European Journal of Applied Business Management, 5(3), 2019, pp. 48-67.

EJABM European Journal of Applied Business and Management

ISSN 2183-5594

Research Paper

Aplicação de Árvore de Decisão Para Adoção de E-commerce B2B por Pontos de Venda

Submitted in July 18, 2019
Accepted in September 24, 2019
Evaluated by a double blind review system

PAULA VEIGA CHEUICHE[1]

DAIANE ROSSI[2]

VITÓRIA ABREU FLORES DA CUNHA[3]

MARCELO NOGUEIRA CORTIMIGLIA[4]

RODRIGO DALLA VECCHIA[5]

RESUMO

Proposta: O presente estudo, conduzido junto ao setor de marketing de uma fabricante de cigarros, tem como objetivo identificar padrões de comportamentos em uma base de varejos com vistas à análise de adesão ao sistema e-commerce no segmento B2B.

Desenho/Metodologia/Abordagem: Esta pesquisa é de natureza aplicada, quantitativa, descritiva e exploratória, com o propósito de criar algoritmos de descoberta. Foram analisadas as características mercadológicas e os volumes comprados pelos varejos a fim de identificar fatores correlacionados com o bom desempenho dos varejos e encontrar suas semelhanças. Consiste na aplicação da técnica de data mining e da árvore de decisão para auxiliar nesta identificação.

Resultados: A partir do algoritmo de árvore de decisão foi possível relacionar as informações mercadológicas dos varejos com as performances de atingimento das metas estabelecidas pela organização, identificando assim duas grandes classes de varejos com bom potencial de adesão ao sistema de comércio eletrônico.

Limitações da investigação/Implicações: Este estudo restringiu-se a uma empresa e seus parceiros de varejo que já estavam cadastrados no programa de relacionamento B2B da empresa. No método, na plotagem da árvore de decisão, o número de nós gerados foi limitado pelo tamanho de área de plotagem, houve sobreposição de nós quando detalhou-se as variáveis.

Originalidade/Valor: O artigo contribui no sentido de apresentar um caso real de aplicação de árvore de decisão com potencial estratégico para a empresa em uma possível expansão da base de varejos para adesão ao sistema de comércio eletrônico.

Palavras-chave: E-commerce; Árvore de Decisão; Machine Learning; Marketing; B2B

[1] Universidade Federal do Rio Grande do Sul, Brasil. E-mail: paulacheuiche@gmail.com
[2] Universidade Federal do Rio Grande do Sul, Brasil. E-mail: dai-rossi@hotmail.com* corresponding author
[3] Universidade Federal do Rio Grande do Sul, Brasil. E-mail: v.abreufc@gmail.com
[4] Universidade Federal do Rio Grande do Sul, Brasil. E-mail: cortimiglia@producao.ufrgs.br
[5] Universidade Federal do Rio Grande do Sul, Brasil. E-mail: rodrigovecchia@gmail.com

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

1. Introdução

A internet trouxe às empresas muitas oportunidades de inovação, tanto na área de desenvolvimento de produtos, como na área de serviços, tais como o marketing. Nesse contexto, as organizações estão adaptando-se ao marketing digital, em que os clientes possuem mais conhecimento e autonomia em suas escolhas e as empresas possuem mais informações sobre o mercado, devido à conectividade (Kotler et al., 2017). Segundo o relatório de 2017 sobre economia digital divulgado pela Conferência das Nações Unidas sobre Comércio e Desenvolvimento (UNCTAD), 59% da população do Brasil tem acesso à Internet, sendo, em número absoluto, o quarto país no ranking de usuários conectados.

Nos últimos anos, o comércio eletrônico (e-commerce) tem ganhado destaque entre as empresas no segmento Business to Business (B2B). O e-commerce é realizado por uma plataforma digital que tornou possível a intensificação da capacidade de pedidos, (Chakraborty et al., 2002), da redução dos custos, da agregação valor ao cliente (Souza, 2016) e da otimização da logística (Kotler et al., 2017), através da eliminação de barreiras geográficas entre cliente e empresa. No Brasil, o comércio eletrônico já está consolidado, representando 76,18% do valor de transações online do mercado total no segmento B2B (FGV-EAESP, 2016).

Com o crescimento acelerado da utilização da internet nos negócios, houve também o aumento do número de bancos de dados de clientes, permitindo que as organizações atingissem o público alvo de forma mais eficaz e obtivessem auxílio em suas tomadas de decisões (Dantas et al., 2008). Um grande banco de dados, o big data, reúne altos volume de informações úteis que agregam conhecimento a respeito de um determinado conteúdo, por meio de interações dos usuários com a Internet (Poongothai et al., 2011). Entretanto, para que essas informações sejam utilizadas de forma eficiente, é necessária a aplicação de técnicas de análise de dados, chamada de Knowledge Discovery in Databases ou KDD (Fayyad et al., 1996).

Uma das etapas do KDD é o data mining, que possibilita a identificação de padrões em um banco de dados por meio da criação de algoritmos de descoberta. Além disso, no marketing, o data mining pode ser utilizado para segmentação de público alvo, análises de sazonalidade e previsões de demanda (James et al., 2013). Além disso, Soltani e Navimipour (2016) afirmam que técnicas de data mining podem ser usadas para descobrir padrões ocultos no comportamento do cliente. Algumas das técnicas para a validação de informações em um banco de dados são a classificação, a análise de associações, o clustering e a análise de outliers (Côrtes et al., 2002).

Para o data mining, é necessário a utilização de recursos de interpretação de dados, que podem englobar a área do machine learning. Segundo Heiler (2017) o machine learning é uma forma de inteligência artificial que permite a busca informações úteis por meio de algoritmos de descoberta. Uma de suas técnicas é a da árvore de decisão, capaz de encontrar semelhanças em dados por uma sequência de testes em diferentes atributos de entrada (Alpaydin, 2016). Com essa técnica, é possível auxiliar no marketing das empresas, como na predição de vendas e na classificação de grupos econômico-sociais por comportamentos (Garcia, 2003).

O avanço da captação de dados de clientes trouxe às empresas, portanto, o desafio de compreender e utilizar informações de maneira proveitosa. Imensos bancos de dados possuem, muitas vezes, enormes variedades de atributos, que dificultam a percepção de análises e resultados (Ling et al., 1998). Para Rocha et al. (2002), a qualidade da captação

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

de dados depende da habilidade da organização de avaliar adequadamente as informações para, assim, gerar conhecimento e sobreviver no mercado.

Este artigo tem como objetivo identificar padrões de comportamentos em uma base de varejos já aderente ao sistema e-commerce de uma empresa de cigarros do segmento B2B, através da técnica da árvore de decisão. O método irá definir a classe mais relevante de clientes, com potencial de compra via comércio eletrônico, para estabelecer as características mercadológicas prioritárias para a empresa. Os resultados encontrados auxiliarão a empresa na sua tomada de decisão em uma possível expansão da base de varejos que utiliza o comércio eletrônico. Serão analisados os volumes comprados e as características mercadológicas de cada ponto de venda, a fim de encontrar as melhores performances e identificar suas semelhanças.

Na seção a seguir será apresentado o referencial teórico. Após, na seção três, será explicado o contexto da situação-problema e os procedimentos metodológicos utilizados. Por fim, na quarta e na quinta seção o leitor poderá, respectivamente, identificar os resultados encontrados e a conclusão do trabalho.

2. Referencial Teórico

2.1 Marketing Digital

Farias (2016) define o marketing digital como um conjunto de informações e ações que podem ser feitos em diversos meios digitais com o objetivo de promover empresas. Ele ainda constata que se difere do tradicional por envolver o uso de diferentes canais online, métodos e ferramentas que permitem a análise dos resultados em tempo real.

Para Kotler (2017), o marketing digital ocasionou uma mudança do poder para os consumidores conectados. O autor afirma que o cenário dos negócios está se transformando de vertical, exclusivo e individual para horizontal, inclusivo e social, devido ao crescimento da competitividade das pequenas empresas pelo progresso da tecnologia, da inclusão dos mercados emergentes à economia e pelo compartilhamento de opiniões e avaliações dos consumidores na internet, respectivamente.

