A ABORDAGEM ANALÍTICA NO CONTEXTO DO VAREJO: UMA PROPOSTA TEÓRICO-METODOLÓGICA

RESUMO

A análise das cestas de compras tem sido vista como uma abordagem que contribui para melhor compreensão do comportamento do consumidor e das suas preferências de compras. Este ensaio teórico atuou na proposição de um modelo baseado na abordagem analítica de dados (do inglês, *business analytics*), com vistas a permitir em etapas empíricas de execução do modelo, a extração de segmentos de visitas de clientes por meio da análise das cestas de compras, no nível individual, bem como caracterizar a visita de compras do cliente, a fim de permitir a identificação da intenção ou missão de compra por trás da visita. Ainda, a partir da revisão de literatura empreendida, notou-se que os trabalhos que utilizaram a mineração de dados no contexto do varejo, não consideraram os efeitos da árvore de taxonomia de produto, os quais influenciaram nos resultados da modelagem. Por esse motivo, o ensaio discute em seu escopo uma abordagem para ajustar a taxonomia do produto. Assim, compreende-se que este estudo traz contribuições teóricas e gerenciais relevantes, dando suporte para a tomada de decisão em diversas áreas, tais como: campanhas de *marketing* por segmento de clientes, redesenho de *layout* de uma loja e, recomendações de produtos personalizados.

Palavras-chave: Abordagem Analítica. Segmentação de Clientes. Ensaio Teórico.

1 INTRODUÇÃO

No ambiente contemporâneo de negócios, os varejistas têm compreendido que a tomada de decisão orientada a dados atua como suporte para a aquisição de conhecimentos úteis, aumento da qualidade decisória e melhoria no desempenho organizacional. Observa-se que os gestores do varejo supermercadista coletam bases de dados extensas sobre clientes diariamente. No entanto, com a grande variedade e volume dos dados disponíveis, é praticamente impossível processar as informações manualmente e, um dos grandes desafios dos gestores está em explorar dados coletados não utilizados (GRIVA et al., 2018).

A partir da evolução do campo de mineração de dados, novas técnicas foram criadas com a finalidade de gerenciar grandes bases de dados e realizar análises profundas, que tem o potencial de melhorar a tomada de decisões nos negócios. Assim, organizações que efetuam análises com base em dados podem facilitar a tomada de decisões e melhorarem seus resultados (PROVOST; FAWCETT, 2013).

Nessa perspectiva, nota-se que as técnicas de mineração de dados aplicáveis no contexto do varejo desempenham papel fundamental na descoberta de conhecimentos ocultos na vasta quantidade de dados disponíveis, permitindo que os gestores adotem uma abordagem mais centrada nos clientes (NGAI; XIU; CHAU, 2009). Dessa forma, as empresas tornam-se capazes de oferecer produtos e serviços personalizados, os quais agregam valor aos clientes (BOONE; ROEHM, 2002).

É importante compreender que a evolução tecnológica tem possibilitado a disseminação de dados e informações úteis para a tomada de decisão gerencial. Nota-se, por exemplo, que com o crescimento atual do uso aplicativos, programas de fidelidade, programas de compras

pessoais, *scanners*, *cookies* e outras metodologias de coletas de dados, as características dos consumidores e os padrões ocultos do comportamento de compra podem ser capturados, o que contribui para a compreensão das intenções futuras de compra dos clientes.

Nessa vertente, este trabalho tem como norte refletir sobre a seguinte questão: **É possível identificar algum padrão de consumo nos *tickets* de compra de um supermercado, que contribua para a caracterização da missão de compra por visita dos clientes?**

Na atualidade, observa-se que varejistas têm buscado maneiras para proporcionar experiências de compras com maior rapidez, fluidez, facilidade, flexibilidade e eficiência. Diante da ampla concorrência, os gestores procuram alternativas para garantir vantagens competitivas e aperfeiçoar o desempenho. Além disso, os varejistas personalizam seus serviços com base na intenção de compras de seus clientes. Nessa perspectiva, o presente trabalho tem como objetivo identificar um modelo capaz de entender a missão de compra do cliente, no nível das lojas supermercadistas.

A temática revela-se importante, pois favorece uma compreensão mais abrangente acerca do alcance da satisfação e do valor agregado dos clientes com os serviços prestados, além de proporcionar ganhos a partir da eficiência e eficácia organizacional.

A literatura na área já vem desenvolvendo estudos para identificar padrões no comportamento de compras dos clientes com dados demográficos, de vendas ou de identificação por rádio frequência (RFID), esses trabalhos perpassam algumas vertentes, como por exemplo: 1) segmentação de clientes com dados referentes a todas as compras efetuadas por um comprador ou componentes demográficos, para caracterizar grupos de clientes; 2) segmentação por cesta de compras, para identificar relações entre itens e categorias de itens adquiridos (tais como: complementar e substituto) e; 3) segmentação por tipo de viagens de compra (exemplo: viagens rápidas ou grandes viagens). Contudo, a literatura ainda carece de maiores investigações que explorem o objetivo ou a intenção de compras de visita de um cliente, por exemplo, considerando a sua cesta de compras de maneira individual.

