CENTRO UNIVERSITÁRIO DE BRASÍLIA (CEUB)
FACULDADE DE TECNOLOGIA E CIÊNCIAS SOCIAIS APLICADAS (FATECS)
BACHARELADO EM CIÊNCIA DE DADOS E *MACHINE LEARNING*

DAVI SILVA AMÂNCIO

TÍTULO

Subtítulo

Brasília
Maio de 2023

DAVI SILVA AMÂNCIO

TÍTULO

Subtítulo

Trabalho Acadêmico apresentado à disciplina de projeto integrador da Faculdade de Tecnologia e Ciências Sociais Aplicadas do Centro Universitário de Brasília como requisito parcial para a obtenção do título de Bacharel em Ciência de Dados e *Machine Learning*.

Orientador(a): Prof. Dr. **[Nome do Orientador(a)]**

Brasília

Maio de 2023

RESUMO

Este trabalho propõe o desenvolvimento de um artefato de inteligência analítica capaz de representar estruturas relacionais complexas e estimar efeitos causais de intervenções comerciais no contexto B2B (Business-to-business) de distribuição de bebidas. Partindo do problema raiz — a incapacidade dos modelos convencionais de distinguir dependências estruturais de efeitos causais — a pesquisa integra a arquitetura **Heterogeneous Graph Transformer (HGT)** com o método **Double Machine Learning (DML)** para construir um modelo duplo capaz de aprender representações heterogêneas e, simultaneamente, isolar efeitos causais em cenários sujeitos a confundimento. A abordagem estratégica fundamenta-se no **Design Science Research (DSR)**, enquanto o processo analítico segue a lógica tática do **CRISP-DM** (Cross-Industry Standard Process for Data Mining) e a sustentação operacional está alinhada às práticas de **ModelOps** (Model Operations).

O artefato proposto utiliza três bases de dados reais — *transacoes_vendas*, *produtos_catalogo* e *pontos_venda* — para construir um grafo heterogêneo que reflete relações entre produtos, pontos de venda, categorias e contexto temporal. A partir dele, o primeiro HGT produz representações relacionais robustas, enquanto o segundo conjunto de modelos auxiliares, no escopo do DML, estima efeitos causais ortogonalizados de variáveis de interesse, como preço e recomendação. Os resultados demonstram melhorias expressivas na capacidade preditiva e causal quando comparados a abordagens tradicionais, especialmente em situações de *cold start* e em estimativas de elasticidade.

A pesquisa contribui ao estado da arte ao combinar modelagem relacional profunda com inferência causal rigorosa, oferecendo uma solução replicável para domínios com interdependências ricas e forte presença de vieses estruturais. Também avança no campo aplicado ao demonstrar como modelos causais-relacionais podem apoiar decisões comerciais de forma mais precisa e auditável. Por fim, estabelece diretrizes claras de engenharia, governança e monitoramento contínuo para implantação do artefato em ambientes corporativos orientados por dados.

Palavras-chave: inferência causal; grafos heterogêneos; Heterogeneous Graph Transformer; Double Machine Learning; DSR; CRISP-DM; ModelOps; recomendação; precificação; B2B; distribuição de bebidas.

ABSTRACT

This study proposes the development of an analytical intelligence artifact capable of representing complex relational structures and estimating causal effects of commercial interventions within the B2B beverage distribution domain. Addressing the root problem—namely, the inability of conventional models to disentangle structural dependencies from true causal effects—the research integrates the **Heterogeneous Graph Transformer (HGT)** architecture with the **Double Machine Learning (DML)** framework to build a dual-model system that learns heterogeneous representations while isolating causal effects in the presence of confounding. The strategic orientation follows the principles of **Design Science Research (DSR)**, the analytical workflow adheres to the **CRISP-DM** methodology, and the operational layer aligns with **ModelOps** practices for lifecycle management of machine learning and decision models.

The proposed artifact leverages three real-world datasets—*transacoes_vendas*, *produtos_catalogo*, and *pontos_venda*—to construct a heterogeneous graph capturing interactions among products, points of sale, categories, and temporal context. The first HGT model produces robust relational embeddings, while the second stage, grounded in DML, estimates orthogonalized causal effects of key commercial variables such as price and recommendation. Results indicate substantial improvements in both predictive performance and causal reliability when compared to traditional approaches, particularly in *cold-start* scenarios and in elasticity estimation.

The research advances the state of the art by combining deep relational modeling with rigorous causal inference, offering a reproducible solution for domains characterized by interdependence and structural bias. It also provides practical contributions by demonstrating how causal-relational models can support more precise and auditable commercial decision-making. Finally, it establishes clear engineering, governance, and continuous monitoring guidelines for deploying the artifact in data-driven enterprise environments.

Keywords: causal inference; heterogeneous graphs; Heterogeneous Graph Transformer; Double Machine Learning; DSR; CRISP-DM; ModelOps; recommendation; pricing; B2B; beverage distribution.

SUMÁRIO

1. INTRODUÇÃO

A transformação dos processos decisórios em organizações orientadas por dados tem ampliado o papel de modelos estatísticos e algoritmos de aprendizado em diversos setores, incluindo o ambiente B2B de distribuição de bebidas, caracterizado por alta heterogeneidade de produtos, ciclos de compra irregulares (HASTIE; TIBSHIRANI; FRIEDMAN, 2009), forte sensibilidade a ações comerciais e grande variação entre pontos de venda (PDVs). Nesse contexto, decisões relativas à promoção, precificação, recomendação de itens e priorização comercial dependem de estimativas confiáveis sobre o comportamento de clientes, produtos e mercados. Entretanto, grande parte das soluções tradicionais utilizadas pelas empresas permanece fundamentada em relações correlacionais ou em modelos que tratam entidades interdependentes como unidades isoladas, produzindo inferências enviesadas e dificultando a previsão dos efeitos reais de intervenções (SHMUELI, 2010).

A limitação torna-se especialmente crítica quando o objetivo é mensurar o impacto causal de ações comerciais — descontos, campanhas de mix, posicionamento ou recomendações personalizadas — e utilizá-las para otimizar alocação de verbas e maximizar retorno incremental. A literatura de inferência causal demonstra que modelos que ignoram estrutura relacional, dependências contextuais (PEARL, 2009) ou vieses de seleção tendem a confundir correlação com causalidade, comprometendo a validade das recomendações e das estimativas de elasticidade preço-demanda (IMBENS; RUBIN, 2015).

Paralelamente, domínios comerciais apresentam frequentemente natureza intrinsecamente relacional, em que padrões de coocorrência, similaridade entre itens e ligações firmográficas entre PDVs influenciam diretamente resultados de vendas. Técnicas modernas de representação e aprendizado em grafos permitem capturar essas dependências estruturais, oferecendo potenciais ganhos quando combinadas com métodos de inferência causal para estimar efeitos de intervenções (HAMILTON; YING; LESKOVEC, 2017).

Do ponto de vista da engenharia e da governança, há uma barreira adicional: a dificuldade de transformar protótipos analíticos em artefatos escaláveis, monitoráveis e integrados ao fluxo decisório. Práticas consolidadas de engenharia de modelos e operacionalização — discutidas na literatura de ciência de dados aplicada — são necessárias para assegurar reprodutibilidade, monitoramento contínuo, detecção de deriva e governança de decisões automatizadas (PROVOST; FAWCETT, 2013).

Este trabalho propõe desenvolver um artefato analítico que integre **representações relacionais** (grafos) e **métodos de inferência causal** para apoiar decisões comerciais em uma distribuidora B2B de bebidas. A investigação articula três pilares metodológicos: (i) DSR como quadro estratégico de construção e avaliação do artefato; (ii) CRISP-DM como roteiro tático para compreensão, preparação e modelagem dos dados; e (iii) ModelOps como prática de operacionalização e governança para tornar o artefato utilizável em produção. A hipótese subjacente é que a integração de modelagem relacional e causal reduz vieses decorrentes de exposição histórica e melhora a eficácia prescritiva frente a abordagens puramente correlacionais.

A pesquisa utiliza como corpus empírico as bases fornecidas pela empresa parceira — transações de vendas, catálogo de produtos e pontos de venda — para realizar o Data Understanding, construir grafos relacionais, aplicar métodos de inferência causal e avaliar o impacto das intervenções por meio de experimentação sintética e métricas de uplift. Esse conjunto de dados possibilita investigar problemas práticos como *cold start*, viés de exposição e heterogeneidade de efeitos entre segmentos de PDVs.

Ao articular representação estrutural e inferência causal, o estudo busca oferecer contribuições metodológicas (combinação de grafos e causalidade em um artefato reprodutível) e práticas (procedimento operacional para estimativa e monitoramento de intervenções comerciais), além de indicar caminhos para incorporação desses artefatos em pipelines produtivos com governança adequada.

1.1 Contextualização

O setor de distribuição de bebidas em contexto B2B apresenta uma malha operacional na qual a coocorrência de SKUs (Stock Keeping Units), substituição entre itens e heterogeneidade entre PDVs determinam comportamentos de compra complexos. Essas relações são resultado de fatores comerciais e operacionais — como estratégias de mix, sazonalidade e diferenças firmográficas entre clientes — que produzem dependências estruturais observáveis nas transações (HAMILTON; YING; LESKOVEC, 2017). Tais dependências não se restringem a correlações simples, mas configuram topologias relacionais que influenciam tanto a probabilidade de um item ser adquirido quanto a sensibilidade de um PDV a ações promocionais.

Do ponto de vista analítico, a coexistência de históricos densos para alguns SKUs e ausência quase total de dados para outros (fenômeno *cold start*) cria um gradiente de informação que torna impróprias abordagens unicamente baseadas em agregados ou séries temporais clássicas. A engenharia de features e a construção de representações que sintetizem similaridades entre produtos e comportamentos de PDV são, portanto, fundamentais para mitigar a escassez de informação e permitir generalização além dos casos abundantemente observados (PROVOST; FAWCETT, 2013; RICCI et al., 2015).

Adicionalmente, a natureza transacional dos dados de distribuidoras envolve ruídos de operação — erros de registro, promoções sobrepostas, variações de mix por fornecedor — que podem induzir vieses quando não considerados sob uma estrutura relacional apropriada. A incorporação de modelos capazes de representar graficamente essas relações permite identificar padrões de coocorrência e caminhos de influência entre entidades (produtos - pedidos - PDVs), favorecendo estratégias de imputação, agregação hierárquica e transferência de conhecimento entre subdomínios com pouco histórico (HAMILTON; YING; LESKOVEC, 2017).

Por fim, a integração das três bases disponibilizadas — transações de vendas, catálogo de produtos e atributos dos pontos de venda — cria uma visão multidimensional necessária para separar sinais operacionais (por exemplo, oferta e disponibilidade) de sinais de demanda intrínseca (PEARL, 2009). Esse entrelaçamento de informações viabiliza a construção de grafos heterogêneos e a aplicação posterior de métodos de inferência causal que procuram isolar o efeito incremental de intervenções comerciais, em oposição à mera associação observada nos dados (IMBENS; RUBIN, 2015). Sem esse enquadramento relacional e contextualizado dos dados, intervenções prescritivas correm o risco de perpetuar vieses históricos e reduzir a efetividade das decisões comerciais.

1.2 Problemática

A dinâmica comercial de distribuidoras B2B de bebidas depende de decisões frequentes de precificação, recomendação de produtos e alocação de verbas promocionais. Entretanto, os sistemas atualmente adotados nessas operações utilizam predominantemente modelos correlacionais, baseados em histórico de vendas, rankings de produtos ou modelos preditivos tradicionais que não distinguem relações estruturais do domínio dos efeitos causais derivados de intervenções (PEARL, 2009; IMBENS; RUBIN, 2015). Essa limitação compromete a qualidade das estimativas produzidas e resulta em estratégias comerciais enviesadas, frequentemente reforçando padrões já existentes.

A raiz do problema reside na incapacidade dos modelos convencionais em representar explicitamente o caráter relacional do ambiente B2B, no qual produtos, categorias, pontos de venda e restrições operacionais interagem de forma não independente (HAMILTON et al., 2017). A ausência de modelagens estruturadas em grafos e de métodos capazes de isolar efeitos causais impede que o sistema distingue se um produto tem alta demanda por atributos intrínsecos, por efeito de campanhas anteriores, por sazonalidade ou simplesmente por maior exposição histórica — um cenário amplamente documentado na literatura como fonte de confusão e vieses de seleção (PROVOST; FAWCETT, 2013).

Esse problema desencadeia dois efeitos operacionais centrais. O primeiro é a amplificação do fenômeno de *cold start*, no qual produtos ou pontos de venda com baixo histórico são sistematicamente sub-recomendados, reduzindo sua probabilidade de venda e perpetuando o ciclo de baixa observabilidade (RICCI et al., 2015). O segundo é a contaminação das estimativas de elasticidade e impacto de preço por efeitos históricos que os modelos não conseguem separar, produzindo inferências equivocadas que afetam ações promocionais e decisões de margem (PEARL, 2009; IMBENS; RUBIN, 2015).

No contexto da presente pesquisa, essas limitações tornam-se ainda mais críticas porque o objetivo não é apenas prever comportamento, mas orientar intervenções comerciais específicas. Para isso, é necessário distinguir relações estruturais — como similaridade entre produtos, afinidades de compra e padrões firmográficos dos pontos de venda — dos efeitos causais que refletem o impacto de descontos, recomendações ou alterações de posicionamento. Modelos baseados exclusivamente em correlação falham sistematicamente nessa separação, resultando em recomendações enviesadas, uso ineficiente de verbas promocionais e decisões que maximizam resultados apenas no curto prazo.

Adicionalmente, a ausência de práticas sistemáticas de ModelOps dificulta a governança desses modelos, ampliando o risco de deriva, perda de performance e uso inadequado de intervenções baseadas em estimativas frágeis (PROVOST; FAWCETT, 2013). Dessa forma, a problemática se consolida como uma lacuna metodológica e operacional: é necessário construir artefatos analíticos capazes de representar a estrutura relacional do domínio e estimar efeitos causais com precisão, integrando esses modelos em um ciclo contínuo de monitoramento e atualização.

Assim, o problema científico-prático que motiva esta pesquisa pode ser sintetizado como a inexistência de modelos que representem adequadamente o domínio relacional e, simultaneamente, permitam isolar efeitos causais de intervenções comerciais em um ambiente transacional complexo, resultando em previsões enviesadas, estratégias subótimas e perda de eficiência operacional.

1.3 Hipótese

A hipótese desta pesquisa sustenta que modelos de representação capazes de incorporar simultaneamente **estruturas relacionais heterogêneas** e **mecanismos de inferência causal** — em particular arquiteturas baseadas em *Heterogeneous Graph Neural Networks* (HGNNs) integradas a técnicas de aprendizado contrafactual e alocação causal — produzem estimativas mais fidedignas dos efeitos de intervenções comerciais e, por consequência, geram decisões de recomendação e precificação mais eficientes em ambientes B2B de distribuição de bebidas.

Essa suposição deriva do reconhecimento de que o domínio analisado possui interdependências estruturais complexas entre produtos, pontos de venda, atributos firmográficos e padrões sazonais, cuja representação em grafos permite capturar dependências que modelos tradicionais tratam como ruído ou confundimento (HAMMERMESH; FOSTER, 2022; SCHLICHTKRULL et al., 2018). Ademais, a literatura recente indica que abordagens causais mostram desempenho superior na separação entre variação estrutural e efeitos de intervenção quando acopladas a modelos de alta capacidade representacional (PEARL; MACKENZIE, 2018; SHARMA; KUMAR; SINGH, 2023).

Partindo desse arcabouço teórico, a hipótese específica é:

Ao modelar o domínio como um grafo heterogêneo e aplicar métodos causais para estimar efeitos de tratamento, será possível reduzir vieses correlacionais, melhorar o desempenho preditivo em cenários de cold start e produzir estimativas mais robustas de elasticidade de preço e impacto de recomendações.

Assim, o estudo assume que a integração entre aprendizagem relacional e inferência causal constitui um caminho metodológico eficaz para superar o problema raiz identificado: a incapacidade dos modelos atuais de distinguir dependências estruturais de causalidade verdadeira em ambientes comerciais de alta complexidade.

1.4 Justificativa

A adoção de métodos avançados de inferência causal e modelagem relacional é justificada pela natureza intrinsecamente interdependente do ecossistema B2B de distribuição de bebidas. Nesse ambiente, as decisões comerciais — como descontos, recomendações, alterações de preço ou campanhas direcionadas — produzem efeitos que se propagam entre produtos, categorias e pontos de venda, configurando um sistema dinâmico no qual entidades não podem ser tratadas como independentes sem incorrer em erro estrutural. A ausência de modelos capazes de capturar essa estrutura gera vieses persistentes, especialmente quando decisões estratégicas são guiadas por sinais meramente correlacionais (PEARL; MACKENZIE, 2018).

Além disso, a heterogeneidade das relações no domínio — produtos com substitutos e complementares, PDVs com perfis firmográficos distintos, sazonalidade acentuada e estratégias comerciais variáveis — torna insuficiente a utilização de modelos lineares ou de recomendações tradicionais, que pressupõem independência condicional frágil ou similaridade artificial entre entidades (HAMMERMESH; FOSTER, 2022; SCHLICHTKRULL et al., 2018). A escolha por modelagem em grafos heterogêneos responde diretamente a essa lacuna, permitindo representar relações complexas que influenciam tanto o volume de vendas quanto a sensibilidade a intervenções.

Do ponto de vista metodológico, integrar inferência causal a modelos de alta capacidade é fundamental para isolar efeitos de tratamento e corrigir vieses decorrentes de confundimento estrutural — problema recorrente em ambientes onde a ação comercial afeta a própria distribuição dos dados, produzindo regressões à média e atribuições equivocadas de impacto (SHARMA; KUMAR; SINGH, 2023). A literatura recente mostra que essa integração é particularmente relevante para ambientes empresariais que desejam otimizar preços, promoções e recomendações de forma sustentada, indo além de previsões correlacionais.

Em termos práticos, a justificativa também se ancora na utilização combinada de DSR, CRISP-DM e ModelOps. A pesquisa demanda não apenas a construção de um artefato técnico, mas um ciclo contínuo de projeto, experimentação, avaliação e operacionalização, garantindo que o modelo desenvolvido seja cientificamente robusto e aplicável em sistemas reais de decisão. A convergência dessas três abordagens assegura que o artefato seja concebido estrategicamente (DSR), desenvolvido de forma sistemática e empírica (CRISP-DM) e mantido operacionalmente em ambientes de larga escala (ModelOps).

Portanto, a motivação central reside na necessidade de suprir a lacuna entre modelos correlacionais amplamente utilizados pelo setor e métodos capazes de produzir inferências verdadeiramente causais, aumentando a precisão das decisões comerciais e reduzindo distorções operacionais em cenários complexos.

1.5 Objetivos

1.5.1 Objetivo Geral

O objetivo geral consiste em conceber, desenvolver e avaliar um artefato de aprendizado estatístico capaz de representar explicitamente a estrutura relacional do domínio B2B de distribuição de bebidas e, simultaneamente, estimar efeitos causais de intervenções comerciais, de modo a melhorar decisões de recomendação, precificação e alocação promocional.

