

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E TRANSPORTES

TRABALHO DE CONCLUSÃO DE CURSO DE GRADUAÇÃO

**APLICAÇÃO DE TÉCNICAS DE *MACHINE LEARNING* NÃO-SUPERVISIONADAS PARA
APOIO À ÁREA DE *CUSTOMER SUCCESS* EM UMA EMPRESA DE CONSULTORIA EM
TECNOLOGIA**

RUIZ LERNER GONÇALVES

Orientador: MICHEL JOSÉ ANZANELLO

PORTO ALEGRE

01/2024

RESUMO

Diante da crescente demanda e expectativas dos clientes, especialmente no mercado *B2B (business to business)* e mais especificamente em empresas que oferecem soluções complexas, compreender as dimensões cruciais da qualidade de serviço torna-se essencial para garantir a satisfação e fidelidade de tais clientes, ressaltando a importância do *CSM (Customer Success Management)*. Nesse sentido, técnicas que possibilitem a categorização dos clientes com base em similaridades e particularidades em seus perfis oferecem *insights* valiosos para a formulação de estratégias personalizadas de atendimento a cada grupo. Assim, o presente artigo propõe a aplicação de duas técnicas de *machine learning* não-supervisionadas para melhor caracterização da base de clientes de uma consultoria em tecnologias para gestão da produção: análise de *cluster* e *PCA (Principal Component Analysis)*. O *PCA* possibilita a identificação das variáveis mais relevantes para caracterização de clientes em relação ao uso das soluções. Por outro lado, a análise de *cluster* gera agrupamentos de clientes com base em suas similaridades, abordando aspectos como qualidade percebida e nível de serviço. Quanto aos resultados, a metodologia reteve 5 componentes principais, explicando cerca de 78,74% da variância entre os clientes, e resultou na formação de 5 *clusters* com um *SI (Silhouette Index)* médio de 0,5328. Esses agrupamentos permitiram o desenvolvimento de estratégias específicas para maximizar o resultado dos clientes em cada grupo, incluindo colaboração e co-criação, desenvolvimento de parcerias estratégicas, mapeamento de necessidades e desenvolvimento de cronogramas de comunicação proativa.

PALAVRAS-CHAVE: *K-Means, Principal Component Analysis, PCA, análise de cluster, Customer Success, consultoria em tecnologia, estratégias de atendimento*

1. INTRODUÇÃO

No atual cenário empresarial, caracterizado pela busca incessante de excelência, inovação e eficiência nos serviços oferecidos, é imprescindível a integração externa e interna das atividades-chave de uma organização. No entanto, a implementação de sistemas integrados em larga escala pode apresentar desafios consideráveis para as organizações em termos de eficiência, custos e tempo (King e Burgess, 2008).

De acordo com Eggert et al. (2020), tanto a implementação quanto a manutenção de *softwares* e ferramentas de visualização integrados são tarefas altamente desafiadoras. Essa complexidade justifica-se na frequente necessidade de adaptações significativas para atender aos requisitos específicos demandados por cada cliente, levando em consideração suas necessidades individuais em relação à coleta e análise contínua de dados relevantes para a tomada de decisões.

Simultaneamente, à medida que o número de opções de empresas e tecnologias disponíveis no mercado aumenta, os clientes se tornam gradualmente mais exigentes, principalmente em relação a serviços de alto contato, em que esperam ter voz ativa durante as diversas etapas que os compõem (Yang et al., 2019). Ulaga (2018) reforça que, nos mercados *B2B (business to business)*, um grande número de clientes está ciente

do direito e da capacidade que possuem de intensificar suas exigências em relação às empresas das quais consomem.

Dessa forma, a fim de explorar novas oportunidades de crescimento, as organizações enfrentam o desafio de identificar possibilidades de aprimoramento do valor de ativos e aumento de produtividade, tanto para os clientes quanto em colaboração com eles (Ulaga, 2018). Gautam (2015) ressalta que, ao identificar e compreender as dimensões da qualidade de serviço mais importantes para os clientes, as empresas podem melhorar sua qualidade de serviço e obter uma vantagem competitiva, o que implica em identificar e priorizar as áreas que têm maior impacto na satisfação do cliente, bem como aprimorar a entrega dos serviços nessas dimensões.

Em complemento, Reinartz et al. (2004) salientam que um processo de *Customer Success* eficaz pode resultar em maior satisfação, fidelidade e retenção dos clientes, levando a um aumento na receita e lucratividade da organização. Todavia, estabelecer e manter uma área dedicada à gestão de valor e sucesso do cliente pode exigir investimentos significativos em expertise e recursos humanos, situação que, geralmente, envolve *trade-offs* com outras atividades e funções-chave (Keränen e Liozu, 2020), tornando essencial a compreensão aprofundada das situações e dos fatores que influenciam a relevância das atividades de *CSM (Customer Success Management)* para definir as estratégias, recursos e abordagens para melhor atender às expectativas e necessidades dos clientes (Kleinaltenkamp et al., 2022).

Rencher (2002) evidencia que a abordagem multivariada permite examinar a interação entre as variáveis de um sistema, possibilitando a identificação de padrões e a investigação da essência do sistema. Dessa forma, a aplicação de técnicas de *ML (machine learning)* não-supervisionadas (como *clustering* e métodos de seleção de variáveis relevantes), permite identificar padrões e tendências sobre o desempenho atual do suporte ao cliente na empresa investigada, bem como as necessidades dos clientes e possíveis áreas de melhoria.

Assim, o objetivo deste estudo é identificar as principais semelhanças e diferenças entre os clientes de uma empresa de consultoria em tecnologia por meio de análise exploratória de dados com base em técnicas de *machine learning* não-supervisionadas. Para tanto, dados contendo informações acerca de atendimentos realizados, as quais incluem avaliações dos clientes ao suporte, percentual de utilização do suporte para resolução de problemas e implementação de melhorias, dentre outros, serão utilizadas como variáveis descritivas dos clientes. Quando aplicadas a tais dados, as técnicas não-supervisionadas permitirão agrupar os clientes de acordo com seus perfis de similaridade, possibilitando o desenvolvimento de estratégias customizadas para cada *cluster* voltadas à criação de estratégias específicas com vistas ao fortalecimento da relação com os clientes, aumentando sua satisfação, fidelidade e retenção, e diminuindo o risco de *churn*. Pode-se ainda impulsionar o crescimento da empresa, maximizando a obtenção de receita recorrente e detectando oportunidades de *cross-sell* e *upsell*.

O presente artigo está estruturado conforme segue. Após esta introdução, a segunda seção apresenta embasamento teórico que aprofunda o conhecimento existente na literatura sobre o tema. A terceira seção detalha a metodologia utilizada para a realização deste estudo. Em seguida, na quarta seção, são apresentados e discutidos os resultados obtidos a partir do estudo e, por fim, na quinta seção, são resumidos os principais resultados obtidos, assim como são apresentadas as conclusões deste trabalho.

2. REFERENCIAL TEÓRICO

2.1 CUSTOMER SUCCESS MANAGEMENT (CSM)

O *CSM* representa uma mudança em relação às práticas tradicionais de gerenciamento de clientes, ao mesmo tempo que se baseia nelas. Seu foco é priorizar proativamente a experiência e o envolvimento dos clientes para alcançar o máximo valor de uso (Hilton et al., 2020).

