

METODOLOGIA DE SEGMENTAÇÃO DE CLIENTES B2B ORIENTADA A LUCRATIVIDADE E OS EFEITOS NO LIFETIME VALUE (LTV)

Resumo: As organizações, atuando em mercados cada vez mais saturados, enfrentam constantemente o desafio de fidelizar seus clientes e gerenciar um processo de venda complexo para a progressão do negócio. Nesse contexto, buscam impedir que seus clientes existentes se tornem inativos, enquanto orientam a priorização de novas empresas a serem prospectadas. Esta pesquisa visa realizar um estudo analítico sobre a importância da metodologia de segmentação de clientes orientada à lucratividade e seus efeitos no Lifetime Value (LTV), tratando da aplicação da descoberta de conhecimento em bancos de dados (KDD). Como limitação, o estudo contempla o cenário de empresas do segmento financeiro nacional, atuando no modelo B2B. Por meio dos resultados obtidos, verifica-se que a metodologia segmentada de clientes orientada ao LTV contribui para as inferências sobre investimentos estruturais, na formação das metas de conversão dos negócios e na definição de projetos comerciais. Diante disso, este trabalho propõe uma metodologia de segmentação de clientes baseada em variáveis de lucratividade, com o uso de algoritmos de aprendizado de máquina, visando aumentar a assertividade da priorização comercial e apoiar a tomada de decisão estratégica com base em dados.

Palavras-Chave: Estratégia Comercial; Inteligência de Mercado; Lucratividade e/ou Indicadores de Lucratividade; Cluster e Negócios; Lifetime Value.

1. Introdução

No atual cenário corporativo, empresas inseridas em mercados saturados enfrentam desafios significativos para manter sua base de clientes ativa e, simultaneamente, expandir suas operações comerciais em um ambiente altamente competitivo. Com a intensificação da concorrência e a crescente exigência dos consumidores, torna-se cada vez mais difícil preservar o engajamento e a lealdade do público-alvo apenas com abordagens tradicionais. Nesse contexto, destaca-se a necessidade de estratégias bem estruturadas que sustentem o relacionamento com os clientes já conquistados, evitando sua inatividade e possível evasão. Para tanto, muitas organizações alocam recursos significativos em seus orçamentos, não apenas para viabilizar campanhas de fidelização, mas também para garantir o cumprimento de metas comerciais e objetivos estratégicos.

É nesse ambiente desafiador que emergem metodologias e técnicas orientadas por dados, com o propósito de auxiliar gestores na identificação das reais necessidades de seus clientes e *prospects* — termo que se refere a potenciais clientes que ainda não realizaram uma compra, mas que apresentam perfil compatível com o público-alvo da organização. Com o auxílio dessas ferramentas, os gestores podem descobrir padrões comportamentais e desenhar ofertas personalizadas que atendam de maneira simultânea às expectativas dos clientes e às metas de rentabilidade das empresas. Entre as ferramentas que vêm se destacando, merece ênfase o LTV (*Lifetime Value*), ou valor do tempo de vida do cliente. Trata-se de uma métrica essencial no monitoramento da rentabilidade de cada cliente ao longo de seu relacionamento com a empresa. Segundo Olnén (2022), o LTV representa o montante total que um cliente pode gerar em receita, sendo especialmente útil para mensurar o sucesso das estratégias de retenção e para orientar investimentos em ações comerciais específicas. Wu et al. (2023) complementam que a aplicação do LTV tem impactos diretos na ampliação da margem de lucro, pois orienta a criação de ofertas mais assertivas e ações proativas de relacionamento.

A proposta metodológica desta pesquisa foi concebida com base em variáveis escolhidas de forma tanto heurística quanto estratégica, levando em consideração as particularidades do produto ou serviço de interesse. O objetivo central é demonstrar a relevância da segmentação inteligente de *leads*, priorizando estrategicamente as ações de prospecção e o gerenciamento da área comercial, com foco na maximização da rentabilidade. Para isso, propõe-se a integração entre o processo de Descoberta de Conhecimento em Bancos de Dados (KDD - *Knowledge Discovery in Databases*) e os sistemas de recomendação. O KDD, ao explorar grandes volumes de dados e extrair padrões relevantes, permite a construção de *insights* mais profundos e personalizados, o que eleva significativamente a qualidade das decisões comerciais. Conforme Fayyad et al. (1996), o KDD assume papel estratégico ao transformar dados brutos em conhecimento aplicável, permitindo que decisões importantes sejam tomadas com base em informações robustas e estruturadas.

O processo de tomada de decisão do planejamento comercial baseia-se em duas etapas interdependentes: uma tática e outra estratégica. Na etapa tática, predomina uma abordagem analítica e racional, baseada em dados concretos e indicadores de desempenho. Na etapa estratégica, por sua vez, o foco se desloca para uma visão mais holística, onde gestores de alto escalão contribuem com interpretações e julgamentos subjetivos baseados em sua vivência e conhecimento do setor. A presente pesquisa está relacionada, principalmente, à etapa tática, e visa gerar uma metodologia que define quais *leads* e clientes, do universo mapeado, necessitam

de priorização de ações de relacionamento ou prospecção. Previsões fundamentadas em indicadores de lucratividade, como o LTV, não apenas ajudam no planejamento financeiro, mas também contribuem para melhores decisões de marketing e orientam o gerenciamento de relacionamento com o cliente (CRM).