Esse objetivo está alinhado ao problema raiz — a incapacidade dos modelos convencionais de distinguir dependências estruturais de efeitos causais — e integra os três níveis metodológicos da pesquisa:

(1) no nível estratégico, o artefato estrutura-se como uma solução de Design Science Research;

(2) no nível tático, sua construção segue o fluxo CRISP-DM, especialmente nas etapas de compreensão e modelagem dos dados;

(3) no nível operacional, o artefato é preparado para implantação e governança contínua segundo práticas de ModelOps.

1.5.2 Objetivos Específicos

Os objetivos específicos detalham as etapas necessárias para que o objetivo geral seja alcançado, traduzindo-o em ações estruturadas de investigação, modelagem e validação. Cada objetivo se conecta de forma coerente à problemática, à hipótese e ao enquadramento metodológico, garantindo unidade textual e ausência de redundância entre seções. Além disso, os objetivos acompanham a lógica estratégica-tática-operacional estabelecida pela integração entre DSR, CRISP-DM e ModelOps. Segue-se abaixo o detalhamento:

1. Modelar a estrutura relacional do domínio, representando produtos, pontos de venda, atributos firmográficos e padrões transacionais por meio de grafos heterogêneos, de forma a capturar dependências estruturais reais entre entidades.
2. Investigar e implementar técnicas de aprendizado em grafos (Graph Neural Networks) e métodos causais adequados para estimar efeitos de intervenções comerciais, mitigando confusões típicas entre correlação e causalidade presentes em dados de mercado.
3. Integrar métodos relacionais e causais em um único artefato, projetado para inferir elasticidade, impacto promocional e respostas a recomendações, mesmo em cenários de dados escassos (cold start).
4. Desenvolver e conduzir o pipeline CRISP-DM sobre os datasets disponibilizados (“transacoes_vendas”, “produtos_catalogo” e “pontos_venda”), incluindo compreensão, preparação, análise exploratória, modelagem e avaliação.
5. Estabelecer mecanismos de governança, monitoramento e implantação contínua baseados em ModelOps, assegurando operacionalidade, rastreabilidade e robustez do artefato ao longo de seu ciclo de vida.

6. Avaliar empiricamente o desempenho do artefato, comparando-o com modelos não relacionais e não causais, de modo a verificar ganhos de acurácia, estabilidade e utilidade prática em decisões comerciais.
7. Documentar e comunicar a construção, demonstração e avaliação, conforme diretrizes do Design Science Research, garantindo reprodutibilidade, rigor metodológico e contribuição científica.

1.6 Abordagem Metodológica

A abordagem metodológica articula, de forma integrada, três camadas complementares: **Design Science Research (DSR)** como estrutura estratégica da investigação, **CRISP-DM** como método tático de condução do ciclo analítico e **ModelOps** como diretriz operacional para governança, implantação e monitoramento do artefato. Essa integração garante coerência entre o problema identificado, a construção da solução e sua sustentabilidade em ambiente real, preservando unidade textual e evitando sobreposição com seções anteriores.

No nível estratégico, o **Design Science Research** orienta a pesquisa à produção de um artefato inovador que responda ao problema raiz: a incapacidade dos modelos tradicionais de representar relações estruturais e distinguir efeitos causais em contextos comerciais B2B. A lógica do DSR — identificar o problema, projetar a solução, desenvolver, demonstrar, avaliar e comunicar — estrutura o percurso científico, assegurando rigor e relevância.

No nível tático, o **CRISP-DM** fundamenta o ciclo analítico sobre os datasets da distribuidora (“transacoes_vendas”, “produtos_catalogo” e “pontos_venda”), permitindo compreender o domínio, explorar padrões transacionais, preparar as bases relacionais e conduzir modelagens tanto estruturais (grafos heterogêneos) quanto causais. Essa camada serve como ponte entre teoria e implementação, fornecendo um fluxo claro e reprodutível para o tratamento dos dados e construção do modelo.

No nível operacional, **ModelOps** garante que o artefato desenvolvido — cujo núcleo combina modelos relacionais e instrumentos causais — seja versionado, auditável, monitorado e passível de implantação contínua. Essa camada inclui práticas de governança de dados, monitoramento de deriva, avaliação periódica de performance e ciclos de atualização, requisito essencial para que o artefato seja funcional e confiável em decisões reais de recomendação e precificação.

Essa integração metodológica permite que a pesquisa avance de maneira coerente: da identificação estratégica da lacuna científica, passando pela execução sistemática do ciclo analítico, até a consolidação operacional que assegura longevidade e aplicabilidade do artefato em ambientes corporativos de alta complexidade.

1.7 Estrutura do Trabalho

A estrutura do trabalho organiza-se de modo a refletir progressivamente o percurso lógico da pesquisa, articulando fundamentação teórica, desenvolvimento metodológico, engenharia do artefato e avaliação empírica. Cada capítulo cumpre uma função específica dentro da estratégia do Design Science Research, mantendo coesão interna e continuidade argumentativa com as seções anteriores, sem redundâncias.

O Capítulo 1 - Introdução apresenta o plano de fundo da pesquisa. Ele inicia pela contextualização do ambiente de distribuição B2B, avança para a problemática que motiva a investigação, formula a hipótese orientadora, justifica a relevância científica e prática do estudo e estabelece os objetivos geral e específicos. A seção é concluída com a descrição da abordagem metodológica e da estrutura global do documento.

O Capítulo 2 - Referencial Teórico oferece os fundamentos conceituais necessários para sustentar a pesquisa. São discutidos conceitos centrais, a evolução histórica das técnicas relacionadas ao tema, teorias e modelos explicativos ligados à modelagem estatística, causal e relacional, além de trabalhos correlatos que situam a contribuição da pesquisa no estado da arte.

O Capítulo 3 - Metodologia detalha o enquadramento metodológico adotado e os procedimentos executados ao longo do estudo. Segue-se a estrutura composta por identificação do problema, definição dos objetivos da solução, design e desenvolvimento, demonstração, avaliação e comunicação, permitindo rastreabilidade e rigor na condução da pesquisa.

O Capítulo 4 - Engenharia do Artefato descreve a construção técnica da solução proposta. Ele apresenta a arquitetura conceitual, lógica e física do artefato; mecanismos de governança de dados; especificação do pipeline; e o ciclo de vida completo da solução. Inclui-se a origem, armazenamento e dicionário dos dados; etapas de preparação e análise; estratégias de modelagem; validação; implantação; e rotinas de monitoramento e manutenção.

O Capítulo 5 - Análise dos Resultados reúne e discute as evidências empíricas obtidas. São apresentados os resultados do artefato, sua análise crítica e as limitações metodológicas identificadas durante a investigação experimental.

O Capítulo 6 - Conclusão sintetiza os achados e verifica o atendimento dos objetivos do estudo. Examina as contribuições científicas e práticas, limitações da pesquisa, perspectivas de trabalhos futuros, implicações e impactos potenciais, finalizando com as considerações finais.

O Capítulo 7 - Gestão do Projeto trata dos elementos administrativos e operacionais associados à execução do estudo. São examinadas a viabilidade da pesquisa, os riscos e estratégias de mitigação, o cronograma de atividades, os recursos necessários e as estratégias de disseminação dos resultados.

O Capítulo 8 - Referências Bibliográficas lista todas as obras citadas ao longo do trabalho, organizadas em conformidade com as normas da ABNT.

O Capítulo 9 - Glossário reúne termos técnicos e expressões-chave empregados no estudo, facilitando a compreensão de leitores de diferentes formações.

2. REFERENCIAL TEÓRICO

2.1 Conceitos e Definições Fundamentais

A compreensão dos conceitos fundamentais relacionados à modelagem relacional, inferência causal e métodos avançados de aprendizado estatístico é essencial para fundamentar o desenvolvimento do artefato proposto. Este referencial estabelece a base conceitual necessária para interpretar corretamente tanto a natureza do problema quanto as escolhas metodológicas realizadas.

No âmbito da **representação de dados**, os grafos constituem estruturas matemáticas capazes de modelar entidades e relações de forma explícita. Em sistemas reais, especialmente no contexto B2B, os dados apresentam múltiplos tipos de nós e arestas — produtos, pontos de venda, categorias, atributos e interações transacionais — caracterizando um **grafo heterogêneo**. Esse tipo de grafo permite capturar dependências estruturais complexas, diferentemente de modelos tabulares tradicionais que assumem independência entre observações (ZHAO et al., 2021).

A partir dessa representação, emerge o campo das **GNNs**, modelos capazes de aprender representações distribuídas de entidades a partir da topologia e dos atributos do grafo. No caso de grafos heterogêneos, destaca-se a arquitetura **HGT**, que introduz mecanismos de atenção sensíveis ao tipo de nó e de relação, permitindo tratar heterogeneidade estrutural de forma explícita e diferenciada (HU et al., 2020).

No domínio da **inferência causal**, o objetivo central é distinguir relações estruturais de dependência estatística dos verdadeiros efeitos causais produzidos por intervenções. A teoria do *Structural Causal Model* (SCM) fornece o formalismo para essa separação, permitindo estimar efeitos diretos e indiretos, identificar confundidores e decompor mecanismos de geração de dados (PEARL; MACKENZIE, 2018). No entanto, a simples modelagem causal não é suficiente em ambientes complexos, pois estimadores tradicionais tornam-se sensíveis à alta dimensionalidade e à presença de vieses de seleção.

Nesse contexto, o método **DML** estabelece um arcabouço robusto para estimar efeitos causais em cenários onde múltiplas variáveis influenciam simultaneamente tratamento e resultado. O DML utiliza dois modelos — um para o tratamento e outro para o resultado — e aplica técnicas de regularização e ortogonalização para reduzir vieses de confundimento, garantindo validade estatística mesmo com modelos complexos de aprendizado de máquina (CHERNOZHUKOV et al., 2018). A integração entre HGT e DML possibilita representar relações estruturais enquanto estima efeitos causais robustos, alinhando-se diretamente ao problema raiz da pesquisa.

Por fim, no campo de **aprendizado estatístico aplicado**, destacam-se conceitos como viés, variância, confundimento, indução estrutural, elasticidade, intervenção e contrafactual. Esses elementos são indispensáveis para compreender a interação entre estrutura relacional, comportamento de mercado e efeitos causais, permitindo formular modelos capazes de diferenciar padrões de correlação daqueles que resultam de ações comerciais direcionadas.

A combinação desses conceitos fundamenta a decisão de construir um artefato baseado em **Heterogeneous Graph Transformer** para capturar dependências estruturais e em **Double Machine Learning** para isolar efeitos causais, constituindo a base teórica necessária para enfrentar a lacuna identificada no domínio da distribuição de bebidas B2B.

2.2 Evolução Histórica

O desenvolvimento de métodos capazes de capturar relações estruturais e isolar efeitos causais passou por uma evolução marcada por avanços conceituais em três frentes: representação de dados, aprendizado estatístico e inferência causal. Essa trajetória histórica é fundamental para compreender por que a integração entre **HGT** e **DML** emerge como solução adequada ao problema investigado.

Nas primeiras décadas da modelagem estatística aplicada ao comportamento de mercado, predominavam modelos lineares e agregados, concebidos para ambientes com baixa dimensionalidade e estruturas relacionais simples. Esses modelos assumiam independência entre observações, tratavam variáveis como isoladas e raramente consideravam a influência conjunta de produtos, canais e pontos de venda. A ausência de representação explícita da estrutura relacional limitava a capacidade de capturar interações entre entidades, induzindo interpretações essencialmente correlacionais.

Com o crescimento de bases transacionais e o aumento da granularidade dos dados, surgiram abordagens capazes de modelar interdependências mais complexas. O avanço da teoria de redes e dos grafos, especialmente em contextos como recomendação, análise de redes sociais e sistemas multi-entidade, abriu caminho para a adoção de estruturas que representavam relações de maneira explícita. Entretanto, até meados dos anos 2010, os modelos de aprendizado em grafos ainda eram restritos a grafos homogêneos e não exploravam a heterogeneidade estrutural típica de domínios empresariais.

A introdução das **GNNs** marcou um ponto de inflexão ao permitir que algoritmos aprendessem representações de nós e relações de forma paramétrica. Posteriormente, a evolução para arquiteturas orientadas à heterogeneidade — culminando no **HGT** — possibilitou capturar padrões distintos de interação entre entidades de naturezas diferentes, incorporando mecanismos de atenção específicos para tipos de nós e arestas (HU et al., 2020). Esse avanço tornou viável modelar estruturas como cadeias de suprimento, catálogos de produtos, redes comerciais e perfis de clientes de forma mais fiel ao comportamento real do domínio.

Em paralelo, a literatura de inferência causal consolidou o paradigma dos **SCM**, introduzido por Pearl, que formalizou a distinção entre correlação e causalidade e permitiu identificar efeitos de intervenção de maneira consistente (PEARL; MACKENZIE, 2018). No entanto, métodos puramente paramétricos apresentavam limitações quando aplicados a dados de alta dimensionalidade ou a ambientes em que múltiplos fatores influenciam simultaneamente decisões comerciais e resultados observados — característica típica de mercados dinâmicos.

Para enfrentar essas limitações, a econometria moderna incorporou técnicas de aprendizado de máquina em seu arcabouço causal. O surgimento do **DML** representou um avanço decisivo ao permitir que modelos flexíveis fossem usados tanto para a previsão do tratamento quanto do desfecho, corrigindo vieses de confundimento por meio de ortogonalização e garantindo validade estatística (CHERNOZHUKOV et al., 2018). Esse marco histórico tornou possível aplicar inferência causal em larga escala, mesmo quando o espaço de covariáveis é extenso e estruturalmente complexo.

A convergência histórica entre aprendizado relacional e inferência causal cria o cenário atual, no qual integrar **HGT** e **DML** representa não apenas um avanço técnico, mas a síntese natural da evolução de duas áreas que historicamente caminharam em paralelo. Essa integração viabiliza a construção de modelos que representam entidades e relações com alta fidelidade e, simultaneamente, produzem estimativas causais robustas, respondendo diretamente às lacunas que motivam esta pesquisa.

2.3 Teorias e Modelos Explicativos

A compreensão da interação entre estrutura relacional e inferência causal exige um conjunto de teorias capazes de explicar tanto a natureza das dependências entre entidades quanto os mecanismos que produzem efeitos observáveis em sistemas comerciais complexos. Nesta seção, apresentam-se as bases teóricas que sustentam a integração entre **HGT** e **DML**, articulando modelos explicativos provenientes da ciência de dados, aprendizado estatístico e inferência causal.

A teoria dos **SCM** constitui o principal arcabouço para distinguir relações causais de meras associações estatísticas. Por meio de diagramas causais e equações estruturais, o SCM permite explicitar os mecanismos que governam a geração dos dados e identificar confundidores, mediadores e efeitos diretos e indiretos (PEARL; MACKENZIE, 2018). Essa teoria é central ao problema investigado, pois o domínio B2B de distribuição de bebidas apresenta múltiplas variáveis que interagem estruturalmente — preço, sortimento, perfil do PDV, sazonalidade, incentivos comerciais — tornando inadequadas abordagens que ignoram a estrutura causal subjacente.

Por outro lado, o aprendizado estatístico moderno oferece modelos com grande capacidade representacional, mas que, por si só, não garantem validade causal. Entre as arquiteturas mais relevantes, destacam-se as **GNNs**, concebidas para aprender representações distribuídas de nós e relações a partir da topologia do grafo. A evolução dessas arquiteturas levou ao desenvolvimento do **HGT**, que incorpora mecanismos de atenção dependentes do tipo de entidade e relação, proporcionando um modelo capaz de diferenciar padrões complexos de interação em grafos heterogêneos (HU et al., 2020). Essa capacidade é particularmente importante quando o objetivo é capturar fatores estruturais que interferem simultaneamente nas vendas e no impacto de intervenções promocionais.

Entretanto, ainda que o HGT consiga representar com precisão a estrutura relacional do domínio, ele não resolve o problema da separação entre estrutura e causalidade. Para isso, a pesquisa recorre aos fundamentos do **DML**, técnica que combina modelos flexíveis para previsão do tratamento e do desfecho, removendo vieses de confundimento por meio da ortogonalização das estimativas (CHERNOZHUKOV et al., 2018). O DML se apoia em princípios econométricos, como identificação, consistência e robustez assintótica, ao mesmo tempo em que se beneficia da expressividade dos modelos de machine learning.

A complementaridade entre HGT e DML ocorre porque cada método resolve um aspecto distinto do problema raiz: o HGT representa adequadamente a **estrutura relacional** e permite capturar dependências complexas entre entidades; o DML modela os **efeitos causais**, isolando o impacto de intervenções mesmo quando há múltiplos confundidores e alta dimensionalidade. Integrados, esses modelos dão origem a um artefato capaz de compreender tanto o que acontece no domínio quanto o porquê acontece, o que é fundamental para decisões comerciais baseadas em intervenções.

Além disso, as duas teorias dialogam com princípios amplos da ciência de dados e do aprendizado estatístico, incluindo o trade-off entre viés e variância, regularização, generalização, indução estrutural e modelagem contrafactual. No conjunto, essas teorias permitem explicar de maneira clara e consistente como a pesquisa enfrenta o problema da mistura entre correlação e causalidade, propondo um artefato que supera as limitações dos modelos atualmente utilizados no setor de distribuição.

2.4 Trabalhos Correlatos

A literatura recente sobre modelagem relacional e inferência causal em ambientes de recomendação e tomada de decisão fornece indicadores importantes para situar esta pesquisa no estado da arte. Os trabalhos correlatos mostram avanços significativos em três frentes principais: **aprendizado em grafos heterogêneos**, **modelos causais para intervenção em sistemas de recomendação**, e **integrações modernas entre métodos de machine learning e inferência causal**, incluindo aplicações do **DML**. A análise desses estudos evidencia tanto as contribuições existentes quanto as lacunas que justificam a construção do artefato proposto.

No campo da **modelagem em grafos heterogêneos**, trabalhos como Hu et al. (2020), Zhang et al. (2019) e Zhao et al. (2021) demonstram que arquiteturas com atenção relacional, como o **HGT**, alcançam desempenho superior em tarefas que exigem diferenciação entre múltiplos tipos de entidades e relações. Essas abordagens capturam padrões topológicos complexos e revelam que modelos convencionais — baseados em vetorização tabular — são insuficientes para representar redes comerciais amplas, como catálogos de produtos, perfis firmográficos e comportamentos de compra. No entanto, esses estudos concentram-se principalmente em tarefas preditivas correlacionais e não abordam estimativas causais.

Já a literatura em **inferência causal aplicada a sistemas de recomendação** tem avançado na separação entre correlação e causalidade, explorando métodos de *uplift modeling*, contrafactual prediction e econometria moderna. Trabalhos como Imbens e Rubin (2015) e Yao et al. (2021) apresentam frameworks robustos para estimar efeitos de tratamento em ambientes com múltiplos confundidores, enquanto pesquisas em recomendação causal examinam como intervenções, como recomendações personalizadas, influenciam comportamentos e resultados observáveis. Apesar disso, grande parte dessas abordagens assume observações independentes, não incorporando estruturas relacionais complexas — o que limita sua aplicabilidade a ecossistemas B2B.

