



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CAMPUS AGRESTE
NÚCLEO DE TECNOLOGIA
CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

MARCELO JANDOY SOUSA COSTA LEMOS

**ANÁLISE DE PADRÕES DE CONSUMO EM SUPERMERCADOS COM BASE EM
REGRAS DE ASSOCIAÇÃO UTILIZANDO O ALGORITMO APRIORI**

CARUARU

2025

MARCELO JANDOY SOUSA COSTA LEMOS

**ANÁLISE DE PADRÕES DE CONSUMO EM SUPERMERCADOS COM BASE EM
REGRAS DE ASSOCIAÇÃO UTILIZANDO O ALGORITMO APRIORI**

Trabalho de Conclusão de Curso apresentado à
Coordenação do Curso de Engenharia de
Produção do Campus Agreste da Universidade
Federal de Pernambuco – UFPE, na modalidade
de monografia, como requisito parcial para a
obtenção do grau de bacharel em Engenharia de
produção.

Área de concentração: Pesquisa Operacional

Orientador (a): Dr. Augusto José da Silva Rodrigues

CARUARU

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Lemos, Marcelo Jandoy Sousa Costa.

Análise de padrões de consumo em supermercados com base em regras de associação utilizando o algoritmo APRIORI / Marcelo Jandoy Sousa Costa
Lemos. - Caruaru, 2025.

42p.

Orientador(a): Augusto José da Silva Rodrigues

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro Acadêmico do Agreste, , 2025.

1. Pesquisa Operacional. 2. Regras de associação. 3. Mineração de dados. 4. Algoritmo Apriori. 5. Análise de cestas. 6. Tomada de decisão. I. Rodrigues, Augusto José da Silva. (Orientação). II. Título.

620 CDD (22.ed.)

MARCELO JANDOY SOUSA COSTA LEMOS

**ANÁLISE DE PADRÕES DE CONSUMO EM SUPERMERCADOS COM BASE EM
REGRAS DE ASSOCIAÇÃO UTILIZANDO O ALGORITMO APRIORI**

Trabalho de Conclusão de Curso apresentado à
Coordenação do Curso de Engenharia de
Produção do Campus Agreste da Universidade
Federal de Pernambuco – UFPE, na modalidade
de monografia, como requisito parcial para a
obtenção do grau de bacharel em Engenharia de
Produção.

Aprovada em: 16/12/2025

BANCA EXAMINADORA

Dr. Augusto José da Silva Rodrigues (Orientador)
Universidade Federal de Pernambuco

Dra. Amanda Carvalho Miranda (Examinadora Interna)
Universidade Federal de Pernambuco

Ma. Lorena Vieira Santos Rodrigues (Examinadora Externa)
Universidade Federal de Pernambuco

Dedico este trabalho à minha mãe, Mércia, que sempre fez de tudo para garantir educação e oportunidades para seus filhos; ao meu irmão, Jandoy, exemplo de integridade e motivo de admiração; ao meu pai, Jandoy, por despertar em mim um olhar mais crítico sobre a vida por meio de nossas conversas e debates; e à minha namorada, Isabelle, pelo apoio constante, incentivo e força ao longo desta caminhada.

AGRADECIMENTOS

Este agradecimento é dedicado a todos que fizeram parte da construção não apenas deste trabalho, mas também do meu desenvolvimento pessoal e acadêmico, desde o início da minha trajetória estudantil até os dias atuais.

À minha mãe, Mércia, por todo amor, dedicação, suporte e perseverança. Obrigado por sempre fazer o impossível para garantir a melhor educação e formação possível, contribuindo com minha evolução pessoal e profissional.

À minha namorada, Isabelle, por estar ao meu lado em todos os momentos, oferecendo conselhos, motivação e paciência. Obrigado por acreditar em mim, incentivar meu crescimento e me ajudar a ser uma pessoa melhor a cada dia.

Ao meu irmão, Jandoy, exemplo de caráter e determinação. Tive o privilégio de seguir seus passos e aprender com sua trajetória, enfrentando desafios que me fortaleceram e me motivaram a persistir sempre. Agradeço por sua integridade e por me mostrar diariamente o valor de ser alguém correto e dedicado.

Ao meu pai, Jandoy, pelos debates enriquecedores e por despertar em mim um olhar crítico sobre o mundo. Seu incentivo constante tornou minha caminhada mais leve e contribuiu diretamente para minha formação intelectual e pessoal.

Ao meu orientador, professor Augusto, pelos valiosos conselhos, tanto nas aulas quanto no desenvolvimento deste TCC e no estágio. Agradeço pelas orientações que sempre buscaram expandir minhas capacidades, pelos desafios propostos e por sua preocupação genuína com o aprendizado e o futuro dos alunos. Sou muito grato por ter encontrado, além de um professor, um amigo que levarei para a vida.

Aos professores, colegas de curso e integrantes da empresa júnior com quem convivi na graduação, meu sincero agradecimento pelo aprendizado e pela convivência que deram sentido à minha trajetória acadêmica e profissional. Em especial, aos professores Osmar Veras, José Leão, Amanda Carvalho, Walton Coutinho, Thalles Garcez, Isaac Pergher, Lucimario Gois, Eduardo Novais, Ramon Swell e Renato Cintra, pelos conhecimentos compartilhados; e aos colegas Matheus Moreira, Hatus William, Gizele Laryssa, Guilherme Leandro, Wesllyanne Carvalho e Sthephanya Vitória, que tornaram essa jornada mais leve e inspiradora.

Agradeço também à Universidade Federal de Pernambuco pela oportunidade de estudar em uma instituição de excelência, com professores qualificados e um ambiente que impulsiona o desenvolvimento acadêmico e pessoal. Tenho orgulho de fazer parte dessa instituição e levarei comigo tudo que ela representa.

“Faça o teu melhor, nas condições que você tem, enquanto você não tem condições melhores para fazer melhor ainda.” (CORTELLA, 2007.)

RESUMO

A crescente disponibilidade de dados no setor varejista tem ampliado as possibilidades de utilizar técnicas analíticas para compreender o comportamento de compra dos consumidores e apoiar decisões estratégicas. Este trabalho investiga a aplicação de métodos de mineração de dados, com foco na extração de regras de associação, para identificar padrões recorrentes em cestas de compras a partir de uma base de notas fiscais de uma rede de supermercados. O algoritmo Apriori foi utilizado para identificar *itemsets* (conjuntos de itens) frequentes e gerar regras avaliadas pelas métricas de suporte (*support*), confiança (*confidence*) e alavancagem (*lift*), permitindo interpretar relações de complementaridade e coocorrência entre produtos. A metodologia envolveu a preparação da base de dados, a transformação das compras em transações, a modelagem das associações e a análise visual dos padrões obtidos. Os resultados revelaram a existência de agrupamentos fortes entre itens relacionados ao preparo de refeições domésticas, como batata inglesa, cenoura, tomate e cebola, que apresentaram altos valores de *lift* e confiança. Esses padrões demonstram que a compra de um item influencia diretamente a probabilidade de aquisição de outro, configurando relações de complementaridade que podem ser exploradas para otimização do layout da loja, criação de promoções cruzadas e planejamento de estoque. As análises gráficas reforçaram tais conclusões ao evidenciar concentrações de regras com elevada relevância estatística. O estudo mostra, portanto, que a identificação de associações entre produtos oferece suporte valioso para a gestão operacional e estratégica do varejo, contribuindo para decisões mais eficientes e alinhadas ao comportamento real do consumidor, além de abrir caminho para investigações futuras que explorem modelos mais avançados ou bases de dados ampliadas.

Palavras-chave: Mineração de dados; Regras de associação; Algoritmo Apriori; Varejo; Análise de cestas; tomada de decisão.

ABSTRACT

The increasing availability of data in the retail sector has broadened the possibilities of using analytical techniques to understand consumer purchasing behavior and support strategic decisions. This work investigates the application of data mining methods, focusing on the extraction of association rules, to identify recurring patterns in shopping baskets from a database of receipts from a supermarket chain. The Apriori algorithm was used to identify frequent itemsets and generate rules evaluated by support, confidence, and leverage metrics, allowing the interpretation of complementarity and co-occurrence relationships between products. The methodology involved preparing the database, transforming purchases into transactions, modeling the associations, and visually analyzing the patterns obtained. The results revealed the existence of strong groupings among items related to the preparation of home meals, such as potatoes, carrots, tomatoes, and onions, which presented high lift and confidence values. These patterns demonstrate that the purchase of one item directly influences the probability of acquiring another, configuring complementarity relationships that can be explored for optimizing store layout, creating cross-promotions, and inventory planning. Graphical analyses reinforced these conclusions by highlighting concentrations of rules with high statistical relevance. The study therefore shows that identifying associations between products offers valuable support for operational and strategic retail management, contributing to more efficient decisions aligned with real consumer behavior, and paving the way for future investigations exploring more advanced models or expanded databases.

Keywords: Data mining; Association rules; Algorithm Apriori; Retail; Basket analysis; Decision making.