O objetivo geral deste trabalho é desenvolver uma metodologia de segmentação de clientes que permita, de forma sistematizada e baseada em dados, a identificação de características específicas e recorrentes de cada grupo formado a partir de variáveis comerciais relevantes. A proposta visa estruturar a definição de segmentos com base em critérios quantitativos e qualitativos, considerando aspectos operacionais, financeiros e comportamentais dos clientes. Além disso, objetiva-se incorporar à metodologia variáveis diretamente relacionadas à lucratividade, como o Faturamento e o Custo de Aquisição de Clientes (CAC), de modo a gerar inferências consistentes e aplicáveis ao indicador LTV.

Para alcançar tal propósito, serão explorados os seguintes objetivos específicos: identificar, a partir da literatura, os principais critérios utilizados na segmentação de clientes no ambiente B2B; analisar o conceito de LTV e sua utilização como métrica orientadora em estratégias de segmentação; investigar modelos teóricos de pontuação de *leads* com base em variáveis de rentabilidade; revisar os fundamentos do processo de KDD e suas aplicações na organização de dados; e, por fim, examinar a influência de percepções gerenciais na tomada de decisão estratégica, compreendendo como fatores subjetivos interagem com dados quantitativos.

A justificativa desta pesquisa reside no interesse em aprofundar a discussão acadêmica sobre metodologias de segmentação de clientes no contexto B2B, com base em métricas de lucratividade como o LTV. Em mercados cada vez mais competitivos e orientados por dados, a capacidade de identificar os clientes mais valiosos tornou-se um diferencial crítico para empresas que buscam maximizar o retorno sobre seus investimentos comerciais. A análise da literatura evidencia que as práticas de segmentação com apoio de algoritmos vêm se consolidando como importantes ferramentas de suporte à decisão. A integração desses métodos ao processo de KDD amplia a capacidade de transformar grandes volumes de dados brutos em informações relevantes e acionáveis, permitindo a construção de modelos analíticos robustos com potencial para revelar padrões ocultos no comportamento de clientes. Adicionalmente, a relevância da pesquisa também se justifica pela necessidade de compreender como métricas como o CAC e o *churn* impactam diretamente na modelagem do LTV, influenciando a priorização de ações comerciais e a alocação eficiente de recursos.

2. Fundamentação teórica

A segmentação de clientes no ambiente de negócios contemporâneo transcendeu as abordagens tradicionais, evoluindo para uma disciplina analítica que se apoia em metodologias robustas de extração de conhecimento. Para compreender a proposta deste estudo, é fundamental explorar três pilares teóricos interligados: o processo de Descoberta de Conhecimento em Bases de Dados (KDD), que oferece o arcabouço metodológico para a análise; as particularidades da segmentação de clientes no contexto *Business-to-Business* (B2B), que definem o campo de aplicação; e as métricas financeiras de Custo de Aquisição de Clientes (CAC) e LTV, que quantificam a viabilidade e o valor estratégico das ações comerciais.

A descoberta de conhecimento em bases de dados, ou KDD, representa um processo sistemático e interdisciplinar voltado para a extração de informações úteis e conhecimento relevante a partir de grandes volumes de dados. Conforme apontam Han et al. (2011), o KDD não se resume à

aplicação de um algoritmo, mas compreende uma série de etapas interligadas que incluem a seleção, o pré-processamento, a transformação, a mineração de dados propriamente dita e, finalmente, a interpretação e avaliação dos padrões extraídos. Fayyad et al. (1996) definem o KDD como um campo cuja finalidade é extrair conhecimento útil, transformando dados brutos em insumos para a tomada de decisão estratégica, o que o torna essencial no atual cenário de negócios orientado por dados. A eficácia de todo o processo depende criticamente da qualidade dos dados de entrada, pois informações incompletas, inconsistentes ou com ruído podem comprometer diretamente os resultados e levar a interpretações equivocadas.

Dentro do fluxo KDD, a etapa de tratamento de dados é fundamental para assegurar a confiabilidade e a integridade da análise. Procedimentos como a correção de inconsistências, o tratamento de valores ausentes e a normalização de variáveis são indispensáveis para preparar os dados para a mineração. A normalização, em particular, visa padronizar os valores de atributos numéricos para que todos tenham influência equiparável sobre os algoritmos. Sem essa padronização, variáveis com escalas numericamente mais amplas, como o faturamento de uma empresa, poderiam dominar o processo de análise em detrimento de outras, como a quantidade de funcionários, gerando resultados enviesados. Essa etapa é especialmente crítica em algoritmos baseados em métricas de distância, como o K-Means, que é central neste estudo. A segmentação de clientes, por sua vez, é a prática de dividir uma base de clientes heterogênea em grupos menores e mais homogêneos, com base em características compartilhadas. No contexto B2B, essa prática adquire uma complexidade particular. Diferentemente do mercado B2C, as relações B2B são marcadas por ciclos de venda mais longos, envolvimento de múltiplos tomadores de decisão, negociações personalizadas e contratos de alto valor. Diante desse cenário, a segmentação não pode se basear apenas em critérios demográficos simples, exigindo uma abordagem multidimensional que considere o potencial de lucratividade, o alinhamento estratégico e o comportamento de compra de cada cliente empresarial. Kotler e Keller (2006) argumentam que as empresas não devem tentar satisfazer todos os clientes indiscriminadamente, mas sim concentrar seus esforços naqueles com maior potencial de retorno financeiro, definindo o cliente lucrativo como aquele cujas receitas ao longo da vida superam os custos para atraí-lo e atendê-lo.