O sistema e-commerce, ou comércio eletrônico, cresce a cada ano em razão das estratégias do marketing digital. Peçanha (2018) observa que o foco do marketing digital não está mais no produto, e sim na experiência do usuário em sua jornada de compra, resultando na necessidade de uma interatividade cada vez mais eficiente entre empresas com seu público. Ryan (2017) ressalta que o marketing digital é feito por pessoas, em que comerciantes se conectam com consumidores para construir um relacionamento favorável para o mercado.

A essência desse novo conceito de marketing é a aproximação entre clientes e empresas, para que se possa compreender suas necessidades, agregar valor aos produtos, aumentar os canais de distribuição e alavancar o volume de vendas, através da interação online (Chaffey et al., 2013). Entretanto, existem riscos, tanto para as organizações, que estão mais expostas a seus concorrentes (Wind et al., 2001) e mais vulneráveis às reclamações de seus clientes (Gubert, 2018), quanto para os consumidores, que sofrem com a insegurança das transações virtuais e do roubo de seus dados pessoais (Chaffey et al., 2013).

É importante ressaltar que o marketing digital trouxe uma nova concepção das empresas em relação a seus clientes. A grande estratégia dos comerciantes é identificar as preferências de seus consumidores e, assim, sugerir os produtos que eles desejam

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

comprar. Farias (2016) observa que é possível segmentar de forma específica as pessoas que a empresa deseja atingir com seu conteúdo, produto ou serviço. Além disso, a conectividade permite que os consumidores personalizem seus produtos de maneira muito mais rápida e pesquisem sobre preços, qualidade e concorrência (Wind et al., 2001).

2.2 E-commerce

Chaffey et al. (2013) definem o e-commerce como qualquer ação que envolva vendas ou transações online. Para Müller (2013), é uma rede em que várias pessoas se conectam e fazem transações que exigem lealdade e transparência. Nakamura (2011) afirma que o e-commerce engloba todos os processos da cadeia de valor realizada em um ambiente eletrônico, com o objetivo de atender as necessidades do mercado.

O comércio eletrônico é realizado através de uma plataforma, a "espinha dorsal" do sistema, que define como a marca irá personalizar a exibição de seus produtos, o catálogo e os itens desejados pelos clientes, para que o consumidor final tenha uma excelente experiência de compra (Iacovone, 2016). Em um estudo sobre a fidelidade do cliente do e-commerce, Srinivasan et al. (2002) identificaram que a qualidade da customização, da interatividade, do capricho, das opiniões de outros consumidores e do seu poder de escolha impacta na fidelidade dos usuários conectados.

Para Chakraborty et al. (2002), a percepção dos usuários conectados sobre as websites das empresas afeta na efetividade do e-commerce. Os autores identificam que personalização (personalization), interatividade das atividades de compra (transaction-related interactivity), interatividade de atividade de relacionamento com usuários (nontransaction-related interactivity), informações (informativeness), organização (organization), privacidade e segurança (privacy and security), acessibilidade (accessibility) e entretenimento (entertainment) são elementos essenciais de websites estratégicos. Quanto maior a percepção desses elementos, maior será a efetividade.

É importante observar que o e-commerce pode trazer vantagens competitivas para as empresas. Tsai et al. (2005) propõem que o comércio eletrônico pode aumentar o volume de vendas e o market share, melhorar o relacionamento com os clientes e melhorar a performance financeira atráves da redução de custos e do crescimento do ROI (Return on Investiment). Além disso, em razão da eliminação das barreiras geográficas, é possível atingir um maior número de consumidores, expandido a capacidade de pedidos (Clarke et al., 2015), e otimizar o sistema de logística e distribuição das organizações (Kotler, 2017).

O progresso do e-commerce e da qualidade na captação de informações sobre clientes, permitiu que as empresas identificassem um perfil de consumidor. Para Rocha et al. (2002), as empresas possuem imensos bancos de dados repletos de informações sobre clientes, em tempo real, que possibilitam a criação e a modificação de produtos ou serviços de forma rápida para satisfazer as necessidades dos usuários. Ainda para Rocha et al., a interatividade e a análise das informações dos clientes conduzem as organizações para a sobrevivência no mercado eletrônico.

A grande quantidade de banco de dados que as empresas tem acesso podem ser chamados de big data. McAfee et al. (2012) caracterizam o big data como a coleta de alto volume de dados, de forma rápida e variada, tanto de forma estruturada como não estruturada. Além disso, McAfee et al. também ressaltam que, com o big data, é possível obter informações úteis e transformá-las em vantagens competitivas no mercado.

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

Assim como no comércio tradicional, o comércio eletrônico também é dividido por segmentos de mercado, definidos como Business to Business (B2B), Business to Consumer (B2C), Consumer to Consumer (C2C), Government to Consumer (G2C) e Government to Business (G2B). Conforme Catalani et al., (2009), os segmentos são definidos na Tabela 1.

Tabela 1 – Segmentos do Mercado Eletrónico

Segmento de Mercado	Definição
B2B	Segmento do comércio eletrônico associado às transações online entre empresas, seja por operação de compra e venda de informações, de produtos e de serviços.
B2C	Segmento do comércio eletrônico associado às transações online entre empresas e consumidores finais, seja por operação de compra e venda de informações, de produtos e de serviços. É o modelo de negócio mais clássico.
C2C	Segmento do comércio eletrônico associado às transações online entre usuários particulares da internet. O comércio envolve apenas consumidores finais, sem intermediários.
G2C	Relação comercial online entre o governo (federal, estadual ou municipal) e consumidores.
G2B	Relação de negócios online entre governo (federal, estadual ou municipal) e empresas.

Fonte: Adaptado de Catalani et al. (2009).

O modelo B2B é definido como um segmento de mercado que envolve transações entre empresas e são, em geral, utilizados para operações de revenda, transformação ou consumo (Nissan, 2014). Uma importante estratégia desse segmento é a qualidade do relacionamento entre as empresas presentes no negócio. Segundo Rauyruen (2007), a satisfação e a qualidade do serviço são itens importantes para a fidelidade do cliente.

Com o crescimento da popularidade e dos bons resultados do e-commerce, a plataforma online tornou-se uma boa opção para potencializar as vendas de uma cadeia produtiva. As indústrias conseguem escoar de forma mais assertiva a sua produção, os distribuidores conseguem otimizar rotas e, consequentemente, aumentar suas capacidades, enquanto os varejistas ganham velocidade e transparência no processo de compra (Chaussard, 2018).

Uma característica do mercado B2B são as transações de altas quantias e altos volumes. Sendo assim, é de grande importância a construção de uma relação de confiança e segurança entre as empresas (Pavlou, 2002). Além disso, com o e-commerce, as empresas vendedoras necessitam manter o processo de compra e venda em constante desenvolvimento (Nakamura, 2011), sempre priorizando o engajamento de seus clientes através de estratégias de comunicação e personalização aos usuários online (Tucunduva, 2018).

2.3 Knowledge Discovery in Databases (KDD)

O KDD, ou a descoberta de conhecimento em banco de dados, tem como objetivo transformar dados puros em informações úteis e compreensíveis (Prass, 2012). Fayyad (1996) define o KDD como um processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados.