LISTA DE FIGURAS

Figura 1 –	Etapas do processo de KDD	18
Figura 2 –	Pseudocódigo do algoritmo Apriori	27
Figura 3 –	Gráfico de Itens mais frequentes	30
Figura 4 –	Gráfico de Correlação entre Confiança e Lift	33
Figura 5 –	Mapa de calor dos produtos antecedentes e consequentes de acordo com o lift médio.	35

LISTA DE TABELAS

Tabela 1 –	Itens mais frequentes	30
Tabela 2 –	Relação de confiança	32

LISTA DE ABREVIATURAS E SIGLAS

Apriori - Algoritmo de Construção de Regras de Associação

KDD - *Knowledge Discovery in Databases*

MBA – *Market Basket Analysis*

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	OBJETIVO GERAL E ESPECÍFICOS.....	14
1.2	JUSTIFICATIVA.....	15
2	REFERENCIAL TEÓRICO.....	17
2.1	MINERAÇÃO DE DADOS E PROCESSO KDD.....	17
2.2	REGRAS DE ASSOCIAÇÃO.....	19
2.3	ALGORITMO APRIORI.....	20
2.4	APLICAÇÃO NO VAREJO.....	22
3	METODOLOGIA.....	24
3.1	TIPO DE PESQUISA.....	24
3.2	DADOS.....	25
3.3	PROCEDIMENTOS.....	26
3.4	PSEUDOCÓDIGO DO APRIORI.....	27
4	RESULTADOS	29
4.1	CARACTERIZAÇÃO DA BASE DE DADOS ANALISADA.....	29
4.2	ITENS E CONJUNTO DE ITENS MAIS FREQUENTES.....	30
4.3	REGRAS DE ASSOCIAÇÃO GERADAS.....	31
4.4	MAPA DE CALOR DO <i>LIFT</i>	34
4.5	DISCUSSÃO E IMPLICAÇÕES GERENCIAIS.....	36
5	CONSIDERAÇÕES FINAIS.....	39
5.1	SUGESTÕES PARA TRABALHOS FUTUROS.....	40
	REFERÊNCIAS.....	41

1 INTRODUÇÃO

O avanço da globalização e o desenvolvimento tecnológico das últimas décadas têm impulsionado as organizações a buscar estratégias baseadas em dados para garantir sua competitividade. Em um mercado dinâmico e altamente competitivo, como o setor supermercadista, a capacidade de transformar dados em informação útil e conhecimento aplicável torna-se um diferencial estratégico essencial (Sacilotti, 2011; Barney; Hesterly, 2011).

Com o aumento do volume de informações disponíveis, o uso de técnicas de análise inteligente, como a mineração de dados, passou a ser fundamental para apoiar a tomada de decisões e a definição de estratégias empresariais (Freitas, 1993; Han; Kamber; Pei, 2012). Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), o processo de descoberta de conhecimento em bases de dados KDD (*Knowledge Discovery in Databases*) consiste em um conjunto de etapas interativas voltadas à identificação de padrões válidos e úteis em grandes volumes de dados, permitindo às empresas compreenderem melhor seus clientes e otimizar suas operações.

Dentre as técnicas de mineração de dados, destacam-se as regras de associação, amplamente aplicadas em contextos de varejo para identificar relações frequentes entre produtos adquiridos conjuntamente. Uma aplicação clássica dessa regra é análise da cesta de compras (*Market Basket Analysis*) que busca revelar padrões de consumo a partir de bases transacionais, possibilitando ações estratégicas como a organização do layout das prateleiras, promoções cruzadas e o planejamento eficiente do estoque (Gonçalves, 2005; Mainali, 2016).

Para que tais padrões possam ser identificados de maneira consistente e em bases de dados de grande escala, torna-se necessário o emprego de algoritmos específicos de mineração de dados, capazes de automatizar o processo de descoberta de conhecimento. De acordo com Castro e Ferrari (2016), o crescimento do volume de dados nas organizações superou a capacidade humana de análise manual, exigindo o uso de algoritmos especializados que possam extrair conhecimento de forma automatizada. Nesse contexto, o algoritmo Apriori é amplamente utilizado por permitir a identificação de conjuntos de itens frequentes e a geração de regras com base em métricas como suporte, confiança e *lift*, que indicam, respectivamente, a frequência, a probabilidade e a força da associação entre produtos (Han; Kamber; Pei, 2012; Castro; Ferrari, 2016).

No cenário dos supermercados, a aplicação dessas técnicas pode gerar insights valiosos. Por exemplo, identificar que consumidores que compram café também adquirem açúcar permite posicionar esses produtos próximos nas gôndolas, aumentando as vendas e melhorando a experiência de compra. Além disso, esse tipo de análise pode apoiar o desenvolvimento de

promoções personalizadas e o controle de estoques, otimizando recursos e reduzindo desperdícios (Semaan; Graça; Dias, 2006).

Considerando esse contexto, a presente pesquisa busca responder à seguinte questão: quais padrões de coocorrência emergem da aplicação do algoritmo Apriori em dados transacionais de supermercados e de que forma esses padrões podem gerar insights valiosos, ao revelar relações não intuitivas e subsidiar decisões gerenciais baseadas em evidências empíricas? Parte-se da hipótese de que a análise de dados transacionais, quando realizada de forma estruturada, pode gerar informações estratégicas que auxiliem gestores na tomada de decisão.

Para responder à questão proposta, este trabalho foi estruturado em cinco seções. A seção atual apresenta a introdução, contextualizando o tema e destacando a importância da análise de padrões de consumo no varejo. A seção 2 reúne o referencial teórico, abordando os principais conceitos relacionados à mineração de dados e processo de descoberta de conhecimento (KDD), às regras de associação, o algoritmo Apriori e a aplicações no varejo. A seção 3 descreve a metodologia adotada, detalhando a base de dados, o tipo de pesquisa e os procedimentos utilizados para a aplicação do método. A seção 4 apresenta a análise dos resultados obtidos e discute suas implicações para a gestão no setor supermercadista. Por fim, a seção 5 reúne as considerações finais e sugestões para estudos futuros.

1.1 OBJETIVOS GERAL E ESPECÍFICOS

O objetivo geral deste trabalho é identificar padrões de compra entre produtos em uma rede de supermercados, por meio da aplicação de regras de associação com o algoritmo Apriori, visando apoiar a tomada de decisões relacionadas à organização do layout, ao controle de estoque e à definição de estratégias promocionais.

Os objetivos específicos incluem:

- a) Revisar a literatura sobre mineração de dados, regras de associação e suas aplicações no varejo supermercadista;
- b) Preparar e organizar os dados de transações de vendas em formato adequado para análise;
- c) Aplicar algoritmo de mineração de dados, o Apriori, para identificar conjuntos de produtos frequentemente adquiridos juntos;
- d) Calcular e avaliar as regras de associação utilizando métricas como suporte, confiança e lift;
- e) Filtrar e ordenar regras segundo limiares analíticos;

- f) Interpretar as regras resultantes e analisar suas implicações gerenciais no contexto do varejo supermercadista.

1.2 JUSTIFICATIVA

A competitividade do setor supermercadista exige que as empresas adotem práticas de gestão cada vez mais orientadas a dados, capazes de sustentar decisões estratégicas que impactem diretamente a eficiência operacional e a satisfação do cliente. Nesse contexto, o uso de informações provenientes das próprias transações de venda possibilita uma visão mais precisa sobre o comportamento do consumidor, permitindo alinhar as operações às demandas reais do mercado (Castro; Ferrari, 2016).

A aplicação de regras de associação na análise de transações de vendas possibilita identificar padrões de consumo que, de outra forma, passariam despercebidos, oferecendo subsídios para a definição de estratégias de layout, estoque e marketing. Conforme destacam Semaan, Graça e Dias (2006), esse tipo de análise é amplamente utilizado no varejo, pois permite descobrir associações entre produtos e otimizar processos de venda e exposição. Ao empregar essas técnicas, torna-se possível reduzir custos operacionais e aumentar a atratividade do ponto de venda, fatores fundamentais para o sucesso competitivo das redes supermercadistas.

Além disso, estudos voltados para Análise de Cesta de Compras em ambientes varejistas têm demonstrado grande relevância competitiva, especialmente em negócios de pequeno e médio porte. Wibowo et al. (2023) evidenciam que o uso de *Association Rule – Market Basket Analysis* (MBA) permite identificar padrões consistentes de interesse dos consumidores durante o ato de compra, fornecendo subsídios diretos para decisões gerenciais.

Esses autores destacam que em um estudo conduzido em um minimercado, as regras com maior confiança possibilitaram “determinar os próximos passos a serem conduzidos, como o ajuste de layout e outras estratégias operacionais”. Esse tipo de evidência reforça a utilidade prática do MBA não apenas como ferramenta analítica, mas como um mecanismo capaz de orientar rearranjos físicos, otimizar exposição de produtos e apoiar ações de *cross-selling*, aspectos diretamente alinhados aos objetivos do presente trabalho.

Outro aspecto importante é o potencial do estudo para apoiar decisões gerenciais a partir de dados concretos. Segundo Gonçalves (2005), as regras de associação, ao revelar relações não triviais entre produtos, constituem uma das técnicas mais úteis para compreender o comportamento do consumidor em ambientes de alta competitividade.

Apesar disso, observa-se que a empresa analisada ainda não realiza qualquer tipo de estudo estruturado baseado em dados, limitando-se a práticas tradicionais de gestão, como reposição empírica, decisões baseadas na experiência dos gerentes e análises pontuais sem suporte estatístico. Essa ausência de análises formais impede a identificação de padrões consistentes de compra, reduz a precisão no planejamento de estoque e limita a efetividade das estratégias de exposição e promoção.

A lacuna existente entre o grande volume de dados disponíveis nas notas fiscais e a ausência de sua exploração analítica configura uma oportunidade relevante para demonstrar o potencial da mineração de dados como ferramenta estratégica para o varejo.

A escolha do algoritmo Apriori justifica-se por sua ampla utilização e consolidação na literatura para a extração de regras de associação em bases transacionais, especialmente no contexto varejista. O método apresenta simplicidade conceitual, robustez na identificação de padrões frequentes e interpretação direta das métricas de suporte, confiança e *lift*, o que facilita a análise e a aplicação prática dos resultados.

Além da relevância prática, este trabalho se justifica academicamente por aproximar conceitos de Engenharia de Produção e Ciência de Dados. Essa integração de áreas promove o desenvolvimento de habilidades analíticas voltadas à melhoria de processos, ao uso de indicadores de desempenho e à tomada de decisão baseada em evidências (Han; Kamber; Pei, 2012). Assim, a pesquisa contribui para a formação de profissionais mais preparados para lidar com os desafios contemporâneos da gestão de operações e do planejamento estratégico no varejo. Assim, este trabalho busca suprir essa deficiência ao oferecer uma abordagem estruturada e replicável para apoiar a tomada de decisão da empresa com base em evidências.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta a base teórica utilizada para a construção do presente trabalho. Inicialmente, aborda-se o conceito de mineração de dados e o processo KDD, que fundamentam o estudo. Em seguida, discutem-se as regras de associação, consideradas o princípio essencial da análise realizada. Posteriormente, descreve-se o algoritmo Apriori, utilizado como metodologia para a aplicação prática. Por fim, apresenta-se a aplicação dessas técnicas no contexto do varejo, ambiente no qual este estudo foi desenvolvido.