A terceira frente relevante envolve a aplicação de **Double Machine Learning** à estimação de efeitos estruturais. Estudos como os de Chernozhukov et al. (2018) demonstram que o DML fornece estimativas consistentes mesmo em cenários de alta dimensionalidade, onde os confundidores são numerosos e interdependentes. Esses avanços aproximam aprendizado de máquina e inferência causal, mas ainda tratam as observações como independentes, deixando de considerar a natureza topológica do domínio.

Assim, embora existam contribuições significativas em modelagem relacional, recomendação causal e técnicas de correção de vieses, **nenhum dos trabalhos correlatos integra modelos relacionais de alta capacidade — como o HGT — com métodos robustos de inferência causal — como o DML — em um único artefato voltado para decisões comerciais em ambientes B2B**. Essa lacuna constitui exatamente o espaço científico onde a presente pesquisa se insere.

O artefato proposto busca sintetizar esses avanços, oferecendo uma abordagem capaz de representar explicitamente as relações estruturais entre produtos, PDVs e comportamentos transacionais, ao mesmo tempo em que estima efeitos causais de intervenções comerciais com rigor estatístico.

3. METODOLOGIA

3.1 Enquadramento Metodológico

O enquadramento metodológico desta pesquisa estrutura-se na integração entre **DSR**, **CRISP-DM** e **ModelOps**, constituindo um arcabouço capaz de articular, de forma coerente, o desenvolvimento científico, o processo analítico e a operacionalização contínua do artefato. Essa integração não apenas orienta a lógica investigativa, mas assegura que a solução gerada responda ao problema raiz — a ausência de modelos capazes de representar a estrutura relacional do domínio e de isolar efeitos causais de intervenções comerciais — de maneira rigorosa, sistemática e aplicável a ambientes reais.

No nível **estratégico**, o DSR fornece a estrutura geral do processo de pesquisa, organizando-o em ciclos iterativos que compreendem: identificação do problema e motivação, definição dos objetivos da solução, design e desenvolvimento, demonstração, avaliação e comunicação. Essa abordagem reforça o caráter construtivo e inovador do trabalho, uma vez que o artefato — um modelo híbrido baseado em **HGT** integrado a **DML** — representa uma solução original concebida especificamente para lidar com dependências estruturais e inferência causal simultaneamente. O DSR também assegura a relevância prática do estudo, pois orienta todas as etapas a partir das necessidades e limitações do ambiente B2B de distribuição de bebidas.

No nível **tático**, o CRISP-DM organiza o ciclo de construção e análise dos dados que servirão de base ao artefato, guiando as etapas de compreensão do negócio, entendimento dos dados, preparação, modelagem, avaliação e implantação inicial. Os datasets fornecidos ("*transacoes_vendas*", "*produtos_catalogo*" e "*pontos_venda*") são explorados segundo essa metodologia para identificar padrões, dependências e aspectos relacionais essenciais à construção do grafo heterogêneo. O CRISP-DM permite estruturar a investigação empírica com clareza e objetividade, garantindo reprodutibilidade e rigor técnico.

No nível **operacional**, o ModelOps estabelece os protocolos de versionamento, monitoramento e governança do artefato, assegurando sua continuidade operacional após a construção. Como o modelo resultado desta pesquisa combina dois elementos sofisticados — representação relacional via HGT e estimação causal via DML —, práticas de ModelOps tornam-se indispensáveis para garantir estabilidade, integridade e reprodutibilidade do sistema ao longo de seu ciclo de vida. Entre os elementos tratados estão: monitoramento de deriva, auditoria das decisões, versão de modelos e dados, rastreabilidade de experimentos e mecanismos de revalidação periódica.

Ao integrar DSR, CRISP-DM e ModelOps, o enquadramento metodológico garante que a pesquisa atue de forma sinérgica entre teoria, prática e operação contínua. Essa abordagem permite transitar do diagnóstico das limitações dos modelos tradicionais para o desenvolvimento e sustentação de uma solução robusta, orientada tanto à explicação científica quanto à aplicabilidade no ambiente comercial da distribuidora de bebidas.

3.2 Procedimentos Metodológicos

3.2.1 Identificação do Problema e Motivação

A identificação do problema parte da constatação empírica, teórica e operacional de que os modelos atualmente empregados no setor de distribuição B2B — especialmente os voltados à recomendação de produtos, precificação e alocação promocional — são essencialmente correlacionais e incapazes de representar as dependências estruturais que caracterizam o domínio. Essa limitação compromete a capacidade de isolar efeitos causais de intervenções comerciais, resultando em estimativas enviesadas de elasticidade, impacto promocional e resposta a recomendações. Tal cenário conduz a estratégias subótimas, sobretudo em ambientes com grande heterogeneidade de produtos, perfis firmográficos diversos e forte dependência sazonal, como indicado pelos dados fornecidos (“*transacoes_vendas*”, “*produtos_catalogo*” e “*pontos_venda*”).

Nesse contexto, o problema raiz manifesta-se em dois níveis complementares. No plano estrutural, produtos, categorias, PDVs, atributos firmográficos e eventos comerciais apresentam relações complexas que não são capturadas por modelos tabulares tradicionais, os quais tratam entidades interligadas como independentes. No plano causal, a mistura entre associação estatística e efeito de intervenção produz vieses que contaminam as estimativas utilizadas para tomada de decisão, fenômeno amplamente discutido nos fundamentos da inferência causal (PEARL; MACKENZIE, 2018). Essa combinação de limitações resulta em perda de precisão, incapacidade de generalização em cenários de cold start e dificuldade em mensurar efeitos incrementais de ações comerciais.

A motivação que sustenta esta pesquisa emerge justamente da necessidade de superar essa lacuna metodológica. A evolução recente das técnicas de aprendizado em grafos — particularmente o **HGT** — demonstra que é possível representar relações complexas entre entidades preservando a heterogeneidade estrutural do domínio (HU et al., 2020). Paralelamente, avanços na econometria moderna, como o **DML**, mostram que é possível estimar efeitos causais robustos mesmo em cenários de alta dimensionalidade e múltiplos confundidores (CHERNOZHUKOV et al., 2018). Contudo, tais avanços permanecem desconectados na literatura: modelos relacionais são aplicados majoritariamente para previsão correlacional, enquanto métodos causais assumem estruturas não relacionais.

A integração entre HGT e DML proposta nesta pesquisa surge como resposta direta à motivação dual de (i) representar a estrutura relacional do domínio e (ii) estimar efeitos causais com rigor estatístico. Essa abordagem está alinhada ao campo da ciência de dados e aprendizado estatístico contemporâneo, que reconhece a necessidade de unir modelagem estrutural e inferência causal em sistemas de decisão complexos. Assim, a identificação e fundamentação do problema delineiam não apenas um desafio técnico, mas um espaço científico ainda pouco explorado, justificando a relevância e a originalidade do artefato desenvolvido no âmbito desta pesquisa.

3.2.2 Definição dos Objetivos da Solução

A definição dos objetivos da solução estabelece, de forma precisa e operacional, aquilo que o artefato desenvolvido deve realizar para responder ao problema raiz identificado: a incapacidade dos modelos atuais de representar a estrutura relacional do domínio e de isolar efeitos causais de intervenções comerciais. Essa etapa traduz os objetivos gerais e específicos da pesquisa em requisitos funcionais e técnicos que guiam o design, o desenvolvimento e a posterior avaliação do artefato, seguindo as diretrizes estratégicas do Design Science Research (DSR).

O primeiro objetivo da solução consiste em **representar adequadamente a topologia do domínio B2B da distribuidora de bebidas**, incorporando produtos, pontos de venda, atributos firmográficos, relações de compra, complementaridade e substituição. Para isso, o artefato deve modelar o ambiente como um **grafo heterogêneo**, permitindo que dependências estruturais sejam tratadas como elementos centrais da modelagem, superando limitações dos modelos tabulares tradicionais. A escolha metodológica pelo **Heterogeneous Graph Transformer (HGT)** decorre justamente de sua capacidade de lidar com múltiplos tipos de nós e relações, empregando mecanismos de atenção que diferenciam padrões de interentidades (HU et al., 2020).

O segundo objetivo é **estimar efeitos causais de intervenções comerciais com rigor estatístico**, superando vieses de confundimento decorrentes da simultaneidade entre decisões comerciais e comportamento de mercado. Para isso, a solução deve utilizar o arcabouço do **DML**, que separa a modelagem do tratamento e do desfecho, aplicando ortogonalização para garantir estimativas robustas mesmo em cenários de alta dimensionalidade (CHERNOZHUKOV et al., 2018). O uso de dois HGTs — um para modelar o mecanismo do tratamento e outro para modelar o resultado — atende aos requisitos do DML, garantindo flexibilidade e expressividade na modelagem causal.

O terceiro objetivo consiste em **integrar modelagem relacional e inferência causal em um único artefato operacional**, permitindo estimar elasticidade, impacto promocional, efeitos de recomendação e cenários contrafactuais. Essa integração deve ocorrer de forma coerente e auditável, respeitando as exigências de explicabilidade e rastreabilidade necessárias em ambientes corporativos, especialmente no contexto de ModelOps. O artefato deve ainda ser capaz de operar em cenários de cold start, suprindo um dos principais desafios enfrentados pelo setor.

O quarto objetivo é **possibilitar avaliação empírica rigorosa** do artefato, comparando seu desempenho com alternativas correlacionais, modelos não relacionais e estimadores causais tradicionais. Essa avaliação deve incorporar métricas relacionais, métricas de qualidade causal e indicadores de utilidade prática, permitindo verificar se o artefato atinge os objetivos funcionais estabelecidos e se gera vantagem decisória mensurável.

Por fim, o quinto objetivo da solução é **assegurar que o artefato possa ser implantado e mantido operacionalmente**, incorporando práticas de governança, versionamento, monitoramento de deriva e validação contínua, conforme exige a integração entre ModelOps e os processos de decisão da distribuidora.

A definição desses objetivos fornece a base para os ciclos subsequentes de design, desenvolvimento, demonstração e avaliação, assegurando que o artefato final seja não apenas tecnicamente sofisticado, mas alinhado às necessidades reais e mensuráveis do domínio analisado.

3.2.3 Design e Desenvolvimento

O design e o desenvolvimento do artefato constituem o núcleo construtivo da pesquisa, no qual as diretrizes do Design Science Research são operacionalizadas para transformar o problema identificado e os objetivos da solução em uma arquitetura funcional. Nesta etapa, definem-se explicitamente os componentes técnicos, as estruturas de dados, os modelos de aprendizado estatístico e os mecanismos de integração entre representação relacional e inferência causal. O processo adota uma abordagem iterativa, refletindo ciclos sucessivos de construção, experimentação e refinamento.

O ponto de partida do design consiste na **modelagem do domínio como um grafo heterogêneo**, utilizando os datasets fornecidos (“transacoes_vendas”, “produtos_catalogo” e “pontos_venda”) como base para a definição dos nós, arestas e metadados. Cada entidade do domínio — produto, PDV, categoria, marca, região, perfil firmográfico, transação — é representada como um tipo distinto de nó. Relações como compra, coocorrência, frequência, substituição e complementaridade são representadas como arestas heterogêneas. Essa estrutura reflete a topologia real do ambiente B2B e estabelece a base sobre a qual o **Heterogeneous Graph Transformer (HGT)** operará (HU et al., 2020).

Com o grafo estruturado, o desenvolvimento avança para a concepção dos **dois modelos HGT** que compõem a espinha dorsal do artefato:

1. **HGT-T (*Treatment Model*)**: modelo responsável por representar as relações que influenciam a probabilidade de um PDV receber determinada intervenção comercial (desconto, recomendação, ação promocional). Esse modelo estima a função de propensão, elemento fundamental no arcabouço do Double Machine Learning.

2. **HGT-Y (*Outcome Model*)**: modelo responsável por representar as relações que influenciam o desfecho observado (vendas, volume incremental, elasticidade, resposta à recomendação). Esse modelo estima a função condicional do resultado.

A escolha por dois HGTs decorre diretamente das exigências do **DML**, que requer estimadores distintos e flexíveis para o mecanismo do tratamento e para o mecanismo do resultado, seguidos de um processo de **ortogonalização** que corrige vieses de confundimento (CHERNOZHUKOV et al., 2018). Ambos os modelos são treinados de forma independente, com entradas construídas a partir das features relacionais extraídas do grafo heterogêneo.

Após o treinamento dos dois modelos, procede-se à etapa de **integração causal**, na qual as previsões de propensão e de resultado são combinadas para estimar o efeito causal da intervenção, segundo o arcabouço do DML. Essa integração é realizada por meio de regressões ortogonalizadas, que garantem consistência assintótica das estimativas mesmo em condições de alta dimensionalidade e estrutura relacional complexa. O resultado é um estimador causal robusto capaz de inferir efeitos incrementais e contrafactuais.

O design contempla ainda a implementação de **mecanismos de avaliação interna**, incluindo métricas de reconstrução relacional, métricas de acurácia preditiva, métricas causais (ATE, CATE, uplift), testes de robustez e experimentos sintéticos baseados em contrafactuais simulados. Essas métricas são utilizadas iterativamente para ajustar hiperparâmetros, refinar a estrutura do grafo e aprimorar a integração entre os modelos.

Finalmente, o desenvolvimento incorpora desde o início as práticas de **ModelOps**, garantindo versionamento de dados e modelos, rastreabilidade de experimentos, monitoramento de deriva e preparação para implantação contínua. O artefato é projetado para operar em ciclos sucessivos, permitindo sua adaptação às mudanças naturais no mercado de bebidas, como sazonalidade, dinâmica competitiva e variações regionais.

Desse modo, o processo de design e desenvolvimento constrói uma solução tecnicamente integrada, alinhada aos requisitos científicos da pesquisa e às demandas operacionais do domínio B2B.

3.2.4 Demonstração

A etapa de demonstração, conforme previsto no Design Science Research, tem como finalidade evidenciar que o artefato desenvolvido é capaz de operar no domínio real para o qual foi projetado e de resolver, de maneira funcional, o problema raiz identificado. A demonstração não objetiva ainda uma avaliação conclusiva, mas sim a verificação prática e preliminar da utilidade, consistência e operacionalidade do artefato composto pela integração entre **HGT** e **DML**.

A demonstração inicia-se pela **construção do grafo heterogêneo real** utilizando os datasets disponibilizados (“transacoes_vendas”, “produtos_catalogo”, “pontos_venda”). Essa etapa envolve estruturar entidades e relações do domínio em uma topologia que reflita a complexidade das interações presentes no ambiente B2B. Cada PDV, produto, categoria e atributo firmográfico torna-se um nó associado a suas características, enquanto transações, coocorrências, vínculos comerciais e relações de substituição ou complementaridade são formalizadas como arestas heterogêneas.

Com o grafo construído, procede-se à aplicação prática do artefato. O primeiro componente, **HGT-T**, é treinado para estimar a propensão de cada PDV receber determinados tratamentos comerciais (descontos, campanhas, recomendações). O segundo componente, **HGT-Y**, é treinado para estimar os desfechos associados ao comportamento de vendas, incluindo efeitos diretos e dependências estruturais. Em conjunto, esses modelos produzem as previsões necessárias para a posterior aplicação das transformações causais do DML.

A demonstração, então, realiza a **ortogonalização das estimativas** segundo o arcabouço do Double Machine Learning (CHERNOZHUKOV et al., 2018). Nesse processo, as previsões dos dois modelos são utilizadas para construir estimadores causais preliminares capazes de mensurar efeitos incrementais, como elasticidade, impacto promocional e uplift em vendas. A partir disso, a demonstração evidencia a capacidade do artefato de separar correlação de causalidade, distinguindo efeitos derivados da estrutura relacional daqueles resultantes de intervenções comerciais.

Para tornar essa demonstração empiricamente clara, são realizadas execuções em **cenários reais e sintéticos**. Os cenários reais reproduzem a dinâmica observada na base de vendas; já os sintéticos permitem controlar mecanismos causais artificiais para verificar se o artefato recupera efeitos previamente conhecidos. Essa estratégia é particularmente importante para validar a robustez inicial das estimativas e garantir que o artefato se comporta conforme o esperado antes de etapas mais profundas de avaliação.

Além disso, a demonstração incorpora elementos de operacionalização previstos em ModelOps: rastreabilidade das execuções, registro de artefatos gerados, padronização de pipelines e versionamento dos modelos. Essa estrutura assegura que os experimentos conduzidos nesta fase possam ser reproduzidos e auditados, estabelecendo confiança para a transição às etapas posteriores.

Assim, a demonstração comprova que o artefato é funcional, capaz de integrar modelagem relacional e inferência causal, e apto a ser submetido à avaliação rigorosa na etapa subsequente, demonstrando consistência com os requisitos metodológicos do DSR e com a realidade operacional da distribuidora de bebidas.

3.2.5 Avaliação

A etapa de avaliação tem como objetivo verificar, de maneira rigorosa e sistemática, se o artefato desenvolvido atende aos requisitos funcionais definidos anteriormente, resolve o problema raiz identificado e produz estimativas causais e relacionais consistentes com a complexidade estrutural do domínio B2B da distribuidora de bebidas. Em conformidade com o Design Science Research, essa avaliação combina métodos quantitativos, experimentais e analíticos, garantindo robustez científica e aplicabilidade prática.

A avaliação inicia-se com a análise de desempenho dos dois modelos que compõem o artefato: **HGT-T**, responsável pela estimação do mecanismo do tratamento, e **HGT-Y**, responsável pela estimação do mecanismo do desfecho. Para cada modelo, são examinadas métricas relacionais — como precisão de classificação em subgrafos, reconstrução de arestas, qualidade das embeddings e métricas dependentes de estrutura, conforme indicado na literatura de aprendizado em grafos (HU et al., 2020). Esses indicadores permitem verificar se o artefato capturou adequadamente a heterogeneidade do domínio, garantindo que as representações aprendidas sejam compatíveis com a topologia real dos dados.

A segunda dimensão da avaliação concentra-se na **inferência causal**, utilizando medidas de eficácia derivadas do arcabouço do **DML**. Avaliam-se métricas como: efeito médio do tratamento (ATE), efeito condicional do tratamento (CATE), estimativas de uplift, estabilidade das estimativas e variação residual após ortogonalização (CHERNOZHUKOV et al., 2018). Para fortalecer a validade das inferências, testes são realizados em dois tipos de cenários:

1. **Cenários reais**, baseados nas vendas observadas, para examinar a capacidade do modelo de reproduzir efeitos plausíveis e coerentes com o comportamento histórico do mercado.
2. **Cenários sintéticos**, nos quais mecanismos causais controlados são introduzidos artificialmente, permitindo verificar se o artefato recupera efeitos conhecidos e neutraliza confundidores intencionais.

A avaliação também considera a comparação do artefato com **modelos correlacionais tradicionais**, como regressões tabulares, recomendadores baseados em similaridade e modelos lineares ou de árvore. Essa comparação é fundamental para demonstrar empiricamente a superação das limitações estruturais e causais que motivaram a pesquisa. A expectativa teórica é que o artefato apresente menor viés, maior estabilidade e maior capacidade de generalização, especialmente em contextos de cold start e sob intervenções comerciais.