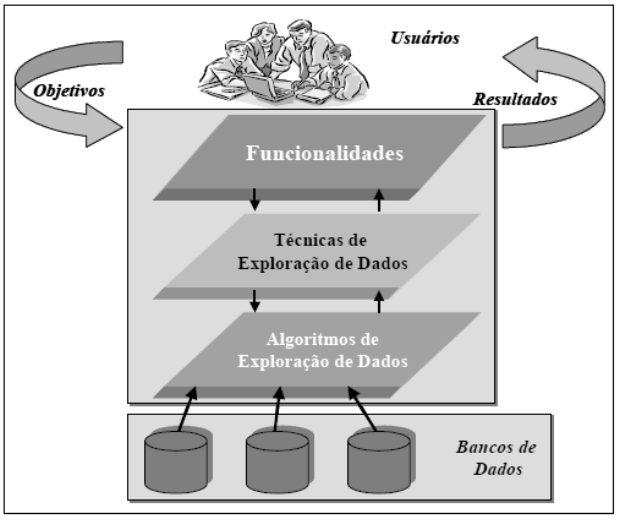
EJABM European Journal of Applied Business and Management
ISSN 2183-5594

Usualmente, um processo KDD visa encontrar, em um dado conjunto de dados: (i) classificadores, (ii) preditores, (iii) agrupamentos, (iv) padrões, (v) anomalias, (vi) associações, e/ou (vii) modelos. Nos últimos anos, tem sido muito utilizado na pesquisa devido aos crescentes volumes de dados disponíveis tanto em bancos dados privados, públicos e da web. Seus campos de aplicação são vastos e em diferentes domínios, tais como: (i) saúde, (ii) análises de mercado (iii) cyber segurança, (iv) bioinformática (Guarascio et al., 2018).

2.3.1 Data Mining

Data Mining é uma recurso de descoberta de informações que podem agregar conhecimento e auxiliar na tomada de decisão das empresas. Segundo Côrtes et al. (2002), essa tecnologia possui aplicações em diversas áreas dos negócios, como marketing, finanças e manufatura, e áreas da saúde, como farmacêutica, hospitalar e biomédica. Além disso, ainda para os autores, a definição da funcionalidade e dos resultados que se deseja conquistar é fundamental para o processo de mineração de dados. Quando se obtém conceitos bem definidos, melhor será a escolha da técnica a ser aplicada e, consequentemente, melhor serão os resultados alcançados. A Figura 1 mostra a relação entre funcionalidade, técnicas e algoritmos, a fim de esclarecer a interatividade entre o objetivo do data mining e as técnicas empregadas.

Figura 1 – Funcionalidades em Mineração de Dados



Fonte: Côrtes et al. (2002).

Para Han et al. (2005), existem variadas funcionalidades de mineração de dados, definidas como análise de associação, clusterização, análise de outliers e classificação. A análise de associação tem como finalidade descobrir regras de associação em um conjunto de dados, visando principalmente a criação de "pacotes" para consumidores. Uma empresa, por exemplo, pode definir itens em uma loja que um determinado padrão de clientes compra e, assim, mantê-los em seções próximas. Silva et al. (2019), utilizaram a análise de associação, por meio da técnica do algoritmo Apriori, que identifica frequentes associações if-then chamadas regras de associação, para classificar clientes com níveis de fidelidade diferentes, o que permitiu à empresa desenvolver estratégias de retenção para seus clientes. A aplicação do algoritmo de associação Apriori no conjunto de transações de cada grupo de clientes permitiu a elaboração de importantes regras de associação com

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

altos níveis de confiança. O estudo de caso demonstrou que, com a mineração de dados, é possível extrair conhecimento útil e contribuir na tomada de decisão da empresa.

A clusterização permite segmentar um conjunto de dados com o objetivo de formar grupos baseados em semelhanças. É importante que esses grupos sejam homogêneos em si e heterogêneos entre si (Han et al., 2005). Huang et al. (2007) utilizaram a clusterização para segmentar o público alvo de uma empresa de bebidas, por meio da técnica do SVC. Foram utilizados dados de clientes como o nível de socialização, tempo de lazer, conhecimento e satisfação para a aplicação da técnica.

A análise de outliers permite a identificação de um conjunto de dados que não obedecem ao modelo. Quando encontrados, podem ser tratados ou descartados para que não ocorram desvios ou riscos que prejudiquem os objetivos traçados no início do data mining (Han et al., 2005). Como exemplo, é possível avaliar as vendas de uma empresa para verificar o comportamento dentro de um período, bem como as vendas de um determinado produto ou de uma região específica (Côrtes et al., 2002).

Por fim, a classificação consiste na análise de uma determinada característica em um banco de dados e na sua atribuição a certa classe previamente definida. As classes, por exemplo, podem ser para itens de compra ou para perfis de consumidores, como os que tendem a comprar muito e os que compram apenas o necessário (Han et al., 2005). Através de Clusters, Rule Mining e Árvore de Decisão, Khalili-Damghani et al. (2018) conseguiram prever potenciais novos clientes em dois estudos de caso através de clusters previamente definidos, identificando grupos de clientes com alto e baixo valor e prevendo o nível de valor de novos clientes.

Após a definição da melhor funcionalidade do banco de dados, são estabelecidas as técnicas de data mining mais adequadas para a extração de conhecimento. As técnicas devem ser aplicadas de acordo com as características do banco de dados e englobam diversas áreas científicas, como estatística e machine learning (Fu, 1997).

2.3.2 Machine Learning

Machine learning é uma inteligência artificial capaz de encontrar informações úteis através de algoritmos de descoberta e surgiu da teoria que as máquinas não precisam ser programadas para aprenderem a realizar tarefas (Heiler, 2017). Por meio dessa tecnologia, computadores são aptos a ajustarem-se quando expostos a mais dados, tornando o "aprendizado" automático e iterativo, sem a necessidade de intervenção humana (Rabelo, 2018). Além disso, difere-se do método estatístico por utilizar "n" atributos ao invés de vetores em "n" dimensões (Fu, 1997).

Um dos objetivos do machine learning é compreender a estrutura do banco de dados para realizar decisões e resultados confiáveis utilizando big data. Para Witten et al. (2016), envolve a descoberta de informação para gerar performance, que comumente é aplicada na área de marketing e vendas. Os autores salientam que é possível personalizar a experiência de compra e, assim, aumentar a fidelização de seus clientes e segmentar o público-alvo, a partir do machine learning.

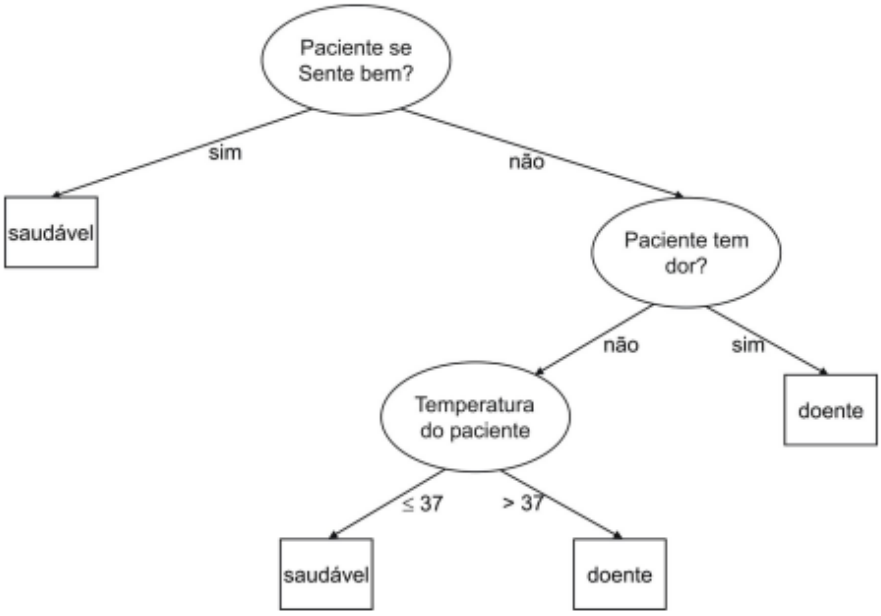
O machine learning é uma ciência que vem ganhando impulso, e uma das técnicas mais comuns é a árvore de decisão. Essa técnica representa um fluxograma em que os nós significam atributos testados, os ramos significam resultados dos testes e as folhas significam as distribuições dos registros (Côrtes et al., 2002). Por meio de seus algoritmos, é possível identificar regras para cada folha e, assim, encontrar padrões de

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

classificação (Witten et al., 2016). Dentre as vantagens da utilização da árvore de decisão para descoberta de conhecimento, para Fonseca (1994), destacam-se a capacidade de ser aplicada a qualquer tipo de banco de dados e a fácil compreensão dos resultados.