2.1 MINERAÇÃO DE DADOS E PROCESSO KDD

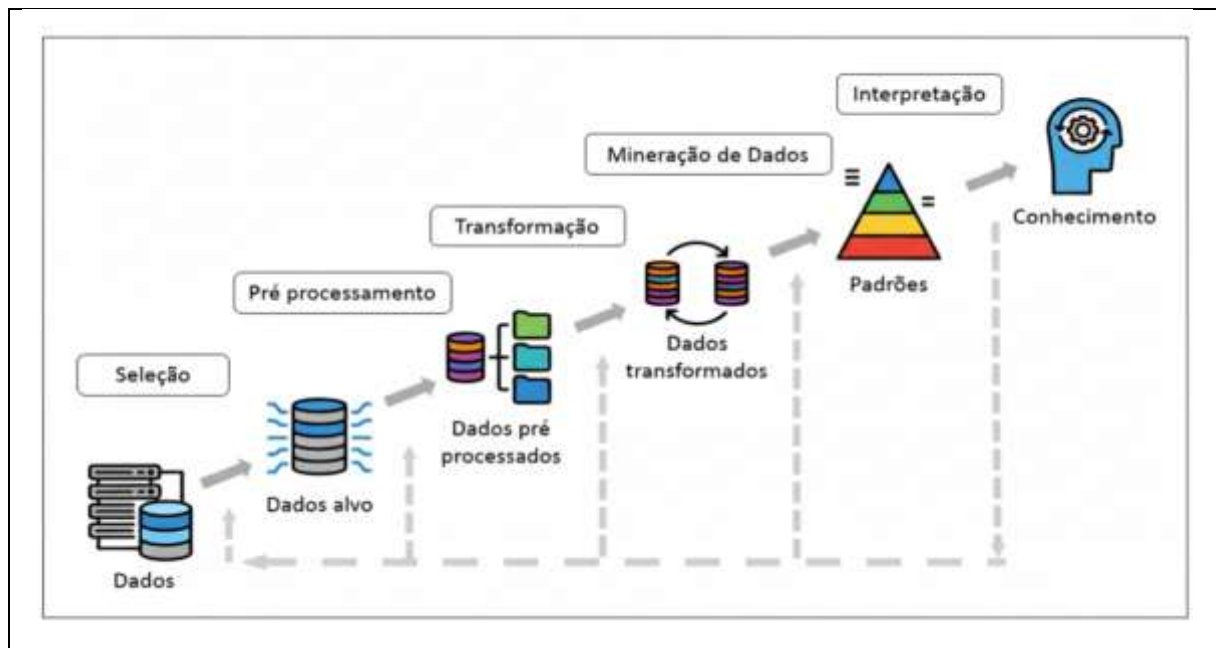
O avanço das tecnologias da informação possibilitou o armazenamento massivo de dados pelas organizações, tornando necessário o uso de técnicas capazes de transformá-los em conhecimento útil para a tomada de decisão. A mineração de dados surge nesse contexto como uma etapa fundamental dentro do processo de *Knowledge Discovery in Databases* (KDD), responsável por extrair padrões significativos a partir de grandes volumes de dados estruturados ou não estruturados. Segundo Ferrari, o processo de KDD envolve etapas sistemáticas, incluindo seleção, pré-processamento, transformação, mineração e interpretação dos resultados, com a mineração de dados desempenhando o papel central na identificação de padrões ocultos e relevantes aos gestores (Ferrari, 2020).

A importância estratégica desse processo está diretamente associada ao cenário competitivo atual, em que decisões baseadas apenas na intuição se tornam insuficientes. Sacilotti destaca que o domínio sobre a informação constitui fator essencial para que as empresas mantenham vantagem competitiva em ambientes dinâmicos e globalizados (Sacilotti, 2011).

Historicamente, o KDD ganhou notoriedade com a definição clássica de Fayyad, Piatetsky-Shapiro e Smyth, que o descrevem como um processo iterativo e interativo destinado à descoberta de padrões válidos, novos, potencialmente úteis e compreensíveis em bases de dados.

Conforme esses autores, o processo envolve a seleção, o pré-processamento e a transformação dos dados, seguidos pela mineração e interpretação dos resultados, de forma a converter informações armazenadas em conhecimento aplicável. As etapas propostas são apresentadas na Figura 1.

Figura 1 – Etapas do processo de KDD (descoberta de conhecimento em bases de dados)



Fonte: Fayyad, Piatetsky-Shapiro e Smyth (1996).

O processo inicia-se com a seleção dos dados que serão analisados, definindo o conjunto de informações relevantes ao problema. Em seguida, realiza-se a limpeza e o pré-processamento, eliminando inconsistências e dados irrelevantes, frequentemente chamados de ruídos. Na sequência, os dados são transformados para adequação ao método analítico, garantindo sua estruturação para a etapa de mineração. A fase de *Data Mining* consiste na escolha e aplicação do algoritmo adequado para extrair padrões e relações ocultas. Posteriormente, os resultados são interpretados à luz do contexto organizacional, o que pode exigir retorno a etapas anteriores para refinamento. Por fim, os conhecimentos descobertos são consolidados para apoiar a tomada de decisão ou ações estratégicas no ambiente empresarial (Fayyad; Piatetsky-Shapiro; Smyth, 1996).

2.2 REGRAS DE ASSOCIAÇÃO

As regras de associação constituem uma das tarefas mais tradicionais da mineração de dados, sendo utilizadas para identificar relações significativas entre itens em grandes bases transacionais. Essa técnica tem como objetivo encontrar padrões do tipo “se X ocorre, então Y tende a ocorrer”, auxiliando empresas na compreensão de comportamentos de consumo e

favorecendo a tomada de decisão orientada por dados. Segundo Han, Kamber e Pei (2012), regras de associação permitem descobrir correlações e dependências entre conjuntos de dados, sendo amplamente aplicadas em contextos de varejo por meio da análise da cesta de compras (*Market Basket Analysis*).

Uma regra de associação é representada na forma:

$$X \Rightarrow Y$$

em que X (antecedente) e Y (consequente) são conjuntos de itens (itemsets) distintos dentro de uma base transacional. Para avaliar a qualidade dessas regras, utilizam-se métricas como suporte, confiança e *lift*, que quantificam a relevância estatística da relação encontrada.

O suporte (*support*) mede a frequência com que um conjunto de itens aparece simultaneamente na base de dados, sendo expressa pela equação 1:

$$Support(X \Rightarrow Y) = \frac{N(X \cup Y)}{N} \quad (1)$$

onde $N(X \cup Y)$ representa o número de transações contendo ambos os itens e N o total de transações. Segundo Ferrari (2020), o suporte garante que apenas padrões suficientemente frequentes sejam considerados, evitando relações irrelevantes.

A confiança (*confidence*) expressa a probabilidade de Y (consequente) ocorrer dado que X (antecedente) ocorreu, sendo definida pela equação 2:

$$Confidence(X \Rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)} \quad (2)$$

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), a confiança representa a força da implicação estatística entre os itens, indicando quão consistente é a relação observada nos dados.

Já o *lift* avalia o grau de dependência entre X (antecedente) e Y (consequente), comparando a probabilidade conjunta com o produto de suas probabilidades individuais, sendo definida pela equação 3:

$$Lift(X \Rightarrow Y) = \frac{Confidence(X \Rightarrow Y)}{Support(Y)} \quad (3)$$

Ou alternativamente, pela equação 4:

$$Lift(X \Rightarrow Y) = \frac{Support(X \cup Y)}{Support(X) * Support(Y)} \quad (4)$$

Um valor de *lift* maior que 1 indica uma associação positiva, menor que 1 sugere correlação negativa e igual a 1 implica independência entre os itens. Conforme Gonçalves (2005), o *lift* é fundamental para distinguir relações verdadeiramente relevantes de correlações ocasionais.

Além dessas métricas, Ferrari (2020) destaca que a análise de regras de associação pode ser aprimorada por meio de indicadores complementares, como *leverage* e *conviction*. A alavancagem (*leverage*) mede o quanto a ocorrência conjunta de dois itens difere do que seria esperado caso fossem independentes, indicando o ganho absoluto de frequência proporcionado pela associação. Já a convicção (*conviction*) avalia o grau de dependência direcional entre o antecedente e o consequente, considerando a frequência com que a regra falha, sendo útil para analisar a consistência da implicação entre os itens e permitindo interpretações mais robustas conforme o contexto de aplicação.

A utilização de regras de associação é especialmente relevante no varejo, onde o grande volume de transações torna inviável a análise manual de padrões. Kerckhoff (2019) afirma que a técnica possibilita identificar perfis de compra, revelar produtos complementares e apoiar decisões estratégicas relacionadas à exposição de produtos, precificação e promoções.

Assim, observa-se que as regras de associação desempenham papel fundamental na extração de padrões úteis e acionáveis em ambientes organizacionais ricos em dados, contribuindo para estratégias mercadológicas e operacionais mais eficientes.

2.3 ALGORITMO APRIORI

O algoritmo Apriori é um dos métodos mais consolidados para a extração de regras de associação em bases de dados transacionais. Desenvolvido por Agrawal e Srikant em 1994, seu objetivo é identificar conjuntos de itens frequentes (*frequent itemsets*) que possam originar regras relevantes para análise de padrões de consumo. Segundo Han, Kamber e Pei (2012), o Apriori tornou-se referência devido à sua eficiência ao reduzir o espaço de busca por combinações de itens, utilizando propriedades matemáticas que evitam processamento desnecessário.