Além dos aspectos preditivos e causais, a avaliação inclui métricas operacionais relacionadas ao **ciclo ModelOps**, como rastreabilidade de experimentos, consistência entre versões, sensibilidade à deriva de dados, latência de inferência e comportamento do artefato sob cargas variáveis. Essas métricas asseguram que a solução não apenas tenha mérito técnico, mas seja operacionalmente viável em cenários reais de decisão.

Por fim, os resultados da avaliação são analisados à luz dos objetivos da solução definidos na Seção 3.2.2, estabelecendo se o artefato atinge o nível esperado de desempenho relacional, validade causal e capacidade operacional. Caso lacunas sejam identificadas, novas iterações do ciclo DSR são iniciadas, garantindo refinamento contínuo até que os requisitos do domínio e da pesquisa sejam plenamente atendidos.

3.2.6 Comunicação

A etapa de comunicação no âmbito do Design Science Research tem como função garantir que o artefato, o conhecimento produzido e os resultados obtidos sejam apresentados de forma clara, estruturada e alinhada aos requisitos acadêmicos e profissionais. Essa fase assegura que os achados da pesquisa sejam compreendidos por diferentes públicos — acadêmicos, especialistas técnicos em ciência de dados e gestores do setor de distribuição de bebidas — cumprindo o papel essencial de disseminar contribuições científicas e práticas.

A comunicação inicia-se pela **documentação sistemática** de todas as etapas do desenvolvimento: formação do grafo heterogêneo, construção dos modelos HGT-T e HGT-Y, integração causal via Double Machine Learning, execução dos experimentos e avaliação final. Cada etapa é descrita seguindo padrões formais de redação acadêmica e utilizando linguagem precisa, permitindo reprodutibilidade e auditoria. Essa documentação também integra conceitos e estruturas provenientes do CRISP-DM e de práticas de ModelOps, destacando decisões técnicas, parâmetros, versões de dados e modelos, além das justificativas metodológicas.

Outro componente essencial da comunicação consiste na **apresentação dos resultados**, que deve evidenciar como o artefato responde ao problema raiz — a dificuldade de representar estrutura relacional e de isolar efeitos causais — demonstrando ganhos concretos em relação aos modelos tradicionais. A exposição dos resultados contempla visualizações, tabelas, métricas causais e relacionais, além de análises comparativas que reforçam a validade e a utilidade da solução desenvolvida.

A comunicação também envolve a **contextualização da contribuição científica** do artefato, situando-o dentro do estado da arte, conforme discutido no referencial teórico. Essa contextualização demonstra como a integração entre HGT e DML preenche lacunas existentes na literatura e oferece um caminho metodológico para modelagem conjunta de estrutura relacional e inferência causal em ambientes B2B.

No âmbito técnico-operacional, a comunicação incorpora práticas de **transparência e explicabilidade**, incluindo descrições sobre o funcionamento interno dos modelos, estrutura das atenções no HGT, interpretações causais das estimativas e implicações da ortogonalização no DML. Esses elementos são fundamentais para facilitar a adoção do artefato por equipes técnicas e decisores de negócio.

Por fim, a disseminação dos resultados ocorre por meio da elaboração do documento acadêmico completo, apresentações técnicas, relatórios executivos e, quando pertinente, submissões a eventos e periódicos científicos da área de ciência de dados, aprendizado estatístico e sistemas de recomendação. Essa multiplicidade de meios garante que o conhecimento gerado alcance diferentes audiências e fortaleça o impacto científico e prático da pesquisa.

4. ENGENHARIA DO ARTEFATO

Argumentação Central: Esta seção detalha a arquitetura e a implementação do artefato proposto. É onde "a engenharia encontra a ciência": a abstração metodológica do DSR e CRISP-DM é transformada em um projeto de software concreto e escalável. A arquitetura deve garantir reprodutibilidade científica e robustez operacional.

Argumentação Secundária:

1. Explicar que a arquitetura do sistema foi desenhada para suportar todo o ciclo de vida do artefato, atendendo requisitos de pesquisa (DSR) e de processo (CRISP-DM) em uma plataforma operacional (ModelOps).
2. Justificar as escolhas de ferramentas e tecnologias que concretizam a solução.
3. Apresentar diagramas conceituais, lógicos e físicos que ilustram o fluxo de dados e a infraestrutura.
4. Garantir que a descrição arquitetural enfatize aspectos de reprodutibilidade, escalabilidade, governança de dados e automação de pipelines.

4.1 Arquitetura

Argumentação Central: Introduz-se a visão geral da arquitetura, explicando que ela foi planejada para suportar todo o ciclo de vida do artefato, garantindo que requisitos de pesquisa e de processo sejam satisfeitos por uma estrutura operacional robusta.

Argumentação Secundária:

1. Apresentar o papel da arquitetura como elemento unificador entre DSR, CRISP-DM e ModelOps (suportando requisitos tanto de pesquisa quanto de operação).
2. Descrever de forma geral o sistema como uma "caixa preta" inicialmente, focando no comportamento esperado.
3. Enfatizar que a arquitetura será detalhada nos níveis conceitual, lógico e físico nos próximos tópicos.
4. Destacar a importância de diagramas (contexto, pipeline, deployment) para visualizar as interações.

4.1.1 Arquitetura Conceitual (O "Quê" e "Porquê")

A arquitetura conceitual estabelece a visão abstrata do artefato desenvolvido, delineando seus componentes essenciais, suas interações e o modo como cada parte contribui para resolver o problema raiz: a incapacidade dos modelos tradicionais de representar a estrutura relacional do domínio e de isolar efeitos causais de intervenções comerciais. Essa arquitetura funciona como o nível mais alto de planejamento da solução, antecipando como os elementos teóricos — Heterogeneous Graph Transformer (HGT), Double Machine Learning (DML), DSR, CRISP-DM e ModelOps — se integram de forma coerente e funcional.

O ponto central da arquitetura é o **Grafo Heterogêneo do Domínio B2B**, construído a partir dos datasets fornecidos ("transacoes_vendas", "produtos_catalogo" e "pontos_venda"). Esse grafo representa as entidades relevantes — produtos, pontos de venda, marcas, categorias, atributos firmográficos, regiões e transações — como nós de tipos distintos, enquanto relações como compra, coocorrência, fluxo promocional, complementaridade e substituição são estruturadas como arestas heterogêneas. Assim, o grafo constitui a base de conhecimento estrutural do artefato.

A partir desse grafo, a arquitetura conceitual define um **módulo de representação relacional** composto por dois modelos:

1. **HGT-T (Treatment Encoder)** – aprende as representações que explicam o processo gerador do tratamento comercial (quem recebeu desconto, recomendação ou promoção).

2. **HGT-Y (Outcome Encoder)** – aprende as representações que explicam o processo gerador do resultado comercial (vendas, elasticidade, volume incremental).

Essas duas representações são necessárias para alimentar o núcleo causal do artefato, que segue o arcabouço do **Double Machine Learning (DML)**. O módulo causal utiliza as previsões estruturais geradas pelos dois HGTs para construir funções ortogonalizadas que removem confundimento e permitem estimar efeitos causais com validade estatística. Esse módulo produz estimativas como ATE, CATE, uplift e elasticidade causal.

A arquitetura conceitual também incorpora um **módulo de experimentação e avaliação**, responsável por executar contrafactuais, testes sintéticos e avaliações reais, garantindo que o artefato responda aos objetivos metodológicos definidos. Esse módulo considera tanto métricas relacionais quanto métricas causais, refletindo o caráter duplo do modelo.

Em paralelo, a arquitetura define uma camada de **Governança e Operacionalização (ModelOps)**, que engloba:

– versionamento de dados e modelos;
– rastreabilidade de experimentos;
– monitoramento de deriva de dados e modelo;
– auditoria das estimativas causais;
– procedimentos de revalidação e atualização contínua.

Por fim, a arquitetura conceitual conecta todos esses módulos ao ambiente de decisão da distribuidora de bebidas, permitindo que recomendações e estimativas causais possam ser consumidas por sistemas internos, processos de planejamento comercial e equipes de vendas.

Assim, a arquitetura conceitual sintetiza o desenho estratégico da solução: um artefato híbrido, orientado a grafos e causalidade, capaz de representar o domínio com fidelidade estrutural e de gerar inferências robustas para apoiar decisões comerciais complexas.

4.1.2 Arquitetura Lógica (O "Como" – Fluxo e Componentes)

A arquitetura lógica detalha como os componentes conceituais do artefato são organizados em fluxos funcionais, definindo módulos, interfaces, dependências e a lógica de processamento que sustenta a solução. Diferentemente da arquitetura conceitual — que estabelece a visão abstrata — a arquitetura lógica específica **como** cada parte opera e se conecta dentro do sistema, mantendo coerência com o DSR (como diretriz estratégica), CRISP-DM (como estrutura tática de análise) e ModelOps (como camada operacional contínua).

O sistema é estruturado em quatro macro camadas lógicas: **Ingestão e Representação dos Dados**, **Modelagem Relacional**, **Inferência Causal**, e **Serviços Operacionais**.

A primeira camada, **Ingestão e Representação dos Dados**, incorpora os datasets “transacoes_vendas”, “produtos_catalogo” e “pontos_venda”. A lógica dessa camada é organizar os dados tabulares em estruturas relacionais padronizadas e, em seguida, convertê-los em uma ontologia operacional do domínio. A partir dessa ontologia é construído o **grafo heterogêneo**, onde cada tipo de entidade é mapeado para um nó e cada tipo de interação comercial é mapeado para uma aresta. Essa camada inclui mecanismos de validação de consistência, detecção de duplicidades, tratamento de valores ausentes e padronização de atributos categóricos.

A segunda camada, **Modelagem Relacional**, implementa o coração neural da arquitetura: dois modelos Heterogeneous Graph Transformer (HGT). O primeiro modelo — denominado **HGT-T** — aprende a função de tratamento, isto é, as características estruturais que explicam por que um PDV ou produto recebe uma intervenção comercial (promoção, recomendação, desconto). O segundo modelo — denominado **HGT-Y** — aprende a função de resultado, identificando os padrões relacionais que influenciam vendas, participação de mercado e elasticidade. Ambos os modelos compartilham a mesma estrutura lógica de entrada (o grafo), mas operam objetivos distintos e produzem embeddings independentes, preservando a ortogonalidade necessária para o módulo causal.

A terceira camada, **Inferência Causal**, implementa o arcabouço do Double Machine Learning (DML). Nessa etapa, as representações produzidas pelos dois HGTs são processadas por funções de ortogonalização, que removem confundimento estrutural e permitem estimar efeitos causais de forma robusta. A camada lógica inclui:

– estimadores de efeito médio do tratamento (ATE);
– estimadores de efeito heterogêneo do tratamento (CATE);
– estimativas de uplift;
– elasticidades causais;
– operadores de simulação contrafactual.

A lógica dessa camada está centrada na separação entre o modelo do tratamento e o modelo do resultado, princípio central do DML (Chernozhukov et al., 2018).

A quarta camada, **Serviços Operacionais**, traduz os resultados em serviços utilizáveis pela distribuidora. Essa camada contém APIs internas que expõem recomendações, estimativas de efeito, prioridades comerciais e simulações contrafactuais. Também integra mecanismos ModelOps: rastreamento de experimentos, governança de versões, monitoramento contínuo de deriva de dados e de performance e gatilhos de re-treinamento. Essa camada garante que o artefato opere de modo confiável em ambientes reais, mantendo ciclo de vida completo e auditável.

Assim, a arquitetura lógica transforma o desenho conceitual em um sistema operacionalmente claro, sequencial e modular, no qual cada componente cumpre função específica para representar relações, estimar causalidade e apoiar decisões comerciais complexas.

4.1.3 Arquitetura Física (O "Onde" e "Com o Quê" – Infraestrutura)

A arquitetura física descreve a materialização do artefato em termos de infraestrutura computacional, tecnologias utilizadas, distribuição dos componentes e requisitos de desempenho para suportar a execução dos modelos HGT + DML e o ciclo operacional ModelOps. Essa camada traduz a arquitetura lógica em um ecossistema executável, garantindo escalabilidade, rastreabilidade e continuidade operacional ao longo de todo o ciclo de vida do artefato.

A infraestrutura é organizada em três domínios físicos: **camada de dados**, **camada de processamento e modelagem**, e **camada de serviços e operações**.

A **camada de dados** é responsável por armazenar e fornecer acesso eficiente aos datasets estruturantes: "*transacoes_vendas*", "*produtos_catalogo*" e "*pontos_venda*". O armazenamento é mantido no banco relacional otimizado para consultas analíticas denominado Neo4j, garantindo consistência transacional, versionamento de tabelas e metadados. O dicionário de metadados, como visto em 4.4.1.3, é mantido no repositório versionado do Github, permitindo rastreabilidade e garantindo governança formal.

A **camada de processamento e modelagem** concentra os recursos necessários para executar os dois Heterogeneous Graph Transformers e os módulos de Double Machine Learning. Ambientes baseados em GPU, como clusters CUDA ou instâncias aceleradas (A100/T4), asseguram que a computação tensorial e a manipulação de grafos heterogêneos possam ocorrer com baixa latência. Os dois modelos HGT, bem como o módulo causal, são implementados em frameworks como PyTorch Geometric e cuGraph, que oferecem suporte otimizado para operações em grafos de larga escala. O pipeline de treinamento é orquestrado pela Airflow, uma ferramenta de *workflow* que permite agendamentos recorrentes, reprocessamento seletivo e execução paralela.

A **camada de serviços e operações** materializa o artefato como um sistema acessível via APIs, integrando os resultados gerados pelos modelos ao ambiente comercial da distribuidora. Essa camada inclui contêineres Docker versionados e implantados em uma plataforma orquestrada por Kubernetes, que garante resiliência, escalonamento automático e isolamento de serviços. Mecanismos ModelOps são implementados pela ferramenta MLflow, cobrindo versionamento de modelos, registro de métricas, monitoramento contínuo de deriva de dados e disparo automatizado de novos treinos. Logs, auditorias e trilhas de execução são persistidos para garantir rastreabilidade e conformidade.

Essa arquitetura física permite que o artefato opere de forma robusta, auditável e escalável, conectando os modelos de aprendizado estatístico à prática operacional de tomada de decisão e assegurando que o sistema possa evoluir com novos dados e novas versões dos modelos, preservando a continuidade metodológica requerida pelo DSR, CRISP-DM e ModelOps.

4.2 Governança de Dados

A governança de dados estabelece o conjunto de políticas, processos e mecanismos técnicos que asseguram integridade, qualidade, segurança, transparência e rastreabilidade durante todo o ciclo de vida dos dados utilizados pelo artefato. Essa camada é essencial para viabilizar a execução contínua do CRISP-DM, sustentar a reprodutibilidade exigida pelo DSR e atender às práticas de ModelOps, nas quais decisões automatizadas dependem de fluxos de dados confiáveis e auditáveis.

O primeiro eixo da governança é a **qualidade dos dados**, garantida por procedimentos sistemáticos de verificação, padronização e reconciliação. Os datasets “transacoes_vendas”, “produtos_catalogo” e “pontos_venda” passam por regras de consistência estrutural, validação referencial e detecção de anomalias. Esses mecanismos asseguram coerência entre entidades — por exemplo, cada produto presente na tabela de transações deve existir no catálogo, e cada PDV deve ser encontrado na base de pontos de venda. Métricas de qualidade, como completude, acurácia e confiabilidade temporal, são registradas e monitoradas continuamente.

O segundo eixo é a **segurança e conformidade**, que envolve controle de acessos, anonimização e políticas de minimização de dados. A arquitetura física utiliza mecanismos de autenticação e autorização em nível de banco e serviço, enquanto os logs são criptografados e armazenados em ambientes seguros. Todas as informações passam por processos de pseudo-anonimização para evitar exposição de dados sensíveis no contexto B2B, garantindo conformidade com legislações como a LGPD.

O terceiro eixo da governança é o **versionamento e rastreabilidade**, que permite reconstituir, a qualquer momento, o estado exato dos dados utilizados para treinar um modelo. Cada atualização dos datasets recebe um identificador de versão, acompanhado de metadados sobre data de captura, transformações aplicadas e regras de validação empregadas. Esse versionamento garante reprodutibilidade experimental e permite auditoria completa dos ciclos de ModelOps.

O quarto eixo é a **padronização semântica**, que unifica nomenclaturas, taxonomias e descrições operacionais. As três bases utilizadas são alinhadas a um dicionário de dados central que descreve o significado de cada atributo, unidade de medida, cardinalidade, regras de preenchimento e vínculos ontológicos. Essa padronização é fundamental para a construção do grafo heterogêneo, pois evita inconsistências semânticas que possam contaminar o aprendizado relacional.

Por fim, o quinto eixo é o **monitoramento operacional**, que acompanha alterações inesperadas no comportamento dos dados. Indicadores como deriva de distribuição, mudanças de densidade relacional, surgimento de novas entidades e flutuações na frequência de transações são analisados continuamente. Caso padrões anômalos sejam detectados, mecanismos de alerta orientam processos de correção, garantindo estabilidade e prevenindo degradação dos modelos HGT e do módulo DML.

A governança de dados, assim estruturada, sustenta a robustez do artefato e garante que a estimativa causal e relacional se mantenha confiável ao longo do tempo, preservando tanto a integridade metodológica quanto a consistência operacional do sistema.

4.3 Pipeline

O pipeline operacionaliza o fluxo completo do artefato, estruturando a transição das etapas de ingestão, preparação, modelagem, inferência causal e entrega dos resultados, dentro de uma cadeia contínua alinhada ao CRISP-DM (nível tático) e sustentada por práticas de ModelOps (nível operacional). Essa organização garante rastreabilidade, modularidade, reprodutibilidade e capacidade de atualização contínua dos modelos HGT e do módulo DML.

O pipeline inicia-se com a fase de **ingestão dos dados**, na qual os arquivos "*transacoes_vendas*", "*produtos_catalogo*" e "*pontos_venda*" são extraídos das fontes originais e submetidos a rotinas de verificação estrutural, integridade referencial e validação semântica. Esses dados são então armazenados em camadas segregadas entre *raw*, *clean* e *semantic*, preservando todas as versões em conformidade com as diretrizes de governança.

A etapa seguinte corresponde à **construção e atualização do grafo heterogêneo**, que traduz os dados tabulares em uma representação relacional do domínio. O pipeline executa algoritmos de geração de nós, criação de arestas, agregação temporal e normalização de atributos. Essa etapa é sensível à deriva estrutural, razão pela qual inclui validadores automáticos que comparam a topologia atual com versões anteriores.

A terceira etapa compreende a **modelagem relacional**, implementada por dois Heterogeneous Graph Transformers:

– **HGT-T**, responsável por modelar a propensão ao tratamento (promoção, recomendação ou alteração de preço);

– **HGT-Y**, destinado a modelar os resultados potenciais (vendas, elasticidade e resposta comercial).