A técnica da árvore de decisão pode ser aplicada em diversas áreas, como marketing, na predição de vendas e na classificação de grupos econômico-sociais por comportamentos, e medicina, na determinação de diagnósticos e tratamentos (Garcia, 2003). A Figura 2 apresenta um exemplo de uma árvore de decisão simples de um diagnóstico de um paciente. Cada elipse, ou nó, representa um atributo testado para um conjunto de pacientes e cada retângulo representa o diagnóstico, ou a folha. Para alcançar o diagnóstico, é necessário realizar o teste em cada nó, até que apenas uma folha seja alcançada (Monard et al., 2003).

Figura 2 – Árvore de Decisão Simples



3. Metodologia

3.1 Descrição do Cenário

Este trabalho foi realizado no setor de marketing de uma empresa de cigarros que atua no segmento B2B, que comercializa tanto para varejos via compra geográfica, como via compra e-commerce. A compra geográfica é caracterizada pela visita de um vendedor que faz a solicitação dos pedidos do varejista. Já a compra e-commerce é representada pela autonomia do varejista, que pode efetuar os pedidos pela plataforma digital da empresa sem a presença de um vendedor.

Cada ponto de venda possui classificações fundamentais para a consolidação da base de varejos da empresa, de acordo com características mercadológicas. Essas características foram as variáveis utilizadas para a análise do trabalho, a fim de encontrar varejos e-commerce com as melhores performances de volumes comprados. Foram então utilizadas 14 variáveis, que podem ser visualizadas na Tabela 2. A amostra utilizada consistiu de 2246 clientes, que são pontos varejos situados nos estados do Rio Grande do Sul e de

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

Santa Catarina, com amostra efetiva após limpeza dos dados de 2040 clientes. Os dados utilizados abrangem o período de setembro de 2017 até julho de 2018.

Tabela 2 – Características dos Pontos de Venda

Variáveis	Definição
Estrutura de Vendas	Informações direcionada para a empresa, como território de vendas, código do ponto de venda e depósito abastecedor
Estratégico	Pontos de vendas considerados estratégicos para o ganho de *market-share*
Endereço	Rua, Bairro, Município e CEP
Número de Habitantes do Município	Número de habitantes do município, de acordo com o IBGE
Atividade Comercial	Mini Mercado; Loja Cnv Independente; Supermercado; Loja Cnv Ka; Bar; Lanchonete/Padaria; Mercearia; Pista Posto De Gasolina; Restaurante; Banca/Loteria/Revistaria; Tabacaria; Cafe
Faixa	Classificação de 1 a 8, sendo 1 varejos com maior média semanal de compra e 8 com menor.
CEV	Perfil do Varejo: A (*Premium)*, B (*Aspirant Premium)*, C (*Value for Money)* e D (*Low Price)*
DV	Dia de Venda: de segunda-feira a sexta-feira
Exclusivo	Sem presença da concorrência: SIM ou NÃO
Contrato	Existência de contrato comercial entre a empresa e o ponto de venda: SIM ou NÃO
Dias Boleto	Prazo de pagamento do boleto: 7, 14 ou 21 dias
Programa Relacionamento B2B	Presença de cadastro do varejista e dos atendentes do varejo no site de relacionamento B2B da empresa (programa de incentivo com premiações)
Meta	Objetivo mensal de compra calculado pela empresa
Volume Comprado	Volume mensal comprado

3.2 Classificação da Pesquisa

A pesquisa pode ser definida como de natureza aplicada, visto que utiliza informações já existentes, com o propósito de criar algoritmos de descoberta. Além disso, a abordagem do trabalho é quantitativa, pois, através de resultados numéricos, tem como intuito determinar semelhanças entre os varejos analisados. Em relação aos objetivos de pesquisa, é classificado como de origem descritiva, em razão de apresentar e analisar os fatores que determinam a performance do sistema de venda e-commerce da empresa. Por fim, o trabalho é caracterizado como exploratório, pois visa avaliar matematicamente o comportamento do banco de dados utilizado.

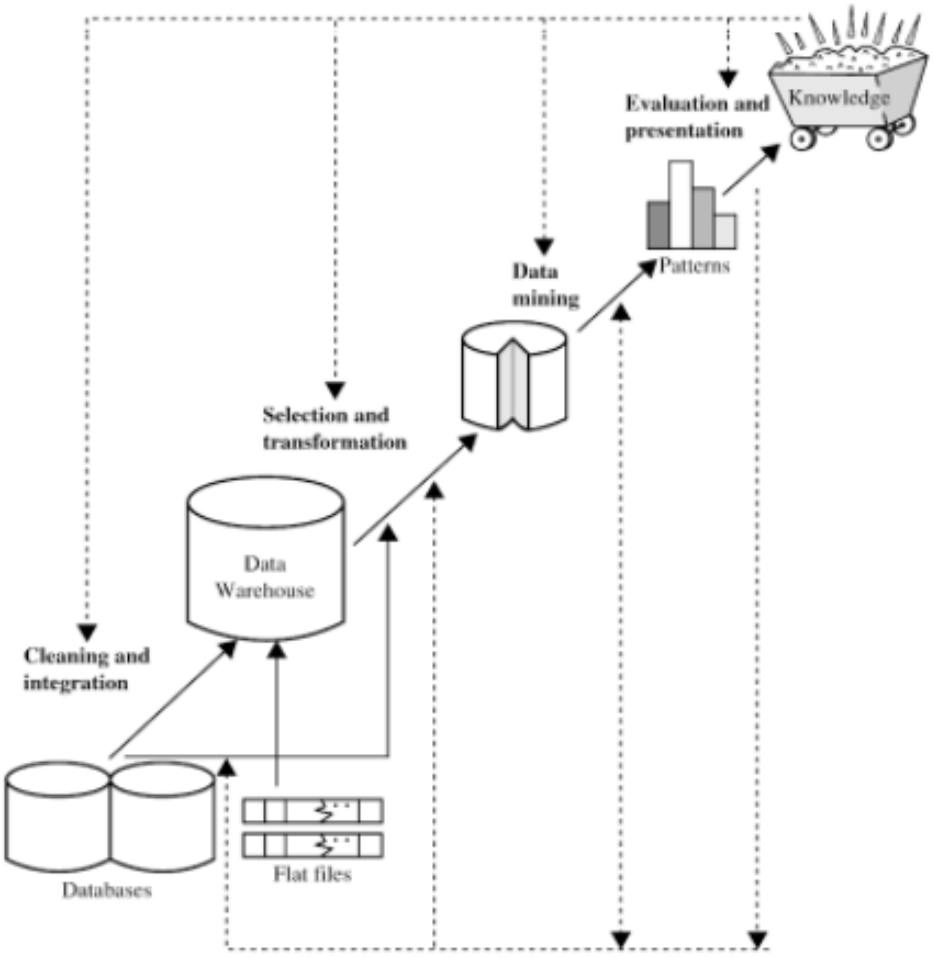
3.3 Método do Trabalho

O método foi inspirado nas sete etapas do KDD (Han et al., 2005). A primeira etapa é a limpeza dos dados, para remover ruídos e inconsistências. Após é realizada a integração, em que ocorre a associação de várias bases de dados em um formato padrão. Na terceira etapa acontece a seleção dos dados que possam ser relevantes na análise. A seguir é feita a etapa de transformação, em que se verifica a transformação e a consolidação dos dados em formatos apropriados para a atividade de garimpagem (mining). A quinta etapa é a da

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

mineração dos dados, que é a fase essencial do processo de KDD. A mineração envolve a identificação dos objetivos e da melhor técnica a ser aplicada. Em seguida é realizada a avaliação de padrões, em que são constatados os padrões e os prognósticos a serem utilizados, sempre com base na análise estatística. Por fim, na última etapa é feita a apresentação dos resultados, no qual todo o processo de mineração de dados é retornado em ações baseadas nos conhecimentos adquiridos. Na Figura 3 é possível verificar as etapas do KDD.

