra, portanto, a aderência do "TabPFN + Prophet"aos dados reais, evidenciando seu bom desempenho na tarefa de previsão.

6.2. Comparação entre Modelos e Implicações Práticas

A análise comparativa entre os modelos testados indicou que abordagens tradicionais, como o Naive (MAPE: 25.327%) e o ARIMA (MAPE: 17.170%), foram significativamente superadas por modelos de aprendizado de máquina avançados, especialmente aqueles que incorporaram *features* de *lag*. Isso sugere que métodos mais simples têm dificuldade em capturar a complexidade e as flutuações rápidas do dinâmico mercado de varejo de moda. O Prophet em sua versão isolada (MAPE 21.610%) não atingiu o desempenho esperado, ficando aquém de modelos de base mais simples em termos de MAPE e exibindo uma tendência a superestimar consideravelmente a demanda, como observado em erros de previsão que chegaram a aproximadamente 47.9% para semanas específicas.

Em contraste, as configurações híbridas demonstraram maior robustez e precisão. O modelo "XGBoost + Prophet", por exemplo, apresentou erros mais equilibrados e consistentes, como um desvio de apenas -5.8% em uma semana de teste particular. O "TabPFN + Prophet"destacou-se ainda mais, alcançando o menor MAPE (5.291%) e, consequentemente, previsões mais próximas da realidade. Essa acurácia e consistência são de imenso valor para o planejamento de estoques e compras, onde a previsibilidade dos erros é fundamental. As versões *standalone* do XGBoost (MAPE 12.693%) e do TabPFN (MAPE 17.823%) apresentaram um desempenho intermediário em relação às suas contrapartes híbridas e ao Prophet isolado.

6.3. Fatores Contribuintes para o Desempenho Superior do TabPFN e XGBoost

O desempenho observado nas abordagens "TabPFN + Prophet"e "XGBoost + Prophet"pode ser atribuído a uma combinação de fatores inter-relacionados. Primeiramente, a capacidade do TabPFN de aprendizado com um volume reduzido de dados (*few-shot learning*) é notável. Sendo pré-treinado em uma vasta gama de conjuntos de dados tabulares sintéticos, ele exibe uma generalização eficaz e consegue realizar previsões acuradas mesmo com históricos de dados relativamente curtos. Esta é uma característica particularmente valiosa no volátil setor de moda, onde os padrões de consumo podem mudar rapidamente. O TabPFN aprende a estrutura dos dados de forma eficiente, sem a necessidade de um ajuste extensivo de hiperparâmetros para cada novo conjunto de dados.

No caso do XGBoost, a incorporação efetiva de *features* temporais, incluindo as defasagens significativas provenientes do Prophet, mostrou-se crucial. O XGBoost foi alimentado com um conjunto diversificado de características, como valores passados da série temporal (*lags* do Prophet e outros *lags* intrínsecos), indicadores de feriados e componentes de sazonalidade. Esta estratégia permitiu ao modelo identificar padrões complexos e relações não lineares que modelos mais simples, ou mesmo o Prophet utilizado isoladamente, não conseguiriam detectar com a mesma eficácia.

Ademais, a robustez a *outliers* e ruídos, uma característica inerente ao XGBoost e, implicitamente, ao TabPFN (devido ao seu treinamento em diversas distribuições de dados), contribui para o desempenho superior. Séries de vendas no varejo frequentemente contêm valores discrepantes devido a promoções, eventos especiais não recorrentes e outros fatores externos. A capacidade de lidar com essas variações sem comprometer significativamente a qualidade da previsão é uma vantagem importante. A sinergia estabelecida entre as *features* de *lag* do Prophet e a capacidade de modelagem avançada do XGBoost e TabPFN parece ser um fator determinante para a superioridade observada nos resultados.

6.4. Implicações para a Gestão de Estoque e Planejamento de Compras

Os resultados deste estudo têm consequências diretas para a gestão operacional no varejo de moda. A utilização de modelos de alta performance, como o "TabPFN + Prophet"(MAPE

de 5.291%) e o "XGBoost + Prophet"(MAPE de 7.620%), proporciona previsões consideravelmente mais exatas do que métodos tradicionais ou as versões isoladas dos algoritmos. Essa precisão aprimorada reflete-se diretamente na redução da necessidade de altos estoques de segurança, diminuindo custos de armazenagem e capital imobilizado, além de minimizar rupturas e perdas de vendas. Adicionalmente, previsões mais acuradas viabilizam um planejamento de compras e produção mais eficiente, crucial para a agilidade do *fast fashion*, e permitem extrair *insights* sobre os fatores de demanda para otimizar estratégias promocionais.

A superioridade preditiva alcançada pela combinação de modelos ou pelo uso de características bem elaboradas (*features*) é um ponto de destaque. De maneira análoga ao estudo de Borges e Nascimento [2022], que observou melhores resultados com uma abordagem híbrida Prophet-LSTM, nosso trabalho demonstrou que o desempenho superior do "TabPFN + Prophet"— enriquecido com *features* de *lag* do Prophet e outros dados contextuais — em comparação com as versões *standalone* e o Prophet isolado, sublinha o potencial de abordagens sofisticadas. Essas estratégias integrativas são mais eficazes para capturar as complexas dinâmicas das séries temporais de vendas, reforçando a ideia de que a sinergia entre modelos é uma via promissora para previsões mais robustas.