Para a realização do presente estudo, foi proposto um modelo teórico que adota técnicas de *cluster* para identificar os segmentos de visitas de clientes. Esse trabalho também considera o efeito da taxonomia do produto na validade dos resultados da análise de *cluster*, dado o seu impacto na geração e descoberta de conhecimentos e, na mineração de dados (ALBADVI; SHAHBAZI, 2009, CHO; KIM; KIM, 2002). Para lidar com problemas de assimetria de dados, o estudo propõe uma abordagem para ajustar a taxonomia, com base na frequência de compras dos produtos.

A principal contribuição do estudo consiste no fato de que ao invés de lidar com o comportamento geral de compra dos consumidores, o modelo proposto permite a visualização do comportamento de compra dos clientes em uma perspectiva alternativa, realizando a caracterização de visita de um cliente pelas categorias de produtos do carrinho de compras e pela identificação da intenção ou missão de compra, dada a visita. Assim, supõe-se que o setor de varejo possa atuar na perspectiva de atender às necessidades dos clientes por visita. Nesse sentido, compreende-se que estudos que contêm uma abordagem que trabalha com todas as vendas por cliente ou são voltados para apenas pares de produtos ou categorias de produtos, são potencialmente capazes de negligenciar conhecimentos acerca de cada visita de compras no nível individual, subestimando as necessidades dos clientes.

O conhecimento adquirido permite o desenvolvimento de uma nova forma de se fazer publicidade, considerando anúncios de categorias de produtos agrupados que representam uma missão de compras, como por exemplo, anúncios de produtos para a realização do café da

manhã. Assim, a abordagem de segmentos de visitas aos clientes pode ditar um novo rumo de *layout* na perspectiva gerencial, no qual categorias de produtos do mesmo segmento de visita se situem em corredores e prateleiras próximas.

Portanto, a partir desse estudo, pode-se gerar uma reflexão conceitual acerca da oportunidade de mudança na maneira de se organizar o *layout*, deixando de ser baseado em categorias de produtos para um pautado na missão de compras, o que favorece a localização de produtos de maneira mais fácil e rápida. Dessa maneira, compreende-se que os conhecimentos extraídos do modelo que está sendo proposto, supostamente serão capazes de dar suporte à tomada de decisão gerencial, seja em relação a campanhas de *marketing* por segmento de clientes, redesenho de *layout* de uma loja ou, recomendação de produtos (CIL, 2012; SARANTOPOULOS et al., 2016). Logo, vislumbrando as consequências positivas da execução do modelo, entende-se que os gestores poderão executar estratégias inovadoras para gerenciar as operações da loja, realizando pedidos de produtos com base nos segmentos de visita identificados.

As demais seções do trabalho foram organizadas da seguinte forma: após esta introdução (seção 1), na seção 2 é empreendida uma revisão de literatura com vistas a caracterizar o fenômeno estudado, assim é analisado o estado da arte referente à segmentação de clientes e a análise da cesta de mercado, apresentando os fluxos, características e atuais limitações da pesquisa na área. Em seguida, na seção 3, o método de pesquisa adotado é descrito, a fim de evidenciar o rigor metodológico empregado. Posteriormente, na seção 4, apresenta-se o modelo conceitual proposto, discutindo-o passo-a-passo e levantando as supostas contribuições para a literatura e para a prática gerencial no contexto supermercadista. Por fim, na seção 5, são tecidas as considerações finais, as limitações e possibilidades de futuras investigações.

2 REFERENCIAL TEÓRICO

2.1 Segmentação de Clientes

A segmentação de clientes tem se mostrado fundamental para lidar com uma grande variedade de clientes, que possuem preferências diversas, com maior eficiência. Por meio da segmentação de clientes, as empresas conseguem atender as necessidades do público-alvo e ter relacionamentos prolongados, o que possibilita o aumento das receitas com a aquisição e manutenção de clientes de alto valor e baixo custo, em um cenário competitivo (NAMIN; DEHDASHTI, 2019).

A segmentação de clientes visa identificar diferentes grupos de clientes que frequentam determinada loja, para compreender necessidades e características específicas de cada segmento, a fim de oferecer serviços personalizados de acordo com as preferências dos consumidores, proporcionando maior satisfação de clientes (BOONE; ROEHM, 2002, HUNG; TSAI, 2008).

É importante que os varejistas segmentem suas bases de clientes, de maneira que existam similaridades dentro e distinções entre cada segmento, para que sejam justificados os gastos com recursos como o tempo, por exemplo, e despesas, na atuação com m*ixes* de *marketing* separados. Acrescenta-se também que uma segmentação inadequada reduz a eficácia organizacional e desperdiça recursos de *marketing*. Na prática, a seleção ou aplicação inadequada de uma técnica de segmentação pode gerar grandes impactos financeiros, dado que

os custos associados a uma campanha de *marketing* direcionada não são triviais (BOONE; ROEHM, 2002).

Para efetuar os estudos de segmentação de clientes, os pesquisadores adotaram diversos tipos de dados, como por exemplo: a) demográficos, b) geográficos, c) psicográficos, d) atitudinais, e) de vendas e f) comportamentais (GRIVA et al., 2018). Além disso, algumas técnicas de mineração de dados foram comumente utilizadas por estudiosos, cita-se por exemplo, modelos de associação, classificação, *cluster*, previsão, regressão, descoberta de sequência e visualização (NGAI; XIU; CHAU, 2009).