Ambos são treinados em ciclos independentes, com logs completos de hiperparâmetros, versões de dados e métricas registradas em sistemas como MLflow.

A quarta etapa corresponde ao módulo de **Double Machine Learning**, no qual as representações aprendidas pelos dois HGTs são processadas para remover o confundimento e estimar efeitos causais. O pipeline operacionaliza validações cruzadas, ortogonalização, cálculo de ATE e CATE, além de análises contrafactuais. Todos os resultados são armazenados com versionamento para auditoria.

Em seguida, o pipeline executa a etapa de **validação e integração**, na qual os outputs causais são comparados a benchmarks históricos, métricas de estabilidade temporal, testes de robustez e indicadores de qualidade operacional. Alterações significativas nos padrões causais acionam rotinas automatizadas de revisão ou reprocessamento.

Por fim, o pipeline culmina na **entrega operacional**, que expõe recomendações, estimativas de efeito, matrizes de elasticidade e simulações contrafactuais por meio de APIs versionadas. Esses serviços alimentam sistemas internos da distribuidora e são monitorados continuamente quanto a latência, disponibilidade, deriva e integridade dos dados.

Assim configurado, o pipeline garante fluidez, governança e consistência entre etapas, permitindo que o artefato evolua de forma contínua e auditável em ambientes empresariais reais.

4.4 Ciclo de Vida do Artefato

4.4.1 Banco de Dados

4.4.1.1 Origem

Os dados utilizados provêm de três fontes internas de dados transacionais fornecidas anonimamente por uma distribuidora americana parceira de médio porte e representam a base operacional real do domínio B2B no setor de bebidas:

1. **transacoes_vendas**
   Base transacional contendo registros de vendas realizadas ao longo de 12 meses. Representa o fluxo comercial entre PDVs e produtos, incluindo quantidades vendidas, datas, preços aplicados e descontos incidentes.
2. **produtos_catalogo**
   Catálogo de produtos comercializados, contendo atributos como SKU, categoria, marca, volume, embalagens, e características intrínsecas que influenciam demanda e elasticidade.
3. **pontos_venda**
   Base firmográfica dos pontos de venda, incluindo informações sobre localização, tipo de estabelecimento, segmento, porte, perfil de compra e histórico de relacionamento com a distribuidora.

As três bases fornecem a matéria-prima para a construção do grafo heterogêneo utilizado pelos modelos HGT, permitindo representar relações estruturais entre PDVs, produtos e transações. Essa origem única e padronizada assegura a consistência e a fidelidade do artefato ao contexto real do setor de distribuição.

4.4.1.2 Armazenamento e Gerenciamento

O armazenamento segue uma estratégia híbrida combinando **Data Lake** e **Data Warehouse**, de modo a maximizar flexibilidade analítica e eficiência transacional. Cada camada cumpre um papel específico:

1. **Camada Raw (Data Lake):**
   Armazena arquivos brutos em formatos como CSV e Parquet, preservando integralmente as versões originais enviadas pela distribuidora. Essa camada garante reprodutibilidade e permite auditorias completas.
2. **Camada Processed (Data Warehouse):**
   Contém tabelas normalizadas, limpas e validadas, utilizadas diretamente na construção do grafo. Essa camada é mantida em um banco relacional como PostgreSQL ou BigQuery, garantindo integridade referencial e desempenho analítico.

3. **Camada Semantic:**
   Estruturas específicas para grafos heterogêneos, incluindo tabelas intermediárias de nós, arestas, atributos enriquecidos e versões temporais do grafo.

O gerenciamento é realizado com ferramentas de versionamento como **Delta Lake** ou **LakeFS**, que permitem registrar modificações linha a linha, garantindo governança plena. Todos os dados possuem metadados associados — data de ingestão, transformações aplicadas, regras de validação, versão do schema — assegurando rastreabilidade histórica.

4.4.1.3 Dicionário de Metadados

O conjunto de dados distribuídos têm origem de operação predominante em *e-commerce*. O *dataset* compreende todas as transações efetuadas durante um período de 12 meses (Janeiro/2022 a Dezembro/2022).

O *dataset* bruto foi fornecido em formato Parquet, e após a finalização da sua ingestão e processamento, contém as seguintes colunas principais para as três tabelas distintas:

Tabela *sales_transactions*

1. ***internal_store_id***: Identificador único da loja (Ponto de Venda) onde a transação ocorreu. Se conecta ao ***pdv_id*** da tabela ***sales_points*** como chave estrangeira.
2. ***internal_product_id***: Identificador único do produto vendido. Se conecta ao ***product_id*** da tabela ***products_catalog*** como chave estrangeira.
3. ***distributor_id***: Identificador do distribuidor associado à venda.
4. ***transaction_date***: A data exata em que a transação foi registrada (dia/mês/ano).
5. ***reference_date***: Uma data de referência, usada por agrupamento pelo primeiro dia do mês da transação baseado no ***transaction_date***.
6. ***quantity***: A quantidade de unidades do produto vendido na transação.
7. ***unit_price***: O preço de uma única unidade do produto. Este valor é obtido dividindo o ***gross_value*** pela ***quantity*** (***gross_value*** / ***quantity***). Ele representa o preço de tabela do produto antes da aplicação de quaisquer impostos ou descontos específicos da transação.
8. ***gross_value***: O valor bruto total da transação (***unit_price * quantity***), antes de impostos e descontos.

9. ***net_value***: O valor líquido da transação. O valor bruto menos descontos e/ou impostos (***gross_value*** - ***discount*** - ***taxes***).
10. ***gross_profit***: O lucro bruto obtido na transação (geralmente, valor líquido - custo do produto).
11. ***discount***: O valor total do desconto aplicado à transação.
12. ***taxes***: O valor dos impostos incidentes sobre a transação.

Tabela *products_catalog*

1. ***product_id***: Identificador único do produto (Chave Primária).
2. ***category***: A categoria principal do produto (ex: *"Distilled Spirits", "Wine", "Non-Alcohol"*).
3. ***description***: O nome ou descrição textual do produto (ex: *"JOSEPH CARTRON CAFÉ LIQUEUR"*).
4. ***type***: Um tipo ou classificação adicional (ex: *"Distilled Spirits", "Draft"*).
5. ***label***: Uma etiqueta ou rótulo de classificação de marketing ou inventário (ex: *"Core", "Specialty", "Private Label"*).
6. ***subcategory***: Uma subdivisão mais específica da categoria (ex: *"Liqueurs & Cordials", "Scotch Whisky"*).
7. ***brand***: A marca comercial do produto.
8. ***manufacturer***: O nome do fabricante ou fornecedor do produto.

Tabela *sales_points*

1. ***pdv_id***: Identificador único do Ponto de Venda (PDV).
2. ***premise***: O tipo de estabelecimento, indicando o modelo de consumo:
    a. ***On Premise***: Consumo no local (ex: restaurantes, bares).
    b. ***Off Premise***: Consumo fora do local (ex: supermercados, lojas de conveniência, liquor stores).
3. ***categoria_pdv***: A categoria específica do ponto de venda (ex: "*Mexican Rest*", "*Convenience*", "*Package/Liquor*", "*Hotel/Motel*").
4. ***zipcode***: O CEP (Código Postal) da localização do ponto de venda.

Os *datasets* contém aproximadamente 6.5 milhões de registros de itens de transação, correspondendo a cerca de 14.419 fornecedores distintos e um catálogo de 7.092 produtos.

4.4.2 Preparação dos Dados

A preparação dos dados constitui uma etapa central do ciclo de vida do artefato, pois estabelece o elo entre os dados brutos armazenados nas camadas bruta e processada e a representação relacional final utilizada pelos modelos HGT e pelo módulo DML. Essa fase segue integralmente as diretrizes táticas do CRISP-DM, articulando limpeza, padronização, enriquecimento semântico e transformação estrutural. O objetivo é assegurar que os dados transacionais, firmográficos e de catálogo sejam convertidos em insumos consistentes, estáveis e informativamente densos para o aprendizado estatístico e causal.

O processo inicia-se com a **limpeza dos dados**, na qual são tratadas ausências, 1473 da coluna label e 32 da subcategoria na tabela ***produtos_catalogo,*** inconsistências como outliers da tabela ***transacoes_vendas***. Para valores ausentes, aplicaram-se métodos distintos conforme o tipo de atributo: imputação determinística para atributos estáticos (por exemplo, volume do produto), imputação baseada em vizinhos ou medianas para atributos semi-variáveis (como preços) e, quando necessário, exclusão criteriosa de registros que não atendiam aos requisitos mínimos de integridade referencial. Nessa etapa, aplicaram-se filtros de plausibilidade temporal e comercial para identificar registros incoerentes, como vendas associadas a produtos inexistentes ou transações registradas fora do período de operação.

A etapa seguinte consiste na **padronização e normalização dos atributos**, garantindo coerência semântica entre bases. A padronização abrange unificação de tipos categóricos (como categorias de produto e perfis de PDV), normalização de unidades (litros, mililitros, embalagens), correção de variações ortográficas e aplicação de taxonomias definidas no dicionário de dados. A normalização numérica, quando exigida pelo modelo, utiliza transformações como *min-max scaling* ou *z-score*, preservando relações estruturais relevantes.

Em seguida, procede-se ao **enriquecimento dos dados**, incluindo a criação de variáveis derivadas que intensificam a expressividade do grafo heterogêneo. Exemplos incluem: frequência de compra por PDV, elasticidade aproximada por categoria, intensidade promocional histórica, sazonalidade temporal e métricas de afinidade entre produtos. Essas features agregadas aumentam a densidade informacional das representações aprendidas pelos HGTs.

Posteriormente é realizada a **construção das tabelas intermediárias de nós e arestas**, que traduzem a estrutura tabular das bases em uma estrutura relacional adequada para a modelagem. As tabelas de nós são geradas a partir das entidades fundamentais — produtos, PDVs e transações — enquanto as tabelas de arestas representam interações comerciais, com atributos temporais e contextuais. Cada aresta é enriquecida com pesos, frequências e atributos derivados, permitindo que o HGT capture padrões de dependência e heterogeneidade.

A etapa final consiste na **validação estrutural e topológica do grafo**, garantindo que a representação seja consistente com as ontologias definidas. São verificadas cardinalidades, conectividade, equilíbrio entre tipos de entidades, formação de componentes desconexos e presença de artefatos que possam enviesar o processo relacional. Caso o grafo apresente distorções — como colapsos estruturais, excesso de isolados ou padrões de transiência — o pipeline aciona rotinas automáticas de correção e reconciliação.

A preparação dos dados, assim estruturada, garante que o artefato receba uma base robusta, semanticamente padronizada, topologicamente consistente e informativamente rica, condições essenciais para o aprendizado relacional profundo do HGT e para a acurácia causal obtida pelo DML.

4.4.3 Análise de Dados

A análise de dados constitui a etapa em que o conjunto preparado passa a ser explorado de modo sistemático, com o objetivo de compreender padrões, verificar coerências estruturais, identificar relações relevantes e orientar o desenho final dos modelos HGT e do módulo DML. Essa fase segue o eixo tático do CRISP-DM, apoiando-se em quatro vetores analíticos: **caracterização descritiva**, **análise relacional**, **análise temporal** e **análises orientadas à causalidade**. Cada vetor cumpre uma função específica na compreensão do domínio e na extração de informações que servirão como insumo para a modelagem relacional e causal.

O primeiro vetor, **caracterização descritiva**, examina as propriedades fundamentais das três bases anexadas — "*transacoes_vendas*", "*produtos_catalogo*" e "*pontos_venda*". São calculadas distribuições marginais de preços, descontos, volumes vendidos, frequência de compra por PDV, composição do portfólio de produtos e segmentação firmográfica dos pontos de venda. Essa caracterização permite identificar assimetrias, caudas pesadas, clusters naturais e padrões de compra diferenciados que influenciam a elasticidade e o comportamento comercial.

O segundo vetor, **análise relacional**, analisa padrões de interação entre entidades com foco na estrutura que posteriormente será convertida em grafo heterogêneo. Avaliam-se métricas como densidade de conexões, frequência transacional entre PDVs e categorias de produto, afinidades condicionais (coocorrência de produtos no mesmo pedido), centralidade de nós e presença de hubs comerciais. Essa etapa é fundamental para identificar a heterogeneidade estrutural, elemento que o HGT é projetado para explorar ao diferenciar tipos de nós e tipos de arestas.

O terceiro vetor, **análise temporal**, examina flutuações sazonais, ciclos de demanda, efeitos de calendário e padrões de recorrência. São avaliadas janelas móveis de vendas, periodicidade de compra, variações de estoque e resposta a intervenções promocionais ao longo do ano. Como as bases representam um período de 12 meses, essa dimensão é essencial para capturar dependências dinâmicas que influenciam estimativas causais e que, quando ignoradas, podem introduzir confundimento sistêmico.

O quarto vetor, **análises orientadas à causalidade**, busca identificar possíveis fontes de vieses e de confundimento estrutural antes mesmo da modelagem DML. São examinadas:
– diferenças sistemáticas entre PDVs que recebem ou não intervenções comerciais;
– correlações espúrias entre preço, demanda e atributos firmográficos;
– indícios de causalidade reversa, especialmente na relação entre descontos e volume vendido;
– padrões que sinalizam seleção endógena (por exemplo, PDVs que recebem mais promoções por histórico de baixa performance).

Essas análises preliminares fundamentam a estrutura do modelo causal-relacional, orientando a necessidade de ortogonalização por meio do DML e mostrando a importância dos dois HGTs operarem funções de tratamento e resultado separadamente.

Por fim, os resultados analíticos são consolidados em relatórios versionados, integrados ao pipeline ModelOps, permitindo que novas ingestões de dados sejam comparadas com padrões históricos e que anomalias sejam detectadas precocemente. Assim, a análise de dados cumpre papel essencial no alinhamento entre o entendimento do domínio e a modelagem relacional profunda, assegurando que o artefato seja fiel às regularidades comerciais observadas.

4.4.4 Modelagem

A modelagem constitui o núcleo técnico do artefato, etapa em que a estrutura relacional do domínio é formalizada matematicamente e traduzida em representações profundas capazes de capturar dependências entre entidades e isolar efeitos causais de intervenções comerciais. Essa fase integra dois componentes centrais — **aprendizado relacional via Heterogeneous Graph Transformer (HGT)** e **inferência causal via Double Machine Learning (DML)** — articulados de forma complementar para resolver o problema raiz identificado: a incapacidade de modelos tradicionais representarem simultaneamente as relações estruturais e os efeitos causais presentes no ambiente B2B da distribuição de bebidas.

A modelagem inicia-se com a **construção das representações relacionais** por meio do grafo heterogêneo derivado das bases “transacoes_vendas”, “produtos_catalogo” e “pontos_venda”. Cada entidade é convertida em um tipo de nó, e cada interação comercial em um tipo de aresta, preservando atributos intrínsecos e contextuais. Essa representação permite que o modelo capture padrões estruturais como afinidade entre produtos, recorrência de compra e características firmográficas que influenciam o comportamento comercial.

Em seguida, são treinados dois modelos HGT independentes, cada qual com objetivo distinto e complementar. O primeiro — **HGT-T (Treatment Model)** — aprende a função que governa a propensão ao tratamento, isto é, os fatores estruturais e relacionais que explicam por que um PDV ou produto recebe uma intervenção comercial, como desconto ou recomendação. Essa função não modela o impacto da intervenção, apenas sua alocação, evitando que informações sobre o resultado contaminem o processo de estimação causal.

O segundo — **HGT-Y (Outcome Model)** — foca na modelagem dos resultados potenciais, aprendendo padrões relacionais que influenciam vendas, elasticidade, variações sazonais e resposta a estímulos comerciais. Ambos os HGTs utilizam mecanismos de atenção heterogênea que permitem ponderar diferentemente tipos de nós e tipos de conexões, extraindo dependências complexas que algoritmos tabulares ou redes neurais convencionais não conseguem capturar.

Uma vez obtidas as representações aprendidas pelos dois modelos HGT, inicia-se a etapa de **inferência causal**, implementada por meio do arcabouço Double Machine Learning (DML). Esse arcabouço utiliza as predições de HGT-T e HGT-Y para remover confundimento estrutural, garantindo que o estimador final capture efeitos causais e não apenas correlações observadas. O processo segue três fases:

1. **Ortogonalização**, na qual erros residuais das funções de tratamento e de resultado são isolados;
2. **Estimativa do efeito causal**, com cálculo de ATE (Average Treatment Effect) e CATE (Conditional Average Treatment Effect);
3. **Simulação contrafactual**, permitindo prever a resposta de PDVs e produtos sob diferentes cenários promocionais.

Os resultados do DML formam a base para recomendações comerciais, estimação de elasticidade causal e identificação de alvos prioritários para intervenções. A combinação dos dois HGTs com o DML elimina vieses oriundos de seleção endógena, causalidade reversa e confundimento estrutural — problemas recorrentes em ambientes de varejo e distribuição.

Toda a modelagem é implementada de forma versionada e rastreável via pipeline ModelOps, garantindo reprodutibilidade, governança de experimentos, comparação entre versões e monitoramento contínuo de desempenho. Essa abordagem articula o caráter exploratório do aprendizado relacional com o rigor estatístico da inferência causal, assegurando precisão e robustez ao artefato.

4.4.5 Validação e Teste

A etapa de validação e teste assegura que o artefato desenvolvido — composto pelos modelos HGT-T, HGT-Y e pelo módulo de Double Machine Learning (DML) — produza resultados confiáveis, interpretáveis e robustos em diferentes condições operacionais. Essa fase cumpre papel essencial no ciclo CRISP-DM (avaliação do modelo) e no fluxo ModelOps (monitoramento, garantia de qualidade e aprovação para produção). O processo é estruturado em três eixos: **validação relacional**, **validação causal** e **teste operacional**.

O primeiro eixo, **validação relacional**, verifica a capacidade dos dois HGTs de capturar padrões estruturais presentes no grafo heterogêneo. Avaliam-se métricas como acurácia de predição de arestas, *mean reciprocal rank* (MRR), *hit rate*, similaridade entre embeddings e estabilidade dos pesos de atenção entre tipos de arestas e nós. Testes adicionais analisam a sensibilidade da arquitetura a perturbações topológicas — por exemplo, remoção parcial de transações ou redução da conectividade de determinados PDVs. Uma performance estável indica que o modelo extrai relações estruturais profundas, não dependendo de artefatos específicos dos dados.

O segundo eixo, **validação causal**, assegura que as estimativas produzidas pelo DML representem efeitos causais e não apenas correlações espúrias. São aplicados testes de robustez, incluindo:

– **Overlap diagnostics**, avaliando se há cobertura adequada entre grupos tratados e não tratados;

– **Placebo tests**, nos quais variáveis fictícias são tratadas como intervenções para verificar ausência de efeitos artificiais;

– **Sensitivity analysis**, medindo a estabilidade do ATE e do CATE diante de perturbações no modelo de tratamento e no modelo de resultado;

– **Double robustness checks**, assegurando que erros moderados em HGT-T ou HGT-Y não comprometam a validade do estimador final.

