

Inteligência Artificial Aplicada à Previsão de *Churn* e Retenção de Clientes

Brenno Rawber Martins Faria – brennorawber2016@ufu.br

Universidade Federal de Uberlândia

Carla Bonato Marcolin – carla@ufu.br

Universidade Federal de Uberlândia

Carolina Coelho da Silveira – carolinasilveira@ufu.br

Universidade Federal de Uberlândia

Resumo

A fidelização de clientes é um dos pilares para o crescimento sustentável das empresas em um mercado cada vez mais dinâmico e competitivo. Retê-los não é apenas uma vantagem estratégica, mas também é importante para a manutenção da rentabilidade e do fortalecimento da marca. Com a transformação digital e o avanço das tecnologias, as empresas precisam começar a adotar práticas inovadoras para se manterem competitivas. Neste contexto, o presente trabalho tem como objetivo aplicar técnicas de Inteligência Artificial Clássica, por meio de algoritmos de *Machine Learning*, para prever o risco de evasão (*churn*) de clientes em uma base comercial. A metodologia envolve o uso de processos estatísticos para segmentação da base, seguida pela aplicação de modelos preditivos como regressão logística e *Random Forest*, com foco na identificação de clientes com maior propensão ao abandono. Os resultados obtidos demonstraram uma acurácia satisfatória, com destaque para a capacidade de identificação precoce de padrões de *churn*, o que permite à empresa agir de forma proativa na retenção. Conclui-se que, ao antecipar comportamentos de evasão, é possível reduzir custos, aumentar a satisfação dos clientes e prolongar seu ciclo de vida, reforçando que investir na fidelização de clientes atuais é mais eficiente e rentável do que focar apenas na aquisição de novos.

Palavras-chave: Churn; Machine Learning; Random Forest; Retensão de Clientes; Inteligência Artificial.

1 Introdução

No ambiente corporativo atual, caracterizado por uma concorrência acirrada, o cliente tende a se tornar cada vez mais importante e a ser alvo de estudos pelas corporações. Mais do que simplesmente satisfazer necessidades imediatas, as empresas buscam agora criar experiências que promovam relações duradouras, conscientes de que a fidelização é um dos pilares para garantir a sustentabilidade de seus negócios (Kotler; Keller, 2017). Reter clientes é um ponto estratégico e, de acordo com Kotler e Keller (2017), os custos de aquisição de novos clientes geralmente superam, e em muito, os custos de manutenção de clientes existentes. Assim, a retenção passa a ser um dos determinantes da rentabilidade de longo prazo.

Frederick Reichheld e Thomas Teal (1996), em seu trabalho seminal intitulado *The Loyalty Effect*, reforçam essa perspectiva ao discutirem sobre a subestimação dos custos de aquisição, demonstrando que pequenas melhorias nas taxas de retenção podem gerar impactos expressivos nos lucros. Segundo os autores, aumentar a retenção de clientes em 5% pode elevar os lucros de uma empresa entre 25% e 95% – essa relação evidencia que a lealdade do cliente não é apenas desejável, mas significativa para o sucesso financeiro. Além disso, Peppers e Rogers (1993) defendem que o desenvolvimento de relacionamentos sólidos com clientes proporciona uma vantagem competitiva sustentável, uma vez que a fidelidade é mais difícil de ser replicada pela concorrência do que outros diferenciais de mercado, como preço ou produto. Pesquisas mais recentes (Cohen *et al.*, 2022) indicam também que a satisfação do cliente tem impacto direto tanto na sua retenção quanto no aumento do volume de compras, o que reafirma a fidelização como um fator a ser considerado para o desempenho financeiro das empresas.

Todavia, promover a retenção em um ambiente de transformação digital apresenta desafios. A era da inovação tecnológica, marcada pela rápida disseminação de novas ferramentas digitais e pela mudança constante nas expectativas dos consumidores, exige que as organizações adotem estratégias mais dinâmicas (Buttle, 2009). Nesse contexto, compreender o *churn* pode ser uma estratégia para manter a sustentabilidade dos negócios. Segundo Buttle (2009), *churn* representa a taxa de clientes que deixam de se relacionar com uma empresa em determinado período. Para Kotler e Keller (2017), a gestão da retenção desses clientes é crítica, uma vez que conquistar novos clientes tende a ser significativamente mais caro do que manter os atuais. Além disso, controlar os índices de *churn* tem impacto direto no *Customer Lifetime Value* (CLV), conceito que reflete o valor financeiro gerado por um cliente ao longo de todo o seu ciclo de relacionamento com a empresa (Gupta; Zeithaml, 2006). No entanto, os efeitos do *churn* vão além da dimensão financeira: altas taxas de cancelamento também afetam a reputação da organização no mercado, uma vez que consumidores insatisfeitos tendem a compartilhar suas experiências negativas. Ainda, a incapacidade de reter clientes pode refletir falhas na proposta de valor e enfraquecer a cultura organizacional voltada à experiência do cliente, dificultando o alinhamento interno e a motivação das equipes (Verhoef *et al.*, 2010; Verhoef *et al.*, 2021).