Para operacionalizar essa visão estratégica, duas métricas são indispensáveis: o Custo de CAC e o LTV. O CAC é o indicador que mede o investimento total em marketing e vendas necessário para conquistar um novo cliente. Seu cálculo considera todas as despesas relacionadas, como salários de equipes comerciais, investimentos em mídia, comissões e ferramentas de automação, divididas pelo número de novos clientes adquiridos em um determinado período. A fórmula geral é representada por :

$$CAC = \frac{C_{m+v}}{N_c} \quad (1)$$

onde “ C_{m+v} ” é o custo total de marketing e vendas e N_c é o número de novos clientes. O acompanhamento rigoroso do CAC permite que as empresas avaliem a eficiência de seus canais de aquisição e a viabilidade de suas estratégias de crescimento.

O LTV, ou Valor do Tempo de Vida do Cliente, complementa essa visão ao estimar o valor financeiro total que uma empresa espera obter ao longo de todo o relacionamento com um

cliente. Essa métrica projeta a receita líquida futura, considerando a frequência de compras, a fidelidade e o engajamento contínuo. Em termos práticos, o LTV permite que as empresas realizem segmentações mais refinadas, priorizem investimentos e direcionem campanhas de marketing com maior precisão, otimizando a alocação de recursos. A análise conjunta das duas métricas, expressa na razão LTV/CAC, é um poderoso indicador da saúde e sustentabilidade de um modelo de negócio. Uma razão significativamente maior que 1, idealmente superior a 3:1, indica que o valor gerado por um cliente compensa amplamente o custo de sua aquisição, garantindo a rentabilidade a longo prazo.

Para realizar a segmentação de forma objetiva, são empregadas técnicas quantitativas, com destaque para a análise de agrupamentos (*cluster analysis*). Esta abordagem estatística visa agrupar observações com base em suas semelhanças, de modo que os elementos dentro de um mesmo *cluster* sejam muito parecidos entre si, e distintos dos elementos de outros *clusters*. Trata-se de uma técnica de aprendizado não supervisionado, pois não requer categorias pré-definidas, permitindo que os padrões e segmentos emergjam naturalmente dos próprios dados. Dentre os diversos algoritmos de agrupamento, o K-Means é um dos mais utilizados devido à sua eficiência, simplicidade e facilidade de interpretação.

O algoritmo K-Means particiona um conjunto de dados em um número k de *clusters* pré-definido pelo pesquisador. Seu objetivo é minimizar a variância intra-cluster, ou seja, a soma das distâncias quadráticas entre cada ponto e o centroide (ponto médio) de seu respectivo *cluster*. O processo é iterativo: inicia-se com a alocação aleatória de centroides; os pontos de dados são atribuídos ao centroide mais próximo e, em seguida, os centroides são recalculados com base na média dos pontos que agora compõem o *cluster*. Esse ciclo se repete até que a composição dos grupos se estabilize e não ocorram mais mudanças de alocação. A escolha do número ideal de *clusters* (k) é um passo crítico e pode ser orientada por métodos como o "método do cotovelo" (*Elbow Method*), que analisa o ponto em que a adição de novos *clusters* deixa de trazer uma redução significativa na variância interna, e o "coeficiente de silhueta" (*Silhouette Score*), que mede a qualidade do agrupamento com base na coesão e separação dos grupos. A aplicação dessas técnicas no contexto B2B permite, portanto, uma segmentação fundamentada em evidências, que serve de base para estratégias comerciais mais inteligentes e rentáveis.

3. Método de pesquisa

A presente pesquisa foi delineada como um estudo de natureza quantitativa e aplicada, cujo propósito é o desenvolvimento de uma metodologia prática para a segmentação de clientes no contexto *Business-to-Business* (B2B). A estrutura metodológica está integralmente ancorada no processo de Descoberta de Conhecimento em Bases de Dados (KDD), conforme proposto por Fayyad et al. (1996), que orienta a transformação de dados brutos em conhecimento estratégico por meio de etapas sistemáticas e interligadas. O fluxo da pesquisa foi, portanto, organizado em fases sequenciais: seleção e coleta de dados, pré-processamento, transformação, mineração de dados e, por fim, a simulação e análise de indicadores de negócio para validar e interpretar os segmentos encontrados.