Por meio da revisão de literatura, observou-se uma corrente de trabalhos que utiliza dados de vendas para segmentar compradores. Para essa finalidade, os pesquisadores examinaram: i) o comportamento de compras (p.ex.: volume de vendas, frequências de visitas, valor monetário e periodicidade da compra) ou ii) a combinação de produtos ou categorias de produtos adquiridos em todo o histórico de compra, ao longo das visitas em uma unidade de loja (GRIVA et al., 2018, AERON; KUMAR; MOORTHY, 2012, CHEN et al., 2009, CHENG; CHEN, 2009, KHAJVAND et al., 2011). A seguir, observa-se ambas as correntes de pesquisas dentro da comunidade científica e suas respectivas contribuições para a literatura.

2.2 Segmentação de Clientes com a Utilização de Dados de Vendas

Sob a perspectiva da realização da segmentação de clientes por dados de vendas, alguns pesquisadores concentraram seus esforços no estudo do comportamento de compras dos consumidores.

Kitts, Freed e Vrieze (2000) estudam o histórico de compras de clientes, com o objetivo de sugerir recomendações e realizar promoções em nível de item. Boone e Roehm (2002) utilizam dados de vendas, com a finalidade de sugerir o uso das redes neurais artificiais (RNAs) para realização de segmentação de clientes. Liao, Chen e Hsieh (2011) segmentaram clientes em grupos de acordo com hábitos de vida e comportamento de compras. Para tal, adotou-se uma abordagem com dados de clientes oriundos das vendas e coletados por questionários. Este trabalho contribuiu para o direcionamento de novos serviços e vendas para cada segmento de cliente. Liao e Chen (2004) empregaram dados do *mix* de categorias de produtos, demografia e dados atitudinais para segmentar clientes. Este estudo visou aumentar a eficácia do *marketing* direto e gerenciamento de vendas, assim como criar catálogos eletrônicos de vendas. Já Larson, Bradlow e Fader (2005) utilizaram dados do histórico de compras capturados por rádio frequência (RFID) e segmentaram os caminhos percorridos por compradores, examinando os comportamentos em uma viagem de compras.

Por outro lado, também existem estudos que se concentraram na compreensão da combinação de produtos ou categorias de produtos adquiridas durante o histórico de compra dos consumidores, ao longo das visitas em uma unidade de loja.

A fim de prever padrões de compra do cliente, Park, Park e Schweidel (2014) estudaram o *mix* de produtos ou categorias de produtos listados no histórico de compras. Miguéis, Camanho e Cunha (2012), por sua vez, realizaram a segmentação por meio do estilo de vida de um cliente, com o suporte de banco de dados secundários. No seu estudo, eles analisaram categorias de produtos, tendo por finalidade direcionar as políticas promocionais para segmentos específicos de clientes. Por conseguinte, Han e colaboradores (2014) estudaram o papel das categorias na segmentação de clientes, a partir de diferentes técnicas. Por meio do algoritmo de *cluster k-means*, identificou-se segmentos de clientes de rotina, sazonais ou de conveniência.

Pode-se notar que esses trabalhos trouxeram contribuições significativas para a literatura de segmentação de clientes por meio de dados de vendas, utilizando métodos como *clustering* e cadeias de *Markov*, por exemplo (CHEN et al., 2009, CHENG; CHEN, 2009, HAN et al., 2014, KITTS et al., 2000, LIAO et al., 2011). No entanto, observa-se que esses estudos possuem limitações, principalmente pelo fato de suas análises serem realizadas com base na totalidade dos dados de vendas, seja em relação ao volume de vendas, frequências das visitas, valor monetário, periodicidade da compra, mix ou categorias de produtos. Em outras palavras, compreende-se que esses estudos originaram segmentos de clientes que podem ser vistos como uma caracterização genérica dos consumidores, visto que eles podem não refletir as suas preferências pelo motivo de não considerarem as particularidades das visitas de compras no nível individual (GRIVA et al., 2018).

2.3 Análise da Cesta de Mercado

Em outra corrente, um grupo de pesquisas foi desenvolvido na perspectiva de identificar associações entre produtos ou categorias de produtos que um comprador adquire durante uma visita de compras. Nota-se que a compreensão da dependência entre categorias e efeitos correspondentes do *mix* de *marketing* são de interesse dos gestores para otimizar a lucratividade do gerenciamento de categorias no varejo. Para essa finalidade, os trabalhos realizam a análise da cesta de compras, geralmente por meio da mineração de regras de associação. Esse método utiliza grandes bases de dados para desvendar as relações entre produtos ou categorias de produtos, além de indicar quais itens ou categorias são mais frequentemente comprados em conjunto (GRIVA et al., 2018). A segmentação por cesta de compras tem sido utilizada em diversas áreas, como finanças, telecomunicações e varejo (CHEN et al., 2005).

