

**ANÁLISE E QUANTIFICAÇÃO DA CANIBALIZAÇÃO EM
PORTFÓLIOS DE VAREJO APÓS IMPLEMENTAÇÃO DE NOVOS
PRODUTOS: ESTUDO DE CASO NO SEGMENTO DE BEBIDAS
ENERGÉTICAS**

**ANALYSIS AND QUANTIFICATION OF CANNIBALIZATION IN
RETAIL PORTFOLIOS FOLLOWING NEW PRODUCT
IMPLEMENTATION: A COMPREHENSIVE CASE STUDY IN THE
ENERGY DRINKS SEGMENT**

**ANÁLISIS Y CUANTIFICACIÓN DE LA CANIBALIZACIÓN EN
PORTAFOLIOS MINORISTAS TRAS LA IMPLEMENTACIÓN DE
NUEVOS PRODUCTOS: UN ESTUDIO DE CASO EN EL SEGMENTO
DE BEBIDAS ENERGÉTICAS**

Rafael Denis de Lima Rodrigues¹
Adriano Maniçoba da Silva²

Artigo recebido em janeiro de 2026
Artigo aceito em abril de 2026

DOI: 10.26853/Refas_ISSN-2359-182X_v12n04_02

RESUMO

Este artigo aborda os efeitos para o lançamento de novos produtos, em que constitui uma estratégia imperativa para um crescimento sustentável e competitivo das empresas, especialmente no dinâmico e saturado mercado de bens de consumo rápido, com destaque para o segmento de bebidas energéticas. No entanto, o motor da inovação é frequentemente acompanhado pelo risco do fenômeno da canibalização, caracterizado pela transferência da demanda de um produto estabelecido (vítima) para um novo produto (canibal) dentro do mesmo portfólio. Este estudo aprofundado tem como foco a análise e a quantificação precisa do efeito da canibalização após a introdução de novos produtos em portfólios de varejo, propondo um framework metodológico robusto e em três estágios. A metodologia proposta integra (1) o arcabouço teórico de desempenho de marca (modelo nbd-dirichlet) para diferenciar a canibalização esperada da excessiva; (2) a análise econométrica da elasticidade de preço cruzada, utilizando regressões log-dinâmicas para detecção precisa e quantificação inicial da substituição; e (3) a aplicação de um *xgboost-based forecasting framework* de três estágios, otimizado por uma função objetivo personalizada (*custom objective function*), desenhada especificamente para isolar, medir e projetar o impacto de canibalização gerado pelas novas ofertas. Os resultados demonstram que a aplicação desta estrutura analítica híbrida fornece insights estratégicos cruciais para a gestão de portfólios, permitindo que os varejistas e fabricantes mitiguem o impacto negativo e desenvolvam estratégias de precificação e mix de produtos que capitalizem a demanda de todo o portfólio.

Palavras-chave: Machine Learning; Xgboost; Modelo Dirichlet; Elasticidade de preço Cruzada; Canibalização de produtos, Vendas orgânicas.

¹ Engenheiro de Produção, IFSP - Campus Suzano. E-mail: rdlr@outlook.com.br.
Lattes: <http://lattes.cnpq.br/3257636365569752>. OrcId: 0009-0008-1805-1945.

² Doutor em Administração, IFSP - Campus Suzano. E-mail: adrianoms@ifsp.edu.br.
Lattes: <http://lattes.cnpq.br/6644799917139100>. OrcId: 0000-0002-0094-404X.

ABSTRACT

This article discusses the effects of launching new products, which constitutes an imperative strategy for sustainable and competitive growth of companies, especially in the dynamic and saturated market of fast-moving consumer goods, with emphasis on the energy drink segment. However, the engine of innovation is often accompanied by the risk of cannibalization, characterized by the transfer of demand from an established product (victim) to a new product (cannibal) within the same portfolio. This in-depth study focuses on the analysis and precise quantification of the cannibalization effect after the introduction of new products in retail portfolios, proposing a robust three-stage methodological framework. The proposed methodology integrates (1) the theoretical framework of brand performance (nbd-Dirichlet model) to differentiate expected from excessive cannibalization; (2) the econometric analysis of cross-price elasticity, using log-dynamic regressions for precise detection and initial quantification of substitution; and (3) the application of a three-stage xgboost-based forecasting framework, optimized by a custom objective function, specifically designed to isolate, measure, and project the cannibalization impact generated by new offerings. The results demonstrate that the application of this hybrid analytical framework provides crucial strategic insights for portfolio management, allowing retailers and manufacturers to mitigate the negative impact and develop pricing and product mix strategies that capitalize on demand across the entire portfolio.

Keywords: Machine Learning; Xgboost; Dirichlet Model; Cross price elasticity; Product cannibalization; Organic sales.

RESUMEN

Este artículo analiza los efectos del lanzamiento de nuevos productos, lo cual constituye una estrategia imperativa para el crecimiento sostenible y competitivo de las empresas, especialmente en el dinámico y saturado mercado de bienes de consumo de rápida rotación, con énfasis en el segmento de bebidas energéticas. Sin embargo, el motor de la innovación a menudo está acompañado por el riesgo de canibalización, caracterizado por la transferencia de demanda de un producto establecido (víctima) a un nuevo producto (caníbal) dentro del mismo portafolio. Este estudio en profundidad se centra en el análisis y la cuantificación precisa del efecto de canibalización después de la introducción de nuevos productos en portafolios minoristas, proponiendo un marco metodológico robusto de tres etapas. La metodología propuesta integra (1) el marco teórico del desempeño de marca (modelo nbd-Dirichlet) para diferenciar la canibalización esperada de la excesiva; (2) el análisis econométrico de la elasticidad cruzada de precios, utilizando regresiones log-dinámicas para la detección precisa y la cuantificación inicial de la sustitución; y (3) la aplicación de un marco de pronóstico de tres etapas basado en xgboost, optimizado mediante una función objetivo personalizada, diseñado específicamente para aislar, medir y proyectar el impacto de la canibalización generada por las nuevas ofertas. Los resultados demuestran que la aplicación de este marco analítico híbrido proporciona información estratégica crucial para la gestión de la cartera, permitiendo a minoristas y fabricantes mitigar el impacto negativo y desarrollar estrategias de precios y combinación de productos que capitalicen la demanda en toda la cartera.

Palabras clave: Aprendizaje automático; Xgboost; Modelo de Dirichlet; Elasticidad cruzada de precios; Canibalización de productos; Ventas orgánicas.