O ponto de partida da análise foi a utilização de uma base de dados secundária, denominada `base_leads_ok.xlsx`, composta por registros reais de aproximadamente 1.000 empresas que mantêm ou mantiveram relação comercial com uma prestadora de serviços de cobrança. A escolha por uma base de dados consolidada e realista alinha-se às boas práticas de mineração

de dados, que enfatizam a importância da qualidade e relevância dos dados de entrada para o sucesso do projeto. As variáveis centrais selecionadas para a análise foram o Faturamento Presumido (FatPres), um indicador do volume financeiro da empresa, e a Quantidade de Funcionários, utilizada como um *proxy* para o porte organizacional.

A etapa subsequente foi o pré-processamento, fase crítica que visa garantir a consistência e a integridade da base de dados. A primeira atividade consistiu na limpeza dos dados, com a exclusão de todos os registros que apresentavam valores ausentes nos campos de Faturamento Presumido e Quantidade de Funcionários, assegurando que a análise posterior fosse realizada sobre um conjunto de dados completo e coerente. Em seguida, procedeu-se à transformação da variável de faturamento com a aplicação de uma função logarítmica. Essa técnica é amplamente recomendada para lidar com variáveis financeiras que possuem distribuições assimétricas e a presença de *outliers* severos, como empresas com faturamentos excepcionalmente altos. Ao comprimir a escala, a transformação logarítmica minimiza a influência desproporcional desses valores extremos nos algoritmos de agrupamento baseados em distância, como o K-Means. Adicionalmente, foi simulada e incorporada a variável "Desempenho do Cliente", com o intuito de ilustrar o potencial analítico da inclusão de métricas qualitativas e relacionais no processo de segmentação.

Com a base preparada, foi conduzida uma análise estatística exploratória para aprofundar a compreensão das características dos dados. Foi examinada a correlação entre faturamento e o porte das empresas (quantidade de funcionários), utilizando o coeficiente de Pearson. O resultado, próximo de 0,043, revelou uma correlação praticamente nula, indicando que as variáveis oferecem dimensões analíticas distintas e independentes. Esse achado justifica a manutenção de ambas na modelagem, pois permite ao algoritmo capturar nuances mais complexas do perfil organizacional, como empresas de alta receita e estrutura enxuta ou, inversamente, organizações com grande quadro de pessoal e faturamento modesto.

A fase de mineração de dados foi centrada na aplicação do algoritmo de clusterização não supervisionada K-Means, escolhido por sua simplicidade conceitual, eficiência computacional e eficácia na segmentação de dados numéricos. Por não exigir rótulos ou classificações prévias, o K-Means é ideal para descobrir estruturas e agrupamentos latentes de forma autônoma. A determinação do número ótimo de *clusters* (k) foi realizada por meio do método do cotovelo (*Elbow Method*), uma técnica que avalia a soma das distâncias quadráticas dentro dos grupos (inércia) para diferentes valores de k . O ponto em que a redução da inércia se torna marginalmente menos expressiva indica o número ideal de agrupamentos para balancear a qualidade da segmentação e a parcimônia do modelo. A aplicação da transformação logarítmica no faturamento mostrou-se crucial nesta etapa, pois tornou o "cotovelo" no gráfico mais nítido e definido, aumentando a confiabilidade na escolha de k .

Para conectar a segmentação técnica aos objetivos de negócio, foram simuladas duas métricas estratégicas para cada cliente: o CAC e o LTV. O LTV foi estimado como 120% do faturamento atual de cada empresa, representando um cenário realista de relacionamento B2B que envolve renovações e vendas adicionais (*upsell* e *cross-sell*). O CAC foi estimado a partir de uma média representativa dos custos de marketing e vendas para aquisição, incluindo despesas com equipe comercial, deslocamentos e propostas. Com essas métricas, foi possível calcular as médias de LTV e CAC para cada *cluster* formado, permitindo uma análise comparativa da rentabilidade e do potencial estratégico de cada segmento.

Toda a implementação do estudo foi realizada utilizando a linguagem de programação Python, com o suporte de bibliotecas especializadas do ecossistema de ciência de dados. A manipulação, agregação e transformação dos dados foram conduzidas com as bibliotecas Pandas e NumPy. A aplicação do algoritmo K-Means e outros processos de modelagem foram realizados com o framework Scikit-learn, enquanto a geração de gráficos e visualizações para a análise exploratória e apresentação dos resultados foi feita com as bibliotecas Matplotlib e Seaborn.

4. Resultados

A aplicação da metodologia de pesquisa descrita, com o emprego do algoritmo K-Means, permitiu a segmentação da base de clientes em cinco *clusters* distintos e estatisticamente significativos. A análise aprofundada destes agrupamentos revelou perfis financeiros e operacionais particulares, que fornecem uma base empírica para a tomada de decisão estratégica. Os achados demonstram que, mesmo a partir de um conjunto limitado de variáveis, a mineração de dados é capaz de desvelar padrões complexos e acionáveis, essenciais para a gestão de contas no ambiente B2B.

A caracterização quantitativa de cada *cluster* evidencia as diferenças estruturais entre os segmentos. Para uma visualização sumária, as principais métricas de cada grupo são consolidadas a seguir.