Em seu trabalho, Cil (2012) buscou identificar associações entre categorias de produtos compradas em um supermercado (categorias cruzadas), a fim de alterar o *layout* dos produtos na loja. São utilizados em seu estudo, dados de vendas e o *mix* de categorias. Hong, Misra e Vilcassim (2016) investigaram se o impacto da variedade de uma categoria de produtos afeta a incidência de compras em outra categoria exibida nas proximidades. Como achado, o trabalho apontou a limitação do gerenciamento de categorias individuais quando os varejistas tomam decisões acerca da variedade de produtos. Ainda, González-Benito, Martínez-Ruiz e Mollá-Descals (2010) estudaram o papel da estrutura competitiva da categoria e propuseram um modelo de decisão capaz de indicar preços ótimos de categoria. Para alcançar essa finalidade o modelo de efeitos cruzados abordado, incluiu efeitos de preço entre pares de categorias separadas.

Já Ghoniem, Maddah e Ibrahim (2016), examinaram o problema da venda cruzada de várias categorias, em que um varejista procura otimizar conjuntamente as decisões de sortimento e preço para uma categoria principal e, várias categorias secundárias relacionadas - cada uma composta por produtos substituíveis, demonstrando a importância do planejamento conjunto de categorias de varejo relacionadas à venda cruzada. Borin, Farris e Freeland (1994), por sua vez, desenvolveram um modelo de gerenciamento de categoria para ajudar os varejistas nas decisões de restrição de espaço de quais produtos estocar (sortimento) e quanto espaço de prateleira deveria ser alocado para esses produtos. Em outra perspectiva, Tang, Chen e Hu (2008) combinaram dados de vendas fornecidos por várias unidades de loja de uma rede de supermercados e realizaram a análise do padrão de compra, em determinado período e local, combinando dias da semana e lojas.

Por fim, os estudos que empregam regras de associação para a mineração de dados em nível da cesta, com a finalidade de extrair pares de produtos que clientes compram com maior frequência em suas compras trouxeram contribuições relevantes. No entanto, esses estudos concentraram esforços apenas em pares de categorias de produtos adquiridos, ou seja, não consideraram a totalidade das categorias de produtos. Por esse motivo, acredita-se que essa abordagem possui limitações e não reflete necessariamente as intenções de compras dos consumidores. Assim como os estudos que focam na totalidade dos dados de vendas para realizar a segmentação de clientes, esses trabalhos que consideram apenas pares de categorias de produtos para estudar relações existentes entre as categorias correm o risco de negligenciar o conhecimento por detrás de cada visita de compras e, subestimar as necessidades de compras que os clientes desejam satisfazer por visita (GRIVA et al., 2018).

3 PROCEDIMENTO METODOLÓGICO

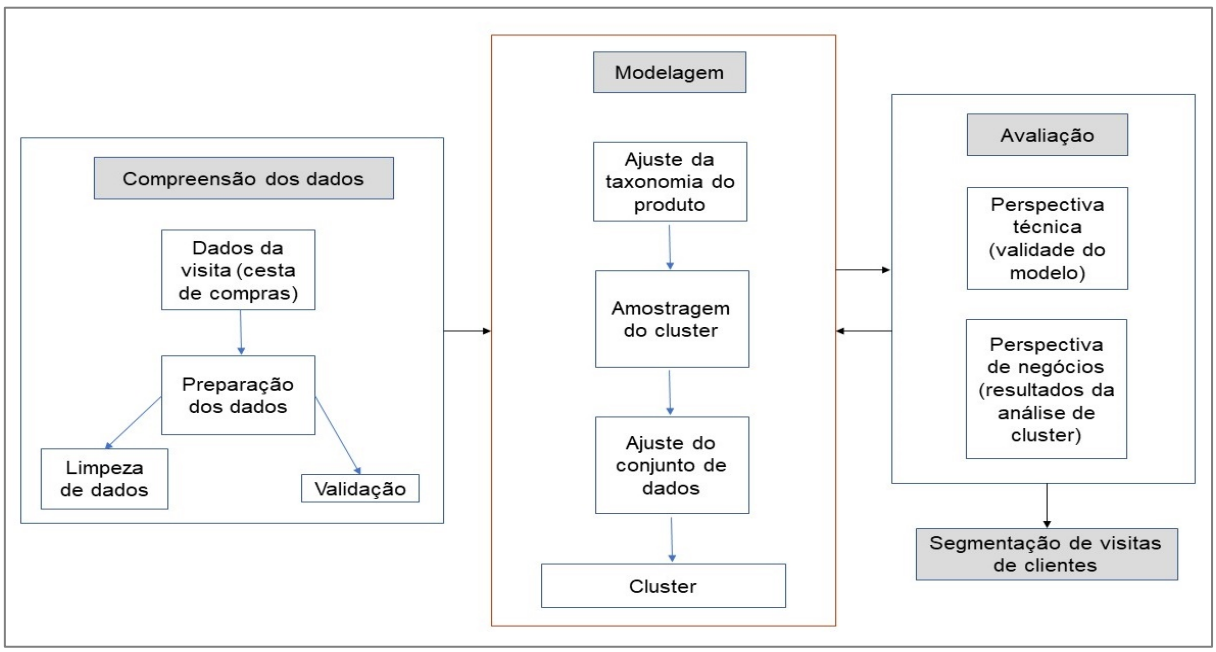
Para o desenvolvimento da pesquisa foi empreendida uma revisão de literatura, por meio do levantamento bibliográfico, tendo por finalidade identificar todas as referências encontradas sobre um determinado tema (CERVO; BERVIAN, 2002). O levantamento foi realizado utilizando-se o Portal de Periódicos Capes, por meio de consulta a periódicos, priorizando artigos científicos publicados em revistas internacionais qualificadas - de alto fator de impacto -, que abordassem os seguintes temas: (i) segmentação de clientes no varejo, (ii) segmentação de clientes por meio de dados de vendas, (iii) análise de cestas de compras, (iv) gerenciamento por categorias, (v) gerenciamento de categorias cruzadas, (vi) gerenciamento entre categorias de produtos e, a (vii) taxonomia de produtos.