Além disso, efeitos estimados são comparados com benchmarks históricos — como elasticidades calculadas manualmente ou respostas comerciais observadas em campanhas anteriores — oferecendo uma triangulação empírica entre modelo e realidade operacional.

O terceiro eixo, **teste operacional**, insere o artefato em condições similares às de produção. Esse teste avalia latência do modelo, comportamento sob cargas variáveis, capacidade de escalar em ambientes distribuídos, robustez frente a entradas inesperadas e compatibilidade com os fluxos de ingestão e atualização do pipeline. Também são monitorados:

– consistência temporal entre versões do grafo;
– impactos de deriva de dados na qualidade das representações;
– integridade dos logs e metadados para auditoria;
– estabilidade das recomendações quando novos dados são incorporados.

Os resultados de validação e teste são consolidados em relatórios versionados integrados ao ecossistema ModelOps. Apenas modelos que atendem a critérios mínimos de qualidade — estatísticos, causais e operacionais — são promovidos para produção. Essa etapa garante que o artefato opere com rigor científico e estabilidade prática, sustentando recomendações comerciais precisas e inferências causais confiáveis.

4.4.6 Implantação

A implantação constitui o momento em que o artefato — composto pelos modelos HGT-T, HGT-Y e pelo módulo de Double Machine Learning (DML) — é transferido do ambiente de desenvolvimento para um ambiente operacional estável, auditável e integrado aos sistemas corporativos da distribuidora. Essa etapa materializa o alinhamento entre o CRISP-DM (fase de implantação), o DSR (momento de comunicação e disponibilização da solução) e o ModelOps (gestão contínua do ciclo de vida do modelo).

A implantação segue um processo estruturado em três camadas: **empacotamento**, **orquestração** e **exposição dos serviços**.

Na camada de **empacotamento**, os modelos são convertidos em artefatos executáveis e versionados. Cada componente — HGT-T, HGT-Y, DML e utilitários de pré-processamento — é encapsulado em contêineres Docker, garantindo reprodutibilidade ambiental e isolamento entre dependências. Os contêineres são registrados em um repositório privado com versionamento semântico, permitindo rastrear com precisão qual versão do modelo está ativa, qual conjunto de dados foi utilizado em seu treinamento e quais parâmetros compõem sua configuração final.

A camada de **orquestração** utiliza plataformas como Kubernetes para gerenciar a implantação, escalonamento e resiliência dos serviços. Essa orquestração garante:

– alta disponibilidade por meio de réplicas automáticas;
– reinicialização de contêineres em caso de falha;
– balanceamento de carga entre instâncias;
– migração suave entre versões (rolling updates).

Durante essa etapa, são ativados módulos de monitoramento que verificam latência, consumo de GPU/CPU, integridade dos logs e estabilidade das predições. Alterações inesperadas de comportamento são detectadas em tempo real e podem acionar mecanismos de *rollback* para versões anteriores.

A terceira camada, **exposição dos serviços**, disponibiliza o artefato por meio de APIs REST internas, capazes de receber requisições operacionais de sistemas de recomendação, planejamento comercial e equipes de inteligência de mercado. Essas APIs entregam:

– recomendações personalizadas por PDV;
– estimativas de elasticidade causal;
– efeitos esperados de campanhas;
– simulações contrafactuais;
– métricas de risco e incerteza associadas à intervenção proposta.

As APIs utilizam autenticação forte (token-based ou OAuth2), auditabilidade completa e versionamento explícito, garantindo que decisões comerciais sejam sempre baseadas em modelos rastreáveis e não ambíguos. Logs de requisição são registrados para fins de governança e compliance.

A implantação é concluída com a execução de testes de produção, que validam a interoperabilidade entre pipeline, banco de dados, modelos implantados e sistemas consumidores. A integração contínua (CI) e a entrega contínua (CD) permitem ciclos rápidos de atualização, assegurando que melhorias nos modelos ou atualizações de dados possam ser incorporadas sem interrupções operacionais.

Com isso, a etapa de implantação assegura que o artefato opere com confiabilidade, escalabilidade e precisão no ambiente real, cumprindo os requisitos de robustez exigidos para suportar decisões comerciais baseadas em inferência causal.

4.4.7 Monitoramento e Manutenção

O monitoramento e a manutenção garantem a continuidade operacional, a estabilidade estatística e a integridade causal do artefato ao longo de todo o seu ciclo de vida. Essa etapa articula os princípios do ModelOps — detecção de deriva, auditoria, reavaliação periódica, readequação de modelos — com a lógica tática do CRISP-DM (avaliação contínua) e com o caráter estratégico do DSR (preservação das contribuições do artefato no tempo). Trata-se de uma fase crítica, pois modelos baseados em estruturas relacionais e inferência causal são particularmente sensíveis a mudanças no comportamento comercial, na conectividade entre entidades e nas políticas de intervenção.

O monitoramento inicia-se com a **supervisão de dados e deriva**, que acompanha alterações nas propriedades estatísticas e topológicas das bases "transacoes_vendas", "produtos_catalogo" e "pontos_venda". São monitoradas distribuições marginais (como preços e volumes), mudanças estruturais (entrada ou saída de produtos e PDVs), variações sazonais inesperadas e alterações na densidade do grafo heterogêneo. A identificação de eventuais anomalias — como compressão topológica, expansão abrupta de arestas ou flutuações anormais de conectividade — aciona processos automáticos de revisão e reconciliação.

Simultaneamente, ocorre o **monitoramento do desempenho dos modelos HGT-T e HGT-Y**, verificando estabilidade de embeddings, robustez de pesos de atenção, performance em métricas relevantes (como MRR, hit-rate e erro residual) e consistência das predições frente a novos dados. Alterações significativas nessas métricas podem indicar necessidade de re-treinamento ou ajuste de hiperparâmetros. As representações aprendidas pelos HGTs também são comparadas a versões históricas, permitindo identificar degradação semântica, fenômeno comum em domínios de alta volatilidade e catálogo dinâmico.

A terceira dimensão envolve o **monitoramento causal**, assegurando que o módulo de Double Machine Learning continue produzindo estimativas confiáveis ao longo do tempo. São acompanhados indicadores como:

– estabilidade do ATE e do CATE;
– consistência entre efeitos estimados e observações recentes;
– detecção de mudanças no mecanismo de tratamento (policy shift);
– aumento de confundimento estrutural;
– redução do overlap entre grupos tratados e não tratados.

Sempre que esses indicadores se movem para fora de faixas aceitáveis, mecanismos automatizados acionam testes de robustez adicionais, reestimativas e auditorias internas.

A quarta dimensão diz respeito à **auditoria completa do ciclo operacional**, que envolve rastreamento de requests de API, integridade dos logs, verificação de compliance e reconstituição de decisões passadas. Sistemas ModelOps permitem recuperar instantaneamente qual versão do grafo, qual versão dos modelos e quais parâmetros alimentaram determinada recomendação comercial.

A manutenção, por sua vez, engloba **rotinas periódicas de reprocessamento**, **re-treinamento automatizado**, **implantação de novos modelos (rolling updates)**, **otimização de performance** e **controle de custos operacionais**, especialmente no uso de GPU. Toda manutenção é registrada com versionamento estrito, garantindo que cada alteração tenha justificativa analítica e evidência experimental.

Por fim, o sistema utiliza mecanismos de **alerta proativo**, que detectam eventos como degradação de performance, falhas de ingestão, inconsistências entre camadas do pipeline ou anomalias no comportamento relacional. Esses alertas orientam intervenções rápidas e preservam a confiabilidade do artefato em ambiente real.

Assim, o monitoramento e a manutenção asseguram que o artefato permaneça robusto, auditável, atualizado e alinhado às dinâmicas reais do mercado, garantindo que sua capacidade de estimar efeitos causais e apoiar decisões comerciais continue válida ao longo do tempo.

5. ANÁLISE DOS RESULTADOS

5.1 Resultados Obtidos

Os resultados obtidos refletem a consolidação do artefato proposto — composto pelos modelos HGT-T, HGT-Y e pelo módulo de Double Machine Learning (DML) — aplicado ao conjunto de dados reais da distribuidora (“transacoes_vendas”, “produtos_catalogo” e “pontos_venda”). A análise se organiza em três eixos: **representação relacional**, **estimativas causais** e **desempenho operacional do artefato**. Cada eixo produz evidências empíricas distintas, mas complementares, permitindo avaliar a eficácia da solução na separação entre relações estruturais e efeitos causais no domínio comercial.

O primeiro eixo, **resultados de representação relacional**, revela que os modelos HGT capturaram com precisão a estrutura heterogênea do domínio. Os nós representando PDVs, produtos e transações foram organizados em embeddings densos, com forte separabilidade entre segmentos de PDV, categorias de produto e padrões sazonais de compra. Métricas como *mean reciprocal rank* (MRR) e *hit-rate* indicaram que o HGT-T modelou adequadamente a propensão ao tratamento, reproduzindo com fidelidade as políticas históricas de alocação de recomendações e descontos. O HGT-Y mostrou capacidade consistente de prever padrões de venda, mesmo em regiões do grafo com conectividade reduzida, evidenciando aprendizagem de dependências que extrapolam relações locais.

O segundo eixo, **resultados de inferência causal**, evidencia a principal contribuição do artefato: a capacidade de isolar o efeito das intervenções comerciais, corrigindo vieses de confundimento estruturais e comportamentais. As estimativas de **ATE (Average Treatment Effect)** revelaram impacto médio positivo de intervenções como descontos direcionados, embora com amplitude inferior aos valores correlacionais originalmente observados, demonstrando a presença de vieses de seleção na política comercial anterior. Os resultados de **CATE (Conditional Average Treatment Effect)** mostraram heterogeneidade substancial entre segmentos de PDV: estabelecimentos com baixa frequência de compra apresentaram maior sensibilidade a promoções, enquanto PDVs de alta recorrência mostraram respostas mais moderadas. Essa heterogeneidade, não capturada por modelos tradicionais, fornece insumos para políticas de segmentação mais precisas.

Além disso, análises contrafactuais realizadas pelo módulo DML demonstraram que a simples intensificação de descontos não necessariamente maximiza vendas, especialmente em categorias cuja elasticidade causal se mostrou baixa ou negativa. O artefato permitiu simular cenários alternativos em que intervenções são redistribuídas entre PDVs, revelando ganhos potenciais advindos de uma política mais alinhada ao efeito esperado — não à mera correlação histórica entre desconto e volume.

O terceiro eixo, **desempenho operacional**, avaliou estabilidade, consistência e integração do artefato no pipeline ModelOps. Os modelos se mostraram robustos a atualizações incrementais nos dados, mantendo coerência topológica nas representações e estabilidade nas estimativas causais. Latência e consumo computacional permaneceram dentro dos limites esperados para ambientes B2B, permitindo execução de simulações e geração de recomendações quase em tempo real. Logs, metadados e versionamento demonstraram conformidade com requisitos de rastreabilidade e auditoria, assegurando que cada resultado pudesse ser reproduzido e validado.

Assim, os resultados obtidos demonstram que o artefato cumpre seus objetivos fundamentais: representar a estrutura relacional do domínio, produzir estimativas causais robustas e operar de forma confiável em ambiente real. Esses achados fornecem base sólida para as análises críticas subsequentes e para a avaliação das limitações metodológicas da abordagem.

5.2 Análise Crítica

A análise crítica examina a profundidade, a consistência e as limitações implícitas nos resultados obtidos, articulando os achados empíricos com a fundamentação teórica e metodológica que sustentou o desenvolvimento do artefato. Essa reflexão é necessária não apenas para avaliar o desempenho do sistema, mas também para compreender os limites estruturais da abordagem HGT + DML no domínio B2B da distribuição de bebidas.

Um primeiro ponto crítico refere-se à **representação relacional aprendida pelos modelos HGT**. Embora os embeddings tenham capturado padrões estruturais relevantes — como segmentação natural de PDVs, afinidade entre categorias de produtos e regularidades sazonais — a qualidade dessas representações depende fortemente da completude e da granularidade dos dados. A base de transações, por exemplo, contém somente 12 meses de histórico, o que restringe a aprendizagem de dinâmicas de longo prazo. Em domínios com alta sazonalidade, como bebidas, essa limitação pode impactar a estabilidade das atenções heterogêneas do HGT e introduzir variações artificiais na conectividade do grafo, especialmente em períodos de baixa atividade comercial.

No campo da inferência causal, o artefato demonstrou eficácia ao corrigir vieses de confundimento e ao produzir estimativas mais realistas de elasticidade e resposta a intervenções. Contudo, a **robustez do Double Machine Learning depende da adequação dos modelos de tratamento e resultado**, o que cria uma relação estrutural delicada: qualquer degradação significativa no HGT-T ou HGT-Y pode repercutir diretamente no módulo causal. Embora o DML ofereça dupla robustez teórica, essa propriedade não elimina a necessidade de alinhamento entre modelos e realidade operacional. Além disso, efeitos causais heterogêneos observados nos resultados precisam ser interpretados com cautela, pois parte dessa heterogeneidade pode refletir diferenças não observáveis nos PDVs, que os dados disponíveis não capturam integralmente.

Outro aspecto crítico envolve a **capacidade dos modelos de lidar com regiões esparsas do grafo**, especialmente em situações de cold start. O HGT mostrou habilidade em extrapolar padrões relacionais, mas sua performance ainda é sensivelmente menor em PDVs com poucas transações ou em produtos de baixa rotatividade. O problema não deriva do modelo, mas da natureza do domínio, no qual regiões esparsas tendem a produzir embeddings menos informativos, afetando tanto predições quanto estimativas de efeito causal. Modelos baseados em grafos mitigam, mas não eliminam completamente esse fenômeno.

Do ponto de vista operacional, os resultados indicaram boa estabilidade e latência adequada, mas o custo computacional do HGT — especialmente em grafos densos com múltiplos tipos de nós e arestas — representa um ponto de atenção. O treinamento dos dois modelos HGT em paralelo, seguido do módulo DML, exige infraestrutura de GPU consistente, o que pode gerar restrições para ambientes com limitações de hardware ou políticas rígidas de custo. O período de reprocessamento também se mostrou sensível à expansão do catálogo ou ao aumento na frequência de ingestão dos dados.

Por fim, a análise crítica evidencia que, embora o artefato represente um avanço substancial sobre abordagens correlacionais tradicionais, ele permanece limitado pela **qualidade e abrangência dos dados disponíveis** e pela **sensibilidade do estimador causal às mudanças estruturais do domínio comercial**. Essas limitações não anulam a contribuição do sistema; ao contrário, orientam ajustes e extensões que podem fortalecer sua aplicabilidade em cenários reais e dinâmicos.

5.3 Limitações Metodológicas

As limitações metodológicas deste estudo decorrem tanto das características intrínsecas dos dados utilizados quanto das escolhas técnicas que estruturam o artefato baseado em HGT + DML. A análise destas limitações permite contextualizar os resultados, qualificar a validade externa do modelo e orientar melhorias futuras em termos de modelagem, coleta e diversidade informacional.

A primeira limitação diz respeito à **restrição temporal e comportamental das bases de dados**. O histórico de 12 meses, embora suficiente para capturar sazonalidade anual, não permite observar ciclos de longo prazo, alterações estruturais em padrões de consumo nem mudanças persistentes na política comercial da distribuidora. Em domínios fortemente influenciados por sazonalidade e variações macroeconômicas, essa limitação pode dificultar a separação entre efeitos causais verdadeiros e flutuações temporárias. Como consequência, estimativas de elasticidade causal ou de resposta a campanhas podem refletir condições específicas do período analisado.

Outra limitação metodológica deriva da **dependência do HGT em estruturas relacionais densas**. Embora o modelo tenha se mostrado competente na extração de padrões estruturais, a performance é sensível a regiões esparsas do grafo, como produtos de baixa rotatividade e PDVs com pouco histórico transacional. Nesses casos, os embeddings tendem a ter menor expressividade, o que afeta tanto o modelo de tratamento (HGT-T) quanto o de resultado (HGT-Y). Essa fragilidade manifesta-se especialmente em fenômenos de cold start, onde as representações aprendidas podem não ser suficientes para sustentar inferência causal robusta.

A terceira limitação refere-se à **suposição implícita do DML de que o mecanismo de tratamento é parcialmente observável nos dados**. Embora o arcabouço seja projetado para corrigir confundimento estrutural, essa correção depende da capacidade do HGT-T de capturar os fatores que determinam a alocação de intervenções comerciais. Se parte relevante da decisão comercial não estiver refletida nas bases disponíveis — por exemplo, acordos específicos com determinados PDVs, metas internas da distribuidora ou fatores humanos não registrados — a estimação causal pode incorporar viés residual. O DML oferece dupla robustez estatística, mas não pode corrigir confundimento totalmente não observado.

Um quarto ponto de limitação envolve a **complexidade computacional da arquitetura**, que exige hardware especializado, especialmente para o treinamento dos dois modelos HGT. Em ambientes corporativos com restrições de infraestrutura ou políticas de contenção de custos, esse requisito pode limitar a frequência de reprocessamento dos modelos e atrasar a resposta a mudanças rápidas na dinâmica comercial. Isso afeta diretamente a capacidade de manter estimativas causais atualizadas, especialmente em contextos altamente dinâmicos.

Há ainda uma limitação relacionada à **interpretação das representações aprendidas**, já que o HGT, apesar de trazer mecanismos de atenção heterogênea, permanece um modelo de difícil interpretabilidade global. A compreensão do porquê determinadas relações recebem mais peso que outras ainda depende de técnicas auxiliares, que podem não capturar integralmente o comportamento interno dos modelos. Isso limita a transparência do sistema e pode dificultar a comunicação dos resultados a equipes não técnicas.

Por fim, reconhece-se que a abordagem adotada — embora avançada do ponto de vista metodológico — foi desenvolvida e avaliada sobre um ecossistema específico de uma distribuidora de bebidas. A **validade externa** do modelo pode variar quando aplicado a outros setores, geografias ou portfólios com dinâmica distinta. A transferência da solução pode exigir adaptações substanciais na construção do grafo, nos atributos firmográficos e nas políticas comerciais modeladas.

Essas limitações, quando consideradas de forma integrada, não invalidam o artefato, mas delineiam o escopo de sua aplicabilidade e destacam áreas onde a pesquisa pode ser aprofundada. Elas fornecem, portanto, uma base concreta para orientar as discussões presentes nas seções conclusivas e o planejamento de trabalhos futuros.

6. CONCLUSÃO

6.1 Síntese

A síntese integra os principais elementos desenvolvidos ao longo da pesquisa, articulando o problema investigado, a solução projetada, os métodos empregados e os resultados alcançados. O estudo partiu do reconhecimento de uma limitação central no setor B2B de distribuição de bebidas: a falta de modelos capazes de representar adequadamente a estrutura relacional do domínio e, simultaneamente, isolar efeitos causais de intervenções comerciais. Essa deficiência metodológica resultava na confusão entre correlação e causalidade, enviesando decisões estratégicas relacionadas a recomendação, precificação e alocação de campanhas.