1 INTRODUÇÃO

No contexto altamente competitivo do varejo moderno, a introdução contínua de novos produtos é essencial para a manutenção da relevância mercadológica, o aumento da participação de mercado (*market share*) e o crescimento da receita. No setor de Bens de Consumo Rápido, particularmente no crescente e dinâmico segmento de bebidas energéticas, as empresas

investem intensamente em inovações de sabor, embalagem e posicionamento de marca para atender às novas tendências de consumo. Contudo, a inovação de portfólio traz consigo o desafio intrínseco da canibalização. Este fenômeno ocorre quando o aumento nas vendas de um novo produto (o "canibal") não representa um crescimento líquido para a empresa, mas sim uma transferência de vendas de um produto já existente (a "vítima") no mesmo portfólio. Em um ambiente onde as margens são apertadas e o gerenciamento de estoque é crucial, a canibalização não planejada pode levar à sub otimização do portfólio, perdas de lucratividade e dificuldades logísticas.

A gestão estratégica de portfólios exige, portanto, a capacidade de analisar e quantificar de forma precisa o impacto da canibalização no momento da implementação de novos produtos. Métodos tradicionais de análise de vendas e previsão de demanda frequentemente falham em isolar o efeito da substituição de produto em meio a variáveis como promoções, sazonalidade e ações da concorrência, levando a avaliações imprecisas do retorno sobre o investimento (ROI) das novas ofertas.

Desta forma, o objetivo central deste estudo é desenvolver e aplicar uma análise robusta para a quantificação da canibalização gerada pela introdução de novos produtos no portfólio de um varejista, utilizando o segmento de bebidas energéticas como estudo de caso compreensivo. Propõe-se uma abordagem híbrida, que combina a fundamentação teórica sobre desempenho de marca com técnicas avançadas de Machine Learning e econometria. Ao quantificar a taxa de canibalização, este estudo visa fornecer ferramentas necessárias para que gestores de portfólio possam planejar estrategicamente o lançamento de produtos, diferenciando a canibalização aceitável e intencional (que visa migrar clientes para uma oferta mais lucrativa) daquela que é prejudicial à saúde financeira da empresa.

2 REFERENCIAL TEÓRICO

Inicia-se o referencial teórico pela canibalização de produtos.

2.1 Canibalização de produtos na gestão de portfólio

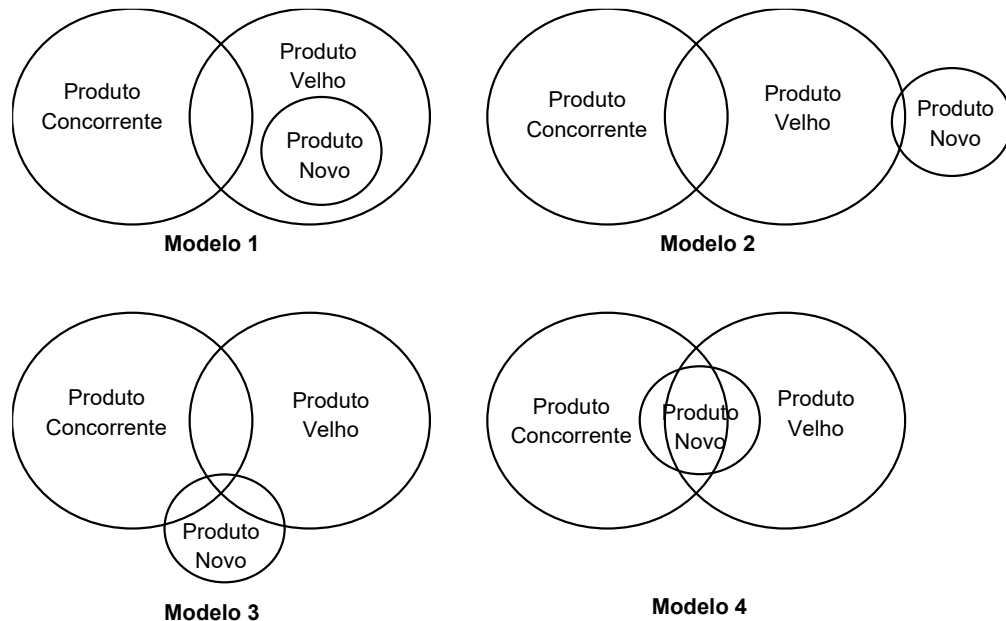
A canibalização de produtos constitui um conceito fundamental na gestão de portfólio para empresas com ofertas diversificadas, sendo definida como a dinâmica de vendas em que um produto recém-lançado (o Canibal) pela mesma organização absorve a demanda de um item já existente (a Vítima) (Novelli, 2015; Laruccia, Loureiro & Lopes, 2011).

Os fatores que estimulam a canibalização estão ligados à pressão competitiva e à política de portfólio. A necessidade incessante de inovar e lançar variações de produtos, em parte devido à pressão dos consumidores e da concorrência, torna o efeito Vítima-Canibal uma consequência direta da estratégia de crescimento e saturação de mercado (Oliveira & Mattar, 2000; Novelli, 2015). Estrategicamente, a canibalização pode ser vista como um risco aceitável para manter a fatia de mercado e gerenciar a competitividade, uma vez que, na ausência do Canibal, o cliente migraria para a concorrência em vez de optar pela Vítima (Novelli, 2015; Oliveira & Mattar, 2000).

O grau do fenômeno está intimamente ligado à sobreposição de mercado (Market Overlap) entre os produtos, ou seja, à similaridade e disputa pelo mesmo segmento de consumidores (target market) (Traylor, 1988), em outras palavras, quanto maior a similaridade e a disputa pelo mesmo segmento de consumidores (o target market), maior será o risco de

canibalização. A gestão estratégica passa, então, por projetar produtos Canibais que maximizem a demanda total (nova demanda + demanda do concorrente) e minimizem o desvio da Vítima. Traylor (1988), em suas análises sobre firmas multimarcas, descreveu quatro cenários distintos por meio do modelo de canibalismo de produtos conforme Figura 1.

Figura 1 – Modelos de canibalismo de produtos



Fonte: Mark B. Traylor (1988)

O Primeiro Modelo descreve uma substituição direta de vendas onde o novo produto ocupa uma posição de mercado equivalente ao item antigo; embora não gere crescimento total, é tolerável se o novo produto apresentar margens de lucro superiores.

O Segundo Modelo ilustra um canibalismo de menor impacto, no qual o produto recém-introduzido desvia parte da demanda do item existente, mas simultaneamente contribui para a expansão do mercado, atribuída a características inovadoras ou a estratégias de posicionamento diferenciadas.