Os principais conceitos na gestão do sucesso dos clientes são o *CRM (Customer Relationship Management)* e o *CEM (Customer Experience Management)*, conforme apontado por Tkachenko et al. (2021). Embora relacionados, eles possuem abordagens distintas. O *CRM* concentra-se na gestão das interações e relacionamentos com os clientes, enquanto o *CEM* se dedica à gestão da experiência geral do cliente. Combinar ambas as abordagens permite às empresas criarem uma estratégia abrangente para o gerenciamento de relacionamentos, impulsionando o crescimento e a lucratividade.

O conceito de *CRM* eleva a interação com o cliente ao próximo nível, utilizando informações abrangentes sobre seus clientes e necessidades. Ao estudar os clientes e atender às suas solicitações, as empresas podem encontrar novas oportunidades para vender bens e serviços, tornando-se um fator-chave no desenvolvimento sustentável e uma fonte de vantagem competitiva a longo prazo (Tkachenko et al., 2021; Sin et al., 2005).

Por outro lado, o *CEM* opera de forma diferente, pois cada interação com o cliente proporciona a eles uma compreensão mais profunda da empresa. Dependendo das informações recebidas, os clientes podem modificar seu comportamento, o que afeta sua própria lucratividade individual. Por meio do *CEM*, as empresas coletam e analisam informações sobre a dinâmica do relacionamento com os clientes, estabelecendo relacionamentos mutuamente benéficos (Tkachenko et al., 2021; Witell et al., 2020).

Prohl-Schwenke e Kleinaltenkamp (2021) destacam que a adoção do *CSM* pelos clientes varia conforme a complexidade e personalização dos produtos ou serviços oferecidos. Quanto maior a complexidade e a personalização, maior é o desejo dos clientes pelas atividades de *CSM*. Portanto, os fornecedores que oferecem produtos e/ou serviços altamente complexos e personalizados devem investir em estruturas adequadas de *CSM*.

No entanto, tanto Prohl-Schwenke e Kleinaltenkamp (2021) quanto Triznová et al. (2015) alertam que implementar medidas de *CSM* de forma seletiva nas equipes de

atendimento ao cliente ou gerenciamento de contas não é suficiente. O *CSM* deve ser aplicado em toda a organização fornecedora, pois é necessário criar uma cultura orientada ao cliente. Isso implica que os funcionários de todos os departamentos compreendam como os clientes utilizam o produto e/ou serviço, e quais são seus objetivos ao optar pelas ofertas dos fornecedores (Guerola-Navarro et al., 2021; Prohl-Schwenke e Kleinaltenkamp, 2021; Triznová et al., 2015).

Além disso, Najafi-Tavani et al. (2022) e Storey e Larbig (2018) enfatizam que gerenciar ativamente o envolvimento do cliente e assimilar seu conhecimento como uma capacidade interconectada coloca a empresa em uma posição vantajosa para criar novos serviços que agreguem valor ao cliente. Empresas com essa capacidade bem desenvolvida tendem a ser mais habilidosas em identificar oportunidades de mercado, entender profundamente seus clientes e suas necessidades, e a forma como os serviços criam valor. Assim, elas são mais capazes de converter oportunidades baseadas em necessidades latentes dos clientes em conceitos bem-sucedidos de novos serviços.

Entretanto, há um risco em envolver os clientes em excesso, uma vez que eles podem sugerir muitas maneiras de melhorar ou alterar o projeto. Isso pode resultar em um novo design de serviço completamente diferente do escopo inicial, tornando investimentos iniciais obsoletos, aumentando o custo de desenvolvimento e afetando a lucratividade. Além disso, a equipe do projeto pode resistir às contribuições dos clientes se perceber as mudanças como prejudiciais à eficiência operacional (Najafi-Tavani et al., 2022; Witell et al., 2020; Storey e Larbig, 2018).

2.2 TÉCNICAS DE *MACHINE LEARNING*

A análise multivariada abrange todas as técnicas estatísticas que analisam simultaneamente múltiplas medições em indivíduos ou objetos sob investigação. Dessa forma, qualquer análise que envolva mais de duas variáveis pode ser considerada como análise multivariada. Muitas das técnicas multivariadas são extensões da análise univariada e bivariada (Hair et al., 2010).

Johnson e Wichern (2008) destacam que estabelecer um esquema de classificação amplamente aceito que indique a adequação das técnicas é uma tarefa desafiadora. Uma possível classificação divide as técnicas em duas categorias: aquelas projetadas para estudar relacionamentos interdependentes (não supervisionadas) e aquelas projetadas para estudar relacionamentos dependentes (supervisionadas). Outra classificação se baseia no número de populações e no número de conjuntos de variáveis em estudo. Independente da classificação, as técnicas multivariadas podem ser usadas em inúmeras aplicações, como redução de dados ou simplificação estrutural, ordenação e agrupamento, investigação da dependência entre variáveis, previsão, e construção e teste de hipóteses. Na sequência, são descritos os fundamentos das técnicas multivariadas utilizadas neste estudo.

2.2.1 *PRINCIPAL COMPONENT ANALYSIS (PCA)*

PCA (Principal Component Analysis) é uma técnica estatística não-supervisionada que busca reproduzir a variabilidade total de um sistema usando um número menor de

componentes principais em relação ao número original de componentes p . Muitas vezes, uma grande parte da variabilidade pode ser explicada por um pequeno número r de componentes principais, o que significa que esses r componentes carregam praticamente a mesma informação que os p componentes originais. Dessa forma, os r componentes principais podem substituir os p componentes iniciais, reduzindo o conjunto de dados original de n medidas em p componentes para um conjunto de dados com n medidas em r componentes principais (Johnson e Wichern, 2008). Da mesma forma, Hair et al. (2010) define *PCA (Principal Component Analysis)* como uma abordagem estatística que pode ser utilizada para analisar as inter-relações entre um grande número de variáveis, explicando-as em termos de suas dimensões em comum. O objetivo desta abordagem é condensar a informação contida em um determinado número de variáveis em um conjunto menor de fatores com mínima perda de informação, facilitando a interpretação desta (Hair et al., 2010; Johnson e Wichern, 2008; Jolliffe, 2002).

Matematicamente, Rencher (2002) explica que o *PCA* visa maximizar a variância de uma combinação linear das variáveis. O primeiro componente principal é a combinação linear que possui a máxima variância, ou seja, representa a direção ao longo da qual as observações estão mais separadas ou distribuídas. O segundo componente principal é a combinação linear com a máxima variância em uma direção ortogonal à do primeiro componente principal, e assim por diante. De acordo com Jolliffe (2002), considerando uma matriz \mathbf{x} composta por n medidas de p componentes, o primeiro componente principal derivado é representado pela equação (1).

$$\alpha'_1 \mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + \dots + \alpha_{1p}x_p \quad (1)$$

Em que α'_1 = vetor transposto de p constantes $\alpha_{11}, \alpha_{12}, \dots, \alpha_{1p}$; e $\alpha'_1 \mathbf{x}$ = função linear dos elementos de \mathbf{x} que maximiza a variância.