Frente a esse desafio, foi concebido um artefato estruturado segundo o Design Science Research (DSR), fundamentado no CRISP-DM (nível tático) e operacionalizado por práticas de ModelOps. O artefato combinou duas arquiteturas Heterogeneous Graph Transformer (HGT) — uma dedicada à modelagem da propensão ao tratamento e outra à modelagem dos resultados — integradas a um módulo de Double Machine Learning (DML), responsável pela estimação causal final. Essa composição permitiu representar o domínio como um grafo heterogêneo, capturando dependências complexas entre PDVs, produtos e transações utilizando dados provenientes de “transacoes_vendas”, “produtos_catalogo” e “pontos_venda”.

A solução desenvolvida produziu resultados expressivos: o HGT demonstrou capacidade de aprender representações relacionais profundas e diferenciadas, enquanto o DML corrigiu confundimentos estruturais, oferecendo estimativas causais robustas e interpretáveis. A análise empírica evidenciou, entre outros aspectos, que muitas relações comerciais historicamente tratadas como causais eram sustentadas apenas por correlações espúrias. Além disso, a heterogeneidade dos efeitos revelou perfis distintos de sensibilidade comercial entre segmentos de PDVs, subsidiando políticas de intervenção mais precisas e eficientes.

A síntese reforça que o artefato atingiu seu propósito fundamental: oferecer uma solução tecnicamente robusta e metodologicamente consistente para o problema raiz, demonstrando que a combinação entre aprendizado relacional e inferência causal é viável, eficaz e superior às abordagens tradicionais baseadas apenas em regressões ou modelos tabulares. Esse resultado constitui a base para a verificação dos objetivos e para as discussões subsequentes acerca das contribuições científicas, limitações e implicações práticas da pesquisa.

6.2 Verificação dos Objetivos

A verificação dos objetivos consiste em avaliar, de forma sistemática, se cada propósito estabelecido no início da pesquisa foi efetivamente alcançado pelo artefato e pelos procedimentos científicos adotados. Essa etapa opera como um fechamento lógico do percurso metodológico, confirmando a coerência entre o problema identificado, os métodos escolhidos e os resultados obtidos.

O objetivo geral — desenvolver um artefato baseado em Heterogeneous Graph Transformer (HGT) combinado a Double Machine Learning (DML), capaz de representar a estrutura relacional do domínio e estimar efeitos causais de intervenções comerciais — foi plenamente atendido. A arquitetura proposta capturou dependências interentidades com granularidade superior aos modelos tabulares convencionais. O módulo DML aplicou correções estatísticas essenciais, mitigando confundimentos e produzindo estimativas causais consistentes, o que confirma a aderência entre o artefato projetado e o problema raiz.

Quanto aos objetivos específicos, cada um foi avaliado à luz dos resultados empíricos:

A caracterização do domínio por meio das bases “transacoes_vendas”, “produtos_catalogo” e “pontos_venda” foi concluída de forma estruturada, alinhada às etapas de Data Understanding do CRISP-DM. Essa análise forneceu insumos para a modelagem gráfica heterogênea e sustentou decisões sobre entidades, atributos e arestas do grafo.

O desenvolvimento da arquitetura dual HGT+HGT atendeu ao objetivo de construir representações distintas para propensão ao tratamento e para resultados observados. Essa separação refletiu os princípios formais do Double Machine Learning, garantindo a ortogonalização entre estimadores e reduzindo o risco de viés simultâneo.

A etapa de implementação, guiada por práticas de ModelOps, demonstrou que o artefato pode ser operacionalizado em um pipeline replicável, versionado e monitorável. Isso cumpre o objetivo específico relacionado à viabilidade prática do modelo em ambientes de produção.

A avaliação demonstrou que a solução superou modelos correlacionais tradicionais, apresentando métricas superiores de inferência causal e maior capacidade de generalização em cenários com histórico parcial, incluindo cold start. Essas evidências confirmam o objetivo de construir um artefato mais robusto e alinhado ao comportamento real das relações comerciais.

Dessa forma, a verificação confirma que tanto os objetivos gerais quanto os específicos foram satisfeitos de maneira rigorosa. Os resultados consolidam a coerência interna do projeto e estabelecem as bases necessárias para as contribuições científicas, práticas e metodológicas discutidas nas seções seguintes.

6.3 Contribuições

As contribuições deste trabalho se distribuem em três dimensões — científica, metodológica e aplicada — e resultam diretamente da integração entre modelagem estatística avançada, arquiteturas de aprendizado profundo para grafos heterogêneos e princípios organizacionais de DSR, CRISP-DM e ModelOps. Cada contribuição emerge da necessidade de corrigir a lacuna estrutural observada no domínio: a ausência de modelos capazes de representar simultaneamente dependências relacionais e efeitos causais.

Na dimensão científica, o trabalho avança ao combinar Heterogeneous Graph Transformer (HGT) com Double Machine Learning (DML), compondo uma arquitetura dual orientada tanto a representação estrutural quanto à identificação causal. Embora a literatura reconheça os ganhos das GNNs heterogêneas na modelagem de interações complexas (Hu et al., 2020), e embora DML represente um marco para correção de confundimento (Chernozhukov et al., 2018), a articulação entre ambas ainda é incipiente. O artefato desenvolvido contribui, assim, para o estado da arte ao demonstrar que a ortogonalização estatística do DML se beneficia de embeddings ricos e informados pelo grafo, ampliando a precisão das estimativas de tratamento.

Na dimensão metodológica, o trabalho contribui ao demonstrar como DSR, CRISP-DM e ModelOps podem ser integrados como camadas complementares de estratégia, tática e operação para projetos de ciência de dados com foco causal. O pipeline estabelecido, sustentado por versionamento, governança de dados e orquestração modular, oferece um referencial sistemático replicável por outros pesquisadores que atuam em contextos industriais com alto nível de dependência relacional.

Na dimensão aplicada, a solução desenvolvida produz benefícios diretos para ambientes B2B de distribuição de bebidas — setor caracterizado por forte heterogeneidade entre pontos de venda, variação de comportamento de compra e grande dependência de intervenção comercial. O artefato melhora a capacidade de prever o impacto real de descontos e recomendações, incluindo cenários com poucos dados históricos, mitigando o problema clássico de cold start. Além disso, auxilia a reduzir vieses que historicamente levam ao uso ineficiente de verba promocional, fornecendo estimativas de elasticidade mais confiáveis e adaptadas ao contexto transacional de cada entidade do grafo.

Por fim, as contribuições se articulam como um conjunto coerente de avanços técnicos e práticos. Elas reforçam o argumento central da pesquisa: a representação explícita da estrutura relacional, combinada a técnicas robustas de inferência causal, constitui um caminho promissor para decisões comerciais mais precisas e sustentáveis no setor de distribuição.

6.4 Limitações da Pesquisa

As limitações desta pesquisa derivam principalmente de três eixos: dependências estruturais dos dados, restrições metodológicas inerentes aos modelos utilizados e limitações operacionais da aplicação em ambiente real. Cada uma dessas dimensões afeta o alcance das conclusões e aponta caminhos importantes para expansão futura do trabalho.

Do ponto de vista estrutural, a qualidade das estimativas causais geradas pelo artefato depende diretamente da representatividade e completude das bases fornecidas — *transacoes_vendas*, *produtos_catalogo* e *pontos_venda*. Embora essas bases capturem elementos essenciais das relações B2B, elas não abrangem todas as potenciais fontes de heterogeneidade e confundimento, como dinâmicas competitivas externas, variação de estoques nos PDVs ou campanhas simultâneas conduzidas por outros fornecedores. A ausência de determinadas variáveis implica a possibilidade de confundidores residuais, que podem limitar a capacidade do DML de produzir estimativas plenamente livres de vieses.

No âmbito metodológico, a combinação HGT + Double Machine Learning introduz limitações próprias de arquiteturas complexas. O HGT demanda quantidade significativa de dados para calibrar adequadamente seus mecanismos de agregação dependentes do tipo de entidade, o que pode reduzir desempenho em segmentos com histórico mais esparso — ainda que o modelo mitigue, porém não elimine, problemas de *cold start*. Da mesma forma, o DML pressupõe a capacidade de estimar modelos auxiliares de alta flexibilidade sem sobreajuste, o que requer cuidados rigorosos de validação cruzada e regularização. A conjugação dessas técnicas aumenta a robustez, mas também a sensibilidade a violações de pressupostos como estabilidade do mecanismo gerador de dados e independência condicional.

No plano operacional, a implementação do artefato em ambiente de produção depende de maturidade em governança, versionamento e monitoramento, conforme delineado pelo arcabouço ModelOps. Mudanças estruturais frequentes no catálogo de produtos, no cadastro de clientes ou nos padrões de vendas podem exigir retreinamentos mais frequentes, elevando custos computacionais e dificultando a manutenção da consistência temporal das estimativas geradas. Além disso, a própria interpretação dos modelos, apesar dos avanços no uso de *attention mechanisms*, ainda traz desafios para equipes comerciais que demandam explicações claras sobre a relevância de cada entidade e relação.

Mesmo com essas limitações, os resultados obtidos são suficientemente robustos para sustentar as conclusões apresentadas, ao mesmo tempo em que fornecem indicação clara das condições sob as quais o artefato opera com maior eficácia. Essas limitações funcionam, portanto, como elementos complementares às contribuições, ajudando a delimitar o escopo da pesquisa e sinalizando direções metodológicas estratégicas para os trabalhos subsequentes.

6.5 Trabalhos Futuros

As direções de aprofundamento natural desta pesquisa emergem da própria arquitetura do artefato e das limitações previamente identificadas. Os trabalhos futuros concentram-se em três frentes: evolução metodológica, ampliação das fontes de dados e avanço na integração operacional com sistemas corporativos de decisão.

No campo metodológico, há espaço para explorar variantes mais recentes de modelos de grafos heterogêneos, como o **Heterogeneous Graph Neural Network with Causal Attention**, que introduz mecanismos de atenção explicitamente orientados à separação entre dependências estruturais e efeitos causais. Outra via promissora consiste em investigar extensões do próprio *Double Machine Learning*, incorporando estimadores ortogonais não lineares e versões tempo-dependentes, capazes de capturar variações sazonais e regimes dinâmicos específicos do setor de bebidas. Complementarmente, a aplicação de *causal discovery* baseado em grafos — como NOTEARS ou variantes direcionais adaptadas a grafos heterogêneos — poderia permitir que parte da estrutura causal fosse aprendida diretamente dos dados, reduzindo a dependência de hipóteses estruturais pré-impostas.

A ampliação das bases de dados também representa um vetor importante de evolução. A inclusão de informações de inventário, elasticidade de oferta, campanhas concorrentes, rotas logísticas e dados meteorológicos pode fortalecer o controle de confundidores e permitir modelagem causal mais refinada. A integração com dados externos, como indicadores socioeconômicos regionais e métricas de presença digital dos PDVs, tornaria a rede heterogênea mais expressiva e aumentaria a capacidade preditiva e explicativa do artefato.

No âmbito operacional, pesquisas futuras podem aprofundar a incorporação do modelo dentro de pipelines de **ModelOps** com monitoramento contínuo de *drift* relacional, detecção automática de mudanças estruturais nos grafos e mecanismos de retreinamento adaptativo baseados em gatilhos estatísticos. Outro caminho consiste na exploração de arquiteturas híbridas em ambientes de recomendação em tempo real, combinando o HGT causal com agentes de *reinforcement learning* orientados por políticas sensíveis ao impacto estimado das intervenções.

Por fim, a evolução interpretativa do artefato é um campo fértil. A criação de ferramentas de explicabilidade específicas para grafos heterogêneos — capazes de detalhar quais relações interentidades sustentam uma estimativa causal — pode facilitar a adoção do modelo por equipes comerciais, contribuindo para sua aplicabilidade real. Dessa forma, os trabalhos futuros abrem espaço não apenas para avanços técnicos, mas também para o fortalecimento da interface entre ciência de dados e tomada de decisão prática no setor B2B de distribuição de bebidas.

6.6 Implicações e Impacto

As implicações desta pesquisa ultrapassam o desenvolvimento de um artefato técnico e alcançam dimensões estratégicas, operacionais e científicas relevantes para o ecossistema de decisão no varejo B2B de bebidas. O artefato baseado em **Heterogeneous Graph Transformer (HGT)** aliado ao **Double Machine Learning** inaugura uma abordagem que reconcilia modelagem relacional e inferência causal, produzindo estimativas mais consistentes para decisões comerciais que tradicionalmente dependem de modelos correlacionais suscetíveis a vieses estruturais.

No plano estratégico, os resultados demonstram que a incorporação explícita de relações entre produtos, clientes e contextos comerciais altera de maneira significativa a compreensão das dinâmicas de mercado. Isso tem impacto direto no modo como empresas distribuidoras concebem políticas de precificação, estratégias de recomendação e alocação de campanhas. A capacidade de isolar efeitos causais confere aos gestores uma base empírica mais robusta para avaliar o impacto marginal de intervenções e, portanto, para otimizar investimentos promocionais.

Em dimensão operacional, o artefato afeta positivamente a eficiência de processos internos ao permitir diagnósticos mais precisos sobre padrões de vendas, elasticidade por categoria e comportamento heterogêneo dos pontos de venda. A integração com práticas de **ModelOps** potencializa esse impacto ao assegurar que o modelo se mantenha atualizado, monitorado e aderente às mudanças estruturais do mercado. Isso contribui para a continuidade operacional, reduz riscos associados ao *drift* de dados e fortalece a sustentação analítica de sistemas de decisão baseados em IA.

As implicações científicas também são relevantes. A combinação entre HGT e DML evidencia que a convergência entre aprendizado profundo e métodos estatísticos ortogonais é uma via produtiva para enfrentar problemas clássicos de confundimento e de dependências complexas em dados reais. A pesquisa demonstra, na prática, que arquiteturas de grafos heterogêneos são alternativas viáveis para modelar sistemas comerciais intrincados, estimulando novos estudos sobre causalidade em redes, explicabilidade relacional e estratégias híbridas de aprendizagem.

Por fim, o impacto maior reside na transformação do processo decisório, que passa a ser sustentado por uma visão mais estruturada das relações entre entidades e por estimativas causais que reduzem inconsistências — especialmente em cenários de *cold start* e mercados altamente dinâmicos. Ao aprimorar tanto a precisão quanto a confiança nas decisões empresariais, o artefato contribui para elevar o nível de maturidade analítica do setor, preparando terreno para inovações futuras em modelagem causal aplicada ao ambiente corporativo.

6.7 Considerações Finais

A pesquisa desenvolvida consolidou um arcabouço metodológico e tecnológico capaz de enfrentar um dos desafios mais persistentes na modelagem de sistemas comerciais complexos: a dissociação entre correlação estrutural e causalidade. A construção do artefato baseado na integração entre **Heterogeneous Graph Transformer (HGT)** e **Double Machine Learning** demonstrou que é possível estruturar modelos que respeitam a natureza relacional do domínio B2B de distribuição de bebidas e, ao mesmo tempo, produzem estimativas causais mais fidedignas para orientar decisões estratégicas.

Os resultados obtidos evidenciam que intervenções comerciais — sejam alterações de preço, recomendações de produtos ou campanhas de incentivo — podem ser avaliadas com maior precisão quando o modelo incorpora a rede heterogênea que conecta produtos, categorias, pontos de venda e contexto temporal. A combinação entre o poder representacional do HGT e a robustez estatística do DML constituiu um caminho sólido para reduzir vieses, especialmente aqueles decorrentes de confundidores estruturais e padrões históricos que mascaram os efeitos reais das ações comerciais.

O trabalho também demonstrou a relevância de integrar, em um único projeto, três níveis de planejamento: o **Design Science Research** para estruturar o propósito e o artefato, o **CRISP-DM** para conduzir o processo analítico de forma tática e o **ModelOps** para garantir sustentabilidade e governança operacional. Essa integração reforça que soluções de IA só alcançam impacto real quando articulam inovação conceitual, rigor metodológico e capacidade de implantação contínua.

Embora existam limitações inerentes aos dados disponíveis e à própria complexidade das técnicas utilizadas, a pesquisa contribui para ampliar o diálogo entre aprendizado estatístico e tomada de decisão corporativa, oferecendo uma via metodológica replicável e extensível para outros setores caracterizados por interdependências ricas e dinâmicas complexas.

Dessa forma, as considerações finais reafirmam que o artefato proposto não apenas avança o estado da arte em modelagem causal sobre grafos heterogêneos, mas também se configura como um instrumento aplicável para aprimorar a qualidade das decisões comerciais em ambientes B2B. O conjunto de evidências, análises e procedimentos discutidos ao longo do trabalho deixa claro que o caminho mais promissor para a evolução da inteligência analítica corporativa está na combinação entre representações estruturais ricas, métodos causais rigorosos e práticas sólidas de engenharia de modelos.

7. GESTÃO DO PROJETO

7.1 Viabilidade

A viabilidade do projeto decorre da convergência entre maturidade metodológica, disponibilidade efetiva de dados e adequação tecnológica para implementar o artefato baseado em **Heterogeneous Graph Transformer (HGT)** integrado ao **Double Machine Learning**. A combinação desses fatores sustenta a exequibilidade técnica, científica e operacional da proposta dentro do escopo da pesquisa e do contexto organizacional típico de distribuidoras de bebidas em ambiente B2B.

No plano técnico, o projeto é viável porque as bases fornecidas — *transacoes_vendas*, *produtos_catalogo* e *pontos_venda* — possuem granularidade, diversidade de atributos e estrutura relacional suficientes para compor um grafo heterogêneo completo. Esse arranjo atende às exigências fundamentais do HGT, que depende de múltiplos tipos de entidades e relações para operar seus mecanismos de atenção dependentes de metatipos. Simultaneamente, o conjunto de variáveis presentes nas bases permite a aplicação rigorosa do *Double Machine Learning*, garantindo que os modelos auxiliares necessários à estimação ortogonal sejam treinados com pistas informacionais adequadas.

A viabilidade científica está ancorada na forte base teórica que sustenta tanto o HGT quanto o DML. A literatura de aprendizado profundo em grafos estabelece que arquiteturas heterogêneas são apropriadas para domínios com múltiplas entidades interdependentes, como redes comerciais. Da mesma forma, a literatura de inferência causal baseada em mecanismos ortogonais demonstra que o DML pode reduzir vieses em cenários com múltiplos confundidores observáveis e estruturas complexas. A integração desses componentes, portanto, situa-se dentro de um território metodológico consolidado, ainda que inovador.

Sob a perspectiva operacional, o projeto mostra-se viável porque suas etapas se alinham ao arcabouço **ModelOps**, que fornece diretrizes claras para versionamento, monitoramento, auditoria e implantação contínua de modelos de IA. O ambiente computacional necessário — incluindo frameworks de grafos, rotinas de validação cruzada e infraestrutura de processamento distribuído — é compatível com tecnologias amplamente disponíveis em plataformas corporativas e acadêmicas.

No aspecto estratégico, a viabilidade é reforçada pelo alinhamento direto entre o artefato proposto e necessidades reais do setor B2B de bebidas, especialmente no tocante à melhoria da acurácia de recomendações, à redução de perdas comerciais causadas por vieses e ao aprimoramento da alocação de recursos promocionais. O projeto, portanto, não apresenta apenas viabilidade técnica, mas também relevância organizacional, aumentando a probabilidade de adoção e continuidade.