Historicamente, as organizações adotaram estratégias de retenção como programas de fidelidade, promoções e atendimento personalizado, baseadas em análises descritivas e segmentações tradicionais. Contudo, essas abordagens não têm se mostrado mais tão efetivas na era digital: a variedade de canais, o empoderamento dos consumidores e o comportamento volátil dificultam prever o momento do *churn* ou identificar os fatores determinantes.

Dessa forma, o uso de técnicas de Inteligência Artificial, como as de *Machine Learning*, apresenta-se como uma possível solução. Segundo Provost e Fawcett (2013), a aplicação de *Machine Learning* em problemas de negócios possibilita uma nova maneira de construir conhecimento, baseada na extração de padrões relevantes a partir dos dados. Assim, as

empresas passam a dispor de ferramentas que ampliam sua capacidade analítica e decisória, tornando suas ações de retenção mais assertivas (Rosário; Boechat, 2024).

Este trabalho, portanto, tem como objetivo a análise e aplicação de modelos de *Machine Learning* para previsão de *churn*. Busca-se, assim, contribuir para a antecipação de comportamentos de evasão e para a construção de uma gestão de clientes orientada por dados, afinal, em um mercado onde a competição por atenção é cada vez mais intensa, a capacidade de reter clientes de forma eficiente e sustentável configura-se como um diferencial para as empresas.

2 Referencial teórico

2.1 Gestão de relacionamento com clientes

A Gestão de Relacionamento com Clientes (CRM, do inglês *Customer Relationship Management*) é uma estratégia empresarial que visa estabelecer conexões duradouras e rentáveis com os clientes, por meio da análise de dados, personalização de interações e otimização da experiência do consumidor (Kotler; Keller, 2017). Segundo Payne e Frow (2013), o CRM não se limita a ferramentas tecnológicas, mas engloba processos, políticas e cultura organizacional voltados para a construção de relacionamentos estratégicos.

Enquanto base do CRM, o marketing relacional surge como contraponto ao marketing transacional, priorizando a fidelização em vez de vendas isoladas (Grönroos, 2015). Para Sheth e Parvatiyar (1995), essa abordagem enfatiza a criação de valor mútuo, onde empresas e clientes se beneficiam de interações contínuas e personalizadas. A segmentação de clientes, o atendimento diferenciado e a comunicação integrada são práticas essenciais nessa abordagem (Pepper; Rogers, 2011).

O *Customer Lifetime Value* (CLV) é uma métrica central no CRM, representando o valor presente líquido dos fluxos de caixa futuros gerados por um cliente ao longo do relacionamento com a empresa (Gupta; Zeithaml; 2006). Payne e Frow (2013) destacam que o CLV permite direcionar investimentos para clientes com maior potencial de retorno, alinhando estratégias de aquisição, retenção e recuperação.

Portanto, a integração entre CRM, marketing relacional e CLV possibilita às empresas não apenas aumentarem a lucratividade, mas também construir vantagem competitiva sustentável (Verhoef *et al.*, 2010).

2.2 Churn

O *churn*, ou taxa de rotatividade de clientes, representa um dos principais indicadores de desempenho em gestão de relacionamento com clientes (CRM) e pode ser definido como a perda de clientes em um determinado período, fenômeno que impacta diretamente a rentabilidade e a sustentabilidade dos negócios (Blattberg; Deighton, 1996). Essa rotatividade se manifesta de duas formas principais: o *churn* voluntário, quando o cliente ativamente decide encerrar o relacionamento com a empresa, e o *churn* involuntário, decorrente de fatores externos como inadimplência ou mudanças regulatórias (Neslin *et al.*, 2006).

Diversos fatores contribuem para a ocorrência do *churn*, incluindo experiência insatisfatória do cliente, qualidade do serviço abaixo das expectativas, preços não competitivos e ofertas mais atraentes da concorrência (Verhoef; Leeflang, 2009). Gupta e Zeithaml (2006) destacam que o valor percebido pelo cliente é um determinante crítico, pois quando a relação custo-benefício se desequilibra, o risco de evasão aumenta significativamente.

Os impactos do *churn* são diversos e podem afetar diferentes áreas organizacionais. Financeiramente, resulta em redução da receita recorrente, aumento do custo médio de

aquisição de clientes (CAC) e diminuição do *Customer Lifetime Value* (Jain; Singh, 2002). Operacionalmente, a perda de clientes reduz a eficiência em escala e sobrecarrega as equipes comerciais (Pfeifer *et al.*, 2005). Além disso, clientes insatisfeitos tendem a compartilhar a experiência negativa, prejudicando a reputação da marca e dificultando novas aquisições (Anderson, 1998).