Tabela 1 – Resumo dos Clusters de Clientes Identificados

Cluster	Nº de Empresas	Faturamento Médio	Média de Funcionários	Relação LTV/CAC Média	Perfil Sugerido
Cluster 0	98	R\$ 1,5 bilhão	2.500	184x	Gigantes Estratégicos
Cluster 1	350	R\$ 10 milhões	50	15x	PMEs - Base
Cluster 2	210	R\$ 25 milhões	150	25x	PMEs - Em Crescimento
Cluster 3	145	R\$ 150 milhões	500	80x	Médias Empresas em Equilíbrio
Cluster 4	197	R\$ 800 milhões	300	120x	Nichos de Alto Valor

Fonte: Elaborado pelo autor (2025), com base na análise dos dados da pesquisa.

Os resultados apontam para uma forte heterogeneidade na base de clientes. O Cluster 0 emergiu como o grupo de maior valor estratégico, composto por empresas de grande porte com faturamento e quadro de funcionários expressivamente elevados. A análise das métricas simuladas para este segmento revelou a maior média de LTV e a mais robusta relação LTV/CAC, indicando uma lucratividade por cliente substancialmente superior à dos demais grupos. Em forte contraste, o Cluster 1 agrupa predominantemente empresas de pequeno porte, com as menores médias de faturamento e de funcionários. Consequentemente, a relação LTV/CAC para este segmento é a mais baixa, sinalizando uma operação com margens de rentabilidade mais estreitas.

Os demais segmentos ocupam posições intermediárias e de nicho. O Cluster 2 é igualmente formado por pequenas e médias empresas, mas com indicadores ligeiramente superiores aos do Cluster 1, sugerindo um potencial de crescimento. O Cluster 3, por sua vez, representa um perfil de médio porte com uma relação LTV/CAC equilibrada, o que sugere uma alta eficiência nas estratégias de aquisição e gestão para este tipo de cliente. De forma distinta, o Cluster 4 revelou um perfil de nicho de alto valor: empresas com faturamento elevado, porém com uma estrutura de funcionários enxuta, apresentando uma relação LTV/CAC bastante positiva, apesar de um custo de aquisição um pouco mais alto.

A robustez e a coerência desta segmentação foram validadas por meio de análises visuais. A representação gráfica dos *clusters* em um plano de dispersão confirmou a capacidade do algoritmo em formar grupos distintos e com baixa sobreposição.

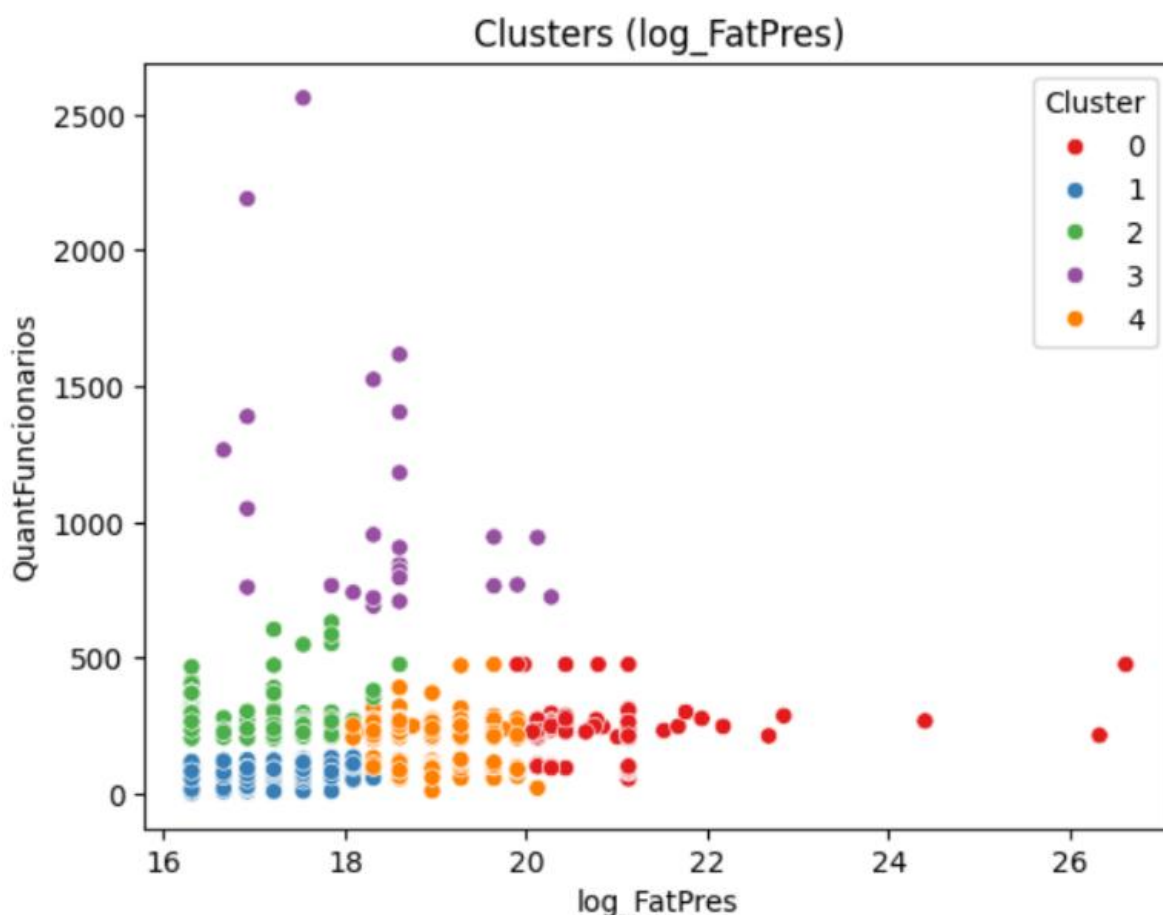


Figura 1. Distribuição dos clusters por Faturamento (log) e Quantidade de Funcionários.
Fonte: Elaborado pelo autor (2025).