Após a seleção dos artigos, fez-se a análise do conteúdo dos textos com o propósito de descartar aqueles que não estavam relacionados com o interesse da pesquisa. Os artigos que por fim, compuseram a base bibliográfica do estudo, contribuíram para responder ao problema de pesquisa; oportunizaram extrair da literatura embasamento para a construção do referencial teórico; bem como permitiram a proposição de um modelo e a discussão teórica de sua aplicação.

4 APRESENTAÇÃO E DISCUSSÃO DO MODELO CONCEITUAL PROPOSTO

Nessa seção é apresentada uma proposta baseada na abordagem analítica de dados (do inglês, *business analytics*) (DELEN; DEMIRKAN, 2013; HOLSAPPLE; LEE-POST; PAKATH, 2014) que emprega a técnica de *cluster* para realizar a segmentação e caracterização das visitas dos clientes, analisando as categorias de produtos que um cliente compra enquanto visita uma loja. Abaixo segue o modelo conceitual (Figura 1), que será discutido passo-a-passo.

Figura 1 - Modelo Conceitual Proposto



Fonte: Adaptado de Griva et al. (2018).

4.1 Compreensão dos Dados

Para realizar a segmentação examinando as categorias de produtos que os clientes compram durante suas visitas em lojas, os dados utilizados deverão ser secundários, fornecidos por gestores de Redes de Supermercados, referentes às compras de produtos por visita (dados da cesta de compras). Os dados de entrada deverão conter a árvore da categoria do produto e o seu código de barras.

Dado o caráter 'ruidoso' dos dados, entende-se que é necessário preparar os dados para a melhor compreensão do contexto e de como eles se originaram. Para encontrar e corrigir ou, remover os erros e inconsistências dos dados, além de executar transformações, a técnica a ser adotada será a limpeza de dados (RAHM; DO, 2000). Por fim, há uma preocupação em validar os dados, ao final de cada etapa, para que a análise feita seja de maneira consistente.

4.2 Modelagem

Nesta etapa são indicados a realização de três procedimentos: (i) ajuste da taxonomia do produto, (ii) amostragem do *cluster* e, (iii) ajuste e *cluster* dos dados de entrada.

4.2.1 Ajuste da Taxonomia do Produto

O varejo possui em sua caracterização uma hierarquia de produtos, também conhecida como árvore de categorias de produtos. Seu papel é importante principalmente no reabastecimento das lojas, alocação de espaço nas prateleiras e composição dos sortimentos de produtos. Essa hierarquia diz respeito à variedade de produtos e subdivisões das categorias desses produtos, com o intuito de auxiliar no desempenho das atividades operacionais. No entanto, pode-se perceber que não é adequado realizar análise de dados com a árvore de categorias de produtos

fornecida por um gestor supermercadista, visto que possui características que prejudicam o desempenho dos algoritmos de mineração de dados (GRIVA et al., 2018).

Algumas das características que podem dificultar a utilização dos algoritmos são: i) os níveis da hierarquia ou da altura da árvore de produtos podem ser distintos, indicando que há alta especialização de produtos em algumas categorias; ii) a árvore ou hierarquia dos produtos pode desenvolver comportamentos de expansão de produtos de forma separada e independente; e iii) variações no nível da unidade de manutenção de estoques (SKU). Assim, nota-se que tais características são capazes de influenciar na mineração dos dados, devido à utilização de dados distorcidos (GRIVA et al., 2018).

Para examinar o conjunto de dados e evitar sinais de distorção, recomenda-se adotar uma abordagem baseada na variedade de produtos, capaz de ajustar a taxonomia original do produto do varejista e produzir uma árvore de categorias de produtos customizada, que por sua vez, possibilita a realização adequada da análise de agrupamento e a identificação dos segmentos de visita dos clientes.

Para equilibrar a taxonomia da árvore de produtos, segue-se alguns passos, tais como: identificar as principais categorias de produtos. Posteriormente, partindo da premissa de que uma pequena quantidade de categorias de produtos contribui para a maior frequência nas cestas de compras, utiliza-se a proposição de Bi, Faloutsos e Korn (2001), de modo que os dados das vendas sejam representados por uma Distribuição Exponencial Gaussiana Discreta. Essa distribuição, geralmente, é adotada na finalidade de dividir as categorias de produtos em dois conjuntos, sendo eles: (i) um conjunto que representa categorias de produtos com alta frequência na cesta e; (ii) o segundo, que monta as demais categorias de produtos. Sendo assim, compreende-se a partir de uma avaliação empírica que a razão ou proporção da frequência encontrada do primeiro conjunto para o segundo é de 1:10.