Figura 3 – Etapas do KDD



Fonte: Adaptado de Han et al. (2005).

Nesse trabalho, inicialmente realizou-se limpeza de dados inconsistentes, para que não houvesse influência na interpretação final da pesquisa. Para a consolidação dos dados, todas as informações foram organizadas em apenas uma base de dados em formato Excel, na qual colunas representaram as variáveis e linhas os diferentes pontos de venda da empresa.

Logo após, foi realizada a data mining propriamente dita. A funcionalidade desejada na pesquisa foi a classificação e, por isso, os varejos foram classificados de acordo com suas performances de atingimento da meta, através de machine learning. A classificação é importante na área de marketing para identificação de classes de clientes, a fim de encontrar um significado estratégico (Linoff et al., 2011). A partir do conjunto de regras definidas pela árvore de decisão, identificou-se padrões entre os varejos do comércio eletrônico, caracterizando a sexta etapa. A técnica de árvore de decisão foi utilizada na análise do banco de dados, por meio da aplicação da linguagem R, uma linguagem capaz

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

de carregar bancos de dados em formato Excel e efetuar modelagens lineares e não lineares, testes estatísticos clássicos e análise de séries temporais (Dionísio, 2015). O pacote rpart foi aplicado para a plotagem da árvore.

Na última etapa do KDD, foi realizada, portanto, a análise dos resultados para definir as características eleitas como essenciais na tomada de decisão da empresa, no que tange à classificação de varejos com potencial de uso do comércio eletrônico. Com isso, foi possível auxiliar nas estratégias de marketing e no direcionamento dos clientes aptos a utilizarem a plataforma digital de forma proveitosa.

4. Resultados

Após a limpeza dos dados inconsistentes na primeira etapa da análise, foram excluídos de 206 varejos com cadastros incompletos. Para a consolidação dos dados, a variável de volume foi mensurada pelo atingimento das metas crivadas pela empresa mês a mês, sendo classificada em uma escala de 0 a 1. Assim, definiu-se 0 o valor para volumes zerados e 1 o valor para metas mensais atingidas. A partir disso, foi calculada a média do atingimento das metas dos meses analisados, resumindo a variável de volume em apenas uma coluna. Para a variável "Atividade Comercial", optou-se por representá-la de forma numérica. Para isso, foi criado um ranking de 1 até 12, de acordo com o volume total comprado no período pelas atividades comerciais. A atividade mais expressiva, portanto, foi considerada a primeira do ranking.

Tabela 3 – Extrato da base de dados analisados

Ponto de Venda	Estrat.	N de Habitantes	Atividade Comercial	Faixa	CEV	DV	Excl.	Contr.	Prog. Relac B2B	Media do Ating. da Meta
1	NÃO	1.4768,6	Restaurante	2	A	2	SIM	SIM	NÃO	0,1
2	SIM	125.975	Lanchonete	4	C	2	NÃO	SIM	NÃO	0,8
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
2039	SIM	205.271	Loja Conveniência KA	4	B	6	SIM	NÃO	SIM	1
2040	NÃO	1.4768,6	Supermercado	8	B	5	NÃO	SIM	SIM	0,6

Para a seleção das variáveis relevantes, eliminou-se informações relacionadas à estrutura de venda da organização, em razão de não apresentarem representatividade para as características mercadológicas dos varejos. O endereço também foi desconsiderado da análise, visto que é uma característica exclusiva, tal que não há pontos de vendas com a mesma localização. Sendo assim, na base de varejos foram utilizadas as variáveis apresentadas na Tabela 3. Na quarta etapa foi realizada a transformação do banco de

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

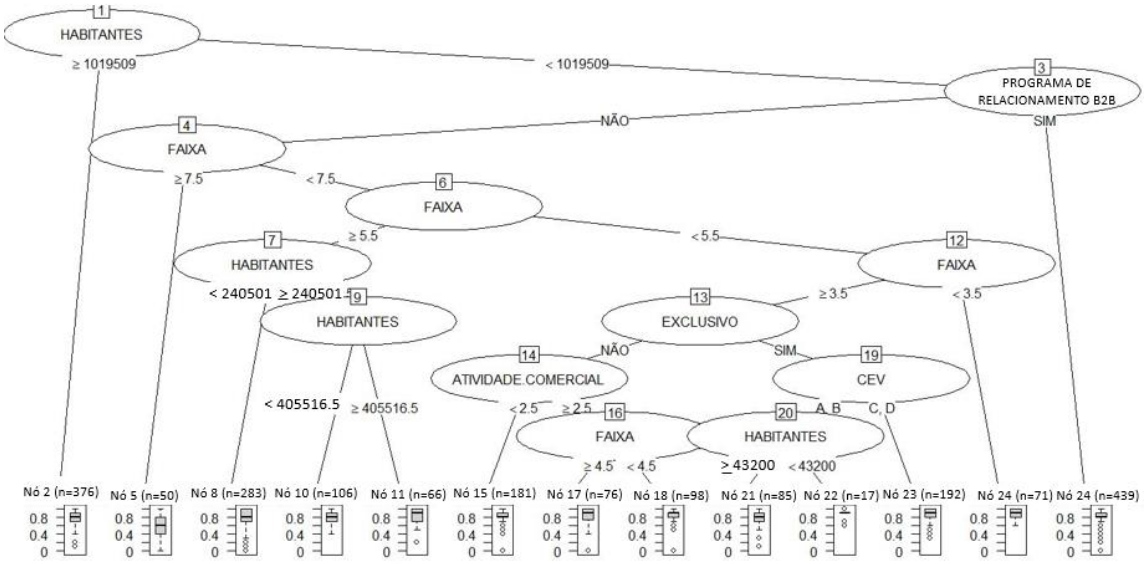
dados em um formato apropriado para a utilização da linguagem R, sendo removidos acentos e sinais gráficos.

A partir do algoritmo de árvore de decisão foi possível relacionar as informações mercadológicas dos varejos com as performances de atingimento das metas estabelecidas pela organização. Como variável resultante da árvore de decisão, as folhas, definiu-se a média de atingimento. Assim, quanto mais próximo de 1 o valor dos registros, melhor a performance do ponto de venda no e-commerce no período avaliado.

Para um melhor desenvolvimento da pesquisa, foram criadas duas árvores de decisão, a fim de compreender de maneira mais aprofundada as correlações entre as variáveis. Para a primeira aplicação, foram consideradas todas as variáveis apresentadas na Tabela 3. Já para a segunda, a variável "Nº de Habitantes" foi desconsiderada.

A primeira árvore de decisão obtida está representada na Figura 4. Foi possível identificar que o algoritmo definiu que as variáveis "DV", "Estratégico" e Contrato" não demonstraram influências relevantes na pesquisa, visto que não foram apresentadas interações com as outras variáveis.

Figura 4 – Árvore de Decisão 1



Fonte: Os autores (2019)

As variáveis resultantes foram divididas pelo algoritmo em 13 folhas, caracterizadas pela média de atingimento da meta. Para cada folha foi gerado um gráfico boxplot da distribuição dos registros, em que foi apresentado o comportamento dos quartis, a sua mediana e o número de varejos representados. No primeiro atributo testado, o "Nº de Habitantes", observou-se a folha resultante nomeada nó 2, em que foi possível verificar 376 varejos localizados em municípios com número de habitantes superior a 1.019.509. A mediana resultante do nó foi de 0,802.