Uma das bases fundamentais do Apriori é o princípio das propriedades anti-monotônicas, também conhecido como propriedade Apriori. Essa propriedade estabelece que, se um conjunto de itens é frequente, então todos os seus subconjuntos também devem ser frequentes. Da mesma forma, caso um subconjunto seja infrequente, nenhum superconjunto que o contenha poderá ser considerado frequente. Essa lógica permite eliminar grandes quantidades de combinações inviáveis logo no início do processo, reduzindo o custo computacional da mineração de dados (Han; Kamber; Pei, 2012).

Diversos algoritmos podem ser aplicados para a geração de regras de associação, entre eles Apriori e FP-Growth. O estudo de Pancho e colaboradores ressalta que o FP-Growth é frequentemente escolhido pela sua maior eficiência computacional, “uma vez que evita a etapa de geração de candidatos presente no algoritmo Apriori”. No entanto, o Apriori permanece amplamente utilizado pela sua simplicidade de implementação e clareza interpretativa, sendo adequado para ambientes acadêmicos e bases de pequeno a médio porte, como a utilizada neste trabalho.

A execução do Apriori ocorre por meio de um processo iterativo denominado busca por níveis (*level-wise search*). Inicialmente, o algoritmo identifica todos os itens individuais que atendem ao suporte mínimo definido, formando o conjunto de candidatos C_1 . A partir desses itens, gera-se o conjunto de combinações de dois itens C_2 , calcula-se o suporte de cada combinação e elimina-se aqueles que não atingem o limiar desejado, formando o conjunto frequente L_2 . Esse procedimento é repetido sucessivamente para k-itens, conforme o esquema:

$$C_k = L_{k-1} \bowtie L_{k-1}$$

$$L_k = \{X \in C_k \mid \text{Support}(X) \geq \text{minsup}\}$$

Ferrari (2020) destaca que esse processo sistemático envolve três ações principais a cada iteração: geração de candidatos, contagem de suporte e eliminação de conjuntos infrequentes. A partir dos conjuntos válidos, o algoritmo expande progressivamente as combinações até que nenhuma nova frequência relevante seja encontrada.

Kerckhoff (2019) observa que a lógica do Apriori minimiza operações redundantes ao evitar o cálculo de suportes para combinações com baixa probabilidade de ocorrência, tornando o método adequado para bases de transações de grande escala, como supermercados. Contudo, a autora ressalta que, embora eficiente, o método pode apresentar limitações quando a

quantidade de itens é muito elevada, levando à necessidade de algoritmos alternativos, como o FP-Growth.

Dessa forma, o Apriori destaca-se como uma abordagem estruturada e fundamentada matematicamente para a descoberta de padrões frequentes, sendo amplamente reconhecido na literatura e aplicado em diversos contextos analíticos, especialmente no varejo e no comércio eletrônico.

2.4 APLICAÇÃO NO VAREJO

A aplicação de técnicas de mineração de dados no varejo tornou-se um elemento essencial para compreender padrões de consumo e otimizar decisões relacionadas ao sortimento, promoções e reposição de produtos. A análise de regras de associação é amplamente utilizada em ambientes com elevado volume de transações, permitindo identificar combinações recorrentes de produtos adquiridos simultaneamente. Segundo Han, Kamber e Pei (2012), a análise de cestas de compra (*Market Basket Analysis*) constitui uma das aplicações mais tradicionais da mineração de dados, auxiliando organizações a identificar relações entre itens e formular estratégias comerciais assertivas.

No contexto de supermercados, essa técnica possibilita a identificação de produtos complementares, favorecendo estratégias como promoções conjuntas e organização eficiente do layout das prateleiras, visando aumentar o ticket médio e melhorar a experiência de compra. De acordo com Gonçalves (2005), a compreensão das combinações mais frequentes permite que varejistas posicionem itens correlacionados próximos uns dos outros, incentivando compras adicionais e promovendo ganhos de desempenho operacional.

A literatura demonstra que a Análise de Cesta de Compras tem papel estratégico na gestão varejista ao permitir identificar relações de coocorrência entre produtos e orientar decisões gerenciais. Como destacado por Patwary et al. (2020), a aplicação do MBA “auxilia os gestores de supermercados a determinar a relação entre os itens adquiridos pelos clientes”. Essa capacidade de revelar padrões ocultos contribui diretamente para melhorias no layout, promoções direcionadas e aumento de rentabilidade, justificando a adoção desse tipo de análise no presente estudo.

Além dos supermercados, o método também encontra aplicação em farmácias, onde padrões de compra podem revelar relações entre medicamentos, vitaminas, produtos de higiene e itens de conveniência. Ferrari (2020) destaca que regras de associação contribuem para

práticas estratégicas no varejo farmacêutico, incluindo o planejamento de mix de produtos e o desenvolvimento de campanhas promocionais segmentadas.

No comércio eletrônico, as regras de associação são amplamente empregadas na construção de sistemas de recomendação. Conforme Kerckhoff (2019), a identificação de itens frequentemente comprados em conjunto permite sugerir produtos complementares ao consumidor, como ocorre em plataformas digitais que utilizam mensagens do tipo “clientes que compraram este produto também compraram...”. Esse processo contribui para o aumento das taxas de conversão e para a fidelização dos clientes, ao ofertar experiências personalizadas ao usuário.

Além de apoiar ações promocionais, a análise de regras de associação auxilia também o planejamento de sortimento e a gestão de estoque. O relatório técnico sobre varejo enviado (Semaan, 2006) reforça que padrões de compra podem ser utilizados para ajustar níveis de estoque, prever demanda e reduzir perdas com produtos de baixa rotatividade, garantindo maior eficiência operacional.

Desse modo, observa-se que o uso de regras de associação no varejo é capaz de gerar inteligência comercial aplicada, beneficiando tanto operações físicas quanto digitais ao promover decisões fundamentadas em dados, melhoria da rentabilidade e maior alinhamento entre oferta e comportamento do consumidor.

3 METODOLOGIA

3.1 TIPO DE PESQUISA

A presente pesquisa pode ser caracterizada sob quatro dimensões metodológicas: abordagem, natureza, objetivos e procedimentos técnicos.

Quanto à abordagem, o estudo é classificado como quantitativo, uma vez que se fundamenta na análise de dados numéricos provenientes de transações comerciais, utilizando métricas estatísticas para identificar padrões de associação entre itens comprados em conjunto. Conforme Gil (2019), pesquisas quantitativas envolvem a coleta e o tratamento de dados numéricos, com o objetivo de verificar relações e padrões mensuráveis. Tal perspectiva se aplica ao presente estudo, pois as conclusões são derivadas da aplicação de algoritmos matemáticos, métricas como suporte, confiança e *lift*, e análise estatística dos resultados.

Em relação à natureza, trata-se de uma pesquisa aplicada, pois visa gerar conhecimento direcionado à solução de um problema prático do ambiente organizacional: a identificação de padrões de compra que possam aprimorar decisões gerenciais, tais como organização de layout, definição de promoções e planejamento de estoque. De acordo com Prodanov e Freitas (2013), pesquisas aplicadas buscam produzir conhecimento voltado ao uso imediato para melhoria de práticas ou processos, alinhando-se aos objetivos deste trabalho.

No que se refere aos objetivos, a pesquisa possui caráter exploratório e descritivo. É exploratória por buscar maior familiaridade com o fenômeno analisado que é a estrutura de compra dos clientes na qual acaba permitindo o aprofundamento sobre relações entre produtos. É também descritiva por analisar e apresentar, de forma sistemática, os padrões encontrados nas transações, sem interferir no ambiente estudado. Gil (2019) destaca que estudos exploratórios são utilizados quando há pouco conhecimento prévio disponível sobre o fenômeno, enquanto pesquisas descritivas objetivam observar, registrar e analisar dados sem manipulá-los, características presentes nesta investigação.

Por fim, quanto aos procedimentos técnicos, este trabalho se caracteriza como um estudo de caso, apoiado por pesquisa documental e pesquisa bibliográfica. Trata-se de um estudo de caso porque analisa detalhadamente dados reais de uma única rede varejista, conforme a definição de Yin (2015) para investigações aprofundadas em contextos específicos. Configura-se como pesquisa documental por utilizar dados operacionais da empresa, de acordo com os critérios apresentados por Gil (2019), e como pesquisa bibliográfica ao fundamentar o

arcabouço teórico em livros e artigos científicos relacionados à mineração de dados, regras de associação e ao algoritmo Apriori.

3.2 DADOS

A base de dados utilizada neste estudo é composta por registros transacionais de vendas provenientes do sistema de emissão de notas fiscais de uma rede supermercadista. Cada registro representa uma compra realizada por um consumidor, contendo dois elementos essenciais para a aplicação de regras de associação: o identificador da transação (número da compra) e o código do produto adquirido. Essas variáveis constituem o requisito mínimo para a aplicação da análise de cestas de compras, pois permitem identificar conjuntos de produtos adquiridos simultaneamente em uma mesma compra.

Os dados analisados abrangem 12.190 transações, com um número variado por compra, contemplando diversos produtos comercializados pela rede. As informações foram obtidas diretamente do sistema operacional da empresa, garantindo sua fidedignidade para fins analíticos. Entretanto, devido a questões de confidencialidade e ética corporativa, não são divulgadas informações sensíveis como o período exato das transações, localização das unidades e identificação de clientes. Todos os registros utilizados encontram-se anonimizados, preservando a privacidade da empresa e dos consumidores.

Assim, a estrutura da base de dados atende aos requisitos necessários para a aplicação do algoritmo Apriori, viabilizando a identificação de padrões recorrentes de compra e contribuindo para a compreensão do comportamento de consumo no ambiente varejista.

