v7.n11.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

INDICADORES ANTECEDENTES NO APOIO À DECISÃO: PROJETO E AVALIAÇÃO DE UM ARTEFATO BASEADO EM REGRAS EM CAMPANHAS B2B COM CONVERSÃO *OFFLINE*

LEADING INDICATORS IN DECISION SUPPORT: DESIGN AND EVALUATION OF A RULE-BASED ARTIFACT IN B2B CAMPAIGNS WITH OFFLINE CONVERSION

INDICADORES ANTECEDENTES EN EL APOYO A LA DECISIÓN: DISEÑO Y EVALUACIÓN DE UN ARTEFACTO BASADO EN REGLAS EN CAMPAÑAS B2B CON CONVERSIÓN OFFLINE

Rafael Rodrigues Flor[1], André Luiz da Silva[2], Fabiana Florian[3]

e7119570
https://doi.org/10.47820/recima21.v7i11.9570

PUBLICADO: 11/2026

RESUMO

Sistemas de apoio à decisão pressupõem que o resultado a ser otimizado esteja registrado pelo sistema operacional e disponível no momento da decisão. Este artigo investiga o projeto de um artefato decisório quando o desfecho ocorre fora do sistema, é raro e só se torna observável após o período de decisão. Adotou-se como método a *Design Science Research*. O artefato é uma árvore de decisão baseada em regras explícitas que combina retorno sobre investimento publicitário, volume e qualidade de *leads*, classificando cada período em escalar, otimizar, testar ou pausar. Foi instanciado sobre quinze meses de uma conta real de anúncios e avaliado por aplicação retroativa. A versão inicial convergiu com as decisões históricas em 27% dos períodos, pois o retorno mensal não discrimina desempenho quando o desfecho é raro. A versão revisada, com retorno acumulado, estado de resultado não observado e prioridade à qualidade, alcançou 47%, abaixo da regra trivial de sempre otimizar, indicando que a convergência mede aderência à prática, e não eficácia. Avaliadas pelo desfecho, as divergências favoreceram o artefato em um período e o desfavoreceram em outro. Conclui-se que a decisão nessas condições depende de indicadores antecedentes, cabendo ao indicador de resultado a função de confirmação posterior.

PALAVRAS-CHAVE: Árvore de decisão. Conversão offline. Design Science Research. Indicadores antecedentes. Qualidade da informação. Sistemas de apoio à decisão.

ABSTRACT

Decision support systems assume that the outcome to be optimized is recorded by the operational system and available when the decision is made. This article investigates the design of a decision artifact when the outcome occurs outside the system, is rare, and only becomes observable after the decision period. Design Science Research was adopted as the methodological framework. The artifact is a rule-based decision tree combining return on ad spend, lead volume and lead quality, classifying each period as scale, optimize, test or pause. It was instantiated over fifteen months of a real advertising account and evaluated through retroactive application. The initial version converged with historical decisions in 27% of the periods, because monthly return does not discriminate performance when the outcome is rare.

[1] Graduando em Sistemas de Informação pela Universidade de Araraquara (UNIARA) e gestor de tráfego pago.
[2] Doutor em Educação Escolar pela Universidade Estadual Paulista (UNESP) e professor da Universidade de Araraquara (UNIARA), com atuação em arquitetura de dados e inteligência artificial.
[3] Doutora em Alimentos e Nutrição pela Universidade Estadual Paulista (UNESP) e professora da Universidade de Araraquara (UNIARA).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

The revised version, with accumulated return, an unobserved outcome state and priority to lead quality, reached 47%, below the trivial rule of always optimizing, which indicates that convergence measures adherence to practice rather than effectiveness. Assessed against later outcomes, the divergences favored the artifact in one period and disfavored it in another. It is concluded that decision making under these conditions depends on leading indicators, with the outcome indicator serving as subsequent confirmation.

KEYWORDS: *Decision support systems. Decision tree. Design Science Research. Information quality. Leading indicators. Offline conversion.*

RESUMEN

Los sistemas de apoyo a la decisión presuponen que el resultado que se desea optimizar está registrado por el sistema operativo y disponible en el momento de la decisión. Este artículo investiga el diseño de un artefacto decisorio cuando el resultado ocurre fuera del sistema, es raro y solo se vuelve observable después del período de decisión. Se adoptó como método la Design Science Research. El artefacto es un árbol de decisión basado en reglas explícitas que combina el retorno de la inversión publicitaria, el volumen y la calidad de los leads, y clasifica cada período en escalar, optimizar, probar o pausar. Se instanció sobre quince meses de una cuenta real de anuncios y se evaluó mediante aplicación retroactiva. La versión inicial coincidió con las decisiones históricas en el 27% de los períodos, pues el retorno mensual no discrimina el desempeño cuando el resultado es raro. La versión revisada, con retorno acumulado, estado de resultado no observado y prioridad a la calidad, alcanzó el 47%, por debajo de la regla trivial de optimizar siempre, lo que indica que la convergencia mide la adherencia a la práctica, y no la eficacia. Evaluadas por el resultado posterior, las divergencias favorecieron al artefacto en un período y lo desfavorecieron en otro. Se concluye que la decisión en estas condiciones depende de indicadores antecedentes, y que al indicador de resultado le corresponde la función de confirmación posterior.