Em consonância, para gerenciar os efeitos negativos da assimetria dos dados, adota-se uma abordagem interativa para a árvore de taxonomia de baixo para cima, visando diminuir a quantidade de categorias de produtos com baixa frequência na cesta de compras, por meio de agrupamentos.

Após a realização desta etapa, é preciso revisar a taxonomia do novo produto, verificando se as categorias agrupadas seriam mantidas conforme o algoritmo indicou ou, divididas para a criação de novas categorias mescladas, a fim de atender às necessidades da mineração de dados. Ademais, compreende-se que se deve ter cuidado para evitar a fusão de categorias não relacionadas, portanto, sendo necessário a realização de uma avaliação qualitativa.

4.2.2 Amostragem do *Cluster*

A técnica indicada para agrupar as amostras é a amostragem por conglomerados, com pesos iguais para eliminar os valores extremos. Além disso, considera-se que as visitas nas quais os consumidores compram muito poucos produtos ou uma quantidade muito grande de produtos, em termos de variedade, culminam por produzir *outliers*. Nota-se que essas visitas podem corresponder a viagens muito concretas ou abstratas (BELL et al., 2011). Assim, observa-se que viagens muito concretas não auxiliam na compreensão da missão das visitas de compras, visto que são direcionadas e não possibilitam a extração de afinidades de produtos. Já as viagens muito abstratas, contêm uma enorme variedade de produtos, que dificultam a compreensão da missão específica das compras.

A técnica de amostragem por conglomerados agrupa uma população finita em conjuntos de sub-populações, para posteriormente selecionar subconjuntos de *clusters* (GRIVA et al., 2018). Para selecionar os *clusters* finais, aponta-se o seguinte critério: a porcentagem pertencente de cada *cluster* na cesta de compras, assim como as receitas por *cluster.* Logo, recomenda-se excluir as cestas com apenas um produto ou com mais de oitenta produtos.

4.2.3 Ajuste do Conjunto de Dados e *Cluster*

Para desenvolver a técnica de *cluster*, é necessário criar uma tabela a partir dos dados, incluindo todas as informações sobre a visita de compras do cliente. Nota-se que as linhas da tabela se referem às informações sobre as visitas (cestas de compras), e as colunas referem-se às categorias de produtos personalizadas. As colunas recebem códigos binários, valores 0 e 1, para indicar respectivamente, a ausência e presença das categorias de produtos na cesta de compras. Os dados presentes na tabela configuram-se como a entrada do modelo de *cluster*.

Para efetuar a modelagem, indica-se a abordagem *k-means*, sendo motivada pelo fato de que visa desvendar *clusters*, de modo que os objetos em cada *cluster* sejam semelhantes entre si e diferentes dos objetos em outros *clusters*. Essa abordagem estabelece um conjunto de *clusters* inicial, que é atualizada, até que não haja oportunidades de melhorias (HUERTA-MUÑOZ; RÍOS-MERCADO; RUIZ, 2017, YOU et al., 2015). No entanto, sua precisão depende do número inicial de *clusters* (MESFOROUSH; TAROKH, 2013). Portanto, deve-se utilizar ferramentas para adotar o número adequado inicial de *clusters*. Alguns métodos conhecidos são: a estatística de 'cotovelo', silhueta e *gap*.

4.3 Avaliação do Modelo

Ainda é preciso avaliar o modelo. Pelo fato da técnica de agrupamento de *cluster* ser pautada na semelhança de objetos contidos, as métricas para avaliar quão relacionados ou separados estão os objetos dentro do mesmo *cluster* precisam ser desenvolvidas e calculadas, para a validação interna dos *clusters* (LIU et al., 2013).

Uma maneira de escolher se os resultados da modelagem de *cluster* são adequados é por meio da comparação da soma do erro quadrático (SSE) para diversos números de *cluster* (diferentes números de K). A SSE é conhecida como uma medida global de erro, sendo definida como a soma das distâncias entre cada objeto de um *cluster* e o seu centroide do *cluster*, elevada ao quadrado (GRIVA et al., 2018).

Nota-se uma relação percebida: quanto maior for o número de *clusters*, menor é o valor do SSE, porque os *clusters* são menores. Uma análise gráfica do SSE em detrimento dos valores de K, pode fornecer uma visualização adequada para escolher um número apropriado de *clusters*. Portanto, um valor ideal para K pode ser visto como aquele que proporciona reduções bruscas no SSE. Fato que é responsável pela produção de um "cotovelo" no gráfico SSE contra as soluções de *cluster* (KETCHEN; SHOOK, 1996, LIKAS et al., 2003).