Para mitigar o *churn*, as organizações podem adotar algumas estratégias baseadas em evidências. Programas de fidelização, como recompensas e benefícios personalizados, demonstraram eficácia em aumentar a retenção (Dowling; Uncles, 1997). A melhoria contínua da experiência do cliente, com foco na redução de atritos e na personalização, também se mostra eficaz (Meyer; Schwager, 2007). Estratégias de recuperação de clientes perdidos, fazem parte dessas iniciativas, buscando reconquistar relacionamentos valiosos (Lam *et al.*, 2004). Mais recentemente, o uso de modelos preditivos para identificar clientes em risco, permitindo intervenções proativas, tem ganhado destaque (Jain; Singh; 2002).

Em síntese, o gerenciamento eficaz do *churn* requer uma abordagem racional que combine análise de dados, entendimento profundo do comportamento do cliente e implementação de ações personalizadas, e empresas que investem sistematicamente na redução do *churn* conseguem não apenas proteger sua base de clientes, mas também construir vantagens sustentáveis no longo prazo (Rust; Zahorik, 1993)

2.3 Modelos estatísticos

A análise estatística é uma disciplina para a compreensão de fenômenos, extração de padrões e suporte à tomada de decisão em diversas áreas do conhecimento (Hair Jr *et al.*, 2019). Desde as ciências sociais até a engenharia e a medicina, o uso de modelos estatísticos permite transformar grandes volumes de dados em informações mensuráveis, e oferece métodos robustos para descrever, inferir e prever comportamentos complexos, proporcionando uma base sólida para a geração de conhecimento (Hair Jr *et al.*, 2019).

Os modelos estatísticos têm como objetivo principal estabelecer relações entre variáveis, identificar tendências e construir previsões baseadas em evidências quantitativas. O desenvolvimento e a aplicação desses modelos são importantes não apenas para explicar o comportamento de variáveis, mas também para antecipar resultados futuros e orientar intervenções estratégicas em ambientes complexos e dinâmicos (Rosário; Boechat, 2024). Segundo Wooldridge (2012), os modelos estatísticos, quando bem especificados e validados, oferecem um instrumento para testar hipóteses e quantificar relações de causa e efeito. Na era de *big data*, esses métodos evoluíram para abordagens mais sofisticadas de aprendizado de máquina, que se dividem em duas categorias principais: aprendizado supervisionado e não supervisionado

No aprendizado supervisionado, os algoritmos são treinados com dados históricos onde a variável alvo (ou resposta) é conhecida, permitindo que o modelo aprenda a relação entre as variáveis preditoras e o resultado desejado (James *et al.*, 2021). Técnicas como regressão linear, regressão logística e máquinas de vetores de suporte (SVM) são utilizadas para problemas de classificação e previsão.

Já o aprendizado não supervisionado lida com dados não rotulados, explorando estruturas e padrões intrínsecos sem a orientação de uma variável alvo (Bishop, 2006). Técnicas desse grupo, como *clustering* (agrupamento) e análise de componentes principais (PCA), são particularmente úteis para segmentação de clientes ou redução de dimensionalidade, ajudando a identificar grupos homogêneos com comportamentos similares (Han *et al.*, 2011).

2.3.1 Regressão linear

A regressão linear é uma técnica estatística que busca modelar a relação entre uma variável dependente contínua e uma ou mais variáveis independentes. Esse modelo assume que existe uma relação linear entre as variáveis explicativas e a variável resposta. Essa relação é expressa através de uma equação que permite prever o valor da variável dependente a partir de valores conhecidos das variáveis independentes, permitindo estimar a relação entre variáveis e inferir tanto a força quanto a direção dessas associações, por meio dos coeficientes ajustados no modelo (Montgomery *et al.*, 2012)

2.3.2 Regressão logística

A regressão logística é um método estatístico utilizado para modelar a relação entre uma variável dependente categórica e um conjunto de variáveis independentes, sendo adequada para situações em que a variável de interesse assume categorias discretas, geralmente codificadas como 0 e 1, representando, por exemplo, a ocorrência ou não de determinado evento (Hair Jr *et al.*, 2019).

O princípio fundamental da regressão logística é que ela modela a probabilidade de um evento ocorrer utilizando a função logística. De acordo com Hosmer Jr *et al.* (2013), a regressão logística transforma a relação linear entre variáveis em uma escala limitada entre 0 e 1, garantindo a interpretação dos resultados como probabilidades. Assim, é possível interpretar os coeficientes do modelo em termos de razão de chances, o que facilita a compreensão do impacto de cada variável independente sobre a probabilidade de o evento ocorrer.