PALABRAS CLAVE: *Árbol de decisión. Calidad de la información. Conversión offline. Design Science Research. Indicadores antecedentes. Sistemas de apoyo a la decisión.*

1. INTRODUÇÃO

Os Sistemas de Apoio à Decisão (*Decision Support Systems*, doravante DSS) auxiliam gestores na resolução de problemas semiestruturados, nos quais dados objetivos se combinam ao julgamento humano (LAUDON; LAUDON, 2001). A disponibilidade de dados e de ferramentas analíticas, contudo, não garante por si só decisões eficazes, sendo necessário estruturar modelos que orientem a interpretação e o uso dessas informações de forma sistemática.

A eficácia de um sistema de apoio à decisão depende de uma condição frequentemente implícita: a de que o resultado que deve orientar a decisão esteja registrado pelo sistema operacional e disponível no momento em que a decisão precisa ser tomada. Existe, contudo, uma classe de problemas organizacionais em que essa condição não se verifica. Nela, o desfecho que efetivamente interessa ocorre fora do sistema que registra as operações, é raro em relação ao volume de eventos observados e só se torna observável depois de encerrado o período em que a decisão foi tomada. Sob essas condições, os indicadores disponíveis no

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

momento da decisão são necessariamente antecedentes ao resultado, e o indicador de resultado propriamente dito só pode cumprir função de confirmação posterior.

Essa configuração não é particular a um único domínio: apresenta-se na análise de crédito, na triagem clínica e na seleção de pessoal, entre outros contextos em que a decisão precede a observação do desfecho. Manifesta-se de forma especialmente nítida na gestão de campanhas de tráfego pago em ambientes B2B, nos quais a conversão se completa fora do ambiente digital, por canais como aplicativos de mensagem ou reuniões presenciais. Nesse contexto, existe desconexão estrutural entre os dados capturados pelas plataformas e os resultados efetivamente obtidos pelas organizações, o que compromete a avaliação de desempenho e a qualidade das decisões (SAURA; RIBEIRO-SORIANO; PALACIOS-MARQUÉS, 2021).

Problemas dessa natureza são semiestruturados, pois envolvem métricas quantitativas e elementos qualitativos, como a percepção sobre a adequação dos contatos recebidos. As árvores de decisão baseadas em regras constituem abordagem adequada a eles, por formalizarem o raciocínio decisório em estruturas lógicas interpretáveis sem depender de grandes volumes de dados ou de algoritmos de aprendizado de máquina, o que as torna pertinentes em contextos com baixo volume de desfechos e alto valor por transação.

Diante do exposto, o presente estudo busca responder à seguinte questão de pesquisa: como projetar um sistema de apoio à decisão quando o sistema operacional não registra o resultado que deveria orientar a decisão, e esse resultado só se torna observável depois do período em que a decisão precisa ser tomada? O contexto de aplicação adotado é o de campanhas de tráfego pago em ambiente B2B com conversão *offline*. Parte-se da hipótese de que a utilização de um modelo baseado em regras, estruturado a partir de indicadores antecedentes e de *feedback* qualitativo, reduz a subjetividade no processo decisório e aumenta a consistência das avaliações realizadas.

O objetivo geral deste trabalho é projetar e avaliar um artefato de apoio à decisão baseado em regras explícitas para contextos em que o indicador de resultado é incompleto e tardio. Os objetivos específicos compreendem: identificar e justificar os indicadores observáveis no momento da decisão; estruturar regras de decisão interpretáveis a partir desses indicadores; instanciar o artefato sobre dados reais de operação; e avaliar o artefato por aplicação retroativa, confrontando suas recomendações com as decisões efetivamente tomadas.

A justificativa do estudo é dupla. No plano teórico, a literatura discute com menor profundidade o projeto de artefatos decisórios para situações em que o indicador de resultado é estruturalmente indisponível no momento da decisão. No plano prático, operações com recursos

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

técnicos limitados tomam decisões recorrentes sobre investimento publicitário sem critérios explícitos, o que dificulta a consistência e a auditoria dessas decisões.