O maior valor encontrado para a mediana foi de 0,878, representado pelo nó 22, e o menor de 0,580 representado pelo nó 5, refletindo um intervalo significativo de 0,298. A Figura 5 mostra o histograma da quantidade de varejos e a distribuição das medianas das folhas, assim como a sua linha de tendência. É possível observar que os registros das folhas dos

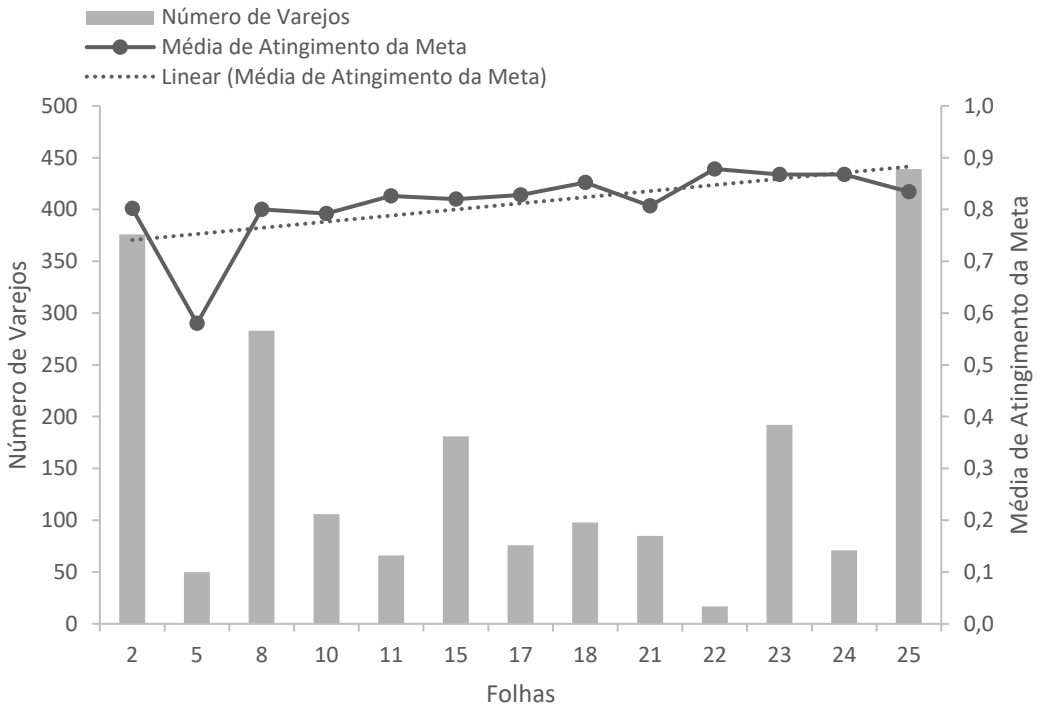
EJABM European Journal of Applied Business and Management
ISSN 2183-5594

nós finais da árvore apresentam valores mais próximos de 1, em comparação às folhas iniciais.

A folha que demonstrou a maior mediana foi o nó 22, com 17 varejos. A partir das características mercadológicas representadas, observou-se que os varejos estão localizados em municípios com número de habitantes inferior a 43.200 e que possuem classificação CEV A e B. Outras características em comum para esses pontos de venda são a exclusividade, a faixa comercial 4 ou 5 e a inexistência de cadastro no programa de relacionamento B2B. Em contrapartida, a menor mediana foi representada pela folha do nó 5, com 50 varejos localizados em munícipios com número de habitantes inferior a 1.019.509, sem cadastro no programa de relacionamento B2B e de faixa comercial 8.

Em relação a folha com maior número de registros, verificou-se o nó 25, com 439 varejos. É valido ressaltar que o valor da mediana da folha é de 0,834, refletindo uma diferença de 0,044 quando comparado com o nó 22, e que as características mercadológicas dos pontos de venda são representadas pela localização em munícipios com número de habitantes inferior a 1.019.509 e pela existência de cadastro no programa de relacionamento B2B.

Figura 5 – Histograma da Árvore de Decisão 1

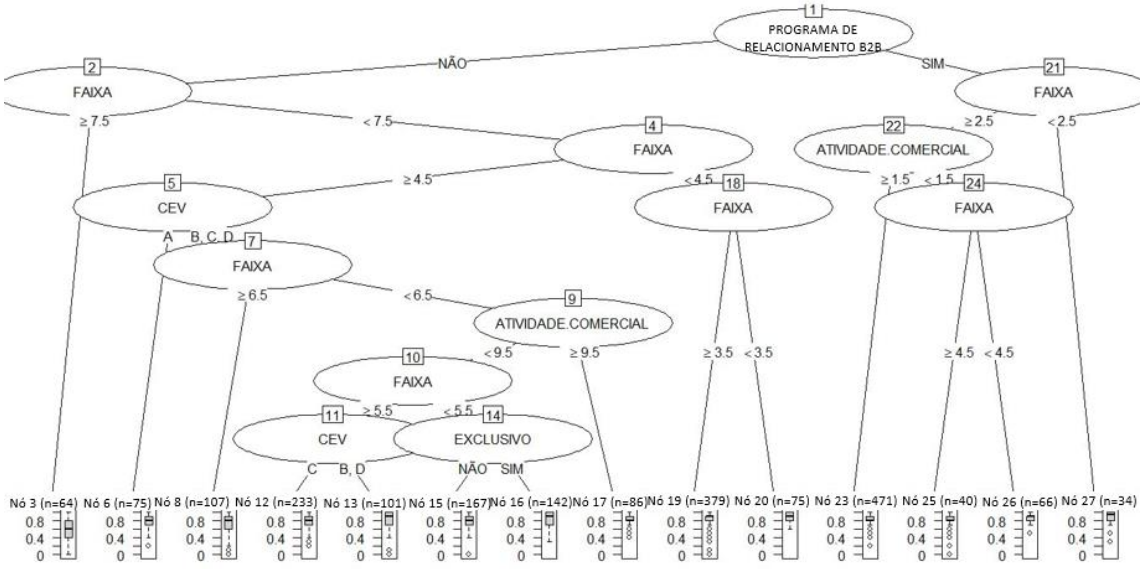


Fonte: Os autores (2019)

A segunda árvore, em que a variável de "Nº de Habitantes" foi desconsiderada, é apresentada na Figura 6. Assim como na árvore anterior, foi definida como variável resultante a média do atingimento da meta e, portanto, os valores do gráfico boxplot variaram de 0 a 1. Além disso, as variáveis "DV", "Estratégico" e Contrato" também não fizeram parte das interações. Foram observadas 14 folhas resultantes e o nó inicial partiu da variável "Programa Relacionamento B2B".

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

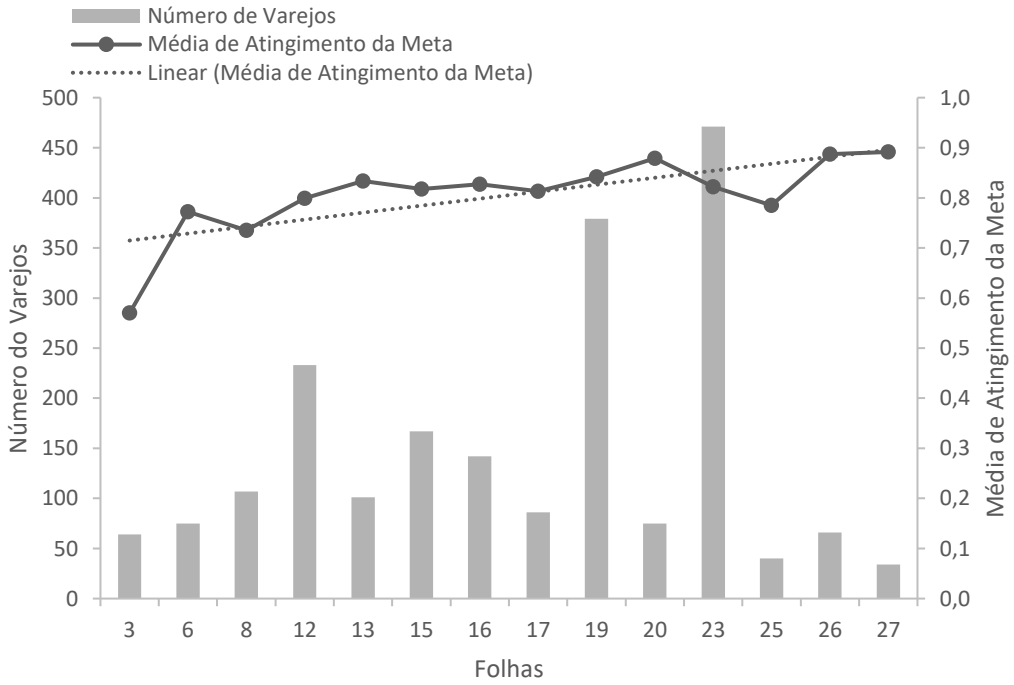
Figura 6 – Árvore de Decisão 2



Fonte: Os autores (2019)