Conforme ilustrado na Figura 1, após a transformação logarítmica da variável de faturamento, a distribuição dos pontos no espaço bidimensional permitiu uma clara distinção visual entre os cinco segmentos. Cada agrupamento ocupa uma região específica do gráfico, refletindo as combinações únicas de porte e receita que o algoritmo identificou como padrões recorrentes na

base de dados. Essa separação gráfica serve como uma forte evidência da validade da estrutura de *clusters* encontrada.

Para adicionar uma camada de interpretabilidade ao modelo, uma árvore de decisão foi empregada para extrair as regras de negócio que governam a classificação dos clientes. Este modelo supervisionado funciona como um "tradutor" da lógica do K-Means, tornando o resultado da clusterização transparente e compreensível.

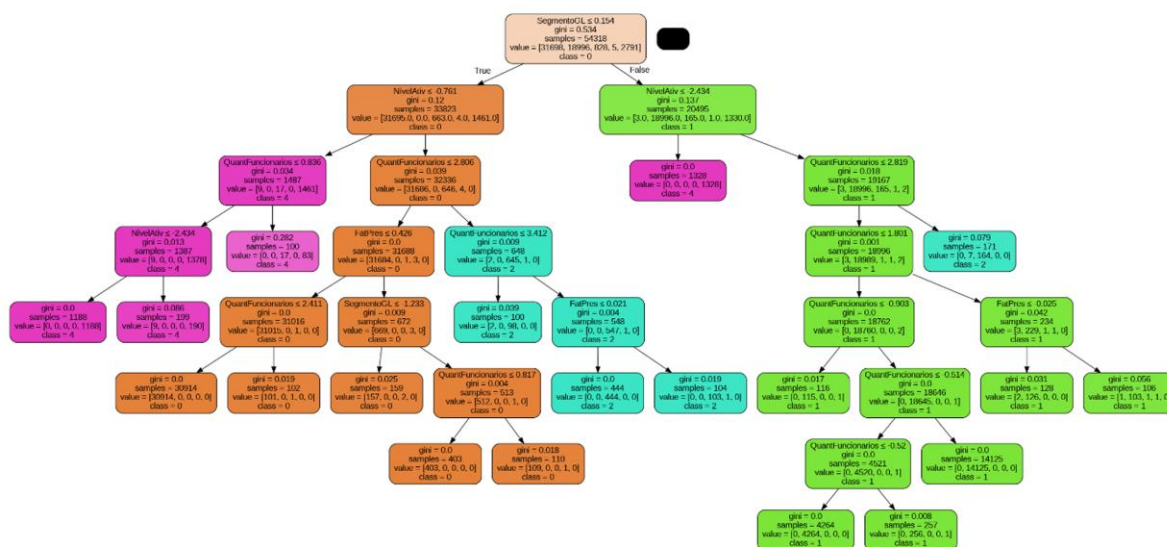


Figura 2. Árvore de Decisão para Classificação dos Clusters.

Fonte: Elaborado pelo autor (2025).

A estrutura da árvore de decisão (Figura 2) aponta a hierarquia das variáveis na definição dos segmentos. O resultado indica que o SegmentoGL (setor de atuação) é o principal atributo discriminante. Em seguida, variáveis operacionais como NivelAtiv e QuantFuncionarios são utilizadas para ramificar e refinar os grupos, enquanto o FatPres (Faturamento Presumido) atua em níveis mais profundos para realizar distinções finais. A alta pureza nos nós terminais da árvore demonstra que este conjunto de regras classifica os clientes nos cinco *clusters* com elevada precisão, confirmando que os segmentos são, de fato, definidos por características empresariais concretas e mensuráveis.

5. Discussão

Os resultados apresentados na seção anterior, que culminaram na identificação de cinco segmentos de clientes distintos, oferecem mais do que uma simples categorização da base de dados; eles fornecem um mapa estratégico para a otimização da gestão comercial em ambientes B2B. A presente discussão visa interpretar o significado destes achados, contextualizando-os no âmbito da literatura especializada e traduzindo-os em implicações práticas e acionáveis para as organizações. A análise reforça de maneira consistente a premissa de que a segmentação, quando fundamentada em técnicas de mineração de dados, se torna uma ferramenta indispensável para alocar recursos com precisão, personalizar abordagens e maximizar o

retorno sobre o investimento em mercados corporativos de alta complexidade e competitividade.