5 CONSIDERAÇÕES FINAIS

Na presente pesquisa, escolheu-se discutir um modelo teórico construído com base na abordagem analítica de dados (BA), com intuito de fornecer subsídios futuros para a área de

pesquisa no setor supermercadista, haja vista que o modelo proposto visa desvendar e apontar o comportamento de compras dos clientes por visita única. É importante notar, que as compras por visitas únicas são capazes de proporcionar informações relevantes para os gestores sobre as necessidades dos clientes, contribuindo para a tomada de decisão varejista, em busca da satisfação dos clientes. Quando se trabalha com a totalidade das visitas de compras, não é possível identificar comportamentos de visitas a nível individual de clientes e, portanto, perde-se informações valiosas, como a missão das compras dos clientes. Por isso, o esforço da presente discussão teórica se concentrou na compra por 'visita única'.

Quanto ao problema invocado na pesquisa, entende-se que a resposta a ser dada é que supostamente é sim possível identificar algum padrão de consumo nos *tickets* de compra de um supermercado, que contribua para a caracterização da missão de compra por visita dos clientes, haja vista se considerarmos a implementação do modelo conceitual proposto pela pesquisa.

Em relação às contribuições do estudo, foi argumentado teoricamente acerca do agrupamento de dados do carrinho de compras com o objetivo de criar grupos de cestas, ou seja, segmentos de visitas de clientes, a partir das categorias de produtos presentes nas cestas. Nota-se que apesar de existir um estudo na Europa sobre o tema (GRIVA et al., 2018), ainda não foi identificado por meio dessa revisão de literatura, obras que estudassem empiricamente segmentos de visitas de clientes no varejo supermercadista brasileiro, a fim de identificar a intenção de compra dos clientes por visita única. Representando, assim, uma lacuna de pesquisa no respectivo campo de estudo. Esse ensaio então, avança em parte com relação a esse *gap*, isto é, ao tratar e analisar por meio de uma revisão de literatura, o modelo de Griva et al. (2018) e sua possível aplicação ao caso brasileiro.

Ressalta-se que a utilização empírica do modelo proposto em pesquisas futuras, não consiste unicamente em trabalhar com os dados da cesta apenas para identificar relações complementares e substitutas entre pares de categorias de produtos que os clientes compram mais frequentemente. Ele visa atuar com todas as categorias de produtos presentes em cada cesta de compras, pois a compreensão das necessidades que levaram o cliente a realizar as compras é altamente relevante. Portanto, parte-se da premissa de que o *mix* resultante das categorias de produtos, reflete o objetivo da compra de cada visita do cliente.

No entanto, pode-se observar ainda que é possível encontrar *clusters* com visitas mais abstratas. Um caminho para compreender melhor as características desse *cluster* é criando subdivisões. Ou seja, o algoritmo pode indicar um *cluster* contendo refrigerantes, sucos, cervejas, salgadinhos, pizza e batata-chips, por exemplo. Assim, uma subdivisão possível seria a separação entre comidas (salgadinhos, pizza e batata-chips) e bebidas (refrigerantes, sucos, cervejas).

Espera-se, então, que a pesquisa futura mediante ao emprego do modelo proposto no artigo, permita à identificação de padrões ocultos na missão das compras. Portanto, uma possibilidade está em executar a função de probabilidade entre *clusters* mais definidos com os abstratos, ainda que, geralmente, os *subclusters* resultantes sejam os mesmos que os originais (GRIVA et al., 2018).

REFERÊNCIAS

AERON, H.; KUMAR, A.; MOORTHY, J. Data mining framework for customer lifetime value-based segmentation. **Journal of Database Marketing & Customer Strategy Management**, v. 19, n. 1, p. 17-30, 2012.

ALBADVI, A.; SHAHBAZI, M. A hybrid recommendation technique based on product category attributes. **Expert Systems with Applications**, v. 36, n. 9, p. 11480-11488, 2009.

BELL, D. R.; CORSTEN, D.; KNOX, G. From point of purchase to path to purchase: How preshopping factors drive unplanned buying. **Journal of Marketing**, v. 75, n. 1, p. 31-45, 2011.

BI, Z.; FALOUTSOS, C.; KORN, F. The DGX distribution for mining massive, skewed data. **Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining**, p. 17-26, 2001.

BOONE, D. S.; ROEHM, M. Retail segmentation using artificial neural networks. **International journal of research in marketing**, v. 19, n. 3, p. 287-301, 2002.

BORIN, N.; FARRIS, P. W.; FREELAND, J. R. A model for determining retail product category assortment and shelf space allocation. **Decision sciences**, v. 25, n. 3, p. 359-384, 1994.

CERVO, A. L.; BERVIAN, P. A. **Metodologia científica**. 5. ed. São Paulo: Prentice Hall, 2002.

CHENG, C.; CHEN, Y. Classifying the segmentation of customer value via RFM model and RS theory. **Expert systems with applications**, v. 36, n. 3, p. 4176-4184, 2009.

CHEN, Y. et al. Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data. **Electronic Commerce Research and Applications**, v. 8, n. 5, p. 241-251, 2009.

CHEN, Y. et al. Market basket analysis in a multiple store environment. **Decision support systems**, v. 40, n. 2, p. 339-354, 2005.

CHO, Y. H.; KIM, J. K.; KIM, S. H. A personalized recommender system based on web usage mining and decision tree induction. **Expert systems with Applications**, v. 23, n. 3, p. 329-342, 2002.