2.3.3 Random Forest (árvore de decisão)

Random Forest é um método estatístico e de *Machine Learning* utilizado para modelar a relação entre uma variável dependente (que pode ser categórica ou contínua) e um conjunto de variáveis independentes, por meio de uma estrutura hierárquica de decisões baseadas em regras simples (Umoh *et al.*, 2022)

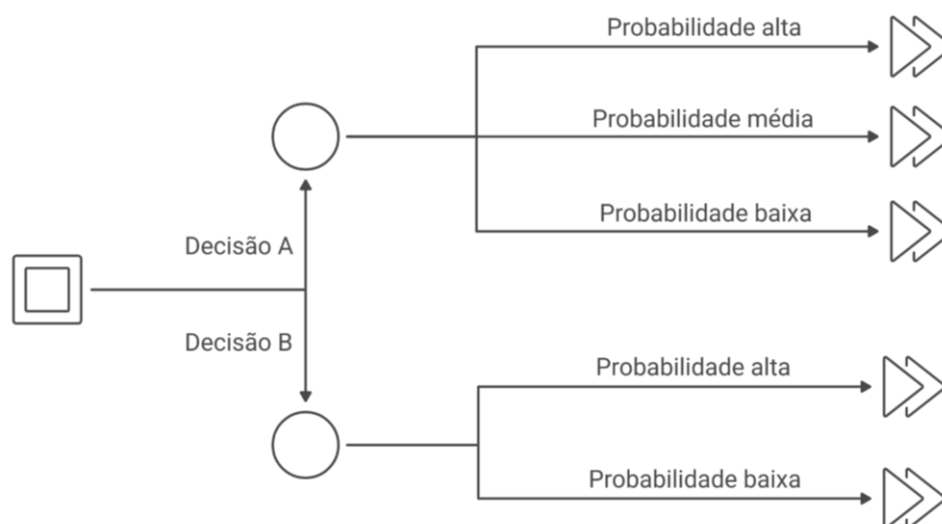


FIGURA 1 - Esquema do funcionamento de uma árvore de decisão. Fonte: dos autores, gerada com o Napkin AI.

O princípio fundamental da árvore de decisão é dividir recursivamente os dados em subconjuntos com base em variáveis preditoras, selecionando em cada nó a divisão que maximiza a pureza das classes (para problemas de classificação) ou reduz a variância (para regressão). O processo de construção da árvore envolve a seleção das melhores divisões a cada nível, até que critérios de parada sejam atingidos, como o número mínimo de observações em

um nó ou a profundidade máxima da árvore. A árvore de decisão pode ser aplicada tanto em problemas de classificação quanto de regressão, sendo uma ferramenta para tarefas de análise exploratória, segmentação e previsão (Breiman, 2001).

3 Procedimentos metodológicos

Este estudo adotou a abordagem de estudo de caso único (Yin, 2015) para investigar os fatores de *churn* e estratégias de retenção de clientes na empresa Martins Atacadista. A pesquisa caracteriza-se como quantitativa e descritiva, com coleta de dados secundários extraídos diretamente do banco de dados corporativo da empresa, abrangendo o período de 2019 a 2025.

A escolha metodológica justifica-se pela oportunidade de investigação profunda de um caso real (Eisenhardt, 1989), permitindo extrair percepções práticas para a gestão de relacionamento no varejo atacadista. O estudo seguiu os princípios éticos de pesquisa com dados corporativos, mantendo o anonimato das informações sensíveis (Brasil, 2018).

Para análise do *churn*, foram processados dados transacionais e comportamentais de 256.423 clientes, utilizando técnicas de mineração de dados e modelos preditivos baseados no modelo de Neslin *et al.* (2006). A análise dos dados foi realizada utilizando as ferramentas Python, Excel, SQL, DataBricks e Tableau, com aplicação de regressão logística, árvore de decisão e outras análises estatísticas.

3.1 Coleta e preparação dos dados

Os dados utilizados para a pesquisa são reais e foram cedidos pelo Martins Atacadista, uma empresa de atacado do Brasil que atua principalmente no segmento de varejo alimentar. A organização atualmente enfrenta um desafio relacionado à alta rotatividade de clientes e à dificuldade em reter sua base ativa, o que tem impactado diretamente seus indicadores de desempenho e sustentabilidade comercial. Diante desse cenário, a análise preditiva de *churn* surge como uma estratégia para auxiliar na identificação precoce de clientes com propensão à evasão.

Os dados foram extraídos diretamente dos sistemas internos da empresa, compreendendo informações cadastrais e históricas de clientes no período de 2019 a 2025, com as seguintes variáveis coletadas:

- Tendência de Vendas;
- Tendência de Produtos;
- Frequência de Compras;
- Tempo Desde a Última Compra;
- Número de Pedidos;
- Segmento do Cliente;
- Tempo de Casa do Representante Comercial;
- Ticket Médio.

O processo de preparação dos dados seguiu as recomendações de Han *et al.* (2011) para mineração de dados, iniciando-se com uma rigorosa limpeza onde foram mantidos apenas clientes com pelo menos três compras. Para tratamento de valores ausentes, adotou-se a substituição por zero, seguida de normalização dos dados e codificação de variáveis categóricas.

Realizou-se uma análise estatística descritiva completa, incluindo cálculo de médias, desvio padrão, valores máximo e mínimo, e variância para todas as variáveis (Hair Jr *et al.*, 2019).