O Terceiro Modelo é caracterizado pela canibalização que afeta não apenas o produto antigo da empresa, mas também a oferta de concorrentes, resultando em expansão de mercado, uma abordagem que pode solidificar a presença no setor.

Por fim, o Quarto Modelo apresenta a situação mais desafiadora, onde o novo produto compete com o antigo e com produtos de rivais, sem, contudo, expandir o mercado consumidor. Traylor (1988) também aborda o canibalismo indireto, como o que ocorre com marcas próprias (private label), e o canibalismo institucional ou pré-canibalismo, que se refere à disputa por recursos internos da empresa (tempo de vendas, gestão e capacidade produtiva) que os novos produtos impõem aos estabelecidos.

Ao se determinar que um produto irá ser canibalizado Kein et al., 1978 propõe dois cenários onde a canibalização pode ser proativa e vantajosa ou desvantajosa. Para a Canibalização ser vantajosa deve-se ser proativa seguindo estratégia de “Dividir para conquistar” que inclui: Lançar o novo produto para canibalizar a si mesmo antes que um

concorrente o faça; atrair novos usuários para a categoria de produtos da empresa; ocupar nichos de mercado ou atender a diferentes necessidades de clientes com SKUs específicos e substituir produtos em fase de declínio por inovações. Já a canibalização desvantajosa é frequentemente originada por decisões da alta administração ou práticas gerenciais específicas (Kerin et al., 1978). Estas práticas podem ser causadas por: excessiva pressão por crescimento via novos produtos; a busca por uma linha de produtos completa para aumentar a participação de mercado; posicionamento inadequado de produtos; segmentação de mercado irrealista que gera sobreposição de demandas; e esforços promocionais que priorizam novos itens em detrimento dos produtos consolidados. Para mitigar o fenômeno, Traylor (1986) e Copulsky (1976) sugerem o lançamento de produtos diferenciados em segmentos de mercado bem definidos.

Portanto, ao se planejar um lançamento de um produto devem-se criar estratégias para evitar ou gerenciar a canibalização, focando em: atacar a concorrência sem expandir o mercado ou expandir o mercado e atacar a concorrência ou então expandir o mercado sem atacar a concorrência (Traylor, 1988). A gestão estratégica busca projetar produtos Canibais que maximizem a demanda total (nova demanda + demanda do concorrente) e minimizem o desvio da Vítima.

O sucesso de um lançamento não é medido pelo volume de vendas brutas do novo produto, mas pelas Vendas Orgânicas (Vendas Líquidas). As Vendas Orgânicas são o volume de vendas do novo produto que veio de clientes que compraram o produto pela primeira vez na categoria, ou de clientes que vieram de produtos concorrentes, ou de um aumento geral no consumo da categoria (Novelli, 2015). A Expansão Orgânica é a métrica que valida o investimento e a estratégia de inovação, ver Fórmula 1.

$$Vendas_{Orgânicas} = D_{Novo,Real} - Canibalização \quad (1)$$

Tabela 1 – Variáveis de Expansão Orgânica

Variável	Legenda	Interpretação
$Vendas_{Orgânicas}$	Volume de vendas de expansão líquida	O indicador de sucesso do lançamento O volume de vendas do novo produto que representa o crescimento real da categoria/portfólio
$D_{Novo,Real}$	Demanda Total Real do Novo Sabor	Volume de vendas total observado do novo sabor
$Canibalização$	Perda total de vendas dos produtos antigos	O valor calculado na Fórmula anterior (Perda Bruta), que é subtraído da demanda bruta para encontrar o valor líquido (Expansão Orgânica)

Fonte: Novelli (2015)

O volume de Vendas Orgânicas deve ser o único volume considerado para o cálculo do Retorno sobre o Investimento (ROI) da inovação e para o planejamento de longo prazo do novo produto. Se este valor for positivo, o lançamento é validado como um vetor de crescimento. Se for negativo, significa que o novo produto não apenas canibalizou o portfólio, mas também pode ter perdido clientes para a concorrência após o lançamento. Este insight direciona as equipes de supply chain a basear seu planejamento de capacidade e investimento de longo prazo

apenas neste volume líquido, evitando o superdimensionamento da capacidade produtiva para um volume que é meramente uma realocação interna.

2.2 Detecção e medição da canibalização

A quantificação do efeito canibal é um desafio metodológico, pois requer o estabelecimento de um cenário contrafactual (o que teria acontecido sem o lançamento do produto canibal), para isso, a análise de Séries Temporais tem sido amplamente empregada para detectar e quantificar a canibalização, identificando "quebras estruturais" (mudanças abruptas ou graduais) nos processos estocásticos de geração de vendas após o lançamento de um novo produto.

De acordo com Novelli (2015), a canibalização é calculada pela diferença entre a demanda que os produtos antigos teriam tido (cenário contrafactual) e a demanda que eles realmente tiveram após o lançamento do novo sabor. O modelo de Machine learning pode ser usado para simular o contrafactual ($D_{Antigo,i,Simul}$), estimando a demanda dos sabores antigos se o novo sabor não existisse (definindo a variável dummy de lançamento como zero), ver Fórmula 2.

$$Canibalização = \sum_i (D_{Antigo,i,Simul} - D_{Antigo,i,Real}) \quad (2)$$

Tabela 2 – Variáveis de Canibalização

Variável	Legenda	Interpretação
<i>Canibalização</i>	Perda total de vendas dos produtos antigos	A soma das vendas perdidas por todos os produtos antigos devido ao lançamento do novo sabor.
$D_{Antigo,i,Simul}$	Demanda Simulada (Contrafactual)	Vendas previstas para o produto antigo <i>i</i> caso o novo produto não existisse (obtida via XGBoost).
$D_{Antigo,i,Real}$	Demanda Real observada	Vendas reais observadas do produto antigo <i>i</i> após o lançamento do novo sabor.

Fonte: Novelli (2015)

2.3 Análise pré-lançamento e sensibilidade

Esta fase foca na identificação da vulnerabilidade à canibalização e na estimativa inicial da demanda utilizando uma regressão log dinâmica com o objetivo de modelar a demanda de um produto antigo ($D_{Antigo,i,t}$) em função de seu próprio preço ($P_{Antigo,i,t}$) e do preço de um novo produto (P_{Novo}), a fim de estimar a canibalização no contexto de lançamento de novos produtos.