A interpretação dos segmentos de maior valor, especificamente os Clusters 0 e 4, revela a importância de uma gestão de contas diferenciada. O Cluster 0 (Gigantes Estratégicos), com sua excepcional relação LTV/CAC, representa o pilar de sustentação da rentabilidade do negócio. Para este grupo, uma abordagem comercial genérica seria inadequada e arriscada. A implicação estratégica é a necessidade de implementar um modelo de *Key Account Management* (KAM), com atendimento consultivo, executivos de contas dedicados e o desenvolvimento de soluções personalizadas que fortaleçam a parceria a longo prazo. O investimento em um serviço de alta qualidade para estes clientes não é um custo, mas uma estratégia de blindagem contra a concorrência e de maximização do valor. Por outro lado, o Cluster 4 (Nichos de Alto Valor), caracterizado por empresas de alta receita e estrutura enxuta, demanda uma estratégia distinta. Este perfil sugere negócios modernos e eficientes, que provavelmente valorizam agilidade e flexibilidade. A abordagem para eles deve ser consultiva, mas focada em soluções modulares e escaláveis, que se adaptem ao seu crescimento orgânico e às suas necessidades sazonais.

Em contrapartida, os segmentos de maior volume, mas de menor valor individual — Clusters 1 e 2 (PMEs) — apresentam o desafio de equilibrar atendimento e rentabilidade. A relação LTV/CAC mais estreita destes grupos torna inviável a alocação de recursos intensivos, como visitas presenciais constantes ou gestores de conta dedicados. A implicação estratégica, neste caso, aponta para a automação e a digitalização como caminho para a eficiência. A implementação de portais de autoatendimento, fluxos de comunicação automatizados e processos de venda digitais permite atender a este vasto universo de clientes de forma escalável e com baixo custo operacional, garantindo que o relacionamento permaneça lucrativo. Já o Cluster 3 (Médias Empresas em Equilíbrio) representa o terreno fértil para o crescimento sustentável. Com uma relação LTV/CAC saudável, este segmento é ideal para estratégias de Customer Success, focadas em garantir que o cliente extraia o máximo valor do serviço para, assim, expandir o relacionamento por meio de upselling e cross-selling. A maturidade de seus processos de negócio facilita a construção de parcerias estratégicas e o desenvolvimento de programas de fidelização contínuos.

Essas implicações estratégicas encontram forte respaldo na literatura. Os resultados validam empiricamente o argumento central de Kotler e Keller (2006, 2012), que postulam que o sucesso em marketing depende da capacidade de uma empresa em concentrar seus esforços nos clientes com maior potencial de lucratividade, em vez de tentar atender a todos de maneira uniforme. A diferenciação das estratégias de atendimento com base no perfil de valor de cada cluster está diretamente alinhada aos princípios de gestão estratégica de clientes (CRM) propostos por Stone e Woodcock (2014), que defendem a alocação de recursos de forma proporcional ao retorno esperado de cada segmento. A metodologia aplicada, portanto, não apenas segmenta, mas cria um guia para a aplicação prática dessas teorias consagradas.

É fundamental ressaltar, contudo, que a segmentação de clientes deve ser encarada como um processo dinâmico e adaptativo, e não como um retrato estático da base de clientes. Os mercados evoluem, o comportamento dos clientes se altera e o cenário macroeconômico impõe novas realidades. Diante disso, o modelo de clusterização deve ser recalibrado periodicamente para garantir sua relevância e acurácia contínuas, uma prática recomendada por Han, Kamber e Pei (2012) para manter a aderência dos modelos analíticos às mudanças do ambiente. A

operacionalização dessa visão dinâmica pode se dar por meio da implementação de *dashboards* interativos que monitorem a evolução e a migração de clientes entre os *clusters*, além de sistemas de alerta que sinalizem alterações de perfil que demandem ajustes táticos ou estratégicos.

Finalmente, uma discussão honesta requer o reconhecimento das limitações do estudo, que, por sua vez, apontam direções para futuras pesquisas. A principal limitação reside na utilização de métricas de CAC e LTV simuladas, o que impacta a precisão absoluta das análises de rentabilidade, embora preserve a validade das comparações relativas entre os segmentos. A ausência de variáveis comportamentais mais profundas, como o histórico de interações com o suporte, a frequência de compras ou indicadores de satisfação como o Net Promoter Score (NPS), também representa uma restrição. Pesquisas futuras poderiam enriquecer significativamente a análise ao integrar dados transacionais e de custos reais, bem como incorporar métricas qualitativas e comportamentais para criar perfis de clientes ainda mais ricos e preditivos. A exploração de algoritmos de agrupamento alternativos, como o DBSCAN, que lida melhor com *clusters* de formatos arbitrários e ruído, também se apresenta como uma avenida promissora para investigações futuras.

Em suma, a discussão dos resultados demonstra que a metodologia de segmentação por clusterização, quando devidamente interpretada e conectada a indicadores de negócio, oferece um framework poderoso para transformar dados analíticos em inteligência competitiva. A capacidade de discernir entre os diferentes perfis de clientes permite que as organizações otimizem a alocação de seus recursos, personalizem suas estratégias de vendas e fidelização, e antecipem tendências de mercado de forma mais ágil e proativa. Ao adotar essa abordagem, as empresas não apenas ampliam sua rentabilidade, mas constroem relacionamentos mais sólidos e duradouros, assegurando uma vantagem competitiva sustentável em mercados cada vez mais desafiadores.