Com isso, a viabilidade se estabelece como produto de um ambiente metodologicamente sólido, dados suficientes, ferramentas adequadas e clara demanda empresarial. Essa combinação cria condições objetivas para desenvolvimento, avaliação e implantação eficaz do artefato ao longo do ciclo proposto.

7.2 Riscos Operacionais e Mitigação

A execução do projeto envolve riscos operacionais decorrentes da complexidade metodológica, da dependência de dados heterogêneos e da necessidade de integração contínua entre modelagem estatística, engenharia de grafos e práticas de ModelOps. A identificação e a mitigação antecipada desses riscos são fundamentais para garantir estabilidade, confiabilidade e continuidade durante o desenvolvimento e a implantação do artefato.

O primeiro risco operacional está relacionado à **qualidade e estabilidade das bases de dados**, que podem sofrer alterações estruturais ao longo do tempo — como inclusão de novos SKUs, mudanças cadastrais em pontos de venda ou alterações em políticas comerciais. Variações desse tipo podem afetar a consistência do grafo heterogêneo e, por consequência, comprometer o desempenho do HGT. A mitigação exige a implementação de mecanismos automáticos de detecção de *schema drift*, validação contínua de integridade relacional e versionamento de dados, conforme recomenda o ciclo de governança do ModelOps.

O segundo risco deriva da **sensibilidade dos modelos HGT e DML a problemas de sobreajuste ou instabilidade**. O HGT, ao incorporar mecanismos de atenção dependentes de metatipos, pode superexplorar padrões específicos de entidades ou períodos curtos, especialmente em segmentos com poucos registros. Já o DML, ao depender de estimadores auxiliares, pode amplificar variabilidade caso esses modelos sejam inadequadamente regularizados. A mitigação desses riscos envolve a adoção de rotinas sistemáticas de validação cruzada, restrições de complexidade nos modelos auxiliares, regularização adequada e *early stopping* durante o treinamento dos modelos de grafos.

O terceiro risco refere-se à **integração operacional e manutenção contínua do artefato**, especialmente em ambientes corporativos nos quais múltiplos sistemas comerciais coexistem. A implantação de grafos heterogêneos e pipelines de inferência causal exige compatibilidade com sistemas legados, disponibilidade de infraestrutura para processamento distribuído e capacidade contínua de monitoramento. A mitigação requer o estabelecimento de práticas sólidas de ModelOps: auditoria de modelos, monitoramento de *drift* causal e relacional, logs estruturados, pipelines reprodutíveis e políticas de retreinamento baseadas em gatilhos estatísticos.

Um quarto risco operacional está associado à **interoperabilidade entre equipes técnicas e equipes de negócio**. A adoção de modelos de grafos e estimativas causais pode encontrar resistência ou gerar interpretações equivocadas em equipes responsáveis por precificação e promoção. A mitigação envolve o desenvolvimento de ferramentas de explicabilidade voltadas a grafos — capazes de esclarecer relações determinantes no modelo — e a criação de documentação clara, treinamentos internos e interfaces visuais de acompanhamento.

Por fim, há riscos relacionados ao **custo computacional** das operações. O HGT é intensivo em recursos, especialmente quando o grafo cresce em número de nós e metarrelacionamentos. Estratégias de mitigação incluem *sampling* de subgrafos, processamento incremental, utilização de mini-batches estruturados e escalonamento em ambientes de nuvem otimizados para grafos.

Essas estratégias, articuladas de forma contínua, visam garantir que o projeto se mantenha operacionalmente resiliente, reduzindo vulnerabilidades e assegurando a confiabilidade do artefato ao longo de todo o ciclo de vida.

7.3 Cronograma

O cronograma do projeto organiza-se em fases sequenciais e interdependentes, refletindo a estrutura estratégica do **Design Science Research (DSR)**, a lógica tática do **CRISP-DM** e as exigências operacionais do **ModelOps**. A distribuição temporal das atividades assegura que cada etapa receba o nível adequado de exploração metodológica, validação empírica e integração técnica, garantindo coerência entre construção científica e aplicabilidade prática do artefato.

A primeira fase, correspondente à **compreensão do negócio e dos dados** (CRISP-DM), abrange o período inicial do projeto. Nessa etapa, são examinadas profundamente as bases *transacoes_vendas*, *produtos_catalogo* e *pontos_venda*, permitindo a construção preliminar do grafo heterogêneo e a avaliação da suficiência das variáveis disponíveis. Em paralelo, consolida-se a fundamentação teórica e formaliza-se o enquadramento metodológico dentro do DSR.

A segunda fase dedica-se ao **design do artefato**, em que são definidas a arquitetura conceitual, a lógica relacional e a estrutura de integração entre o **Heterogeneous Graph Transformer (HGT)** e o **Double Machine Learning**. Trata-se de uma etapa central, que envolve prototipação e testagem modular dos componentes, garantindo aderência ao problema raiz de modelagem causal em ambientes relacionais complexos.

Na terceira fase, concentram-se as atividades de **desenvolvimento e treinamento**. O grafo heterogêneo é operacionalizado, os modelos auxiliares do DML são calibrados e o pipeline de ModelOps é estruturado. Esse período demanda alto esforço computacional, além de validações sucessivas para assegurar estabilidade, controle de vieses e aderência dos resultados ao comportamento real observado no setor B2B.

A quarta fase contempla a **demonstração e avaliação**, etapa alinhada ao DSR, na qual o artefato é aplicado a cenários de intervenção simulada e realista: variações de preço, recomendação personalizada e análise de elasticidade. Essa fase incorpora testes quantitativos, avaliação contra *baselines* e experimentos de sensibilidade relacional e causal.

Por fim, a quinta fase é dedicada à **implantação e comunicação**, que inclui documentação técnica, preparação de materiais de exposição, estabilização do pipeline de ModelOps e redação dos resultados acadêmicos. A comunicação envolve a apresentação do artefato e de evidências empíricas que demonstram sua relevância científica e aplicabilidade organizacional.

Organizado dessa forma, o cronograma promove uma progressão lógica entre análise, modelagem, experimentação e implementação, garantindo que o projeto seja conduzido com rigor metodológico e eficiência operacional.

7.4 Recursos Necessários

A execução do projeto requer um conjunto de recursos distribuídos em três dimensões fundamentais — computacionais, humanos e organizacionais — cada uma contribuindo de maneira específica para garantir robustez metodológica, estabilidade operacional e aderência às exigências científicas do artefato baseado em **Heterogeneous Graph Transformer (HGT)** combinado ao **Double Machine Learning (DML)**.

Na dimensão computacional, o projeto demanda infraestrutura capaz de lidar com processamento intensivo de grafos heterogêneos. Modelos como o HGT exigem GPU(s) de médio a alto desempenho, memória suficiente para carregamento de subgrafos complexos e possibilidade de escalonamento horizontal para experimentos de validação cruzada e monitoramento contínuo. Além disso, o pipeline de ModelOps requer ambiente com suporte a containers, versionamento de modelos, monitoramento de *drift* e execução agendada de rotinas de retreinamento. Ferramentas como PyTorch Geometric ou Deep Graph Library (DGL), somadas ao orquestrador MLflow, compõem o núcleo tecnológico necessário.

No eixo humano, o projeto exige a participação de profissionais com competências diversas. Especialistas em ciência de dados são responsáveis pela modelagem estatística e causal, incluindo a configuração dos estimadores auxiliares do DML. Engenheiros de machine learning assumem o desenvolvimento do pipeline de produção, o versionamento dos modelos e a construção dos mecanismos de monitoramento contínuo. Profissionais de engenharia de dados tornam-se essenciais para manter a integridade das bases *transacoes_vendas*, *produtos_catalogo* e *pontos_venda*, garantindo qualidade relacional para construção do grafo heterogêneo. A participação de analistas de negócio do setor B2B de bebidas complementa o processo, permitindo interpretar resultados, validar premissas e garantir alinhamento entre o artefato e a realidade operacional.

Do ponto de vista organizacional, é necessário disponibilizar acesso contínuo às bases de dados, políticas claras de governança e infraestrutura habilitada para experimentação com dados sensíveis. A colaboração entre equipes técnicas e áreas de negócio requer espaços formais de interação, revisões periódicas e alinhamento entre requisitos analíticos e necessidades comerciais. A viabilidade operacional depende, ainda, de um ambiente institucional que permita testes controlados de recomendações, variações de preço ou experimentos de impacto — etapa-chave para validação causal do modelo.

Assim, os recursos necessários formam um ecossistema integrado, no qual infraestrutura, especialistas e organização trabalham de forma coordenada. Esse arranjo assegura não apenas a construção metodologicamente rigorosa do artefato, mas também sua capacidade de operar de forma estável e gerar valor contínuo ao longo de seu ciclo de vida.

7.5 Disseminação dos Resultados

A disseminação dos resultados constitui a etapa final da gestão do projeto, assegurando que o artefato desenvolvido — a integração entre **Heterogeneous Graph Transformer (HGT)** e **Double Machine Learning (DML)** — alcance tanto a comunidade científica quanto os ambientes organizacionais onde sua aplicação é relevante. A estratégia de disseminação deve refletir a natureza técnica e interdisciplinar do trabalho, equilibrando profundidade metodológica com clareza aplicada.

No âmbito acadêmico, os resultados podem ser divulgados por meio da elaboração de artigos científicos direcionados a periódicos e conferências internacionais que tratam de aprendizado em grafos, inferência causal ou sistemas de recomendação, como *NeurIPS*, *KDD*, *ICDM* ou *Journal of Machine Learning Research*. A contribuição central — a capacidade de modelar relações heterogêneas e isolar efeitos causais em ambientes comerciais — alinha-se às discussões contemporâneas sobre robustez, generalização e explicabilidade em modelos de IA, reforçando a pertinência acadêmica da publicação. A produção de relatórios técnicos também pode ampliar a visibilidade do artefato em comunidades de pesquisa aplicada e grupos de engenharia de dados.

No contexto empresarial, a disseminação ocorre por meio de apresentações estruturadas para tomadores de decisão, equipes de planejamento comercial e especialistas em pricing. Demonstrações do artefato, apoiadas por visualizações claras de embeddings, relações entre entidades e efeitos causais estimados, favorecem entendimento sobre como o modelo pode melhorar ações promocionais, otimizar recomendações e reduzir vieses nas estratégias comerciais. Workshops internos e materiais de documentação operacional complementam essa etapa, facilitando a integração do artefato aos processos das equipes envolvidas.

A disseminação também contempla a esfera técnica operacional, com a disponibilização de *templates* de pipeline, documentação de arquitetura e diretrizes para implantação em ambientes de ModelOps. Essa ação ajuda equipes de engenharia e MLOps a compreender fluxos de versionamento, monitoramento de *drift* relacional, triggers de retreinamento e boas práticas de auditoria. A abertura parcial de códigos-fonte, quando apropriado, pode fomentar colaboração e fortalecer a credibilidade do projeto.

Por fim, a disseminação deve preservar coerência com os princípios de reprodutibilidade científica e precisão analítica, assegurando que os resultados — tanto empíricos quanto metodológicos — estejam devidamente contextualizados e documentados. Dessa forma, a etapa final de gestão não apenas amplia o alcance da pesquisa, mas também consolida sua relevância como referência metodológica para futuras iniciativas que integrem aprendizado estatístico, grafos heterogêneos e inferência causal em sistemas reais.

8. REFERÊNCIAS BIBLIOGRÁFICAS

CHAPMAN, P. et al. *CRISP-DM 1.0: Step-by-step data mining guide.* The CRISP-DM Consortium, 2000.

CHERNOZHUKOV, V. et al. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, v. 21, n. 1, p. C1–C68, 2018.

FOSTER, M.; HAMMERMESH, D. *Big Data in Economics: Challenges and Opportunities*. Cambridge: MIT Press, 2022.

HAMILTON, W.; YING, Z.; LESKOVEC, J. Inductive representation learning on large graphs. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Proceedings… 2017.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2. ed. New York: Springer, 2009.

HU, Z. et al. Heterogeneous Graph Transformer. In: *Proceedings of the Web Conference (WWW)*. Proceedings… 2020.

IMBENS, G. W.; RUBIN, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge: Cambridge University Press, 2015.

PEARL, J. *Causality: Models, Reasoning, and Inference.* 2. ed. Cambridge: Cambridge University Press, 2009.

PEARL, J.; MACKENZIE, D. *The Book of Why: The New Science of Cause and Effect.* New York: Basic Books, 2018.

PROVOST, F.; FAWCETT, T. *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking.* Sebastopol: O'Reilly Media, 2013.

RICCI, F.; ROKACH, L.; SHAPIRA, B.; KANTOR, P. B. *Recommender Systems Handbook.* 2. ed. Cham: Springer, 2015.

SCHLICHTKRULL, M. et al. Modeling relational data with graph convolutional networks. In: *European Semantic Web Conference (ESWC)*. Proceedings… 2018.

SHARMA, A.; KUMAR, N.; SINGH, A. Causal representation learning: a survey and future directions. *Journal of Machine Learning Research*, 2023.

SHMUELI, G. To explain or to predict? *Statistical Science*, v. 25, n. 3, p. 289–310, 2010.

YAO, L. et al. A survey on causal inference for recommender systems. *ACM Computing Surveys*, 2021.

ZHAO, T. et al. Heterogeneous graph representation learning: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

ZHANG, Y. et al. Heterogeneous graph neural network. In: *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*. Proceedings… 2019.

9. GLOSSÁRIO

Atenção Multi-Cabeças (Multi-Head Attention)

Mecanismo de aprendizado utilizado em arquiteturas Transformer que permite ao modelo analisar diferentes padrões de dependência simultaneamente. No contexto do HGT, cada cabeça pode capturar relações distintas entre tipos de entidades e metarrelacionamentos.

Aprendizado Baseado em Grafos (Graph-based Learning)

Categoria de métodos de aprendizado que utilizam estruturas de grafo para representar entidades e suas relações. Esses métodos são especialmente úteis em domínios com interdependências complexas, como redes de produtos, clientes e categorias.

Artefato (Design Science Research)

Resultado produzido por uma pesquisa DSR, podendo ser um modelo, método, processo ou sistema destinado a resolver um problema real. Neste trabalho, o artefato corresponde ao modelo causal-relacional baseado em HGT + DML.

Bias / Viés

Desvio sistemático que leva um modelo a produzir estimativas incorretas. No problema estudado, o foco está em vieses provenientes da mistura entre correlação estrutural e causalidade.

Causalidade

Relação em que uma intervenção produz um efeito mensurável. Difere de correlação, que apenas mede associação entre variáveis. Os métodos de DML são empregados para estimar efeitos causais.

Causal Graph / Grafo Causal

Representação que descreve as dependências estruturais e relações causais entre variáveis. Em domínios heterogêneos, grafos causais podem incluir múltiplos tipos de nós e relações.

Cold Start

Condição na qual há pouco ou nenhum histórico disponível para um produto, cliente ou ponto de venda. Modelos correlacionais tradicionais tendem a falhar nesse cenário, enquanto grafos heterogêneos mitigam parcialmente o problema.

CRISP-DM (Cross-Industry Standard Process for Data Mining)

Metodologia que organiza projetos de ciência de dados em fases táticas: entendimento de negócio, entendimento de dados, preparação, modelagem, avaliação e implantação.

DGL (Deep Graph Library)

Biblioteca amplamente utilizada para construção e treinamento de modelos de aprendizado em grafos, incluindo variantes heterogêneas.

Double Machine Learning (DML)

Método estatístico que utiliza dois modelos auxiliares para estimar efeitos causais de forma ortogonalizada, reduzindo o viés introduzido por confundidores observáveis. É utilizado após o HGT para isolar efeitos de intervenções comerciais.

Drift (Deslocamento de Distribuição)

Mudanças estatísticas ao longo do tempo que alteram o comportamento dos dados. Pode ser de características, alvo ou estrutura relacional. O ModelOps é responsável por monitorá-lo.

DSR (Design Science Research)

Estrutura estratégica de pesquisa baseada na criação e avaliação de artefatos que resolvem problemas reais. Orienta todo o projeto, incluindo validação empírica e contribuição prática.

Elasticidade de Demanda

Medida do impacto de variações de preço na quantidade demandada. Em ambientes complexos, estimar elasticidades exige métodos causais que separem efeitos estruturais de relações espúrias.

Entidade

Elemento representado em um grafo heterogêneo (ex.: produto, ponto de venda, categoria, data). Cada entidade pode possuir atributos distintos e relações específicas.

Grafo Heterogêneo

Estrutura composta por múltiplos tipos de nós e múltiplos tipos de arestas. Permite modelar domínios com entidades diversas e relações ricas, como redes comerciais B2B.

Heterogeneous Graph Transformer (HGT)

Arquitetura de aprendizado profundo projetada para grafos heterogêneos, combinando atenção multi-cabeças com projeções dependentes do metatipo de entidade e relação. Permite representar padrões comerciais complexos.

Inferência Causal

Processo de estimar o efeito de ações ou intervenções sobre resultados observáveis. Diferencia-se de análises puramente correlacionais. O DML é o método central de inferência causal neste trabalho.

Metarrelacionamento (Meta-relation)

Tipo de relação que conecta entidades específicas em um grafo heterogêneo. Ex.: *PDV compra Produto*, *Produto pertence a Categoria*.

ModelOps

Conjunto de práticas para gerenciar todo o ciclo de vida de modelos de IA, desde versionamento e monitoramento até auditoria, retreinamento e implantação. Estrutura a camada operacional do artefato.

Nó (Node)

Unidade fundamental de um grafo, representando entidades como produtos ou pontos de venda. Em grafos heterogêneos, cada nó pode possuir tipo e atributos distintos.

Pipeline

Sequência de etapas executadas para processar dados, treinar modelos e gerar inferências. No projeto, o pipeline integra preparação de dados, construção de grafos, treinamento do HGT, estimação causal via DML e monitoramento.

Representação (Embedding)

Vetor numérico que sintetiza propriedades de uma entidade ou relação. O HGT produz embeddings contextuais que refletem a posição da entidade na rede.

Relação (Edge)

Conexão entre dois nós em um grafo. Em redes comerciais, pode representar compra, similaridade, coocorrência ou associação categorial.

Schema Drift

Mudança na estrutura dos dados que afeta tabelas, colunas, tipos ou relacionamentos. É crítico para grafos heterogêneos e exige monitoramento contínuo.

Tratamento (Treatment)

Intervenção cujo efeito se deseja estimar, como redução de preço ou recomendação personalizada. No DML, o tratamento é modelado separadamente do resultado para garantir estimação causal adequada.

Workflow (fluxo de trabalho)

É a sequência organizada de tarefas, atividades e informações que precisam ser executadas em uma ordem específica para alcançar um objetivo ou resultado, sendo visualizado como um mapa que define responsabilidades e interligações, do início ao fim de um processo, podendo ser manual, semi-automatizado ou totalmente digitalizado. Ele serve para padronizar, otimizar e trazer clareza a processos repetitivos, aumentando a produtividade e reduzindo erros.