O desafio de quantificar o fenômeno da canibalização requer uma abordagem analítica que estabeleça a relação de substituição causal entre os produtos. A econometria de demanda, particularmente o uso de Elasticidades de Preço, oferece a fundação para a análise de canibalização, convertendo dados de transação em insights de sensibilidade de mercado. A utilização da Regressão Log-Dinâmica é metodologicamente eficaz, pois a transformação logarítmica permite que os coeficientes do modelo sejam interpretados diretamente como elasticidades, facilitando a mensuração precisa da canibalização (Silva, et al., 2017).

2.4 Fórmula do modelo de elasticidade log-dinâmica

A Fórmula (3) que integra a elasticidade direta e a elasticidade cruzada para o produto antigo (i) no tempo (t) é:

$$\log D_{Antigo,i,t} = \beta_0 + \beta_1 \log P_{Antigo,i,t} + \beta_2 \log P_{Novo} + \sum_{k=1}^K \gamma_k X_{k,t} + \epsilon_t \quad (3)$$

Tabela 3 – Variáveis de Elasticidade Log-Dinâmica

Variável	Legenda	Interpretação
$\log D_{Antigo,i,t}$	Logaritmo da Demanda do produto antigo i no período t	Variável dependente. Volume de vendas cujo comportamento está sendo modelado.
β_0	Intercepto (Base da Demanda)	Demanda base do produto antigo na ausência de todas as variáveis explicativas e <i>dummies</i> .
$\beta_1 \log P_{Antigo,i,t}$	Logaritmo do Preço do produto antigo i no período t	Variável independente utilizada para calcular a Elasticidade de Preço Direta (β_1).
$\beta_2 \log P_{Novo}$	Logaritmo do Preço do novo produto/sabor	Variável independente utilizada para calcular a Elasticidade de Preço Cruzada (β_2).
$\sum_{k=1}^K \gamma_k X_{k,t}$	Variáveis de Controle (X)	Capturam o efeito de fatores exógenos (sazonalidade, promoções, feriados) que influenciam a demanda.
ϵ_t	Termo de erro aleatório no período t	Variações na demanda não explicadas pelo modelo.

Fonte: Silva, et al. (2017)

2.5 Elasticidade de preço direta

A Elasticidade de Preço Direta (β_1) mede a variação percentual na demanda do produto antigo i causada por uma variação percentual no preço do próprio produto i . É o coeficiente da variável $\log P_{Antigo,i,t}$ no modelo de regressão, ver Fórmula 4.

$$\beta_1 = \% \Delta D_{Antigo,i} \backslash \% \Delta P_{Antigo,i} \quad (4)$$

O coeficiente β_1 é classicamente esperado ser negativo ($\beta_1 < 0$), refletindo a lei da demanda: um aumento no preço de 1% resulta em uma queda de $|\beta_1|$ na sua demanda. A riqueza da interpretação reside na magnitude desse valor: se $|\beta_1|$ for alto, indica que os consumidores desse sabor antigo são extremamente sensíveis ao preço, o que implica que qualquer estratégia promocional deve ser cuidadosamente calibrada (Silva, L. L. R. e. 2017). Gerencialmente, este insight define as margens de manobra de pricing para o SKU antigo; uma elasticidade direta muito alta sugere que a empresa deve buscar maior eficiência de custo ou priorizar o volume em vez da margem para manter a competitividade contra produtos entry-level do mercado (Oliveira, B. A. C., 2000). Além disso, a comparação da elasticidade direta do produto antigo antes e depois do lançamento do novo produto pode indicar uma mudança no status de preço do produto antigo dentro da percepção do consumidor.

2.6 Elasticidade de preço cruzada

A Elasticidade de Preço Cruzada (β_2) mede a variação percentual na demanda do produto antigo i causada por uma variação percentual no preço do novo produto (canibal). É o coeficiente da variável $\log P_{Novo}$ no modelo, ver Fórmula 5.

$$\beta_2 = \% \Delta D_{Antigo,i} \backslash \% \Delta P_{Novo} \quad (5)$$

Este coeficiente é o indicador econométrico primário da canibalização. Um resultado positivo ($\beta_2 > 0$) e estatisticamente significativo confirma que os produtos são substitutos dentro do portfólio da empresa (Laruccia, Loureiro & Lopes, 2011). A magnitude de β_2 é um insight estratégico inestimável: um β_2 alto sugere uma canibalização forte, onde os produtos são percebidos como quase idênticos pelos consumidores; ou seja, um aumento de 1% no preço do novo sabor leva a um aumento de 1.8% nas vendas do sabor antigo, confirmando a transferência de demanda.

Gerencialmente, o valor de β_2 é utilizado para análise ponderada, pois os SKUs antigos com maior β_2 são aqueles que exigem a maior precisão na previsão contrafactual, pois são os mais vulneráveis à substituição interna e também é utilizada para validar a diferenciação, pois se o valor de β_2 for baixo ou próximo de zero, isso sugere que o novo produto conseguiu atingir um segmento de mercado único (expansão orgânica pura) ou que a canibalização é insignificante, validando a estratégia de inovação do produto (Novelli, 2015).

A análise de β_2 transforma a gestão de pricing de uma decisão isolada para uma otimização de portfólio interdependente, onde o preço de um SKU deve ser fixado considerando o impacto na demanda de outro.

Em resumo, os coeficientes betas medem a sensibilidade e caracteriza a canibalização conforme exemplificado na Tabela 4.

Tabela 4 – Coeficientes beta e suas implicações

Coeficiente	Legenda	Fórmula	Sinal Esperado	Implicação
β_1	Elasticidade de Preço Direta	$\beta_1 = \% \Delta D_{Antigo,i} \backslash \% \Delta P_{Antigo,i}$	Negativo ($\beta_1 < 0$)	Mede a sensibilidade do SKU ao seu próprio preço. Determina a elasticidade de volume em estratégias promocionais focadas
β_2	Elasticidade de Preço Cruzada	$\beta_2 = \% \Delta D_{Antigo,i} \backslash \% \Delta P_{Novo}$	Positivo ($\beta_2 > 0$)	Confirma a canibalização. A magnitude indica o grau de substituição percebida entre o novo e o antigo produto

Fonte: Novelli (2015)

2.7 Machine Learning aplicado à caniblicação

A previsão de demanda, especialmente em cenários voláteis como o lançamento de um novo produto, historicamente dependia de métodos estatísticos tradicionais como ARIMA ou regressões lineares (Wanke & Julianelli, 2006). Contudo, a ascensão do Machine Learning (ML), particularmente o XGBoost (eXtreme Gradient Boosting), em que oferece maior capacidade de capturar a não-linearidade e as interações complexas inerentes ao comportamento do consumidor e ao preço (Tewogbade & Akinlose, 2023).

O ML se torna importante neste estudo pois é encarregado de construir o cenário contrafactual com a máxima acurácia possível.