A maior mediana foi apresentada na folha denominada nó 27, com 34 pontos de venda e com o valor de 0,892. As características mercadológicas em comum dessa folha foram a presença de cadastro no programa de relacionamento B2B e classificação de faixa 1 ou 2. Já o valor da menor mediana foi 0,570, apresentada pela folha denominada nó 3, com amostra de 64 pontos de venda sem cadastro no programa de relacionamento B2B e com faixa de mercado 8. O intervalo entre essas medianas foi de 0,321.

Figura 7 – Histograma da Árvore de Decisão 2



Fonte: Os autores (2019)

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

A folha com maior representatividade em números de varejos foi a 23, com uma amostra de 471 pontos de venda com cadastro no programa de relacionamento B2B, com faixa comercial a partir de 3 e com atividade comercial maior que 1. Observou-se um valor de 0,822 para a folha, refletindo uma diferença de 0,070 em comparação com a maior mediana. Na Figura 7 é possível identificar a linha de tendência das variáveis resultantes.

Sendo assim, analisando os resultados das duas árvores de decisão, compreendeu-se que tanto o tamanho da amostra como o resultado da mediana são importantes para a pesquisa. As variáveis consideradas como as mais representativas para a performance e-commerce foram "Programa Relacionamento B2B" e "Faixa", em razão de seus diferentes caminhos nas árvores influenciarem no valor das variáveis resultantes. Para varejos com a presença de cadastro no programa de relacionamento B2B, verificou-se a tendência de uma melhor performance de metas atingidas. Isso pode ser explicado pela digitalização do programa, visto que o cliente já está habituado a acessar o site da empresa. Para as faixas comerciais, foi observado que quanto menor a faixa, melhor foi a performance do ponto de venda.

Em relação ao número de habitantes, verificou-se que municípios com mais habitantes tendem a uma melhor performance que municípios menores. Para as atividades comerciais, foi observado a atividade 1 apresentou melhores resultados para pontos de vendas com cadastro no programa de relacionamento B2B e de faixa 1, 2, 3 ou 4. Entretanto, as demais atividades comerciais apresentaram correlações semelhantes para a mesma interação. Por fim, para "CEV" e "Exclusivo" foi verificado essas variáveis não influenciaram em grande parte das interações das árvores, sendo decisivas apenas para situações pontuais.

5. Conclusão

O trabalho foi realizado no setor de marketing de uma empresa de cigarros que atua no segmento B2B e teve como objetivo identificar padrões de comportamento de clientes que são aderentes ao sistema e-commerce. A partir da metodologia do KDD, definiu-se a funcionalidade de classificação para a identificação de classes de clientes, a fim de encontrar um significado estratégico para a empresa em uma possível expansão da base do comércio eletrônico. Na análise, foram selecionadas variáveis mercadológicas, presentes no banco de dados da empresa, para aplicação da árvore de decisão.

Para os resultados, foram avaliados os valores das medianas dos gráficos boxplot e o tamanho da amostra de cada folha, ambos definidos pelo algoritmo das árvores de decisão. Observou-se, portanto, que a classe de clientes mais relevante para a utilização do e-commerce é a que possui cadastro no programa de relacionamento B2B, que pertence a faixa de 1 até 4, que está localizado em um munícipio com mais de 1.019.509 habitantes e que possui atividade comercial classificada como 1. Além disso, para a classe de clientes que não possui cadastro no programa de relacionamento B2B foi possível identificar que, correlacionadas com a variável "Faixa", os pontos de venda de faixa 1 até 3 também demonstraram uma boa performance. Sendo assim, esses dois grupos de clientes apresentaram bom potencial de compra via comércio eletrônico. Entretanto, o grupo que manifestou o pior resultado foi o de varejos de faixa 8 e sem cadastro no programa de relacionamento B2B.

Com os aspectos encontrados nos resultados da pesquisa, compreende-se que a empresa adquiriu um melhor direcionamento estratégico para a expansão do seu sistema de venda e-commerce. A presença do programa B2B trouxe uma relevância positiva para a performance dos varejos, visto que, de maneira geral, o cadastramento dos clientes

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

contribuiu para o bom desempenho de atingimento das metas. Em adição, para faixas menores e para municípios com mais habitantes, verificou-se também a tendência de resultados mais satisfatórios. Logo, a definição das classes de clientes com melhor e pior desempenho foi importante para estabelecer as características mercadológicas prioritárias para a empresa manter-se competitiva no mercado. Esses dados permitem que a empresa alinhe sua estratégia de e-commerce para contemplar melhor as classes de clientes com maior potencial de sucesso na plataforma, bem como direcione sua estratégia de expansão a pontos de varejo com características similares aos com bom desempenho.

A plotagem da árvore de decisão pode ser considerada uma limitação do trabalho, visto que o número de nós gerados foi limitado pelo tamanho de área de plotagem. Houve sobreposição de nós quando buscou-se correlacionar as variáveis forma mais detalhada, deixando a árvore ilegível. Em relação a trabalhos futuros, podem ser consideradas para a análise varejos de outros estados do Brasil, a fim de compreender se as mesmas classes de clientes têm desempenho semelhante ao longo do território nacional. Além disso, pode-se incluir variáveis externas ao banco de dados da empresa, como dados do IBGE, possibilitando assim, a construção de uma estratégia unificada de expansão de e-commerce.

Referências bibliográficas

Alpaydin, E. (2016), Machine Learning: The New A.I, The MIT Press, pp. 67.

Catalani, L. Kischinevsky, A., Ramos, E. A. D. A., Junior, H. N. S., (2009), *E-commerce*, FGV, São Paulo.

Chaffey, D., Smith, P. (2013), *Emarketing Excellence: Planning and Optimizing your Digital Marketing*, Routledge, Londres.

Chakraborty, G., Lala, V., Warren, D., (2002), "An Empirical Investigation of Antecedents of B2B Websites' Effectiveness", *Journal of Interactive Marketing*, Vol. 16, No. 4, pp. 51-72.

Chaussard, C., (2018), *E-commerce Descentralizado para Indústrias: Integrando as vendas da distribuição, representantes, revendas até o consumidor,* Flexy.

Clarke, G., Thompsom. C., Birkin. M., (2015), "The emerging geography of e-commerce in British retailing", *Regional Science*, Vol. 2, No. 1, pp. 371-391.

Côrtes, S. D. C., Porcaro, R. M., Lifschitz, S., (2002), *Mineração de Dados – Funcionalidades, Técnicas e Abordagens,* PUC - Rio Inf. MCC 10112, Rio de Janeiro.

Dantas, E. R. G.; Júnior, J. C. A. P.; Lima, D. S. De; Azevedo, R. R. De., (2008), "O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de Decisões", *V Simpósio de Excelência em Gestão e Tecnologia*, p. 1–10.

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

Dionísio, E. J. (2015). "Trabalhando com a Linguagem R.", disponível em : https://www.devmedia.com.br/trabalhando-com-a-linguagem-r/33275 (acesso em: 15 de setembro de 2019).