Devido a restrições internas da empresa e a políticas de confidencialidade, a base de dados utilizada neste estudo não pode ser disponibilizada publicamente. Todas as informações analisadas foram cedidas exclusivamente para fins acadêmicos, mediante compromisso de uso restrito. Da mesma forma, o código desenvolvido para o processamento e análise dos dados constitui um artefato de natureza particular, elaborado especificamente para este trabalho, não sendo disponibilizado integralmente por se tratar de uma ferramenta de caráter pessoal e de uso limitado. Ainda assim, todos os procedimentos metodológicos empregados bem como os critérios adotados para a extração e interpretação dos resultados estão descritos de forma detalhada neste capítulo, garantindo a transparência e a reprodutibilidade conceitual da pesquisa.

3.3 PROCEDIMENTOS

A condução da análise seguiu um conjunto de etapas estruturadas, alinhadas às boas práticas de mineração de dados e especificamente ao processo de descoberta de padrões associado ao algoritmo Apriori. Inicialmente, realizou-se a preparação das transações, etapa na qual os registros de venda foram organizados de modo que cada nota fiscal representasse uma transação única, contendo a lista de produtos adquiridos conjuntamente. Esse agrupamento permite identificar conjuntos frequentes de itens e caracteriza o primeiro passo para a aplicação da técnica de análise de cesta de compras.

Em seguida, procedeu-se com a codificação binária (*one-hot encoding*), transformando a estrutura de listas de produtos em uma matriz binária transação-por-item. Nessa matriz, o valor 1 indica que determinado item esteve presente na transação, enquanto 0 indica sua ausência. Essa representação é necessária para a aplicação do algoritmo Apriori, conforme preconizado por Han, Kamber e Pei (2012), uma vez que possibilita a mensuração da frequência de ocorrência dos itens e combinações.

Após a construção da matriz, foi realizada a extração dos *itemsets* (conjunto de itens) frequentes, aplicando-se o algoritmo Apriori com um valor mínimo de suporte $\alpha = 0,02$. Essa etapa permitiu identificar os conjuntos de produtos que ocorrem simultaneamente em um número significativo de transações, filtrando combinações possivelmente irrelevantes ou aleatórias. Na sequência, efetuou-se a geração das regras de associação, calculando-se métricas de confiança e *lift* a partir dos *itemsets* identificados, possibilitando avaliar a probabilidade condicional de compra conjunta e a força das associações.

Com as regras identificadas, procedeu-se à filtragem mediante valores mínimos de confiança $\beta \geq 0,1$ e *lift* $\gamma \geq 1,2$; garantindo que apenas relações estatisticamente relevantes fossem consideradas na análise. Posteriormente, foi realizada a ordenação das regras, priorizando-as segundo critérios de interesse analítico, com destaque para a confiança, a fim de privilegiar associações com maior probabilidade de ocorrência.

A definição dos valores mínimos de suporte, confiança e *lift* adotados neste estudo baseou-se em referências consolidadas na literatura sobre regras de associação aplicadas ao varejo. Trabalhos anteriores indicam que limiares muito elevados tendem a eliminar padrões relevantes, enquanto valores excessivamente baixos podem gerar um volume excessivo de regras pouco significativas ou espúrias.

Por fim, realizou-se a interpretação dos resultados, examinando-se as regras mais expressivas e discutindo sua utilidade prática no contexto varejista. Essa etapa envolveu a

identificação de padrões potencialmente úteis para a organização de layout de gôndolas, definição de promoções combinadas e gerenciamento de estoques, permitindo compreender como os achados podem subsidiar decisões organizacionais fundamentadas no comportamento real dos consumidores.

3.4 PSEUDOCÓDIGO DO APRIORI

Figura 2 – Pseudocódigo do algoritmo Apriori

Algorithm 1 Mineração de Regras de Associação com o Algoritmo Apriori

Entrada: Planilha contendo as colunas *Nº da Compra* (identificador da transação) e *Código do Produto* (identificador do item); parâmetros α (suporte mínimo), β (confiança mínima) e γ (lift mínimo).

Saída: Tabela de regras ($A \rightarrow B$) com *suporte*, *confiança* e *lift* após filtragem.

Definições iniciais

- 1: $D \leftarrow \text{CARREGARPLANILHA}(\text{caminho})$ // Conjunto original de dados
- 2: $\mathcal{T} \leftarrow$ transações agrupadas por *Nº da Compra*, cada $\tau \in \mathcal{T}$ é um conjunto de produtos
- 3: $M \in \{0, 1\}^{|\mathcal{T}| \times |\mathcal{I}|}$ // Matriz binária transação \times produto

Etapa 1 — Preparação das transações

- 4: $D \leftarrow \text{SELECIONARCOLUNAS}(D, \{\text{Nº da Compra}, \text{Código do Produto}\})$
- 5: $\mathcal{T} \leftarrow \text{AGRUPARPOR}(D, \text{Nº da Compra}) \Rightarrow$ lista de conjuntos de produtos

Etapa 2 — Codificação binária (One-Hot Encoding)

- 6: $M \leftarrow \text{CODIFICARBINÁRIO}(\mathcal{T})$ // $M_{ij} = 1$ se o produto j aparece na transação i , caso contrário 0

Etapa 3 — Identificação dos conjuntos frequentes (Apriori)

- 7: $\mathcal{F} \leftarrow \text{APRIORI}(M, \text{suporte_mnimo} = \alpha)$
- 8: $\mathcal{F} \leftarrow \text{ORDENARPOR}(\mathcal{F}, \text{suporte}, \text{decrecente})$

Etapa 4 — Geração das regras de associação

- 9: $\mathcal{R} \leftarrow \emptyset$
- 10: **for all** $X \in \mathcal{F}$ com $|X| \geq 2$ **do**
- 11: **for all** partições não vazias $A \cup B = X$ com $A \cap B = \emptyset$ **do**
- 12: $\text{suporte} \leftarrow \frac{\#\{\tau \in \mathcal{T} : A \cup B \subseteq \tau\}}{|\mathcal{T}|}$
- 13: $\text{confiana} \leftarrow \frac{\text{suporte}(A \cup B)}{\text{suporte}(A)}$
- 14: $\text{lift} \leftarrow \frac{\text{confiana}}{\text{suporte}(B)}$
- 15: $\mathcal{R} \leftarrow \mathcal{R} \cup \{(A \rightarrow B, \text{suporte}, \text{confiana}, \text{lift})\}$
- 16: **end for**
- 17: **end for**

Etapa 5 — Filtragem e ordenação das regras

- 18: $\mathcal{R} \leftarrow \{r \in \mathcal{R} \mid \text{confiana}(r) \geq \beta \wedge \text{lift}(r) \geq \gamma\}$
- 19: $\mathcal{R} \leftarrow \text{ORDENARPOR}(\mathcal{R}, \text{confiana}, \text{decrecente})$
- 20: **Retornar** \mathcal{R}

A Figura 2 apresenta o pseudocódigo do algoritmo Apriori utilizado neste trabalho, descrevendo de maneira estruturada o processo operacional empregado para a extração de conjuntos frequentes. Sua inclusão tem o objetivo de explicitar as etapas lógicas fundamentais do método, facilitando a compreensão do fluxo de processamento aplicado aos dados. O pseudocódigo evidencia como o algoritmo realiza, de forma iterativa, a geração de candidatos, o cálculo do suporte, a eliminação de combinações infrequentes e a expansão dos conjuntos até que não haja mais *itemsets* que atendam ao suporte mínimo estabelecido. Dessa forma, a figura sintetiza, em nível conceitual, o procedimento utilizado na etapa de mineração de dados, servindo como ponte entre a teoria apresentada no referencial teórico e a implementação prática realizada em Python.

4 RESULTADOS

Este capítulo apresenta e discute os resultados obtidos a partir da aplicação das técnicas de mineração de dados propostas neste estudo. Inicialmente, são descritas as características gerais da base analisada, de modo a contextualizar o volume e a estrutura das transações avaliadas. Em seguida, são apresentados os principais produtos e *itemsets* frequentes identificados, bem como as regras de associação extraídas pelo algoritmo Apriori, analisadas a partir das métricas de suporte, confiança e *lift*. Posteriormente, os resultados são explorados por meio de análises gráficas, que auxiliam na visualização das relações mais relevantes entre os itens. Por fim, os padrões identificados são interpretados à luz do contexto varejista, discutindo-se suas implicações práticas para a gestão, o planejamento operacional e a tomada de decisão estratégica.

4.1 CARACTERIZAÇÃO DA BASE DE DADOS ANALISADA

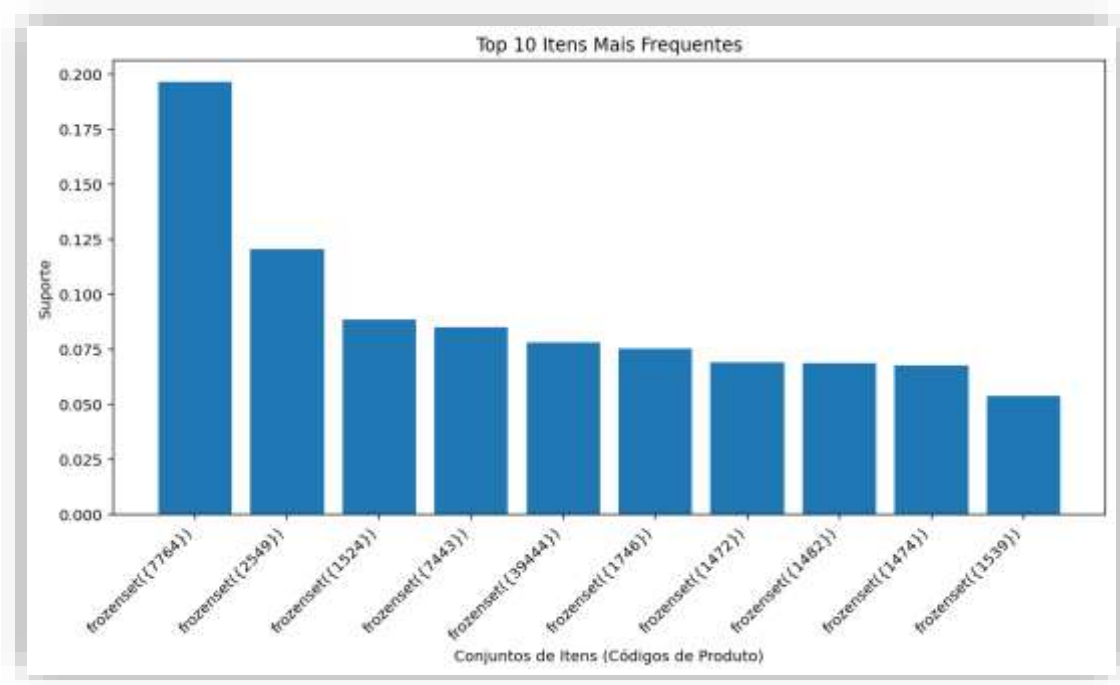
A base de dados utilizada neste estudo contém 12.190 transações, cada uma representando uma compra individual realizada em supermercados analisados. No total, foram identificados 7.884 produtos distintos, o que evidencia uma ampla diversidade de itens comercializados. A quantidade média de produtos por compra foi de 8 itens, indicando um comportamento de consumo caracterizado, em grande parte, por compras de reposição de curto prazo.