O modelo XGBoost, sendo um algoritmo que utiliza a técnica de Gradient Boosting de forma otimizada, robusto para séries temporais. Sua arquitetura aditiva, onde cada nova árvore corrige os erros residuais da árvore anterior, permite que ele modele eficientemente as complexas relações de caniblicação detectadas pela Elasticidade de Preço Cruzada visto na sessão anterior, superando as limitações dos modelos tradicionais na previsão de longo prazo (Bekal & Bari, 2021).

2.8 Função objetivo ponderada (Loss Function)

A inovação metodológica reside em adaptar este preditor a um problema de negócio específico: a priorização do erro de previsão para os produtos que são mais sensíveis à caniblicação. O principal desafio na caniblicação é prever a demanda dos produtos antigos caso o novo produto não existisse ($D_{Antigo,i,Simul}$). O XGBoost, ao ser treinado com o histórico de vendas (incluindo o período pré-lançamento) e variáveis exógenas (como dummies de lançamento), é capaz de simular o comportamento de demanda. O problema surge na acurácia dessa previsão, especialmente porque o erro de previsão dos produtos antigos (as "vítimas") tem um impacto desproporcional no cálculo final da caniblicação.

A função objetivo padrão do XGBoost trata todos os erros de previsão da mesma forma. No entanto, para o problema de caniblicação, a precisão da previsão nos produtos "vítimas" é mais importante do que nos produtos com alta Expansão Orgânica, pois o erro na previsão do contrafactual propaga-se diretamente para a métrica de sucesso líquido (Vendas Orgânicas). Para resolver o problema da caniblicação, é preciso priorizar a acurácia na previsão de vendas dos produtos antigos (as "vítimas"). O erro de previsão nos produtos antigos é crucial para calcular o cenário contrafactual e, conseqüentemente, a Expansão Orgânica.

A solução reside na introdução de um peso ponderado ($w_{i,t}$) na função de perda (Loss Function) do XGBoost, transformando a função objetivo padrão ($L(\phi)$) em Fórmula 6:

$$L(\phi) = \underbrace{\sum_i \sum_t w_{i,t} \cdot l(y_{i,t}, \hat{y}_{i,t})}_{\text{Termo de Perda Ponderado}} + \underbrace{\sum_{k=1}^K \Omega(f_k)}_{\text{Termo de Regularização}} \quad (6)$$

Tabela 5 – Variáveis da função Objetivo Ponderada e suas interpretações

Variável	Legenda	Interpretação
$L(\phi)$	Função Objetivo Ponderada (<i>Loss Function</i>)	A função que o modelo XGBoost busca minimizar durante o treinamento para encontrar o conjunto ideal de árvores (f_k).

i	Índice do Produto (SKU)	Representa cada produto individual dentro do portfólio.
t	Índice do Tempo	Representa o período (dia, semana, mês) da observação.
$w_{i,t}$	Peso Ponderado	Um peso ajustável dado à perda do produto i no tempo t . É maior para os produtos canibalizados (vítimas), forçando o modelo a priorizar a precisão da previsão desses SKUs.
$l(y_{i,t}, \hat{y}_{i,t})$	Função de Perda (<i>Loss Function</i>)	Mede a diferença entre a demanda real (y) e a demanda prevista \hat{y} tipicamente o Erro Quadrático Médio (MSE).
$y_{i,t}$	Demanda Real	O volume de vendas real observado para o produto i no tempo t .
$\hat{y}_{i,t}$	Demanda Prevista	O volume de vendas previsto pelo modelo XGBoost para o produto i no tempo t .
$\sum_{k=1}^K \Omega(f_k)$	Termo de Regularização	Penalidade aplicada à complexidade do modelo (árvores f_k) para evitar <i>overfitting</i> e garantir a generalização.

Fonte: Bekal & Bari (2021)

Exemplificando a Fórmula, o primeiro termo é o "Placar do Erro" que o XGBoost, tenta zerar durante o treinamento. Imagine que o modelo está fazendo uma prova: a primeira parte da fórmula (Primeiro termo) é a nota que ele tira, ou seja, o erro de previsão, que mede a diferença entre o que ele previu (\hat{y}) e o que realmente aconteceu (y). O fator principal é o $w_{i,t}$ (o "peso"). Em um teste normal, todos os erros valem o mesmo. Mas no nosso caso de canibalização, nós personalizamos a prova: dizemos ao modelo que errar a previsão de vendas de um produto antigo (que está sendo canibalizado) é muito mais grave do que errar a previsão de um produto novo. Ao dar um "peso" maior a esses erros críticos, forçamos o modelo a se esforçar ao máximo para ser precisa na previsão dos produtos antigos, garantindo que o cálculo final das "Vendas Orgânicas" (o lucro líquido real da inovação) seja o mais confiável possível. A segunda parte da fórmula ($\Omega(f_k)$) é apenas uma "penalidade", que impede o modelo de tentar decorar as respostas em vez de realmente aprender o padrão. Em resumo, esta fórmula é a nossa regra para dizer ao robô: "Seja preciso onde a imprecisão custa mais dinheiro."

2.8 Termo perda ponderada

O primeiro termo da função Objetivo Ponderada (*Loss Function*) é a Perda Ponderada que representa a diferença (o erro) entre as vendas reais e as vendas previstas ao longo do tempo t para todos os produtos i . A Função de Perda de uma única observação (produto i no tempo t) com peso (w_i) é:

$$l(y_i, \hat{y}_i) = w_i \cdot (y_i - \hat{y}_i)^2 \quad (7)$$

Onde:

y_i é o valor real (demanda observada).

\hat{y}_i é a previsão do modelo até o passo anterior ($\hat{y}_i^{(t-1)}$).

w_i é o peso estratégico (alto para produtos canibalizados).

O Peso Ponderado $w_{i,t}$, ajusta o valor para ser significativamente maior para os produtos antigos que demonstram alta Elasticidade de Preço Cruzada, ou que historicamente têm seu volume de vendas mais volátil devido à introdução de novos produtos. Essa personalização força o algoritmo a gastar mais capacidade preditiva e esforço de otimização minimizando o erro de previsão para os SKUs que são cruciais para o cálculo do cenário contrafactual, garantindo assim que a desagregação final das Vendas Orgânicas. Sua definição matemática pode ser apresentada como:

$$w_{i,t} = 1 + \alpha \cdot I_{Canibal}(t) \cdot |\beta_{2,i}| \quad (8)$$

Onde:

$I_{Canibal}(t)$ é uma função indicadora que é 1 se o novo produto está disponível no período t e 0 caso contrário.