Após encontrar o primeiro componente, busca-se uma função linear $\alpha'_2 \mathbf{x}$ não correlacionada com $\alpha'_1 \mathbf{x}$, de modo que, na “ k -ésima” etapa, seja encontrada uma função linear $\alpha'_k \mathbf{x}$ que tenha a máxima variância, sujeita a não estar correlacionada com $\alpha'_1 \mathbf{x}, \alpha'_2 \mathbf{x}, \dots, \alpha'_{k-1} \mathbf{x}$. A “ k -ésima” variável derivada é o “ k -ésimo” componente principal (Jolliffe, 2002).

Ainda, Jolliffe (2002) aponta que, para reduzir a dimensionalidade substituindo as p variáveis pelos m primeiros componentes principais, o critério mais comum é selecionar uma porcentagem cumulativa da variação total que se deseja que os componentes selecionados contribuam. Nesse sentido, o número de componentes retidos será o menor valor de m para o qual essa porcentagem é excedida. Esta condição pode ser verificada por meio de recursos visuais como o *scree graph* ou pela equação (2).

$$t_m = \frac{100}{p} \sum_{k=1}^m l_k \quad (2)$$

Em que I_k = variância do “k-ésimo” componente principal; m = número de componentes que se deseja reter; p = número original de variáveis do sistema; e t_m = variabilidade explicada pelos componentes retidos.

2.2.2 CLUSTER ANALYSIS

Cluster Analysis é uma técnica analítica não-supervisionada para geração de agrupamentos de indivíduos ou objetos. O objetivo da análise é juntar uma amostra de entidades em um pequeno número de grupos mutualmente exclusivos baseado nas similaridades entre as entidades (Hair et al., 2010; Rencher, 2002).

Hair et al. (2010), Johnson e Wichern (2008) e Rencher (2002) apontam que, para realizar o agrupamento das observações em clusters, a maior parte das técnicas utilizam medidas de similaridade entre pares de observações. Em muitos casos, essa similaridade é baseada em alguma medida de distância. Outros métodos de *cluster* utilizam uma escolha preliminar para os centros dos *clusters*, medidas de correlação ou uma comparação da variabilidade dentro e entre os *clusters*. A função de distância mais comumente utilizada é a euclidiana (3).

$$d(x, y) = \sqrt{(x - y)'(x - y)} = \sqrt{\sum_{j=1}^p (x_j - y_j)^2} \quad (3)$$

Em que $d(x, y)$ = distância entre os vetores x e y ; j = variável medida; e p = número de variáveis do sistema.

Hair et al. (2010), Johnson e Wichern (2008) e Rencher (2002) particionam as técnicas de clusterização em dois grandes tipos: hierárquicas e não hierárquicas. Procedimentos hierárquicos envolvem uma série de $n-1$ decisões de agrupamento que combinam observações em uma hierarquia ou uma estrutura em forma de árvore. Em contraste com os métodos hierárquicos, procedimentos não hierárquicos não envolvem o processo de construção em forma de árvore. Em vez disso, eles atribuem objetos a clusters assim que o número de clusters é especificado (Hair et al., 2010).

Os dois tipos básicos de procedimentos de agrupamento hierárquico são aglomerativos e divisivos. Nos métodos aglomerativos, cada objeto ou observação começa como seu próprio cluster e é sucessivamente unido aos dois clusters mais similares até restar apenas um único cluster. Nos métodos divisivos, todas as observações começam em um único cluster e são sucessivamente divididas até que cada uma seja um cluster com apenas um membro. Não há um procedimento padrão para a definição do número ótimo de *clusters* formados, fazendo com que comumente, utilize-se alguma regra de parada que sugira duas ou mais soluções que possam ser comparadas antes de uma decisão final. Uma abordagem possível seria o uso de técnicas matemáticas e/ou visuais para escolher o número de *clusters* com a máxima distância entre si, como dendrogramas (Hair et al., 2010; Johnson e Wichern, 2008; Rencher, 2002).

Rencher (2002) aponta que, dentre os métodos de agrupamento hierárquico aglomerativos, o método de Ward é um dos que possuem os melhores desempenhos

em geral. No método de Ward, a seleção de quais dois *clusters* combinar é baseada na combinação que minimiza a soma dos quadrados dentro do *cluster* em relação ao conjunto completo de *clusters* unidos ou separados. Em cada etapa, os dois *clusters* combinados são aqueles que minimizam o aumento na soma total dos quadrados em todas as variáveis em todos os *clusters* (Hair et al., 2010; Johnson e Wichern, 2008; Rencher, 2002).

A abordagem não hierárquica mais comum é a do *K-Means*. Nesta abordagem, o processo inicia-se com a seleção inicial de g itens como *seeds*, que serão substituídos, posteriormente pelos centroides dos *clusters*. Os itens restantes são atribuídos aos *clusters* com base na distância euclidiana para a *seed* mais próxima. Após essa etapa, os centroides são recalculados e os itens são movidos para outros *clusters* se estiverem mais próximos dos centroides desses *clusters*. Esse processo é repetido até que não seja possível obter mais melhorias no agrupamento (Hair et al., 2010; Johnson e Wichern, 2008; Rencher, 2002).

Hair et al. (2010), Johnson e Wichern (2008) e Rencher (2002) indicam que, para verificar a validade de uma solução de *cluster*, uma possível abordagem consiste em testar a hipótese de que não existem *clusters* ou grupos na população da qual a amostra foi obtida. Além disso, a validade ou estabilidade de um resultado de clusterização pode ser verificada por meio da aplicação de técnicas de validação cruzada. Nesse método, os dados são divididos aleatoriamente em dois subconjuntos, geralmente denominados A e B, e a análise de *cluster* é realizada separadamente em cada um desses subconjuntos. Se os *clusters* forem válidos, os resultados obtidos em A e B devem ser semelhantes.

Rousseeuw (1987) propõe o método *Silhouette* para a interpretação e validação de análises de *cluster*. O indicador *Silhouette Index (SI)* mede o quão bem um objeto i se encaixa no *cluster* em que foi inserido através da comparação entre a diferença entre a distância média do objeto em relação às outras do mesmo *cluster* e a mínima distância média em relação às observações de outros *clusters*, e o máximo entre estas duas medidas, através da equação (4).

$$SI = \frac{a_i - b_i}{\max(a_i, b_i)} \quad (4)$$

Em que a_i = distância média do objeto em relação às outras do mesmo *cluster*; e b_i = mínima distância média em relação às observações de outros *clusters*.

Em relação à subseção anterior, Jolliffe (2002) expõe que a análise de *cluster* é um dos contextos mais frequentes nos quais os componentes principais são derivados para reduzir a dimensionalidade antes do uso de uma técnica multivariada diferente. Existem duas principais formas em que os componentes principais são utilizados na análise de *cluster*: para construir medidas de distância ou para fornecer uma representação gráfica dos dados.

2.3 APLICAÇÃO DE TÉCNICAS DE MACHINE LEARNING NO CONTEXTO DE CUSTOMER SUCCESS MANAGEMENT

Małecki e Wątróbski (2017) utilizaram técnicas de *clustering* para desenvolver um sistema de apoio à decisão com o objetivo de segmentar clientes de lojas virtuais com base nos *sites* que visitavam. Os resultados revelaram diferentes grupos de clientes e forneceram informações para uma ampla variedade de aplicações práticas, como identificar consumidores dispostos a trocar qualidade por preços reduzidos.