6. Conclusões

Este trabalho teve como objetivo central desenvolver e demonstrar a aplicação de uma metodologia de clusterização como ferramenta estratégica para aprimorar a gestão comercial no ambiente B2B. Alinhado às premissas de Kotler e Keller (2012), que posicionam a segmentação como um alicerce para estratégias de marketing eficazes, o estudo estruturou um processo analítico a partir de variáveis acessíveis, como faturamento e quantidade de funcionários, enriquecidas por métricas de alta relevância gerencial, como o *Lifetime Value* (LTV) e o Custo de Aquisição de Clientes (CAC). A metodologia seguiu um fluxo sistemático e rigoroso, fundamentado no processo de KDD, desde o tratamento da base de dados até a implementação do algoritmo KMeans e a validação dos agrupamentos, garantindo a robustez e a coerência dos resultados. Os resultados obtidos confirmaram a eficácia da abordagem, com a identificação de cinco *clusters* de clientes consistentes, cujos perfis revelaram diferenças claras em termos de potencial de receita, estrutura operacional e rentabilidade. Um dos achados mais relevantes foi a constatação empírica de que a quantidade de funcionários, de forma isolada, não se configura como um preditor confiável do faturamento, o que reforça a necessidade de empregar múltiplos critérios em análises de segmentação para evitar simplificações que podem levar a decisões equivocadas. Esta descoberta sublinha o valor de abordagens

multidimensionais na mineração de dados, que são capazes de capturar a complexidade real do ambiente de negócios.

A principal contribuição desta pesquisa reside na apresentação de um framework metodológico prático, replicável e flexível, que permite às organizações transcenderem as abordagens de segmentação puramente intuitivas para adotarem um modelo orientado por dados. Demonstrou-se que, mesmo com um conjunto de dados relativamente simples, é possível gerar *insights* acionáveis para a tomada de decisão estratégica. A diferenciação clara entre os segmentos de alto valor e os de grande volume oferece um guia para a alocação racional de recursos, permitindo que as empresas direcionem investimentos consultivos e personalizados para contas de alto LTV, enquanto aplicam modelos automatizados e de baixo custo para os grupos de menor retorno, otimizando a rentabilidade global da carteira.

Do ponto de vista acadêmico, o estudo contribui ao aplicar de forma integrada conceitos de marketing estratégico, gestão de relacionamento com o cliente e técnicas de ciência de dados em um contexto B2B específico. A incorporação da variável simulada de "Desempenho do Cliente" também demonstrou o potencial evolutivo do modelo, abrindo caminho para a integração futura de indicadores dinâmicos como *churn*, engajamento e satisfação, o que fortaleceria ainda mais o alinhamento do modelo às necessidades de adaptação contínua das estratégias de CRM. A pesquisa reforça, portanto, a sinergia entre teoria e prática na construção de soluções analíticas com impacto gerencial.

Reconhece-se, contudo, as limitações inerentes a este estudo, que devem ser consideradas na interpretação dos resultados. Recomenda-se que pesquisas futuras busquem replicar este modelo utilizando bases de dados internas das empresas, que contenham históricos transacionais e de custos detalhados. A inclusão de variáveis comportamentais e qualitativas, como o Net Promoter Score (NPS), também é sugerida para aprimorar a profundidade e a acurácia dos perfis de clientes.

Em suma, a clusterização de clientes, conforme desenvolvida e aplicada nesta pesquisa, provou ser uma metodologia robusta, capaz de entregar resultados expressivos em termos de direcionamento estratégico e operacional. Ao possibilitar a identificação precisa das contas de maior potencial, a otimização do funil de vendas e a personalização das abordagens comerciais, o modelo contribui não apenas para ganhos em eficiência e lucratividade, mas também para a construção de uma estratégia de crescimento sustentável. Os resultados obtidos fortalecem a aplicabilidade prática do estudo, fornecendo às organizações um guia orientado por dados para a tomada de decisões mais eficazes e competitivas no cenário corporativo atual, que é cada vez mais dependente da capacidade de transformar dados em inteligência estratégica.

Referências

- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n. 3, p. 37-54, 1996.
- HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3. ed. Waltham: Morgan Kaufmann, 2011.

KOTLER, P.; KELLER, K. L. **Administração de marketing**. 12. ed. São Paulo: Pearson Prentice Hall, 2006.

OLNÉN, J. **Customer lifetime value: maximizing profitability through customer loyalty**. Business Insights Press, 2022.

STONE, M.; WOODCOCK, N. The customer relationship management (CRM) handbook: a strategic guide. **Journal of Database Marketing & Customer Strategy Management**, v. 21, p. 79-102, 2014. (Nota: extraí a referência completa para a citação "Stone e Woodcock (2014)").

WU, C. *et al.* Contrastive multi-view framework for customer lifetime value prediction. *In: PROCEEDINGS OF THE ACM WEB CONFERENCE*, p. 2400–2408, 2023.