$|\beta_{2,i}|$ é o valor absoluto da Elasticidade de Preço Cruzada do SKU i (calculada anteriormente). Quanto maior a elasticidade, maior o peso β_2 e, portanto, maior o peso $w_{i,t}$.

α é um fator de escala ajustável para controlar o impacto do peso na função de perda.

Quando a Função de Perda é personalizada para incluir o peso $w_{i,t}$, o gradiente e a Hessiana devem ser recalculados para incluir esse peso. Este processo garante que, a cada iteração de construção da árvore, o algoritmo utilize o método de descida de gradiente para convergir para o conjunto de árvores que minimiza o erro ponderado, cumprindo o objetivo estratégico de priorizar a acurácia no cenário contrafactual dos produtos canibalizados.

O Gradiente é a derivada de primeira ordem da função de perda em relação à previsão (\hat{y}_i). Ele indica a direção e a magnitude do erro que a próxima árvore precisa corrigir conforme Fórmula 9.

$$g_i = -2 \cdot w_i \cdot (y_i - \hat{y}_i) \quad (9)$$

O Gradiente é diretamente proporcional ao peso w_i . Se a observação é de um produto canibalizado e o peso w_i é alto (por exemplo, 10 vezes maior), o Gradiente g_i será 10 vezes maior. Isso faz com que o algoritmo perceba o erro dessa observação como muito mais grave, forçando a próxima árvore de decisão a se concentrar agressivamente em reduzir o erro daquele produto específico, garantindo a prioridade na acurácia da previsão dos produtos-vítimas.

Já a Hessiana é a derivada de segunda ordem da função de perda (derivada do Gradiente). Ela mede a curvatura da função e é usada para determinar o tamanho ideal do passo de otimização (taxa de aprendizado), garantindo que o modelo não "salte" demais nem demore a convergir conforme Fórmula 10.

$$h_i = 2 \cdot w_i \quad (10)$$

Assim como o Gradiente, a Hessiana também incorpora o peso w_i . a Hessiana customizada assegura a estabilidade numérica e a convergência eficiente do algoritmo. Ao incluir o peso, a Hessiana garante que o processo de otimização (a forma como o modelo aprende) seja dimensionado corretamente de acordo com a importância estratégica de cada erro, mantendo a coerência do treinamento.

Aplicando o gradiente e hessiana temos:

$$L^{(t)} \approx \sum_{i=1}^n \left[\text{constante} + g_i f_t(x_i) + \frac{1}{2h_i f_t^2(x_i)} \right] + \Omega(f_t) \quad (11)$$

Porém, a Fórmula $L(\phi)$ define o que é a perda já o Gradiente e a Hessiana definem como a perda será minimizada. A fórmula final do loss não precisa incluir as ferramentas de otimização. A única maneira de o XGBoost utilizar g_i e h_i é através da Expansão de Taylor de segunda ordem da função $L(\phi)$ em cada passo. Após essa expansão, a função $L(\phi)$ é reformulada temporariamente para um formato quadrático que explicitamente contém g_i e h_i . Portanto, o Gradiente e a Hessiana são derivados de $L(\phi)$ mas não são componentes de $L(\phi)$ em sua forma original.

2.9 Termo de regularização

Já o segundo Termo é o de Regularização $\sum_{k=1}^K \Omega(f_k)$ é aplicado ao modelo (XGBoost) para impor uma penalidade sobre a complexidade das árvores de decisão. Ao aplicar esse termo no modelo, estamos efetivamente dizendo ao algoritmo: "Minimize o erro, mas apenas se o ganho de precisão compensar o custo da complexidade." A aplicação deste termo é crucial e se manifesta diretamente no critério de divisão (ou split) dos nós de cada árvore de decisão. O processo de construção de cada árvore no XGBoost busca a melhor divisão possível em um nó para maximizar o ganho (redução da perda). A regularização entra nesse cálculo, pois o Ganho de Otimização em cada divisão de nó (L_{split}) é definido pela redução na perda, subtraída pela penalidade de regularização. A otimização para a escolha da melhor divisão (split) em um nó é dada pela Fórmula 12.

$$L_{split} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (12)$$

Tabela 6 – Componentes do termo de Regularização e suas implicações

Componente	Variáveis	Função Matemática	Implicação no Treinamento
Ganho de Redução de Perda (Ponderado)	O primeiro colchete [...]		Representa a redução total na função de perda (erro) obtida ao dividir o nó atual (I) em um nó filho esquerdo (I_L) e um nó filho direito (I_R). Como g_i e h_i já incorporam o peso $w_{i,t}$, esta redução já é ponderada .
Penalidade de Estrutura	γT	γT	(Onde T é o número de nós folha da árvore f) Controle da Profundidade da Árvore: Impede que as árvores se tornem muito profundas e detalhadas. O modelo só realiza uma nova divisão se o ganho na redução do erro for maior do que o custo γ , garantindo que apenas <i>splits</i> significativos sejam criados.

Penalidade L2 (λ no denominador)	$\frac{1}{2}\lambda \sum w_j^2$	$\frac{1}{2}\lambda \sum_{j=1}^T w_j^2$	(Onde w_j é o peso de <i>output</i> da folha j) Suavização dos Resultados: Penaliza grandes pesos nos nós folha. Isso impede que uma única regra de decisão tenha uma influência desproporcionalmente grande na previsão final, promovendo uma função de previsão mais suave e resistente a ruídos nos dados.
Função Completa	$\Omega(f)$	$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^T w_j^2$	Prevenção do Overfitting: Assegura que o modelo XGBoost, mesmo sendo pressionado pelo peso $w_{i,t}$ a ser extremamente preciso nos produtos canibalizados, não decore os dados de treinamento, mantendo a generalização e a confiabilidade para previsões futuras (cenário contrafactual).

Fonte: Bekal & Bari (2021)

Este termo garante que, mesmo com a priorização agressiva de erro nos produtos canibalizados (introduzida pelo peso $w_{i,t}$), o modelo não crie uma árvore excessivamente detalhada (super ajustada) que funcionaria bem apenas nos dados de treinamento, mas falharia na previsão de longo prazo em um cenário real.

3 MÉTODO

O estudo proposto é uma abordagem híbrida de *analytics* de demanda, projetado para maximizar a acurácia do cenário contrafactual e, conseqüentemente, a precisão da métrica de Vendas Orgânicas.