De maneira semelhante, Hartoyo et al. (2023) realizaram um estudo para entender como mais de quinhentos consumidores eram segmentados com base na admiração por marcas de carros e as motivações por trás dessa admiração. Eles também investigaram as características de cada segmento de clientes. Os resultados permitiram concluir que a segmentação pode ser uma ferramenta poderosa para a indústria automobilística compreender os fatores que influenciam o comportamento do consumidor e identificar os grupos mais propensos a ter uma forte conexão com a marca. Além disso, após a segmentação, as empresas podem obter *insights* sobre os clientes e desenvolver estratégias direcionadas para atender às suas necessidades específicas, gerando diferenciação no mercado e lealdade à marca através da inovação na oferta de produtos e/ou serviços.

Ainda, Hosseini et al. (2022) propuseram uma abordagem bidimensional para examinar o comportamento dos clientes em canais bancários eletrônicos e tradicionais. Utilizando algoritmos de agrupamento *RFM (Recency, Frequency and Monetary Value)* e *K-Means*, analisaram os comportamentos de compra dos clientes e permitiram que diferentes bancos classificassem melhor seus clientes e compreendessem mais profundamente o comportamento de cada grupo, possibilitando a oferta de serviços personalizados, a fim de reter os clientes, assim como o planejamento de estratégias e políticas de retenção de clientes, visando a redução de custos e aumento de receitas.

Em relação à seleção de variáveis relevantes, Al-Mashraie et al. (2020) empregaram abordagens e modelos estatísticos como regressão logística, *SVM (Support Vector Machine)* e *Random Forest* para identificar os fatores que mais afetam o *churn* em empresas de telecomunicações. Os resultados auxiliaram as empresas a gerar *insights* para elaborar campanhas de marketing mais eficientes e estratégias de incentivo para reter os clientes.

Xie et al. (2023) elaboraram modelos de previsão de telemarketing bancário para aumentar a eficiência e a taxa de sucesso no processo de captura de depósitos como fonte estável de financiamento, a fim de reduzir as apreensões dos depositantes em relação à solvência e liquidez do banco, tornando-o menos vulnerável a crises financeiras. Os modelos foram elaborados utilizando os métodos de *ML Random Subspace (RS)*, *Multi-Boosting (MB)* e *Random Subspace-Multi-Boosting (RS-MB)*, sendo que a performance destes foi estressada em dois cenários: empregando as variáveis independentes originais e selecionando as variáveis independentes mais relevantes através do método *Random Forest*. Através da análise dos resultados, concluíram que os modelos alimentados apenas com variáveis relevantes tiveram resultados

excepcionalmente melhores e a classificação da importância das variáveis independentes proporcionou informações valiosas aos bancos para a implementação de um *marketing* de precisão, potencialmente reduzindo os custos e aumentando os retornos das campanhas.

Similarmente, Samunderu e Farrugia (2022) desenvolveram um modelo preditor para distinguir passageiros de companhias aéreas entre viajantes à trabalho e à lazer, a fim de proporcionar ofertas personalizadas para cada segmento de passageiros. O modelo, baseado em *K-Means* (classificador que demonstrou melhor performance), foi testado com quatro diferentes métodos de seleção de variáveis: regressão logística, árvores de decisão, *SVM* e *Random Forest*. Tais autores salientam que a inclusão e exclusão de variáveis teve um impacto significativamente maior do que o tipo de classificador usado. Sete das dez principais variáveis estão disponíveis no momento da reserva, antes mesmo que o cliente tenha comprado o voo, o que significa que, quando um passageiro está envolvido com a companhia aérea, o fornecedor pode tomar decisões em tempo real para adaptar a oferta de produtos e possivelmente o preço, de acordo com o propósito da viagem, sendo um resultado muito poderoso do ponto de vista de vendas. Além disso, a categorização do propósito da viagem abre várias outras possibilidades de experimentação que podem ser conduzidas durante o processo de venda, sendo muito útil para adaptar dinamicamente a oferta de produtos enquanto o cliente está envolvido com a companhia aérea.

A partir das consultas realizadas para o desenvolvimento da presente subseção, pode-se inferir que a literatura existente tem se concentrado principalmente na utilização de técnicas de *machine learning* supervisionadas ao se analisar dados no contexto de sucesso do cliente. Dessa forma, a adoção de métodos não-supervisionados permitirá uma exploração mais ampla e abrangente do assunto, possibilitando a identificação de padrões e tendências anteriormente não detectados, e enriquecendo o campo de *CSM* com novas perspectivas analíticas.

3. PROCEDIMENTOS METODOLÓGICOS

3.1 DESCRIÇÃO DO CENÁRIO

A empresa estudada nesse artigo, fundada no ano de 2013, é uma consultoria especializada em tecnologia para gestão da produção que conta com mais de trinta funcionários distribuídos entre três sedes, sendo duas localizadas no Brasil, nas cidades de Porto Alegre e São Paulo, e uma em Portugal, na cidade de Lisboa. Em parceria com uma empresa multinacional de tecnologia, a consultoria implementa três principais *softwares*: *Advanced Scheduling (AS)*, que permite a programação da produção de fábricas em capacidade finita, levando em consideração a disponibilidade de recursos, o sincronismo entre operações e demais restrições; *Advanced Planning (AP)*, que gera os planos de produção de médio e longo prazo com base nas políticas de estoques, restrições produtivas, demandas futuras e regras de negócio específicas; e *Distribution Requirements Planning (DRP)*, que realiza o planejamento de transferências entre fábricas e centros de distribuição respeitando as restrições logísticas, o limite de carga dos veículos, a capacidade de armazenagem e diferentes estratégias de abastecimento.

As três principais atividades da empresa são: (i) venda e manutenção de licenças dos *softwares*; (ii) execução de projetos de implementação dos *softwares*; e (iii) atuação no suporte e acompanhamento do uso dos *softwares*. No suporte, atuam tanto a equipe de Operações, monitorando o desempenho das soluções e oferecendo suporte técnico contínuo, quanto a de CS, trabalhando em estreita colaboração com os clientes para entender suas experiências, necessidades, sugestões, metas e desafios para garantir que estes recebam o máximo valor possível. Dessa forma, o presente artigo concentra-se na atuação da equipe de CS em relação aos clientes de suporte.

3.2 CLASSIFICAÇÃO DA PESQUISA

De acordo com o método de classificação de pesquisa proposto por Gil (2010), em relação à natureza, o presente estudo assume o caráter aplicado, concentrando-se na produção de conhecimentos práticos relativos a um problema específico. No que tange à abordagem, assume uma natureza quantitativa, uma vez que é caracterizado pela análise de variáveis para a realização da análise de *cluster*. Em relação aos objetivos, assume a classificação de pesquisa explicativa, uma vez que se baseia na formação de grupos para determinação de estratégias mais apropriadas a cada conjunto. Quanto ao procedimento, assume a classificação de pesquisa-ação, pois incorpora a colaboração dos profissionais da empresa durante o processo de pesquisa.

3.3 ABORDAGEM PROPOSTA

O desenvolvimento de estratégias customizadas de gerenciamento do sucesso de uma gama variada de clientes se apoia no agrupamento destes com base em suas similaridades. Para isto, a abordagem proposta baseia-se em seis etapas: (i) adoção de premissas para realização da análise; (ii) extração e tratamento dos dados; (iii) definição do número de *clusters* a serem formados; (iv) aplicação do *PCA* para o ranqueamento da relevância das variáveis; (v) formação dos *clusters*, e (vi) análise e discussão dos resultados para a formação de planos de ação específicos para cada grupo.