CIL, I. Consumption universes based supermarket layout through association rule mining and multidimensional scaling. **Expert Systems with Applications**, v. 39, n. 10, p. 8611-8625, 2012.

DELEN, D.; DEMIRKAN, H. Data, information and analytics as services. **Decision Support Systems**, v. 55, n. 1, p. 359–363, abr. 2013.

GHONIEM, A.; MADDAH, B.; IBRAHIM, A. Optimizing assortment and pricing of multiple retail categories with cross-selling. **Journal of Global Optimization**, v. 66, n. 2, p. 291-309, 2016.

GONZÁLEZ-BENITO, Ó.; MARTÍNEZ-RUIZ, M. P.; MOLLÁ-DESCALS, A. Retail pricing decisions and product category competitive structure. **Decision Support Systems**, v. 49, n. 1, p. 110-119, 2010.

GRIVA, A. et al. Retail business analytics: Customer visit segmentation using market basket data. **Expert Systems with Applications**, v. 100, p. 1-16, 2018.

HOLSAPPLE, C.; LEE-POST, A.; PAKATH, R. A unified foundation for business analytics. **Decision Support Systems**, v. 64, p. 130–141, jun. 2014.

RAHM, E.; DO, H. H. Data cleaning: Problems and current approaches. **IEEE Data Eng. Bull.**, v. 23, n. 4, p. 3-13, 2000.

HAN, S. et al. Category role aided market segmentation approach to convenience store chain category management. **Decision Support Systems**, v. 57, p. 296-308, 2014.

HONG, S.; MISRA, K.; VILCASSIM, N. J. The perils of category management: The effect of product assortment on multicategory purchase incidence. **Journal of Marketing**, v. 80, n. 5, p. 34-52, 2016.

HUERTA-MUÑOZ, D. L.; RÍOS-MERCADO, R. Z.; RUIZ, R. An iterated greedy heuristic for a market segmentation problem with multiple attributes. **European Journal of Operational Research**, v. 261, n. 1, p. 75-87, 2017.

HUNG, C.; TSAI, C. Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand. **Expert systems with applications**, v. 34, n. 1, p. 780-787, 2008.

KETCHEN, D. J.; SHOOK, C. L. The application of cluster analysis in strategic management research: an analysis and critique. **Strategic management journal**, v. 17, n. 6, p. 441-458, 1996.

KHAJVAND, M. et al. Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. **Procedia Computer Science**, v. 3, p. 57-63, 2011.

KITTS, B.; FREED, D.; VRIEZE, M. Cross-sell: a fast promotion-tunable customer-item recommendation method based on conditionally independent probabilities. **Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining**, p. 437-446, 2000.

LARSON, J. S.; BRADLOW, E. T.; FADER, P. S. An exploratory look at supermarket shopping paths. **International Journal of research in Marketing**, v. 22, n. 4, p. 395-414, 2005.

LIAO, S.; CHEN, Y. Mining customer knowledge for electronic catalog marketing. **Expert Systems with Applications**, v. 27, n. 4, p. 521-532, 2004.

LIAO, S.; CHEN, Y.; HSIEH, H. Mining customer knowledge for direct selling and marketing. **Expert Systems with Applications**, v. 38, n. 5, p. 6059-6069, 2011.

LIKAS, A.; VLASSIS, N.; VERBEEK, J. J. The global k-means clustering algorithm. **Pattern recognition**, v. 36, n. 2, p. 451-461, 2003.

LIU, Y. et al. Understanding and enhancement of internal clustering validation measures. **IEEE transactions on cybernetics**, v. 43, n. 3, p. 982-994, 2013.

MESFOROUSH, A.; TAROKH, M. J. Customer profitability segmentation for SMEs case study: network equipment company. **International Journal of Research in Industrial Engineering**, v. 2, n. 1, p. 30-44, 2013.

MIGUÉIS, V. L.; CAMANHO, A. S.; CUNHA, J. F. Customer data mining for lifestyle segmentation. **Expert Systems with Applications**, v. 39, n. 10, p. 9359-9366, 2012.

NAMIN, A.; DEHDASHTI, Y. A "hidden" side of consumer grocery shopping choice. **Journal of Retailing and Consumer Services**, v. 48, p. 16-27, 2019.

NGAI, E. W. T; XIU, L.; CHAU, D. C. K. Application of data mining techniques in customer relationship management: A literature review and classification. **Expert systems with applications**, v. 36, n. 2, p. 2592-2602, 2009.

PARK, C. H.; PARK, Y.; SCHWEIDEL, D. A. A multi-category customer base analysis. **International Journal of Research in Marketing**, v. 31, n. 3, p. 266-279, 2014.

PROVOST, F.; FAWCETT, T. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. O'Reilly Media, Inc., 2013.

SARANTOPOULOS, P. et al. Shopping missions: An analytical method for the identification of shopper need states. **Journal of Business Research**, v. 69, n. 3, p. 1043-1052, 2016.

TANG, K.; CHEN, Y.; HU, H. Context-based market basket analysis in a multiple-store environment. **Decision Support Systems**, v. 45, n. 1, p. 150-163, 2008.

YOU, Z. et al. A decision-making framework for precision marketing. **Expert Systems with Applications**, v. 42, n. 7, p. 3357-3367, 2015.