A metodologia para quantificação e previsão da canibalização de produtos foi estruturada em quatro fases de forma sequencial, os modelos matemáticos discutidos na sessão anterior foram introduzidos no ambiente *Python* através do *Google Collab*, em que foram utilizadas as bibliotecas *numpy* para regressão e *xgboost* para *Machine Learning*. O estudo concentra-se na categoria Energéticos em lata para consumo individual, analisando um período de 13 meses (Setembro de 2024 a Setembro de 2025) para quantificar o impacto do SKU canibalizador (A300 - Energético Red Bull Zero 250ml) sobre 21 SKUs vítimas (de A100 a A2200 – Toda Família de Red Bull). O processo de modelagem é dividido em três fases sequenciais conforme Tabela 7.

Tabela 7 – Fases do estudo

Fase	Foco Principal	Técnica Chave	Objetivo Estratégico
Fase 1: Detecção	Relação de Substituição e Contexto Teórico	Elasticidade de Preço Cruzada (Regressão Log-Dinâmica) & Modelo Dirichlet	Identificar pares de SKUs canibalizáveis, quantificar a relação e calcular a canibalização esperada.

Fase 2: Previsão Contrafactual	Previsão da Demanda Futura dos Produtos Antigos	XGBoost com <i>Custom Objective Function</i>	Gerar a previsão de demanda mais precisa possível para os produtos "vítimas" ($D_{Antigo,i,Simul}$).
Fase 3: Desagregação	Cálculo do Sucesso Líquido	Álgebra de Decomposição de Vendas	Desagregar as vendas brutas do novo produto em Canibalização e Vendas Orgânicas.

Fonte: Autores (2025)

A Fase 1, Detecção e Quantificação, utiliza a Regressão Log-Dinâmica por OLS (*ols_numpy* em *Python*), onde o coeficiente (Elasticidade de Preço Cruzada) é estimado para medir a vulnerabilidade de cada vítima ao preço do SKU A300.

A Fase 2, Previsão Contrafactual, emprega o algoritmo XGBoost com uma Função Objetivo Personalizada (*custom_obj_weighted*), cujo Gradiente e Hessiana são ponderados por um peso derivado da Elasticidade de Preço Cruzada da Fase 1, garantindo que o modelo priorize a acurácia na previsão da demanda histórica dos SKUs mais vulneráveis, gerando o cenário contrafactual, simulado ao zerar as variáveis de preço e promoção do A300.

A Fase 3, Cálculo da Canibalização, mensura as Vendas Perdidas através da diferença em que finalmente, calcula o sucesso líquido do lançamento do A300 subtraindo a Canibalização total das Vendas Brutas do novo produto, fornecendo a métrica definitiva para avaliação estratégica.

O estudo avalia o impacto do lançamento do SKU canibalizador A300 (presumidamente, ENER G RED BULL ZERO 250ML) sobre 21 SKUs vítimas do mesmo portfólio de energéticos.

4 RESULTADOS E DISCUSSÃO

Iniciam-se os resultados e discussão pelo diagnóstico econométrico.

4.1 Diagnóstico econométrico – Elasticidade cruzada

A Tabela 8 apresenta os resultados completos da Regressão Log-Dinâmica (OLS) para todos os 21 SKUs vítimas, destacando a Elasticidade de Preço Cruzada (β_2), o principal indicador de vulnerabilidade.

Tabela 8 – Elasticidades de Todos os SKUs Vítimas (Fase 1)

SKU Vítima	Nome/Descrição do Produto	β_1 (Elasticidade Direta)	β_2 (Elasticidade Cruzada)	R ²
A100	Energ Red Bull Lata 250ml (C/ Açúcar, Tradicional)	-1.634	-0.539	0.740

A200	Energ Red Bull Lata 473ml (C/ Açúcar, Tradicional)	-0.076	0.007	0.279
A400	Energ Red Bull Lata 355ml (C/ Açúcar, Tradicional)	0.264	-0.027	0.323
A500	Energ Red Bull Sugar Free 250ml (S/ Açúcar)	0.449	0.199	0.590
A600	Energ Red Bull Summer Edition Fruts Tropicis 250ml	0.627	0.024	0.233
A700	Energ Red Bull Summer Melancia 250ml (Saborizado)	0.060	0.000	0.443
A800	Energ Red Bull Summer Lt 250ml (Saborizado)	0.207	0.009	0.297
A900	Energ Red Bull Pomelo Sugar Free 250ml	-0.395	-0.022	0.803
A1000	Energ Red Bull The Amora Edition Sugarfee 250ml	1.025	0.386	0.279
A1100	Energ Rb Summer Edit Mel Marac 250ml	1.462	-13.637	0.999
A1200	Energ Red Bull Tropical 250ml (Saborizado)	-3.436	-0.893	0.850
A1300	Energ Red Bull Sugar Free 355ml (S/ Açúcar)	-1.131	-5.051	0.485
A1400	Energ Red Bull Summer Edicao Melancia Lt 355ml	-0.320	0.002	0.751
A1500	Energ Red Bull Coco E Acai Lt 250ml (Saborizado)	0.347	0.043	0.187
A1600	Energ Red Bull Pêra Canela Lt 250ml (Saborizado)	-0.550	0.003	0.435
A1700	Energ Red Bull Coco E Acai Lt 355ml (Saborizado)	-1.704	-0.001	0.754
A1800	Energ Red Bull Lata 250ml - Kit (C/ Açúcar, Tradicional)	-1.559	0.043	0.760
A1900	Kit Energ Red Bull 2pc 250ml (C/ Açúcar, Tradicional)	0.985	0.000	0.935
A2000	Energ Red Bull Maca/Figo Lata 250ml (C/ Açúcar, Saborizado)	0.353	0.505	0.627
A2100	Energ Red Bull Tropical 355ml (Saborizado)	0.106	0.000	0.730
A2200	Energ Red Bull Melancia 355ml (Saborizado)	-0.261	-0.002	0.867

Fonte: Autores (2025)

4.2 Quantificação da canibalização por SKU

A Tabela 9 lista o volume de canibalização quantificado para cada SKU vítima pela diferença entre a previsão contrafactual do XGBoost (Fase 2) e o volume real.

Tabela 9 – Canibalização Total em Unidades por SKU Vítima

SKU Vítima	Canibalização Total (Unidades)	% sobre Total Canibalizado
A500	1,761.30	85.53%
A1200	52.81	2.56%
A1800	52.73	2.56%
A600	41.33	2.01%
A400	36.66	1.78%
A1500	8.69	0.42%
A100	1.94	0.09%
A1100	1.12	0.05%
A1400	0.66	0.03%
A1000	0.42	0.02%
A2000	0.25	0.01%
A1300	0.22	0.01%
A1600	0.09	0.00%
A200	0.02	0.00%
A700	0.00	0.00%
A800	0.00	0.00%
A900	0.00	0.00%
A1700	0.00	0.00%
A1900	0.00	0.00%
A2100	0.00	0.00%
A2200	0.00	0.00%
Total Agregado	2,059.20	100.00%

Fonte: Autores (2025)

4.3 Expansão orgânica

A Tabela 10 resume a métrica de sucesso final do lançamento do SKU A300.