Com apoio dos profissionais da equipe de *CSM* da empresa investigada, adotou-se a premissa de que seriam analisados somente os dados dos últimos seis meses de suporte de clientes com suporte ativo por pelo menos doze meses. Isto se dá pelo fato de que deseja-se entender apenas a situação recente dos clientes e, de maneira geral, os primeiros seis meses de suporte após a finalização do projeto são mais focados na correção de problemas do que no desenvolvimento de melhorias, podendo distorcer os resultados da análise.

Em seguida, os dados utilizados na análise foram extraídos de dois módulos do sistema de gestão empresarial da consultoria investigada, sendo estes o módulo de resumo dos chamados de suporte e o de pesquisa de satisfação do suporte. O primeiro módulo refere-se a um resumo mensal dos chamados de suporte dos clientes, contendo 9 variáveis quantitativas relacionadas à natureza (melhoria; solicitação; problema e problema emergencial) e indicadores de atendimento dos chamados. O segundo módulo refere-se à pesquisa de satisfação do suporte respondida semestralmente pelos clientes, contendo 6 variáveis quantitativas e 3 variáveis qualitativas. Após a coleta, realizaram-se transformações na base de dados com propósito de identificar e tratar

eventuais observações faltantes; também se normalizou as variáveis, tendo em vista que a magnitude das mesmas poderia afetar a consistência dos clusters no momento do cálculo da Distância Euclidiana.

Por meio da aplicação de técnicas hierárquicas de agrupamento, juntamente com a contribuição dos especialistas da empresa sob estudo, foi definido o número de *clusters* a serem formados. Essa determinação desempenha um papel significativo na análise, uma vez que o número de grupos estipulado tem uma influência direta tanto nas estratégias de gestão elaboradas quanto na organização dos funcionários encarregados da administração dos clientes agrupados. A metodologia empregada incluiu a utilização de um dendrograma, recurso gráfico que permite a visualização da evolução na formação dos agrupamentos, bem como a detecção de *outliers* (valores discrepantes capazes de impactar a consistência dos grupos formados). O dendrograma foi avaliado juntamente aos membros da equipe de *CSM* da empresa, os quais propuseram um intervalo adequado de número de *clusters* a serem formados com vistas à análise subsequente e desenvolvimento de estratégias.

Na sequência, utilizou-se o *PCA* para ranquear as variáveis com base na sua importância na explicação da variabilidade dos dados. Para tanto, gerou-se um índice de relevância da variável i , denominado P_i , o qual resulta da interseção entre o peso atribuído a essa variável no componente principal e a variância explicada por esse mesmo componente através da equação (5).

$$P_i = (\lambda t_1 |w_{i1}|) + (\lambda t_2 |w_{i2}|) + \dots + (\lambda t_p |w_{ip}|) \quad (5)$$

Em que P_i = peso ponderado da variável i ; λt_p = variância do componente principal p e w_{ip} = peso da variável original i na formação do componente principal p . Valores elevados de P_i denotam variáveis mais relevantes.

Com o número de *clusters* definido e as variáveis ordenadas pelo seu índice de relevância, realizou-se o processo iterativo de clusterização através da técnica não hierárquica *K-Means* e da remoção das variáveis menos relevantes. Tal processo consiste na geração sucessiva de agrupamentos, armazenando o *Silhouette Index* médio e removendo a variável com menor índice de relevância a cada iteração até restar apenas uma variável ou um determinado número mínimo de variáveis necessárias para manter um nível razoável de detalhamento do conjunto de dados. O conjunto de variáveis tido como mais relevante foi aquele responsável pelo maior *SI* médio.

Por fim, os *clusters* gerados foram analisados qualitativamente junto aos profissionais da empresa, identificando as similaridades dos clientes com apoio no ranqueamento dos componentes principais retidos e na reinserção das variáveis qualitativas na análise, possibilitando o esboço do perfil médio dos clientes alocados em cada um dos agrupamentos. A partir disso, foi possível desenvolver planos personalizados de sucesso e monitoramento de resultados baseados nos desafios e objetivos específicos dos clientes de cada segmento, definindo metas, marcos e métricas de sucesso distintas e priorizando a alocação de recursos e esforços com base nas demandas e no potencial de cada grupo.

4. RESULTADOS E DISCUSSÃO

O banco de dados utilizado no estudo continha 35 observações (clientes) e 16 variáveis, descritas na Tabela 1.

Módulo	Sigla	Unidade de Medida	Variável
Descritivo	D1	-	Cliente
Pesquisa	P1	Escala de 1 a 5	Frequência de contato com a equipe de suporte
Pesquisa	P2	Escala de 1 a 5	Prontidão da equipe de suporte
Pesquisa	P3	Escala de 1 a 5	Qualidade da equipe de suporte
Pesquisa	P4	Sim ou Não	Problemas significativos com a equipe de suporte
Pesquisa	P5	Escala de 1 a 5	Uso do suporte para desenvolvimento de melhorias
Pesquisa	P6	Escala de 1 a 10	Probabilidade de recomendação da consultoria
Resumo	R1	Chamados	Chamados abertos
Resumo	R2	%	Chamados encerrados
Resumo	R3	%	Chamados sob garantia
Resumo	R4	%	Chamados de solicitação
Resumo	R5	%	Chamados de melhoria
Resumo	R6	%	Chamados de problema
Resumo	R7	%	Chamados de problema emergencial
Resumo	R8	%	Chamados respondidos no prazo
Resumo	R9	%	Chamados solucionados no prazo
Resumo	R10	%	Chamados avaliados

Tabela 1: Variáveis utilizadas na geração dos *clusters*

A fim de definir o número ideal de *clusters* a serem utilizados nas técnicas de agrupamento não hierárquicas, gerou-se um dendrograma através do método de Ward. O dendrograma da Figura 1, no qual o eixo x representa as observações e o eixo y representa a distância de Ward entre os grupos, recomenda a formação de 5 *clusters*. Esse número é compatível com o desejado pela empresa, que havia estimado um número ótimo entre 3 e 5 *clusters* tendo em mente um conjunto condizente de estratégias para cada cluster.

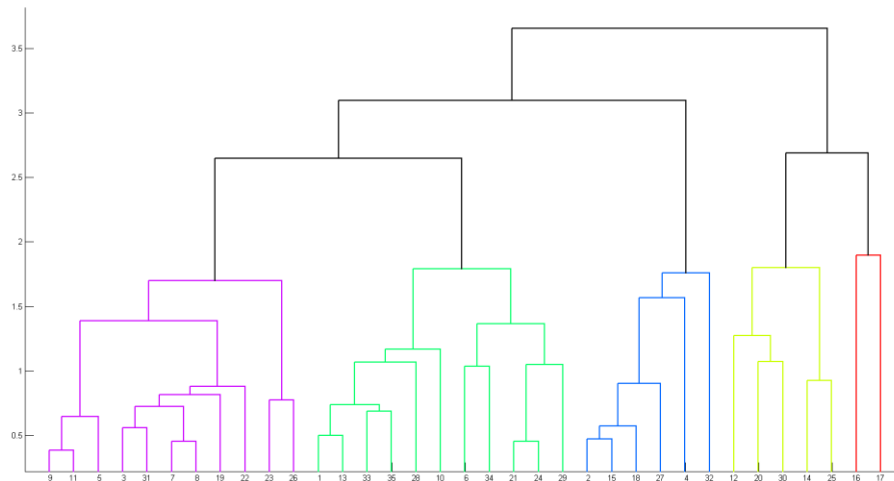


Figura 1: Dendrograma gerado pelo método de Ward

Após determinar uma faixa para o número ideal de *clusters*, realizou-se a análise de componentes principais. Na Figura 2, pode-se observar o percentual da variância explicada por cada componente principal, nas colunas, assim como a variância explicada acumulada, na curva, sendo necessário reter, por exemplo, 5 componentes principais para explicar cerca de 78,74% da variância total dos dados.