2. REFERENCIAL TEÓRICO

Os sistemas de informação desempenham papel essencial nas organizações ao transformar dados em informações relevantes para a gestão. Destacam-se os sistemas de apoio à decisão, projetados para auxiliar gestores na resolução de problemas complexos e semiestruturados. Segundo Laudon e Laudon (2001, p. 318), um sistema de suporte à decisão "assiste a tomada de decisão gerencial pela combinação de dados, modelos e ferramentas analíticos sofisticados e softwares amigáveis ao usuário em um único e poderoso sistema que pode suportar tomadas de decisões semi-estruturadas ou não estruturadas".

A tomada de decisão orientada a dados tornou-se elemento central na gestão organizacional. Organizações que utilizam dados de forma intensiva apresentam melhor desempenho (MCAFEE; BRYNJOLFSSON, 2012), e o *Business Intelligence and Analytics* emergiu como área fundamental tanto na academia quanto na prática organizacional (CHEN; CHIANG; STOREY, 2012). O valor dos dados, porém, depende da capacidade de transformá-los em ações práticas por meio de processos estruturados de análise (PROVOST; FAWCETT, 2013). Troisi *et al.* (2020) demonstram, a partir do estudo de três organizações, que estratégias orientadas a dados aceleram ciclos decisórios e aumentam a assertividade das ações.

A eficácia dessas abordagens, contudo, está condicionada à qualidade da informação disponível. Wang e Strong (1996) demonstram que a qualidade de dados não se reduz à acurácia do registro, compreendendo também a adequação do dado ao uso pretendido pelo consumidor da informação. Um sistema operacional pode, assim, registrar corretamente aquilo que foi configurado para registrar e, ainda assim, produzir indicador inadequado à decisão, sem que a inadequação seja visível na saída.

No marketing digital, métricas como cliques, impressões e taxas de conversão permitem acompanhamento em tempo real, mas não devem ser analisadas isoladamente (JÄRVINEN; KARJALUOTO, 2015). Estratégias orientadas a dados impactam o crescimento e a lucratividade quando estruturadas de forma coerente com os objetivos do negócio (GRANDHI; PATWA; SALEEM, 2021), e a análise de desempenho tornou-se atividade central e contínua, e não mais processo pontual de avaliação (SHAH; MURTHI, 2021).

A limitação das métricas analisadas isoladamente torna-se mais evidente em ambientes B2B, nos quais os processos de venda envolvem múltiplos pontos de contato e as interações iniciadas em canais digitais podem ser concluídas em ambientes *offline*. Saura, Ribeiro-Soriano

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

e Palacios-Marqués (2021) evidenciam que essa desconexão é estrutural e demanda modelos analíticos capazes de integrar dados de diferentes origens. Karczmarczyk, Jankowski e Wątróbski (2018) argumentam que avaliar campanhas por métrica única não corresponde à prática, que pondera cobertura, custo e duração, e propõem um sistema de apoio à decisão multicritério baseado no método PROMETHEE II.

A literatura de gestão por processos acrescenta que a formalização de atividades recorrentes reduz erros e aumenta a consistência, sobretudo quando as decisões precisam seguir critérios objetivos e reproduzíveis (DUMAS *et al.*, 2018).

Os trabalhos que modelam o atraso da conversão (CHAPELLE, 2014) ou pontuam *leads* por aprendizado de máquina (NYGÅRD; MEZEI, 2020) pressupõem grande volume de desfechos registrados pela plataforma, condição ausente quando o desfecho é raro e ocorre fora dela. O trabalho que estrutura a decisão em múltiplos critérios explícitos (KARCZMARCZYK; JANKOWSKI; WĄTRÓBSKI, 2018) opera sobre desfechos observáveis de imediato, e os demais descrevem o uso de dados na decisão sem propor artefato avaliado contra o desfecho.

Persiste, portanto, uma lacuna: a literatura consolidada trata da integração entre fontes de dados e da qualidade do registro, mas discute com menor profundidade o projeto de artefatos decisórios para situações em que o indicador de resultado é estruturalmente indisponível no momento da decisão. É essa lacuna que o presente trabalho aborda.