A variabilidade entre as compras também se mostrou significativa, enquanto algumas notas fiscais apresentaram apenas um produto, outras ultrapassaram trinta itens. Essa heterogeneidade demonstra a presença de diferentes perfis de clientes, que variam desde compras rápidas até compras completas de abastecimento semanal. Esse comportamento diversificado favorece a aplicação do algoritmo Apriori, pois tende a gerar uma matriz de transações rica, com combinações variadas de itens.

Além disso, a estrutura da base composta por itens discretos, categorizáveis e com diferentes frequências de ocorrência confirma sua adequação para o processo de mineração de dados. A presença de produtos com alta e baixa frequência de compra, somada à amplitude de combinações possíveis, cria condições ideais para a identificação de padrões de coocorrência, característica essencial para a análise de regras de associação.

4.2 ITENS E CONJUNTO DE ITENS MAIS FREQUENTES

Figura 3 – Gráfico de Itens mais frequentes



Fonte: O autor (2025)

A figura 3 apresenta os dez itens com maior suporte na base analisada. O suporte representa a proporção de transações em que um determinado produto aparece. Os itens mais frequentes foram:

Tabela 1 – Itens mais frequentes

Produto	%
Pão Francês Kg (7764)	19,64%
Queijo Muçarela Kg (2549)	12,03%
Cebola Kg (1524)	8,85%
Tomate Kg (7443)	8,49%
Sacola Atacarejo M (39444)	7,79%
Ovos Branco (1746)	7,53%
Banana Pacovan Kg (1472)	6,89%
Batata Inglesa (1482)	6,87%
Banana Prata Kg (1474)	6,78%
Cenoura Kg (1539)	5,36%

Fonte: O autor (2025)

Observa-se que o Pão Francês aparece em quase uma em cada cinco compras, destacando-se como produto-âncora da operação. Outro ponto relevante é a forte presença de hortifrúti entre os itens mais frequentes, o que indica que esses produtos possuem grande rotatividade e são essenciais no perfil de compra dos consumidores.

A aplicação do algoritmo Apriori resultou na identificação de 57 *itemsets* (Conjunto de itens) frequentes com suporte igual ou superior a 2%. Desse total:

- 38 correspondem a *itemsets* unitários;
- 16 representam combinações de dois itens;
- 3 são formados por três itens.

A existência de combinações frequentes reforça a hipótese de que determinados produtos são adquiridos em conjunto de forma recorrente, especialmente itens relacionados ao preparo de refeições.

4.3 REGRAS DE ASSOCIAÇÃO GERADAS

Após a aplicação dos critérios de filtragem confiança mínima de 10% e *lift* mínimo de 1,2 foram identificadas 50 regras de associação. As principais regras encontradas destacam relações fortes e estatisticamente relevantes entre os itens.

As regras de maior confiança também revelam padrões consistentes de compra conjunta. A relação entre Queijo Muçarela e Presunto apresenta confiança de 77,13%, indicando que, sempre que o cliente compra queijo, há grande probabilidade de adquirir presunto no mesmo momento, comportamento típico de produtos consumidos de forma complementar. Entre os itens de hortifrúti, observa-se que combinações como Cebola e Cenoura levam à compra de Batata Inglesa em 71,71% das transações, enquanto Tomate e Cenoura aparecem associados à Batata Inglesa em 71,46% dos casos. Também se destaca o fato de que a presença conjunta de Tomate e Batata Inglesa resulta na aquisição de Cebola em 70,87% das compras, e que Cebola e Batata Inglesa levam ao acréscimo de Tomate em 69,03% das transações. Esses valores evidenciam conjuntos de produtos que tendem a ser adquiridos em conjunto com elevada regularidade.

Tabela 2 – Relação de confiança

%	Relação de confiança
77,13%	Queijo muçarela → Presunto
71,71%	Cebola + Cenoura → Batata Inglesa
71,46%	Tomate + Cenoura → Batata Inglesa
70,87%	Tomate + Batata Inglesa → Cebola
69,03%	Cebola + Batata Inglesa → Tomate

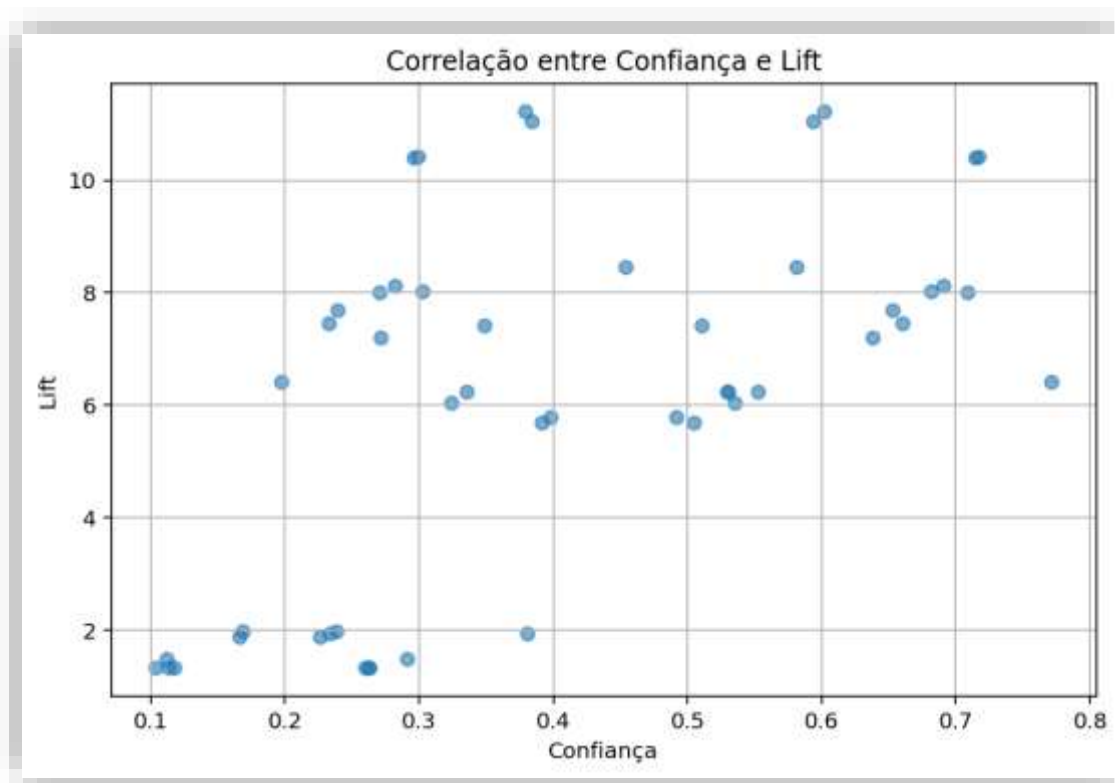
Fonte: O autor (2025)

As maiores forças de associação (*lifts*) foram observadas em combinações de grande intensidade entre itens de hortifrúti, com valores superiores a 10, um indicativo claro de interdependência de compra. Quando Batata Inglesa e Tomate aparecem juntos na mesma transação, a probabilidade de que Cenoura também seja adquirida aumenta em cerca de 11,21 vezes, evidenciando um padrão de compra bastante consistente entre esses três produtos. O comportamento inverso também é observado: a presença de Cenoura em uma compra eleva de forma equivalente, aproximadamente 11,21 vezes, a chance de que Batata Inglesa e Tomate apareçam em conjunto. Além disso, Cenoura também se destaca como antecedente relevante em outra combinação relevante, pois sua presença aumenta em torno de 11,06 vezes a probabilidade de aquisição simultânea de Batata Inglesa e Cebola. Esses valores mostram que tais itens tendem a compor cestas semelhantes, sinalizando relações de complementaridade típicas de produtos utilizados no preparo de refeições do dia a dia.

De forma geral, os padrões revelados pela análise evidenciam a existência de estruturas de compra bem definidas, especialmente entre produtos do setor de hortifrúti. Observou-se a formação de um conjunto de itens altamente complementares como batata, cenoura, cebola e tomate que apresentam alta coocorrência e compõem as regras mais fortes do estudo.

Esses resultados indicam que os consumidores tendem a adquirir esses itens de forma conjunta, possivelmente relacionados ao preparo de refeições básicas. As métricas de suporte, confiança e lift mostram-se consistentes e reforçam a relevância dos padrões identificados.

Figura 4 – Gráfico de Correlação entre Confiança e Lift



Fonte: O autor (2025)

O gráfico de dispersão de correlação entre confiança e lift apresentado na figura 4 permite visualizar o comportamento das regras de associação geradas a partir da base de dados analisada. Observa-se que as regras se distribuem majoritariamente entre valores de confiança de aproximadamente 0,10 a 0,77 e valores de lift entre 1,2 e 11, indicando a existência de padrões de compra com diferentes intensidades e níveis de relevância.