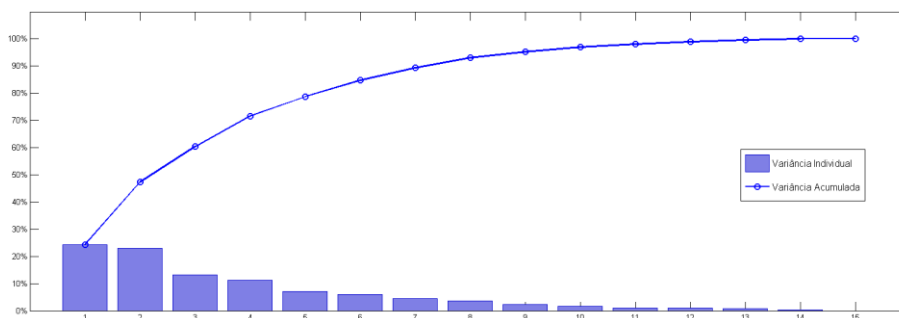


Figura 2: Gráfico do percentual da variância explicada por cada componente principal e acumulada

Com a análise de componentes principais realizada, foi possível realizar o ranqueamento das variáveis com base na sua importância na explicação da variabilidade dos dados conforme a equação (5). Na Tabela 2, pode-se observar as variáveis originais, ordenadas da mais relevante (maior valor de P_i) para a menos relevante (menor valor de P_i).

Índice	Variável	Sigla	P_i
1	Frequência de contato com a equipe de suporte	P1	0,3466
2	Problemas significativos com a equipe de suporte	P4	0,3163
3	Expectativa de aumento do uso do suporte para desenvolvimento de melhorias	P5	0,3129
4	Qualidade da equipe de suporte	P3	0,2606

5	Chamados de melhoria	R5	0,2559
6	Chamados respondidos dentro do prazo	R8	0,2527
7	Chamados solucionados dentro do prazo	R9	0,2426
8	Prontidão da equipe de suporte	P2	0,2410
9	Chamados de solicitação	R4	0,2194
10	Chamados abertos	R1	0,2067
11	Probabilidade de recomendação da consultoria	P6	0,1900
12	Chamados encerrados	R2	0,1421
13	Chamados de problema	R6	0,1011
14	Chamados sob garantia	R3	0,0841
15	Chamados de problema emergencial	R7	0,0511

Tabela 2: *Ranking* das variáveis por peso ponderado

Ao analisar o ranking P_i , percebeu-se a formação de blocos de 2 a 3 variáveis similares. As variáveis P1 (Frequência de contato com a equipe de suporte) e P4 (Problemas significativos com a equipe de suporte), ao liderar o primeiro bloco, indicam que a variabilidade observada nos clientes está fortemente influenciada pela natureza das interações diretas. A frequência de contato e a ocorrência de problemas significativos são fatores-chave que contribuem para a diversidade nas experiências dos clientes, sugerindo que a forma como os clientes interagem com a equipe de suporte tem um papel significativo na variabilidade do conjunto de dados.

O segundo bloco também se destacou como um impulsionador substancial da variabilidade. A expectativa do aumento do uso do suporte para desenvolver melhorias (P5), percepção da qualidade da equipe de suporte (P3), e solicitação de chamados específicos de melhoria (R5) contribuem para diferentes trajetórias para cada cliente. Essas variáveis refletem a diversidade de abordagens dos clientes em relação ao suporte, desde os que buscam soluções inovadoras até aqueles que priorizam a qualidade percebida.

O terceiro bloco, centrado no nível de serviço, adiciona uma camada de complexidade à variabilidade dos clientes, mesmo com um menor *ranking* de importância. A prontidão da equipe de suporte (P2), juntamente com chamados respondidos e solucionados dentro do prazo (R8 e R9), introduz diferentes dimensões na experiência do cliente. A rapidez e eficiência na entrega do serviço contribuem para a variabilidade, evidenciando como diferentes clientes podem ter expectativas distintas em relação ao tempo de resposta e resolução de requerimentos.

Por fim, o quarto bloco foca em problemas específicos relacionados à solução e funcionalidades, adicionando nuances à análise. As variáveis R6 (Chamados de Problema), R3 (Chamados Sob Garantia) e R7 (Chamados de Problema Emergencial) sugerem que a variabilidade entre clientes relacionada aos desafios enfrentados, seja em questões emergenciais, problemas recorrentes ou funcionalidades fora do escopo inicial, não impactam tanto na explicação da variabilidade total quanto as demais variáveis.

Em síntese, a análise dos blocos indicou que a variabilidade nos clientes está intrinsicamente ligada às suas interações, abordagens de uso do suporte e percepções de qualidade, e mais superficialmente relacionada às expectativas de nível de serviço e desafios enfrentados. Cada bloco contribui de maneira única para a diversidade observada no conjunto de dados, fornecendo uma visão abrangente das diferentes dimensões que moldam a experiência dos clientes da consultoria.

Além da análise da relevância das variáveis na explicação da variabilidade do sistema, também se avaliou a relevância dessas variáveis na formação de agrupamentos. Esta avaliação foi realizada através do processo de 11 iterações de clusterização não-hierárquica pelo método *K-Means*, em que em cada iteração foi removida uma variável, de acordo com o *ranking* estabelecido na Tabela 2, até que restassem pelo menos as 5 variáveis mais relevantes, a fim de melhor controlar a quantidade de informação retida no conjunto de dados. Para cada iteração, o valor do *Silhouette Index* médio do agrupamento foi computado, a fim de identificar o agrupamento e conjunto de variáveis mais informativas, conforme apresentado na Tabela 3.