O Quadro 1 sintetiza os trabalhos relacionados quanto ao objetivo, ao método e à forma de avaliação, ao resultado e à limitação em relação ao problema investigado neste artigo.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Quadro 1. Matriz comparativa dos trabalhos relacionados

Trabalho	Objetivo	Método e avaliação	Resultado principal	Limitação para o problema deste artigo
Järvinen e Karjaluoto (2015)	Compreender como empresas industriais usam a análise *web* para medir o desempenho do marketing digital	Estudo de caso em profundidade em empresas industriais (B2B), com análise qualitativa do uso das métricas	Os benefícios dependem da escolha das métricas, do tratamento dos dados e do contexto organizacional	Não propõe artefato decisório nem trata desfecho ocorrido fora do sistema
Karczmarczyk, Jankowski e Wątróbski (2018)	Apoiar o planejamento e a avaliação de campanhas de marketing viral com múltiplos critérios	PROMETHEE II em rede sintética, verificado em rede real de 7.610 nós, com sensibilidade aos pesos	Recomendações da rede sintética mantidas na rede real (correlação de 0,76 entre os rankings)	Desfecho simulado e observável de imediato; sem conversão *offline*
Chapelle (2014)	Estimar a taxa de conversão quando a conversão ocorre dias depois do clique	Modelo probabilístico de conversão e atraso, comparado em dados de anúncios de *display* a modelos que ignoram o atraso	Modelar o atraso reduz o viés de tratar como negativas conversões ainda não ocorridas	Exige grande volume de conversões registradas pela plataforma
Nygård e Mezei (2020)	Automatizar a pontuação de *leads* pela probabilidade de compra	Floresta aleatória sobre dados comportamentais históricos, com teste de formas de agregação das séries	O aprendizado supervisionado estima a probabilidade de compra dos *leads*	Depende de histórico volumoso de compras; não orienta a decisão sobre a campanha
Troisi *et al.* (2020)	Investigar como análise de dados e computação cognitiva apoiam decisões de marketing B2B	Pesquisa-ação em três empresas B2B com o modelo de *growth hacking* e observação participante	Identifica estratégias e táticas de uso de dados na decisão de marketing	Não formaliza regras nem confronta decisões com o desfecho
Saura, Ribeiro-Soriano e Palacios-Marqués (2021)	Mapear o uso de CRM baseado em inteligência artificial no marketing digital B2B	Revisão da literatura com análise de correspondência múltipla	Classifica tipos de CRM e técnicas de inteligência artificial e propõe agenda de pesquisa	Trata a integração de dados, não a decisão sob desfecho raro e tardio
Este trabalho	Apoiar a decisão periódica sobre campanhas quando o desfecho é *offline*, raro e tardio	Árvore de regras sob *Design Science Research* em quinze meses de uma conta real; convergência, kappa, regras de referência e leitura das divergências pelo desfecho	Convergência de 47% na versão revisada, abaixo da regra de referência; mérito parcial pelo desfecho	Instanciação única; qualidade dos *leads* subjetiva; série curta

Fonte: elaborado pelos autores (2026).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

3. METODOLOGIA

3.1. Delineamento da pesquisa

A pesquisa é aplicada, de natureza quali-quantitativa e de caráter exploratório e descritivo, e adota a *Design Science Research*, na qual o artefato construído e avaliado para resolver um problema organizacional relevante constitui a contribuição científica (HEVNER *et al.*, 2004). Seguiram-se as etapas propostas por Peffers *et al.* (2007): identificação do problema, definição dos objetivos da solução, projeto e desenvolvimento, demonstração, avaliação e comunicação. A contribuição pretendida é a instanciação de um artefato em contexto real, nível que Gregor e Hevner (2013) admitem como válido desde que declarado como tal.

3.2. Caso e dados

O artefato foi instanciado sobre quinze meses consecutivos de operação de uma única conta de Google Ads, de junho de 2025 a agosto de 2026, com dados coletados mediante autorização formal do cliente responsável pelas campanhas. Os dados são agregados por mês e não incluem informações pessoais dos contatos. O primeiro autor atuou como gestor da conta durante todo o período analisado, responsável pelas decisões sobre as campanhas, de modo que a avaliação foi conduzida pelo próprio praticante.

A estrutura de campanhas variou ao longo da série. Em cinco dos quinze meses houve duas campanhas simultâneas: na substituição de uma campanha de pesquisa por outra, em outubro de 2025 e em abril de 2026, e na operação de uma campanha do tipo Performance Max em paralelo à principal, de janeiro a março de 2026. Considerou-se substituída a campanha com gasto residual inferior a R$ 50,00 no mês. Por essa variação, a unidade de análise é o período mensal, e não a campanha, pois o objeto de decisão do artefato é a conta em operação ao longo do tempo.