Tabela 10 – Resumo do Impacto do Lançamento do SKU A300

Métrica	Valor Total (Unidades)	% sobre Venda Bruta A300	Implicação Estratégica
A) Venda Bruta SKU A300	38,449.61	100.0%	Total de vendas do novo produto.
B) Canibalização Agregada (Soma das 21 Vítimas)	2,059.20	5.4%	Vendas desviadas de produtos existentes.
C) Expansão Orgânica (A - B)	36,390.41	94.6%	Crescimento real de mercado.

Fonte: Autores (2025)

Os resultados detalhados em todas as fases confirmam o sucesso do lançamento do A300. A Tabela 9 demonstra que a canibalização é altamente concentrada no SKU A500 (85.53% do total da perda), que é um produto similar na linha "sem açúcar" de 250ml. A Tabela 10 mostra que, em termos de impacto estratégico, o efeito canibal é marginal (5.4%), resultando em um forte crescimento do mercado total para a empresa (94.6% de Expansão Orgânica).

5 CONSIDERAÇÕES FINAIS

O estudo do impacto do lançamento do SKU canibalizador A300 (presumidamente, Energ Red Bull Zero 250ml) demonstrou um sucesso estratégico, sustentado por um framework de análise rigoroso. Os resultados agregados da Fase 3 indicam que a Canibalização foi marginal, totalizando apenas 2.059,20 unidades, o que representa 5.4% das Vendas Brutas do novo produto. Conseqüentemente, a maior parte do volume de 38.449,61 unidades do A300 traduziu-se em uma robusta Expansão Orgânica de 94.6% (36.390,41 unidades), validando o lançamento como um gerador de crescimento de mercado. A análise detalhada da Fase 3 revelou que a canibalização está altamente concentrada no SKU A500 (Sugar Free 250ml), que absorveu 85.53% de toda a perda, confirmando a substituição esperada entre produtos similares do segmento "sem açúcar" e minimizando o impacto nos demais 20 SKUs do portfólio.

A eficácia do processo analítico reside na abordagem faseada, que integrou modelos econométricos e de Machine Learning. A Fase 1 (Elasticidades) forneceu o diagnóstico causal, quantificando a vulnerabilidade de cada SKU vítima ao preço do A300 através do coeficiente β_2 . Essa métrica foi crucial para calibrar o modelo XGBoost na Fase 2 (Previsão Contrafactual), garantindo que o algoritmo de Machine Learning priorizasse a acurácia na previsão de demanda justamente dos produtos mais sensíveis (aqueles com o $|\beta_2|$ mais alto). Essa sinergia entre o diagnóstico causal (econometria) e a quantificação preditiva (XGBoost) assegura que o volume de Vendas Perdidas calculado na Fase 3 seja uma estimativa acurada e robusta, superando as limitações dos modelos tradicionais que negligenciam a não-linearidade e a complexidade das relações de preço.

Em termos de oportunidades futuras, os resultados apontam para dois caminhos principais. Primeiro, a alta concentração da perda no SKU A500 sugere a necessidade de uma análise mais profunda no posicionamento de preço entre o A300 e o A500 para minimizar a substituição interna, talvez focando em diferenciação de valor ou promoção conjunta. Segundo,

há presença de elasticidades cruzadas negativas extremamente altas em magnitude (notavelmente, o SKU A1100 com β_2 -13.637) levanta a hipótese de que, para esses produtos, o SKU A300 atua como um complemento ou âncora de preço em vez de um substituto. Pesquisas futuras devem focar na expansão do framework para modelar ativamente essas relações de complementaridade e ancoragem, permitindo o desenvolvimento de políticas de precificação dinâmica que capitalizem a demanda de todo o portfólio.

6 REFERÊNCIAS

- BEKAL, G., & BARI, M. *An XGBoost-Based Forecasting Framework for Product Cannibalization*. arXiv:2111.12680v1, 2021.
- CASTRO, R., & BELO, O. Análise de Situações de Canibalização de Produtos em Sistemas de Retalho. Centro de I&D ALGORITMI, Universidade do Minho.
- COPULSKY, W. *Cannibalism in the Marketplace*. *Journal of Marketing*, 40(4), 103-105, 1976.
- KERIN, R. A., HARVEY, M. G., & ROTHE, J. T. *Cannibalism and New Product Development*. *Business Horizons*, 21(5), 25-31, 1978.
- KOTLER, P. *Administração de Marketing*. Prentice Hall, 1998.
- LARUCCIA, M., LOUREIRO, S. M. C., & LOPES, R. *A Conceptual Approach for Cannibalism Between Goods*. *Conference Paper*, 2011.
- NOVELLI, F. *Detection and Measurement of Sales Cannibalization in Information Technology Markets*. *Dissertation, Technische Universität Darmstadt*, 2015.
- OLIVEIRA, B. A. C. de; MATTAR, F. N. Canibalismo entre produtos: um estudo de múltiplos casos na indústria alimentícia brasileira. Doutorado em Administração na FEA/USP, 2000.
- SHAKIR, T., & MODUPE, A. *Enhancing Demand Forecasting in Retail Supply Chains: A Machine Learning Regression Approach*. *Global Journal Of Management And Business Research*, 2023.
- SILVA, L. L. R. e. Aplicação de Elasticidade de Preço e Canibalização de Smartphones Utilizando Séries Temporais. Dissertação de Mestrado, PUC-Rio, 2017.
- TEWOGBADE, S.; AKINLOSE, M. *Enhancing Demand Forecasting in Retail Supply Chains: A Machine Learning Regression Approach*. *Global Journal of Management and Business Research*, v. 23, n. A8, p. 1-15, 2023.
- TRAYLOR, M. B. *Cannibalism in multibrand firms*. *Journal of Consumer Marketing*, Santa Barbara, v. 3, n. 2, p. 69-75, 1988.
- WANKE, P.; JULIANELLI, L. Previsão de vendas: Processos organizacionais & métodos qualitativos e quantitativos. São Paulo: Atlas, 2006.
- XU, L. *A Clustering Method for Product Cannibalization Detection Using Price Effect*. *Electronics*, 2025.