Variáveis Retidas	Variável Removida	Sigla	P_i	$k=5$	$k=4$	$k=3$
				\overline{SI}	\overline{SI}	\overline{SI}
15	-	-	-	0,3200	0,3445	0,3121
14	Chamados de problema emergencial	R7	0,0511	0,3226	0,3469	0,3138
13	Chamados sob garantia	R3	0,0841	0,3315	0,3535	0,3191
12	Chamados de problema	R6	0,1011	0,3409	0,3623	0,3263
11	Chamados encerrados	R2	0,1421	0,3618	0,3703	0,3290
10	Probabilidade de recomendação da consultoria	P6	0,1900	0,3890	0,3539	0,3575
9	Chamados abertos	R1	0,2067	0,4189	0,3713	0,3734
8	Chamados de solicitação	R4	0,2194	0,4253	0,3781	0,3904
7	Prontidão da equipe de suporte	P2	0,2410	0,4343	0,4193	0,4351
6	Chamados solucionados dentro do prazo	R9	0,2426	0,4548	0,4301	0,4264
5	Chamados respondidos dentro do prazo	R8	0,2527	0,5328	0,5022	0,4878

Tabela 3: *Silhouette Index* médio dos *clusters* e variáveis removidas por iteração

Ao analisar a Tabela 3, foi possível observar que o agrupamento gerado utilizando 5 como o número pré-definido de *clusters* e as 5 variáveis melhor posicionadas no *ranking* foi o que apresentou o melhor índice *SI* médio (0,5328). Os agrupamentos gerados, o número de observações em cada cluster e a média das observações referentes a cada banco de dados podem ser vistos na Tabela 4.

Cluster	Clientes	P1	P4	P5	P3	R5
A	9	1	0	0,33	0,89	0,30
B	5	0,80	0	0,60	1	0,92
C	6	0,42	1	0,50	0,67	0,55
D	9	0,61	0	1	1	0,30

E	6	0,50	0	1	0,50	0,16
---	---	------	---	---	------	------

Tabela 4: *Clusters* gerados e nota média por variável

Ao analisar o *cluster* A, percebeu-se que clientes inseridos neste *cluster* estão totalmente satisfeitos (1) com a frequência de contato (P1) e não relataram problemas significativos com a equipe de suporte (P4) (0). Isso sugere uma relação de suporte sólida e uma comunicação constante, resultando em uma qualidade percebida da equipe de suporte (P3) alta (0,89). No entanto, a expectativa de utilização do suporte para desenvolvimento de melhorias (P5) e a demanda por chamados de melhoria (R5) são relativamente baixas (0,33 e 0,30, respectivamente), indicando uma possível oportunidade de explorar mais o potencial de melhoria e inovação com esses clientes. Sessões personalizadas de mapeamento de necessidades podem ser implementadas, aproveitando a relação sólida existente e não apenas capitalizando a confiança estabelecida, mas também estimulando uma participação mais ativa nos processos de desenvolvimento.

Os clientes do *cluster* B também mantêm uma frequência de contato (P1) elevada (0,80) e não relatam problemas significativos com a equipe de suporte (P4) (0). A expectativa de aumento da utilização do suporte para desenvolvimento de melhorias (P5) é significativa (0,60) e a demanda por chamados de melhoria (R5) é alta (0,92), indicando que esses clientes estão ativamente envolvidos em melhorias. A qualidade percebida da equipe de suporte (P3) é máxima (1). Esse cluster revela uma relação positiva, em que os clientes estão não apenas satisfeitos, mas também engajados no desenvolvimento contínuo, indicando uma parceria estratégica.

Dessa forma, a implementação de estratégias centradas em desenvolvimento de *cases* e realização de *benchmarkings* e de eventos como *webinars*, *workshops* e conferências pode elevar ainda mais o envolvimento e beneficiar não apenas esses clientes, mas toda a base de usuários. Enquanto os *cases* destacam as conquistas dos clientes do *cluster* B, inspirando outros clientes e promovendo a adoção de boas práticas, a realização de eventos e *benchmarkings* oferece uma plataforma para o compartilhamento direto de experiências, fortalecendo a comunidade e fomentando a colaboração entre os clientes.

O *cluster* C apresenta a menor frequência de contato (P1) (0,42) entre todos os clusters, sugerindo uma comunicação mais esporádica. Além disso, esses clientes foram os únicos a relatar problemas significativos com a equipe de suporte (P4) (1), indicando desafios na relação. A expectativa de utilização do suporte para desenvolvimento de melhorias (P5) e a demanda por chamados de melhoria (R5) é moderada (0,50 e 0,55, respectivamente). A qualidade percebida da equipe de suporte (P3) é razoável (0,67). Neste cluster, a estratégia principal é melhorar a comunicação e resolver problemas percebidos. Sessões regulares de feedback estruturado e um cronograma regular de comunicação proativa, oferecendo atualizações, dicas e suporte técnico personalizado oferecem um canal para abordar preocupações específicas de clientes dentro deste grupo.

Clientes no *cluster* D mantêm uma frequência moderada de contato (P1) (0,61) e não relatam problemas significativos com a equipe de suporte (P4) (0). A expectativa de uso do suporte para desenvolvimento de melhorias (P5) é máxima (1), indicando que esses clientes têm altas expectativas de utilizar o suporte para inovação. A qualidade percebida da equipe de suporte (P3) é alta (1), evidenciando uma relação sólida. No entanto, a demanda por chamados de melhoria (R5) é baixa (0,30), indicando uma desconexão entre as expectativas e a demanda real por melhorias. Esse *cluster* destaca uma oportunidade para a consultoria colaborar de forma mais eficaz com esses clientes, transformando suas altas expectativas em iniciativas concretas de desenvolvimento. Programas de co-criação e *beta testing* são formas eficazes de envolver ativamente esses clientes, transformando suas expectativas em ações tangíveis e promovendo uma participação mais ativa em melhorias.

De forma similar ao *cluster* D, os clientes no *cluster* E também evidenciam um entusiasmo para utilizar o suporte como ferramenta para a melhoria contínua (P5) (1). No entanto, este cluster se destaca pela manutenção de uma frequência de contato (P1) moderada (0,50) e pela menor percepção da qualidade do suporte (P3) (0,50) entre todos os clientes da amostra, sugerindo possíveis barreiras na entrega do serviço. A nota moderada para frequência de contato e a menor nota para qualidade percebida podem ter impacto direto no uso atual do suporte para desenvolvimento de melhorias (R5) (0,16), que também figura como a menor nota dentre todos os clientes da amostra. Essa dinâmica ressalta a necessidade de estratégias focalizadas em aprimorar a qualidade percebida e em promover uma comunicação mais eficaz para alinhar as expectativas dos clientes do *cluster* E com a oferta efetiva de suporte. Assim, a estratégia inicial seria similar à do *cluster* C, transformando estes clientes em algo similar ao que é hoje o *cluster* D, possibilitando a replicação das estratégias específicas.

5. CONCLUSÃO

O estudo proposto teve como objetivo identificar semelhanças e diferenças entre os clientes de uma consultoria em tecnologia por meio de análise exploratória de dados, utilizando técnicas de *machine learning* não-supervisionadas. Com uma base de dados contendo 35 observações (clientes) e 16 variáveis, foram empregadas técnicas de agrupamento e análise de componentes principais para formar *clusters* de clientes com características semelhantes.

Inicialmente, o método de Ward foi utilizado para gerar um dendrograma, indicando a formação ideal de 5 clusters, valor esse corroborado pela empresa. Em seguida, a análise de componentes principais revelou que reter 5 componentes seria apropriado, explicando cerca de 78,74% da variância total dos dados.

A análise das variáveis indicou quatro blocos de importância, destacando a influência significativa da natureza das interações, abordagens de uso do suporte, percepções de qualidade e nível de serviço na variabilidade dos clientes. Cada bloco contribuiu de maneira única para a diversidade observada no conjunto de dados, oferecendo uma visão abrangente das diferentes dimensões que moldam a experiência dos clientes da consultoria.