Foram criadas duas variáveis derivadas, Tendência de Venda e Tendência de Produtos, utilizando regressão linear simples (James *et al.*, 2021) para capturar padrões históricos de comportamento dos clientes.

A variável alvo (*churn*) foi operacionalizada como binária, seguindo a abordagem de Neslin *et al.* (2006), sendo que o valor 1 foi atribuído a “clientes que cancelaram ou deixaram de comprar por um período superior ao percentil 90 do intervalo de compras” (análise descritiva) e o valor 0 a “clientes ativos”.

3.2 Construção e validação do modelo

A metodologia adotada neste trabalho baseia-se em técnicas de aprendizado de máquina supervisionado e não supervisionado, com o objetivo de identificar padrões de comportamento de clientes e prever a probabilidade de *churn*. Para isso, foram selecionados os algoritmos *Random Forest*, regressão logística e regressão linear, sendo os dois primeiros para geração da previsão do *churn* e o último para criação de variáveis derivadas.

Inicialmente, para estabelecer o tempo de *churn* de cada cliente, foi calculado o valor do percentil 90 do tempo entre os pedidos de cada cliente. Assim, o tempo alvo para um cliente ser considerado *churn* é dinâmico e personalizado de acordo com o histórico de compras e perfil individual.

Depois, foi utilizado o algoritmo de regressão linear para capturar o valor do coeficiente angular de dois indicadores: Venda e Produto. Para isso, é visto o valor da venda (em reais) no tempo e os produtos comprados no tempo, a fim de capturar tendências no comportamento de compra de cada cliente. O valor desse resultado foi empregado como variável nos outros dois modelos (Tendência de Venda e Tendência Produtos).

Em seguida, foram aplicados algoritmos de aprendizado supervisionado para a previsão de *churn*, sendo testado primeiramente o modelo de regressão logística. A escolha desse modelo se justifica por ser um modelo linear interpretável e adotado em problemas de classificação binária, como a previsão de *churn* (Neslin *et al.*, 2006). O modelo foi treinado com os dados históricos de cada consumidor, utilizando com as seguintes variáveis:

TABELA 1 - Descrição das variáveis utilizadas

Variável	Descrição
TENDENCIA_VENDA	Coeficiente Angular da Regressão Linear com as variáveis VENDA (Eixo Y) e DATA DA COMPRA (Eixo X).
TENDENCIA_PRODUTOS	Coeficiente Angular da Regressão Linear com as variáveis QUANTIDADE DE PRODUTOS (Eixo Y) e DATA DA COMPRA (Eixo X).
DATA_ÚLTIMO_PEDIDO	Data da última compra do cliente.
DIAS_SEM_COMPRA	Quantidade de dias comparando a data do dia atual com a data da última compra do cliente.
CRÉDITO_CLIENTE	Situação de crédito do cliente, se está apto para compras no crédito ou está bloqueado.
DURACAO_RELACIONAMENTO	Tempo desde a primeira compra do cliente, em dias.
TOTAL_PEDIDOS	Desde a primeira compra do cliente, quantidade total de pedidos que foram feitos.

SEGMENTO	Classificação feita pela própria empresa com base na duração do relacionamento e o volume de compras do cliente.
TEMPO_DE_CASA_RCA	Tempo em que o Representante que atende o cliente trabalha na empresa.
TICKET_MEDIO	Valor médio gasto pelo cliente em cada compra.

O *Random Forest*, por sua vez, foi escolhido por sua capacidade de capturar relações não lineares entre as variáveis e bom desempenho em bases desbalanceadas (Burez; van den Poel, 2009). Após o treinamento, o modelo também foi validado com as métricas de avaliação e variáveis.

Para validação dos modelos de regressão logística e *Random Forest*, foram utilizadas as métricas de Área Sob a Curva Característica de Operação do Receptor (*ROC AUC*, do inglês *Receiver Operating Characteristic Area Under the Curve*), *Recall*, Precisão e Acurácia (Tabela 2).

Os resultados de todo esse processo metodológico serão apresentados e analisados na próxima seção.

TABELA 2 – Descrição das métricas utilizadas

Métrica	Descrição
Acurácia	Proporção de previsões corretas sobre o total de casos.
Precisão	Proporção de positivos previstos que realmente são positivos.
Recall	Proporção de positivos reais corretamente identificados.
ROC AUC	Mede a capacidade do modelo em distinguir entre as classes.

4 Apresentação e análise dos resultados

O modelo desenvolvido por meio do algoritmo *Random Forest* apresentou resultados satisfatórios quanto à sua capacidade de previsão e utilidade prática para a empresa. Após o processo de treinamento e validação, os principais resultados obtidos estão apresentados na Tabela 3.

TABELA 3 – Métricas obtidas com a aplicação do algoritmo *Random Forest*

Métrica	Descrição
Acurácia	81,2%
Precisão	85,4%
Recall	85,2%
ROC AUC	90,9%

Essas métricas indicam que: em 81,2% das vezes, o modelo acerta se o cliente se tornará *churn* ou não; de todas as vezes em que o modelo aponta um cliente como *churn*, ele acerta 85,4% das vezes; em 85,2% das vezes em que um cliente se torna *churn*, o modelo consegue identificá-

lo; e em 90,9% das vezes, quando o modelo analisa um cliente que vai sair e um que vai ficar, ele atribui a pontuação de risco corretamente.