7. Conclusão

Este estudo concentrou-se na previsão de demanda para a categoria de produto *Manga Curta Masculina* em uma empresa de *fast fashion*, utilizando um horizonte de quatro semanas. Ao avaliar uma diversidade de modelos, desde abordagens mais simples até técnicas avançadas como ARIMA, Prophet, XGBoost e TabPFN, tanto isoladamente quanto em configurações híbridas com *features* de *lag* do Prophet, os resultados foram reveladores. Conforme detalhado na Tabela 1, a combinação "TabPFN + Prophet"demonstrou um desempenho superior em precisão. Essa eficácia decorre da capacidade desses modelos mais sofisticados de integrar as *features* de *lag* e outras variáveis pertinentes, permitindo capturar a complexa dinâmica da demanda no volátil setor de *fast fashion*.

A análise aprofundada indicou que, enquanto o Prophet utilizado de forma isolada tendeu a superestimar a demanda no cenário específico deste estudo, as abordagens híbridas "TabPFN + Prophet"ofereceram previsões mais equilibradas e consistentes. Essa estabilidade é de grande valor para o planejamento operacional diário, onde erros menores e mais previsíveis são preferíveis. As implicações práticas desses achados são consideráveis, pois uma maior acurácia nas previsões pode levar à otimização de estoques, redução de rupturas, planejamento de compras mais eficiente e melhores estratégias de precificação, elementos cruciais para o sucesso no *fast fashion*. O êxito dessas variantes híbridas em nosso estudo reforça a ideia, também explorada por Borges e Nascimento [2022], de que a sinergia entre diferentes modelos ou componentes pode resultar em previsões mais apuradas.

Para trabalhos futuros, vislumbra-se a continuidade na exploração de sinergias entre modelos, como a otimização mais detalhada de hiperparâmetros para as configurações híbridas e a incorporação de fontes de dados externas — por exemplo, tendências de mídias sociais ou indicadores macroeconômicos — que possam influenciar a demanda. Além disso, propõe-se investigar a aplicabilidade da engenharia de *features* de *lag* nos modelos clássicos de séries temporais, especialmente os baseados na metodologia de Box & Jenkins (ARIMA - Box e Jenkins [1970]), a fim de avaliar se tais transformações podem ampliar sua capacidade preditiva. A busca incessante por aprimoramento nos modelos de previsão é vital para otimizar a gestão de estoques, reduzir perdas e, consequentemente, ampliar a competitividade da empresa no dinâmico mercado da moda.

Agradecimentos

Este trabalho foi apoiado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) Bolsa de Pesquisa **408313/2023-4**.

Referências

Borges, D. e Nascimento, M. C. (2022). Covid-19 icu demand forecasting: A two-stage prophet-lstm approach. *Applied Soft Computing*, 125:109181. ISSN 1568-4946. URL https://www.sciencedirect.com/science/article/pii/S1568494622004276.

Box, G. E. P. e Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Chopra, S. e Meindl, P. (2007). *Supply chain management. Strategy, planning & operation*. Springer.

Fildes, R., Ma, S., e Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, 38(4):1283–1318.

Hollmann, N., Müller, S., Eggensperger, K., e Hutter, F. (2023). Tabpfn: A transformer that solves small tabular classification problems in a second. URL https://arxiv.org/abs/2207.01848.

Huang, T., Fildes, R., e Soopramanien, D. (2019). Forecasting retailer product sales in the presence of structural change. *European Journal of Operational Research*, 279(2):459–470.

Hyndman, R. J. e Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Masini, R. P., Medeiros, M. C., e Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of economic surveys*, 37(1):76–111.

Nenni, M., Giustiniano, L., e Pirolo, L. (2013). Demand forecasting in the fashion industry: A review. *International Journal of Engineering Business Management*, 5.

Oblizanov, A., Shevskaya, N., Kazak, A., Rudenko, M., e Dorofeeva, A. (2023). Evaluation metrics research for explainable artificial intelligence global methods using synthetic data. *Applied System Innovation*, 6(1):26.

Petropoulos, F., Apiletti, D., e Assimakopoulos (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3):705–871. ISSN 0169-2070. URL http://dx.doi.org/10.1016/j.ijforecast.2021.11.001.

Sean J. Taylor, B. (2017). Forecasting at scale. https://facebook.github.io/prophet/static/prophet_paper_20170113.pdf. Acessado em 2025-05-12.

Svetunkov, I. e Boylan, J. E. (2020). State-space arima for supply-chain forecasting. *International Journal of Production Research*, 58(3):818–827.

Tugay, R. e Oguducu, S. G. (2020). Demand prediction using machine learning methods and stacked generalization. *arXiv preprint arXiv:2009.09756*.

Wirth, R. e Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining (PADD 2000)*, p. 29–39, Manchester, UK. The Practical Application Company Ltd.

Zhang, G., Patuwo, B. E., e Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1):35–62.

https://proceedings.science/p/212266?lang=pt-br