Nota-se claramente, na região superior direita do gráfico, um agrupamento de regras que combinam valores elevados de ambos os indicadores. Esse agrupamento sinaliza a presença de associações fortes e consistentes, já que regras com alta confiança representam relações frequentes e previsíveis, enquanto valores elevados de lift evidenciam dependência estatística positiva entre os itens, ou seja, a compra conjunta ocorre muito acima do que seria esperado ao acaso. Nesse grupo encontram-se, por exemplo, combinações envolvendo batata inglesa (1482), cenoura (1539), tomate (7443) e cebola (1524), que se destacaram como alguns dos pares e trios com maior força associativa em toda a base.

Por outro lado, regras posicionadas em regiões inferiores do gráfico apresentam lift mais próximo de 1, sinalizando que, apesar de possuírem confiança razoável, não representam

associações realmente relevantes no contexto varejista. Essas regras tendem a refletir apenas a popularidade individual de determinados produtos, e não necessariamente um padrão de compra conjunta.

A análise do gráfico também evidencia que confiança e lift não crescem de forma proporcional: algumas regras possuem confiança relativamente alta, mas lift moderado, indicando consequentes que são comprados com frequência geral elevada; enquanto outras possuem lift muito alto, mas confiança intermediária, sugerindo associações de nicho fortes, porém menos recorrentes.

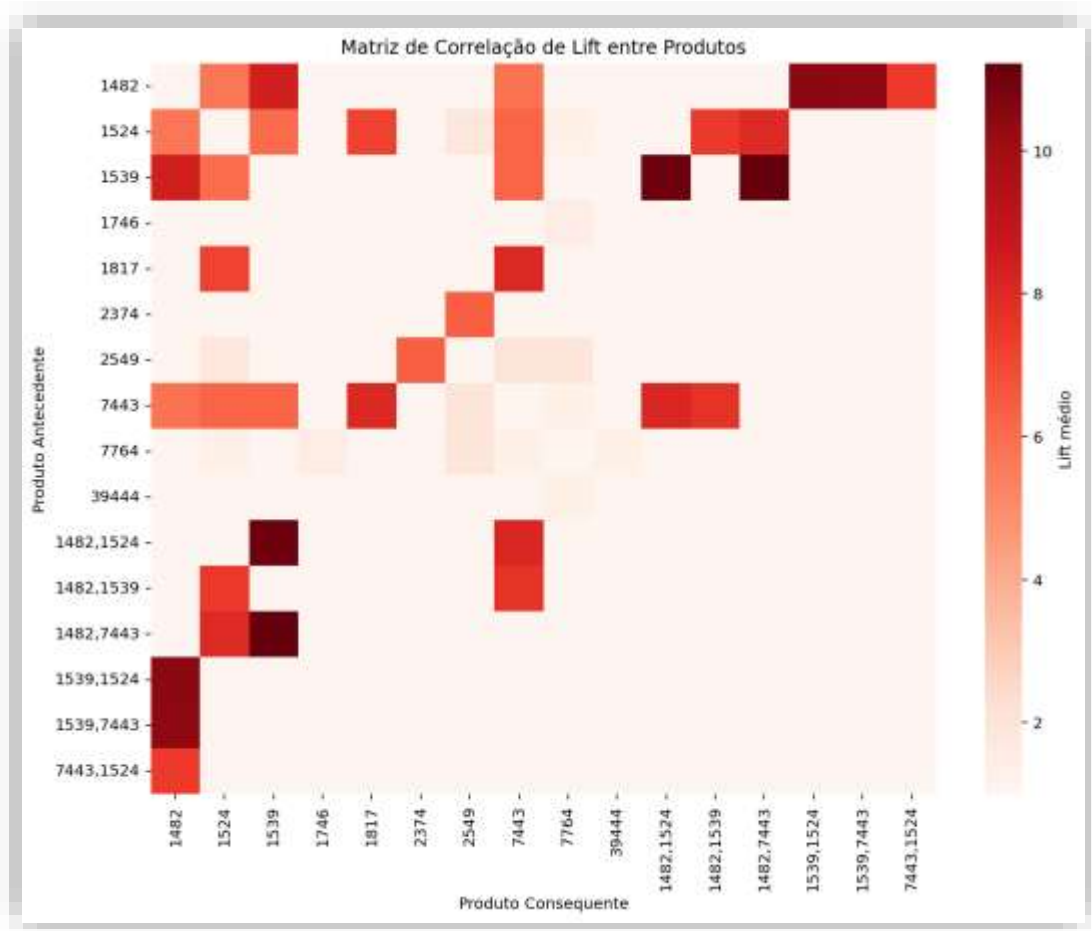
4.4 MAPA DE CALOR DO *LIFT*

A fim de complementar a análise das regras de associação e facilitar a identificação de padrões globais de relacionamento entre os produtos, elaborou-se uma matriz de *lift* representada por meio de um mapa de calor. Esse tipo de visualização permite observar não apenas regras individuais, mas grupos de itens que apresentam comportamentos semelhantes, destacando núcleos de produtos com forte coocorrência.

Cada célula da matriz indica o valor médio de *lift* entre um item antecedente e um consequente, de modo que cores mais escuras representam associações mais intensas, enquanto tons claros sugerem relacionamentos fracos ou inexistentes.

Essa abordagem permite identificar clusters naturais de compra e compreender como determinados produtos se conectam dentro das cestas dos consumidores. Assim, o *heatmap* funciona como uma ferramenta visual complementar à tabela de regras, facilitando a interpretação das dependências entre itens e revelando padrões.

Figura 5 – mapa de calor dos produtos antecedentes e consequentes de acordo com o lift médio.



Fonte: O autor (2025)

O mapa de calor acima representado pela figura 5, revela três aspectos principais:

- 1- Identificação de núcleos de compras: Existe um cluster de produtos fortemente associados, formado especialmente por:
 - Batata Inglesa (1482)
 - Cenoura (1539)
 - Cebola (1524)
 - Tomate (7443)

Esses itens aparecem repetidamente em regras de alta força e compõem o principal núcleo de consumo identificado.

- 2- Produtos âncora: Cenoura e Batata Inglesa se destacam por aparecerem com frequência como antecedentes nas regras, indicando que são itens que "puxam" outros produtos na cesta.
- 3- Produtos com baixa associação: alguns itens apresentam baixa interação com outros, sugerindo compras mais pontuais ou sem padrão definido. Esses itens aparecem em regiões claras do mapa de calor.

De modo geral, o mapa de calor reforça visualmente a existência de relações fortes entre determinados grupos de produtos, confirmando os resultados obtidos na análise das regras.

4.5 DISCUSSÃO E IMPLICAÇÕES GERENCIAIS

Os resultados obtidos por meio da análise de regras de associação forneceram um conjunto robusto de evidências capazes de orientar decisões operacionais e estratégicas no ambiente varejista. A identificação de padrões de compra permitiu compreender como os consumidores organizam suas escolhas no ponto de venda, revelando relações que não seriam perceptíveis apenas por análises tradicionais de volume ou faturamento. Com base nesses achados, diversas aplicações práticas podem ser delineadas.

A primeira delas refere-se à otimização do layout da loja. A análise mostrou a existência de um núcleo de produtos altamente coocorrentes, especialmente cenoura, batata inglesa, cebola e tomate, que apresentaram valores de *lift* superiores a 10 em diversas combinações. Esse comportamento sugere que tais itens são adquiridos de forma complementar e estão associados ao preparo de refeições básicas. A reorganização do espaço físico, aproximando esses produtos ou criando ilhas temáticas sazonais, pode reduzir o tempo de busca do consumidor, aumentar a conveniência e estimular compras adicionais por associação natural (*cross-merchandising*). Essa estratégia tende a elevar o ticket médio e melhorar a experiência de compra.

Outra aplicação relevante diz respeito à criação de promoções cruzadas. As regras de maior confiança revelaram combinações com elevada probabilidade condicional, como a relação entre batata inglesa, cenoura e cebola, cuja confiança ultrapassa 70%, e a complementaridade entre muçarela e presunto, com confiança superior a 77%. Essas relações permitem estruturar promoções direcionadas, como combos de hortifrútis, descontos condicionados à compra de itens complementares ou cupons personalizados. A oferta de

promoções ancoradas em padrões reais de comportamento tende a aumentar a percepção de valor do consumidor e estimular o consumo conjunto.

A formação de kits de produtos representa outra possibilidade derivada das associações identificadas. Kits como “Kit Sopa” (batata, cenoura, cebola, tomate), “Kit Café da Manhã” (pão francês, presunto e muçarela) ou “Kit Salada” (cenoura, cebola e tomate) podem ser estruturados tanto em exposição física quanto em canais digitais. Além de facilitar a decisão de compra, esses agrupamentos aumentam o volume adquirido por transação e fortalecem o posicionamento comercial da loja.

Os resultados também oferecem subsídios para o planejamento avançado de estoque. Como os itens fortemente associados apresentam correlação direta de demanda, uma ruptura em um produto pode reduzir a venda dos seus complementares. O conjunto formado por cenoura, batata inglesa e cebola é um exemplo claro dessa dinâmica, uma vez que esses itens compõem cestas de compra recorrentes. Uma falha de abastecimento em cenoura, por exemplo, pode gerar perda indireta de demanda nos demais itens, visto que o consumidor tende a adquiri-los de forma conjunta. Dessa forma, a sincronização de estoques e a reposição coordenada tornam-se essenciais para evitar prejuízos e maximizar o atendimento à demanda real.