Logo, os indicadores demonstram que o modelo é eficaz na identificação dos clientes com maior probabilidade de evasão, apresentando um bom equilíbrio entre falso-positivos e falso-negativos, que é identificado através da métrica de Acurácia do modelo.

Já na regressão logística, os resultados foram fracos para implementação, o que ocorreu devido ao desbalanceamento de classes (clientes *churn* majoritários). Os indicadores para este modelo são apresentados na Tabela 4 a seguir.

Essas métricas indicam que: em 51,3% das vezes, o modelo acerta se o cliente se tornará *churn* ou não; de todas as vezes em que o modelo aponta um cliente como *churn*, ele acerta 52,5% das vezes; em 49,1% das vezes em que um cliente se torna *churn*, o modelo consegue identificá-lo; e em 54,6% das vezes, quando o modelo analisa um cliente que vai sair e um que vai ficar, ele atribui a pontuação de risco corretamente.

TABELA 4 – Métricas obtidas com a aplicação do algoritmo de regressão logística

Métrica	Descrição
Acurácia	51,3%
Precisão	52,5%
Recall	49,1%
ROC AUC	54,6%

Essas métricas indicam que: em 51,3% das vezes, o modelo acerta se o cliente se tornará *churn* ou não; de todas as vezes em que o modelo aponta um cliente como *churn*, ele acerta 52,5% das vezes; em 49,1% das vezes em que um cliente se torna *churn*, o modelo consegue identificá-lo; e em 54,6% das vezes, quando o modelo analisa um cliente que vai sair e um que vai ficar, ele atribui a pontuação de risco corretamente.

Quanto à aplicação prática do modelo, seus resultados poderão ser utilizados principalmente em três frentes:

- Melhoria conceitual do tempo de *churn*, que antes era considerado 90 dias de maneira fixa, e agora será considerado de forma dinâmica de acordo com o comportamento de cada cliente (percentil 90 do intervalo entre compras);
- Gamificação da predição de clientes *churn* em aplicação interna, onde os clientes com alto risco de *churn* serão apontados em destaque de forma personalizada na carteira do gestor responsável para que ele tenha ciência e atenda os clientes em risco, aumentando a saudabilidade da carteira e evitando a evasão;
- Ações segmentadas no programa de relacionamento interno, oferecendo ofertas e ações personalizadas para clientes em risco (como *cashback* mais atrativo e cupons de desconto).

É importante pontuar que a implementação do modelo de previsão de *churn* na empresa analisada representa um avanço significativo em relação às abordagens tradicionais utilizadas até então, que se baseavam predominantemente em análises descritivas e em decisões empíricas dos gestores comerciais. Enquanto os métodos anteriores dependiam de regras fixas ou percepções subjetivas sobre o comportamento dos clientes, o modelo atual utiliza algoritmos

de aprendizado de máquina para identificar padrões nos dados e antecipar com maior precisão os clientes com maior risco de evasão.

A performance do modelo demonstra que a solução possui boa capacidade preditiva, o que permite direcionar ações proativas com mais foco e eficiência, otimizando os recursos destinados à retenção. Comparado aos métodos tradicionais, que não consideravam múltiplas variáveis simultaneamente nem apresentavam uma estrutura de priorização clara dos clientes em risco, o novo modelo apresenta superioridade técnica e operacional.

Em relação à manutenção do modelo, a empresa já considera a necessidade de um plano de monitoramento contínuo, especialmente devido à possibilidade de mudanças no comportamento dos clientes ao longo do tempo. A proposta é realizar atualizações periódicas do modelo, utilizando dados mais recentes e reavaliando as variáveis preditoras conforme novas tendências ou produtos sejam introduzidos no mercado.

Quanto à viabilidade econômica, os custos de desenvolvimento e implementação do modelo foram considerados baixos, especialmente pelo uso de ferramentas já disponíveis na infraestrutura analítica da empresa (como Databricks e Tableau). No entanto, há a preocupação com a adoção prática do modelo pelas áreas comerciais, o que exige capacitação das equipes e integração dos resultados aos processos de CRM e atendimento. A geração de valor dependerá não apenas da qualidade técnica do modelo, mas da capacidade da empresa e seus funcionários em agir com base nas previsões geradas.

Por fim, é importante destacar algumas limitações do modelo. Uma delas é a dependência da qualidade dos dados históricos, o que pode comprometer a robustez da previsão caso haja inconsistências ou ausência de variáveis relevantes. Além disso, o modelo atual, embora eficaz, não considera aspectos qualitativos como o relacionamento interpessoal com o cliente, que podem influenciar sua decisão de permanência.