Farias, F. (2016), *Série Épicos: O Guia Definitivo do Marketing Digital*, Resultados Digitais, São Paulo.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., (1996), From data mining to knowledge discovery: An overview. In Advances in knowledge discovery and data mining, *AAAI/MIT Press*, Vol. 17, No.3, pp 37- 54.

FGV-EAESP, (2016). *Pesquisa Comércio Eletrônico no Mercado Brasileiro edição 18*, FGV, São Paulo.

Fonseca, J. M. M. R. D. (1994). "Indução de Árvores de Decisão: HistClass - Proposta de um algoritmo não paramétrico", 140p. Dissertação (Mestrado em Engenharia Informática), Universidade Nova de Lisboa, Lisboa.

Fu, Y., (1997), "Data mining. Potentials", *IEEE*, Vol. 16, No. 4, pp. 18-20.

Garcia, S. C. (2003), "O Uso de Árvores de Decisão na Descoberta de Conhecimento na Área da Saúde", 88p. Dissertação (Mestrado em Ciência da Computação), Universidade Federal do Rio Grande do Sul, Porto Alegre.

Guarascio, M., Manco, G., & Ritacco, E., (2018), *Knowledge Discovery in Databases. Reference Module in Life Sciences*, Elsevier, The Institute of Calculation and High Performance Networks, Cosenza.

Gubert, F., (2018), "Marketing Digital: 4 riscos que sua marca corre na internet e como evita-los" *Deen Marketing Digital*., disponível em: http://deen.com.br/blog/marketing-digital-4-riscos-que-sua-marca-corre-na-internet-e-como-evita-los/ (acesso em: 13 de maio de 2018).

Han, J., Kamber, M., (2005), *Data mining: Concepts and techniques*, Morgan Kaufmann Publishers Inc., Burlington.

Heilier, L., (2017), "Difference of Data Science, Machine Learning and Data Mining", *Data Science Central*, disponível em: https://www.datasciencecentral.com/profiles/blogs/difference-of-data-science-machine-learning-and-data-mining (acesso em: 23 de junho de 2018).

Huang, J., Tzeng, G., Ong, C., (2007), "Marketing segmentation using support vector clustering.", *Expert Systems with Applications*, Vol. 32, No. 2, pp. 313-319.

Iacovone, C. (2016). "Comércio Eletrônico: Como Funciona?" *Administradores.com* disponível em: http://www.administradores.com.br/artigos/tecnologia/comercio-eletronico-como-funciona/99259/ (acesso em: 15 de maio de 2018).

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R.,* Springer, Nova York.

Khalili-Damghani, K., Abdi, F., Abolmakarem, S., (2018), "Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries", *Applied Soft Computing,* Vol. 73, pp. 816-828.

Kotler, P., Kartajaya, H. & Setiawan, I., (2017), *Marketing 4.0: Moving from Tradicional to Digital,* John Wiley & Sons, Inc., Hoboken.

Ling, C. X., Li, C. (1998), "Data Mining for Direct Marketing: Problems and Solutions", in *Proceedings of the International Conference on Knowledge Discovery and Data MiningAmerican - KDD98,* Nova York, 1998, pp. 73-79.

Linoff, G. S., Berry, M. J. A. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management,* Wiley-Blackwell, Hoboken.

McAfee, A., Brynjolfsson, E. (2012). "Big Data: The Management Revolution. Spotlight on Big Data". *Harvard Business Review,* disponível em: https://hbr.org/2012/10/big-data-the-management-revolution. (acesso em: 13 de maio de 2018).

Monard, M. C., Baranauskas, J. A. (2003). "Indução de Regras e Árvores de Decisão", *Sistemas Inteligentes - Fundamentos e Aplicações,* Manole Ltda, Barueri, pp. 115-139.

Müller, V. N. (2013), *E-commerce: Vendas pela Internet,* FEMA, Assis.

Nakamura, A. M. (2011), *Comércio Eletrônico: Riscos nas Compras pela Internet*, FATECSP, São Paulo.

Nissan, M., (2014), "Qual a diferença entre B2B e B2C?", *E-Commerce News,* disponível em: https://ecommercenews.com.br/artigos/cases/qual-e-a-diferenca-entre-b2b-e-b2c/ (acesso em 17 de maio de 2018).

Ochi, L. S., Dias, C. R., Soares, S. S. F. (2004), "Clusterização em Mineração de Dados", *Instituto de Computação - Universidade Federal Fluminense,* Niteroi.

Pavlou, P. A. (2002), "Institution-based trust in interorganizational exchange relationships: the role of online B2B marketplaces on trust formation", *Journal of Strategic Information Systems* Vol 11, pp. 215 – 243.

Peçanha, V. (2018), "Marketing Digital", *Marketing de Conteúdo – o Blog da Rock Content,* disponível em: https://marketingdeconteudo.com/marketing-digital/ (acesso em: 13 de maio de 2018).

EJABM European Journal of Applied Business and Management
ISSN 2183-5594

Poongothai, K., Parimala, M. and Sathiyabama, S. (2011), "Efficient Web Usage Mining with Clustering", *IJCSI International Journal of Computer Science Issues*, Vol. 8, No. 6.

Rabelo, A. (2018), "Machine Learning: o que é e qual sua influência no marketing digital?" *Marketing de Conteúdo – o Blog da Rock Content*, disponível em: https://marketingdeconteudo.com/machine-learning/ (acesso em 20 de setembro de 2018).

Rauyruen, P., Miller, K. E. (2007), "Relationship quality as a predictor of B2B customer loyalty", *Journal of Business Research,* Vol. 60, pp. 21-31.

Rocha, R. A., Bortoluzzi, A. C., Zanini, M. R. K., Júnior, N. J. Z. (2002), "A Internet e a Reinvenção dos Negócios.", in *Proceedings of Encontro Nacional de Engenharia de Produção - ENEGEP,* Curitiba.

Ryan, D. (2017), *Understanding Digital Marketing,* Kogan Page, Londres.

Silva, J., Varela, N., López, L., Millán, R. (2019), "Association Rules Extraction for Customer Segmentation in the SMEs Sector Using the Apriori Algorithm", *Procedia Computer Science,* Vol. 151, pp. 1207-1212.

Soltani Z., Navimipour N. (2016), "Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research", *Computers in Human Behavior*, Vol. 61, pp. 667-688.

Souza, B., Bremgartner, V. (2016), Evolução das modalidades B2B e B2C em e-business no Brasil", in *Proceedings of Congresso de Administração do Sul de Mato Grosso – CONASUM*, Mato Grosso.

Srinivasan, S. S., Anderson, R., Anderson, K., (2002), "Customer loyalty in e-commerce: an exploration of its antecedents and consequences", *Journal of Retailing*, Vol. 76, No. 1, pp. 41-50.

Tsai, H., Huang, L. (2005), "Emerging e-commerce development model for Taiwanese travel agencies", *Tourism Management*, Vol. 26, No. 5, pp. 787-796.

Tucunduva, R. (2018), "Os 3 tipos de fluxos de automação de e-mail marketing para e-commerce", disponível em: https://ecommercenews.com.br/artigos/dicas-artigos/os-3-tipos-de-fluxos-de-automacao-de-e-mail-marketing-para-e-commerce/ (acesso em: 19 de maio de 2018).

UNCTAD – United Nations Conference of Trade and Development, (2017), "World Investment Report 2017 – Investment and the Digital Economy". UNITED NATIONS PUBLICATION, disponível em: https://unctad.org/en/PublicationsLibrary/wir2017_en.pdf. (acesso em: 17 de maio de 2018).

ISSN 2183-5594

Wind, J., Mahajan, V. (2001). *Digital Marketing: Global Strategies from World's Leading Experts*, John Wiley & Sons, Inc., Hoboken.

Witten, I. H., Frank, E., Hall, M. A., Pal, C. J. (2016), *Data Mining: Practical Machine Learning Tools and Techniques,* Morgan Kaufmann, Burlington.