A previsão de demanda também pode ser aprimorada a partir das regras de associação. Como o aumento nas vendas de um produto pode indicar elevação proporcional em seus complementares, modelos preditivos podem utilizar a estrutura de lifts e probabilidades condicionais para projetar variações sazonais ou comportamentais. Por exemplo, um aumento de 10% nas vendas de tomate tende a impactar diretamente os volumes de batata e cenoura, dada a relação forte entre esses itens. Essa informação pode alimentar modelos estatísticos, algoritmos de previsão e planejamentos semanais de abastecimento.

Por fim, os resultados oferecem suporte às estratégias de precificação. Produtos âncora, como pão francês ou hortifrúti de alta rotatividade, podem ser utilizados com margens menores para atrair consumidores. O efeito compensatório ocorre nos itens complementares de alto lift, cujas vendas aumentam proporcionalmente quando o produto âncora é adquirido. Assim, a margem reduzida em um item pode ser recuperada pelo volume adicional gerado nos produtos associados. Essa lógica também pode ser aplicada em negociações com fornecedores, uma vez que itens de alta elasticidade de demanda possuem maior relevância estratégica.

A literatura também demonstra que a análise de cestas de compra tem sido utilizada como ferramenta fundamental para otimização comercial. Conforme Patwary et al. destacam, a técnica compõe um dos principais pilares analíticos do varejo ao auxiliar na “definição da disposição dos produtos e no desenvolvimento de promoções direcionadas a diferentes

segmentos de consumidores”. Essa perspectiva reforça a relevância prática da análise aplicada neste trabalho, uma vez que os padrões identificados podem subsidiar rearranjos de gôndolas e estratégias de *cross-selling*.

De forma geral, os achados deste estudo demonstram que a análise de regras de associação oferece suporte direto à gestão varejista ao revelar padrões ocultos que influenciam o comportamento de compra. A partir das evidências identificadas, é possível orientar ações de layout, estoque, precificação, promoção e abastecimento de maneira mais integrada, eficiente e alinhada às necessidades dos consumidores. Dessa forma, o uso do algoritmo Apriori se confirma como uma ferramenta acessível, prática e de elevado impacto para supermercados que buscam aprimorar sua tomada de decisão baseada em dados.

5 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo identificar padrões de compra em um ambiente varejista por meio da aplicação da metodologia de regras de associação, utilizando o algoritmo Apriori para analisar uma base real de notas fiscais. A partir do processamento e da mineração dos dados, foi possível extrair informações relevantes sobre o comportamento dos consumidores, evidenciando produtos com alta frequência de coocorrência e relações estatísticas consistentes entre diferentes itens.

Os resultados demonstraram que itens pertencentes ao segmento de hortifrúti, como batata inglesa, cenoura, tomate e cebola, formam um núcleo de compra recorrente, apresentando valores elevados de suporte, confiança e *lift*. Essas associações reforçam que os consumidores tendem a adquirir tais produtos de maneira conjunta, o que pode estar relacionado ao preparo de refeições cotidianas. Regras fortemente significativas, com *lifts* superiores a 10, evidenciaram a robustez das relações encontradas.

A análise gráfica e estatística permitiu ainda visualizar a intensidade e a distribuição dessas relações, consolidando evidências de que os padrões identificados não são aleatórios, mas refletem comportamentos reais e consistentes. Esses achados oferecem suporte direto à gestão varejista, indicando oportunidades para reorganização de layout, criação de promoções cruzadas, planejamento de kits, tomada de decisão comercial e estratégias de reposição e gestão de estoque.

Dessa forma, conclui-se que o uso de técnicas de mineração de dados, aliado a ferramentas de análise estatística, mostra-se altamente eficaz para transformar grandes volumes de dados transacionais em conhecimento aplicável. A aplicação do Apriori se revelou adequada ao contexto estudado, permitindo extrair insights que podem contribuir para a melhoria da eficiência operacional, aumento do ticket médio e aprimoramento da experiência de compra.

O estudo reforça a importância crescente da tomada de decisão baseada em dados no varejo e evidencia o potencial de métodos computacionais acessíveis para apoiar estratégias comerciais. Assim, a pesquisa cumpre seu propósito ao demonstrar, de forma prática, como informações ocultas em bases transacionais podem ser identificadas e utilizadas como vantagem competitiva.

5.1 SUGESTÕES PARA TRABALHOS FUTUROS

Embora os resultados obtidos tenham sido satisfatórios, algumas oportunidades de aprofundamento e ampliação podem ser consideradas em investigações futuras. Uma primeira possibilidade consiste na incorporação de variáveis adicionais, como categorias mercadológicas, sazonalidade, localização das lojas ou dados socioeconômicos agregados. A expansão das variáveis permitiria análises segmentadas e comparativas, contribuindo para uma visão mais abrangente dos padrões de consumo.

Outra sugestão reside na aplicação e comparação de diferentes algoritmos de mineração de dados, como FP-Growth ou Eclat, que podem apresentar maior eficiência em bases volumosas. Essa comparação pode revelar melhorias de desempenho e qualidade das regras extraídas. Além disso, sugere-se integrar modelos preditivos ou técnicas de aprendizado de máquina, de modo a permitir a previsão de demanda conjunta a partir das associações identificadas. Essa integração pode ampliar o valor estratégico das análises, especialmente para o planejamento de estoque. Recomenda-se ainda que pesquisas futuras investiguem o impacto real das ações sugeridas neste estudo. A implementação experimental de promoções cruzadas, reorganizações de layout ou kits temáticos poderia ser avaliada por meio de testes controlados, medindo seus efeitos no ticket médio, conversão e comportamento do consumidor.

Por fim, estudos que combinem dados de compras físicas e digitais configuram-se como um campo promissor, permitindo analisar o comportamento *omnichannel* e identificar padrões híbridos de consumo. Assim, os caminhos sugeridos abrem espaço para novas investigações que possam ampliar o alcance e a aplicabilidade dos resultados, fortalecendo a adoção de abordagens analíticas avançadas no setor varejista.

REFERÊNCIAS

- BARNEY, J. B.; HESTERLY, W. S. **Administração estratégica e vantagem competitiva**. 3. ed. São Paulo: Pearson, 2011.
- CASTRO, L. N. de; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva, 2016.
- FAYYAD, U.; PIATETSKY S. G.; SMYTH P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2., 1996, Portland. **Anais eletrônicos...** Portland: Association for the Advancement of Artificial Intelligence, 1996. Disponível em: . Acesso em: 21 out. 2025.
- FREITAS, H. **Processamento de dados e sistemas de informação**. São Paulo: Atlas, 1993.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 7. ed. São Paulo: Atlas, 2019.
- GONÇALVES, E. C.. Regras de associação e suas medidas de interesse objetivas e subjetivas. **INFOCOMP: Journal of Computer Science**. Rio de Janeiro, v. 4, n. 1., p. 27-36, 2005. Disponível em: <<http://www.dcc.ufla.br/infocomp/index.php/INFOCOMP/article/view/79>>. Acesso em: 22 out. 2025.
- GONÇALVES, L. P. F. **Mineração de Dados em Supermercados: O Caso do Suoermercado "Tal"**. 1999. 36 f. Dissertação (Mestrado em Administração) - Programa de Pós-Graduação em Administração, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 1999. Disponível em: <http://www.ufrgs.br/gianti/files/orientacao/mestrado/proposta/pdf/24_mest_proposta_goncalves.pdf>. Acesso em: 22 out. 2025.
- HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3. ed. San Francisco: Morgan Kaufmann, 2012.
- KURNIAWAN, F.; Umayah, B.; HAMMAD, J.; NUGROHO, S. M. S.; HARIADI, M. **Market Basket Analysis to Identify Customer Behaviours by Way of Transaction Data**. Knowledge Engineering and Data Science, v. 1, n. 1, p. 20-25, 2017.
- MAINALI; Sanjeev. **Market Basket Analysis**. 2016. 45 f. Trabalho de Conclusão de Curso (Bacharel em Ciência da Computação e Tecnologia da Informação) - Department of Computer Science and Information Technology, Deerwalk Institute of Technology, Tribhuvan University, Nepal, 2016.
- MARTINEZ, M.; ESCOBAR, B.; GARCÍA D., María E.; PINTO R., Diego P. **Market basket analysis with association rules in the retail sector using Orange: Case Study: Appliances Sales Company**. CLEI Electronic Journal, v. 24, n. 2, 2021. DOI: 10.19153/cleiej.24.2.12.

PATWARY, A. H.; ESHAN, T.; DEBNATH, P.; SATTAR, A. **Market Basket Analysis Approach to Machine Learning**. In: *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021. DOI: 10.1109/ICCCNT51525.2021.9580175.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013.

SACILOTTI, A. C. **A Importância da Tecnologia da Informação nas Micro e Pequenas Empresas: Um Estudo Exploratório na Região de Jundiaí**. 2011. 116 f. Dissertação (Mestrado em Administração) - Programa de Mestrado em Administração, Faculdade Campo Limpo Paulista (FACCAMP), São Paulo, 2011. Disponível em: <http://www.faccamp.br/new/arq/pdf/mestrado/Documentos/producao_discente/2011/04abril/AdaniCusinSacilotti/dissertaCAo.pdf> Acesso em: 22 out. 2025.

SEMAAN, G. S.; GRAÇA, A. D. A.; DIAS, C. R.. Extração De Associações Em Bases De Dados De Varejo. Simpósio Brasileiro de Pesquisa Operacional, 38., 2006, Goiânia. **Anais eletrônicos...** Juiz de Fora: Faculdade Metodista Granbery, 2006. p. 1312–1322. Disponível em: <<http://www.din.uem.br/sbpo/sbpo2006/pdf/arq0197.pdf>>. Acesso em: 22 out. 2025.

WIBOWO, A. H.; WINDYATRI, H.; MARULAN, A.; HARITS, D.; NUGROHA, D. **A pattern analysis on goods purchase relationship for minimarket's customers by using association rule – Market Basket Analysis Method (AR-MBA)**. *Journal of Information Technology Ampera*, v. 4, n. 3, p. 195-206, 2023.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 5. ed. Porto Alegre: Bookman, 2015.