5 Considerações finais

Este trabalho teve como objetivo aplicar técnicas de *Machine Learning* à previsão de *churn* (evasão de clientes), com o intuito de apoiar uma empresa do setor varejista na identificação precoce de comportamentos de risco e, conseqüentemente, no desenvolvimento de estratégias mais eficazes de retenção. Ao longo do estudo, foi possível demonstrar como algoritmos de aprendizado podem contribuir para a tomada de decisões orientadas por dados.

O modelo *Random Forest*, escolhido como abordagem principal, mostrou-se eficaz na tarefa de classificação, alcançando bons resultados em métricas como acurácia, precisão e *recall*. Os resultados obtidos indicam que é possível prever, com razoável grau de confiança, quais clientes têm maior propensão a abandonar uma empresa, o que abre espaço para ações preventivas, aumentando as chances de retenção e reduzindo os custos associados à perda de clientes.

Do ponto de vista acadêmico e técnico, este trabalho contribuiu ao demonstrar, na prática, como métodos de *Machine Learning* podem ser efetivamente aplicados em contextos reais, considerando as restrições, limitações e desafios organizacionais. Para pesquisas futuras, existe a oportunidade de testar novas variáveis (por exemplo, região) e outros algoritmos (por exemplo, o *XGBoost*), que podem oferecer melhorias em desempenho. Além disso, modelos explicáveis, baseados em abordagens interpretativas, podem ser incorporados na tentativa de se obter maior transparência nos fatores que influenciam o *churn*. Também é possível enriquecer o modelo com dados qualitativos ou externos, como informações de satisfação do cliente ou indicadores econômicos.

Conclui-se, portanto, que o uso de *Machine Learning* é uma forma promissora para mitigar a evasão de clientes. Mais do que uma solução tecnológica, trata-se de um recurso estratégico

que, quando bem implementado, fortalece a relação com o cliente e amplia a competitividade das organizações. Espera-se que este trabalho contribua para futuras pesquisas e aplicações práticas, incentivando o uso inteligente de dados na construção de negócios mais sustentáveis e centrados no consumidor.

6 Referências

- ANDERSON, E. W. Customer satisfaction and word of mouth. **Journal of Service Research**, v. 1, n. 1, 1998. Disponível em: <https://doi.org/10.1177/109467059800100102>.
- BISHOP, C. M. **Pattern recognition and machine learning**. New York: Springer, 2006.
- BLATTBERG, R. C.; DEIGHTON, J. Manage marketing by the customer equity test. **Harvard Business Review**, v. 74, n. 4, 1996. Disponível em: <https://hbr.org/1996/07/manage-marketing-by-the-customer-equity-test>.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5-32. 2001. Disponível em: <https://doi.org/10.1023/A:1010933404324>
- BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, 2018. Disponível em: www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.
- BUREZ, J.; VAN DEN POEL, D. Handling class imbalance in customer churn prediction. **Expert Systems with Applications**, v. 36, n. 3, p. 4626–4636. 2009. Disponível em: <https://doi.org/10.1016/j.eswa.2008.05.027>.
- BUTTLE, F. **Customer Relationship Management: Concepts and Technologies**. 2. Ed. Burlington: Butterworth-Heinemann, 2009. Disponível em: https://repository.dinus.ac.id/docs/ajar/customer_relationship_management.pdf.
- COHEN, Y.; AMORIM, M.; REIS, J. Artificial Intelligence Trends and Applications in Service Systems. **Applied Sciences**, v. 12, n. 24, eLoc 13032. 2022. Disponível em: <https://doi.org/10.3390/app122413032>
- DOWLING, G. R.; UNCLES, M. D. Do customer loyalty programs really work? **Sloan Management Review**, v. 38, n. 4, 1997. Disponível em: <https://sloanreview.mit.edu/article/do-customer-loyalty-programs-really-work/>.
- EISENHARDT, K. M. Building theories from case study research. **Academy of Management Review**, v. 14, n. 4, 1989. Disponível em: https://josephmahoney.web.illinois.edu/BADM504_Fall%202019/Eisenhardt1989.pdf.
- GRÖNROOS, C. **Service management and marketing: managing the service profit logic**. 4. ed. Chichester: Wiley, 2015. Disponível em: <https://tashfeen.pbworks.com/f/Book%204%20-%20Service%20Management%20and%20Marketing.pdf>
- GUPTA, S.; ZEITHAML, V. A. Customer metrics and their impact on financial performance. **Marketing Science**, v. 25, n. 6, p. 718-739. 2006. Disponível em: <https://doi.org/10.1287/mksc.1060.0221>
- HAIR JR, J. F.; BLACK, W. C.; BABIN, C. J.; ANDERSON, R. E. **Multivariate data analysis**. 8. ed. Cengage, 2019. 813 p.
- HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3. ed. Burlington: Morgan Kaufmann, 2011.
- HOSMER JR, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied Logistic Regression**. 3. Ed. John Wiley & Sons, 2013. 528 p.