Ao utilizar a técnica *K-Means* e o *Silhouette Index* para avaliar a formação de agrupamentos, foi identificado que a combinação de 5 variáveis mais relevantes resultou no agrupamento mais informativo, com um *SI* médio de 0,5328. Isso permitiu a formação de cinco clusters (A, B, C, D, E) com características distintas, revelando nuances nas interações, expectativas e demandas dos clientes.

A análise dos clusters forneceu subsídios claros para a formulação de estratégias personalizadas. O *cluster A*, por exemplo, sugere uma relação sólida, mas com oportunidades para explorar mais o potencial de melhoria e inovação. Já o *cluster B* representa clientes altamente envolvidos em melhorias contínuas, indicando parcerias estratégicas. Por outro lado, o *cluster C* evidencia a necessidade de aprimorar a comunicação e resolver desafios percebidos, o *cluster D* revela clientes com expectativas máximas e uso baixo do suporte em melhorias, indicado estratégias de colaboração, e o *cluster E* sugere estratégias para aprimorar a qualidade percebida e alinhar expectativas, elevando-os ao patamar do que hoje é o *cluster D*.

Em síntese, os resultados obtidos fornecem uma base sólida para a consultoria tomar decisões sobre estratégias específicas para cada *cluster* de clientes, visando o fortalecimento da relação, aumento da satisfação e retenção, além da identificação de oportunidades de inovação e crescimento. Essas conclusões podem orientar ações direcionadas e eficazes, maximizando o valor entregue aos clientes e impulsionando o sucesso organizacional.

A fim de aprimorar a pesquisa, recomenda-se a aplicação de técnicas mais robustas para a análise de agrupamentos em conjuntos de dados em que a subjetividade desempenha um papel significativo, como é o caso de uma pesquisa de satisfação. O algoritmo *Fuzzy C-Means*, por exemplo, representa a similaridade que um ponto compartilha com cada *cluster* por meio de uma função de pertinência, permitindo uma abordagem mais flexível na atribuição de pontos a *clusters* e, por consequência, uma representação mais realista do agrupamento dos dados.

BIBLIOGRAFIA

Al-Mashraie, M., Chung, S.H., & Jeon, H.W. (2020). Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach. *Computers & Industrial Engineering*, 144, 106476.

Bezdek, J.C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10, 191-203.

Eggert, A., Ulaga, W., & Gehring, A. (2020). Managing Customer Success in Business Markets: Conceptual Foundation and Practical Application. *Journal of Software Maintenance and Evolution: Research and Practice*, 4, 121-132.

Gautam, V. (2015). Service Quality Perceptions of Customers About Mobile Telecommunication Services: A Case of India. *Journal of Global Marketing*, 28, 19 - 31.

Gil, A.C. (2017). *Como Elaborar Projetos de Pesquisa*, Sexta Edição. São Paulo: Atlas.

Guerola-Navarro, V., Oltra-Badenes, R., Gil-Gómez, H., & Fernández, A.I. (2021). Customer relationship management (CRM) and Innovation: A qualitative comparative analysis (QCA) in the search for improvements on the firm performance in winery sector. *Technological Forecasting and Social Change*, 169, 120838.

Hair, J., Black, W., Babin, B., & Anderson, R. E. (2010). *Multivariate Data Analysis*, Seventh Edition. London: Pearson.

Hartoyo, H., Manalu, E.P., Sumarwan, U., & Nurhayati, P. (2023). Driving Success: A Segmentation of Customer Admiration in Automotive Industry. *Journal of Open Innovation: Technology, Market, and Complexity*, 9, 100031.

Hilton, B., Hajihashemi, B., Henderson, C.M., & Palmatier, R.W. (2020). Customer Success Management: The next evolution in customer management practice? *Industrial Marketing Management*, 90, 360 - 369.

Hosseini, M.S., Abdolvand, N., & Harandi, S.R. (2022). Two-dimensional analysis of customer behavior in traditional and electronic banking. *Digital Business*, 2, 100030.

Johnson, R.A., & Wichern, D.W. (2008). *Applied Multivariate Statistical Analysis*, Sixth Edition. Upper Saddle River: Pearson.

Jolliffe, I.T. (2002). *Principal Component Analysis*, Second Edition. New York: Springer.

Keränen, J., & Liozu, S.M. (2020). Value champions in business markets: Four role configurations. *Industrial Marketing Management*, 85, 84-96.

King, S.F., & Burgess, T.F. (2008). Understanding success and failure in customer relationship management. *Industrial Marketing Management*, 37, 421-431.

Kleinaltenkamp, M., Prohl-Schwenke, K., & Keränen, J. (2022). What drives the implementation of customer success management? Antecedents of customer success management from suppliers' and customers' perspectives. *Industrial Marketing Management*, 102, 338-350.

Małeckki, K., & Wątróbski, J. (2017). The Classification of Internet Shop Customers based on the Cluster Analysis and Graph Cellular Automata. *Procedia Computer Science*, 112, 2280-2289.

Najafi-Tavani, S., Zaefarian, G., Robson, M.J., Naudé, P., & Abbasi, F. (2022). When customer involvement hinders/promotes product innovation performance: The concurrent effect of relationship quality and role ambiguity. *Journal of Business Research*, 145, 130-143.

Prohl-Schwenke, K., & Kleinaltenkamp, M. (2021). How business customers judge customer success management. *Industrial Marketing Management*, 96, 197-212.

Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The Customer Relationship Management Process: Its Measurement and Impact on Performance. *Journal of Marketing Research*, 41, 293-305.

Rencher, A.C. (2002). *Methods of Multivariate Analysis*, Second Edition. New York: Wiley.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

Samunderu, E., & Farrugia, M. (2022). Predicting customer purpose of travel in a low-cost travel environment – A Machine Learning Approach. *Machine Learning with Applications*, 9, 100379.

Sin, L., Tse, A., & Yim, F. (2005). CRM: conceptualization and scale development. *European Journal of Marketing*, 39, 1264-1290.

Storey, C., & Larbig, C. (2018). Absorbing Customer Knowledge. *Journal of Service Research*, 21, 101 - 118.

Tkachenko, K. I., Faizullin, R. V., Zolkin, A. L., & Sychanina, S. N. (2021). Customer Relationship Management In The Modern Economy. *European Proceedings of Social and Behavioural Sciences*, 116, 1711-1718.

Triznová, M., Maťová, H., Dvořáček, J., & Sadek, S. (2015). Customer Relationship Management Based on Employees and Corporate Culture. *Procedia Economics and Finance*, 26, 953-959.

Uлага, W. (2018). The journey towards customer centricity and service growth in B2B: a commentary and research directions. *AMS Review*, 8, 80-83.

Xie, C., Zhang, J., Zhu, Y., Xiong, B., & Wang, G. (2023). How to improve the success of bank telemarketing? Prediction and interpretability analysis based on machine learning. *Computers & Industrial Engineering*, 175, 108874.

Witell, L., Kowalkowski, C., Perks, H., Raddats, C., Schwabe, M., Benedettini, O., & Burton, J. (2020). Characterizing customer experience management in business markets. *Journal of Business Research*, 116, 420-430.

Yang, A.J., Huang, Y., & Chen, Y.J. (2019). The importance of customer participation for high-contact services: evidence from a real estate agency. *Total Quality Management & Business Excellence*, 30, 831 - 847.