- JAIN, D.; SINGH, S. S. Customer lifetime value research in marketing: a review and future directions. **Journal of Interactive Marketing**, v. 16, n. 2, 2002. Disponível em: <https://doi.org/10.1002/dir.10032>
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning: with applications in R**. 2. Ed. New York: Springer, 2021. 607 p.
- KOTLER, P; KELLER, K. L. **Administração de marketing**. 15. ed. São Paulo: Pearson Universidades, 2017.
- LAM, S. Y.; SHANKAR, V.; ERRAMILI, K.; MURTHY, B. Customer satisfaction, loyalty, and switching costs: a longitudinal analysis. **Journal of the Academy of Marketing Science**, v. 32, n. 3, 2004, p. 293-311. Disponível em: <http://dx.doi.org/10.1177/0092070304263330>.
- MEYER, C.; SCHWAGER, A. Understanding customer experience. **Harvard Business Review**, v. 85, n. 2, 2007. Disponível em: <https://hbr.org/2007/02/understanding-customer-experience>
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 5. ed. Hoboken: Wiley, 2012.
- NESLIN, S.; GUPTA, S.; KAMAKURA, W.; LU, J.; MASON, C. Defection detection: measuring and understanding the predictive accuracy of customer churn models. **Journal of Marketing Research**, v. 43, n. 2, 2006. Disponível em: <https://doi.org/10.1509/jmkr.43.2.204>.
- PAYNE, A.; FROW, P. **Strategic customer management: integrating relationship marketing and CRM**. Cambridge: Cambridge University Press, 2013. Disponível em: <https://doi.org/10.1017/CBO9781139057417>
- PEPPERS, D.; ROGERS, M. **The One-to-One Future: Building Relationships One Customer at a Time**. New York: Currency Doubleday, 1993. 443 p.
- PEPPERS, D.; ROGERS, M. **Managing customer relationships: a strategic framework**. 2. ed. Hoboken: Wiley, 2011. 512 p.
- PFEIFER, P. E.; HASKINS, M. E.; CONROY, R. M. Customer lifetime value, customer profitability, and the treatment of acquisition spending. **Journal of Managerial Issues**, v. 17, n. 1, p. 11-25. 2005. Disponível em: <https://www.jstor.org/stable/40604472>.
- PROVOST, F.; FAWCETT, T. **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**. Sebastopol: O'Reilly Media, 2013.
- REICHHELD, F. F. TEAL, T. **The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value**. Boston: Harvard Business School Press, 1996. 323 p.
- ROSÁRIO, A. R.; BOECHAT, A. C. How Automated Machine Learning Can Improve Business. **Applied Sciences**, v. 14, n. 19, eLoc 8749. 2024. Disponível em: <https://doi.org/10.3390/app14198749>
- RUST, R. T.; ZAHORIK, A. J. Customer satisfaction, customer retention, and market share. **Journal of Retailing**, v. 69, n. 2, 1993. Disponível em: <https://link.gale.com/apps/doc/A14520380/AONE?u=anon~d5d50ec5&sid=googleScholar&xid=67f0e940>.
- SHETH, J. N.; PARVATIYAR, A. The evolution of relationship marketing. **International Business Review**, v. 4, n. 4, p. 397-418. 1995. Disponível em: [https://doi.org/10.1016/0969-5931\(95\)00018-6](https://doi.org/10.1016/0969-5931(95)00018-6)

UMOH, U. A.; EYOH, I. J.; MURUGESAN, V. S.; NYOHO, E. E. Fuzzy-machine learning models for the prediction of fire outbreaks: A comparative analysis. In: PANDEY, R.; KHATRI, S. K.; SINGH, N. K.; VERMA, P. (Eds). **Artificial Intelligence and Machine Learning for EDGE Computing**. O'Reilly Academic Press, 2022. P. 207-233.

VERHOEF, P. C.; REINARTZ, W. J.; KRAFFT, M. Customer engagement as a new perspective in customer management. **Journal of Service Research**, v. 13, n. 3. 2010. Disponível em: <https://doi.org/10.1177/1094670510375461>.

VERHOEF, P. C.; BROEKHUIZEN, T.; BART, Y.; BHATTACHARYA, A.; DONG, J. Q; FABIAN, N.; HAENLEIN, M. Digital Transformation: A Multidisciplinary Reflection and Research Agenda. **Journal of Business Research**, v. 122, p. 889-901. 2021. Disponível em: <https://doi.org/10.1016/j.jbusres.2019.09.022>.

VERHOEF, P. C.; LEEFLANG, P. S. H. Understanding the marketing department's influence within the firm: a customer-centric perspective. **Journal of Marketing**, v. 73, n. 2. 2009. Disponível em: <https://doi.org/10.1509/jmkg.73.2.14>

WOOLDRIDGE, J. M. **Introductory econometrics: a modern approach**. 5. ed. Mason: South-Western Cengage Learning, 2012.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 5. ed. Porto Alegre: Bookman, 2015. 320 p.