3.3. Procedimento de avaliação

A avaliação aplicou retroativamente as regras descritas na seção 4.2 aos quinze meses da série e comparou a classificação do artefato com a decisão tomada pelo gestor em cada período. As decisões reais foram reconstituídas a partir de duas fontes: a alteração da estrutura de campanhas registrada nos relatórios mensais, evidência documental que cobre quatro períodos, e o relato do gestor sobre a prática de rotina nos onze meses sem alteração estrutural, correspondente ao refinamento de palavras-chave e segmentação.

Adotou-se como medida a convergência com decisões históricas, definida como a proporção de períodos em que a classificação do artefato coincide com a decisão efetivamente

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

tomada. A medida expressa aderência à prática do gestor, e não acerto, uma vez que a decisão real não constitui gabarito: o artefato pode divergir de uma decisão que o desfecho posterior mostrou equivocada. Reporta-se também o coeficiente kappa de Cohen (COHEN, 1960), que desconta a concordância esperada por acaso, e as divergências são examinadas à luz do desfecho na seção 4.8.

Como o gestor da conta é o primeiro autor, a referência de avaliação foi produzida por quem formulou o artefato. As regras da versão inicial foram escritas após abril de 2026, de modo que as decisões de junho de 2025 a abril de 2026, que incluem os quatro períodos com decisão documentada e a interrupção da campanha Performance Max, não poderiam ter sido influenciadas por elas. De maio a agosto de 2026, essa influência não pode ser descartada, mas os dados não a indicam: o gestor manteve a otimização nos quatro meses, enquanto a versão inicial convergiu em um e a revisada, em dois. Restrita aos onze períodos anteriores às regras, a convergência foi de três períodos na versão inicial e de cinco na revisada, com kappa de 0,05 e 0,22, respectivamente, valores próximos aos da série completa.

O Quadro 2 distingue as etapas de definição e de avaliação do artefato. Apenas a versão inicial foi avaliada com parâmetros fixados antes do exame da série. As etapas seguintes utilizaram os mesmos quinze períodos, e seus resultados devem ser lidos como ajuste ou demonstração dentro da amostra; a divisão temporal e a calibração expansível, descritas na seção 4.7, mantêm fora da amostra apenas os limiares de volume.

Quadro 2. Etapas de calibração e avaliação do artefato

Etapa	O que foi definido	Períodos usados para definir	Períodos usados para avaliar	Natureza da evidência
Versão inicial	Regras, piso de ROAS 2, limiares 8 e 11	Nenhum: percepção do gestor, antes da análise	jun/2025 a ago/2026	Avaliação retroativa com parâmetros prévios
Calibração dos limiares	Limiares 9 e 12	jun/2025 a ago/2026	Nenhum	Ajuste dentro da amostra
Versão revisada	Regras R1 a R4, janela de 12 meses, defasagem de 1 mês	jun/2025 a ago/2026	jun/2025 a ago/2026	Demonstração dentro da amostra
Divisão temporal	Limiares 6 e 11	jun/2025 a jan/2026	fev/2026 a ago/2026	Limiares fora da amostra; regras não
Calibração expansível	Limiares de cada período	Períodos anteriores ao avaliado	out/2025 a ago/2026	Limiares sem informação futura; regras não

Fonte: elaborado pelos autores (2026).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

4. RESULTADOS E DISCUSSÃO

4.1. Arquitetura e indicadores do artefato

O artefato é um sistema de apoio à decisão orientado a dados, organizado em quatro camadas. Na coleta, o gestor exporta a cada mês, em formato PDF, o relatório de uma plataforma de monitoramento vinculada à conta de anúncios, etapa que permanece manual e restrita a uma única plataforma. No tratamento, uma rotina automatizada converte os relatórios em registros mensais rastreáveis até o arquivo de origem e verifica a plausibilidade das entradas, requisito cuja origem é discutida na seção 4.3. Na visualização, uma planilha calcula cada indicador por fórmula explícita, o que privilegia a auditabilidade do cálculo. Na decisão, uma árvore de regras explícitas classifica cada período em escalar, otimizar, testar ou pausar.

A opção por regras explícitas, e não por algoritmos de aprendizado de máquina, decorre do contexto: o baixo volume de desfechos, o alto valor por transação e a necessidade de que o artefato seja auditável pelo gestor favorecem regras interpretáveis, uma vez que a escolha do método deve considerar o volume de dados disponível, a interpretabilidade necessária e as condições operacionais do ambiente de aplicação (PROVOST; FAWCETT, 2013). Karczmarczyk, Jankowski e Wątróbski (2018) oferecem precedente no mesmo domínio, ao avaliar campanhas por critérios explícitos que refletem as preferências do decisor.

O artefato opera com três indicadores de decisão: retorno sobre investimento publicitário (*Return on Ad Spend*, doravante ROAS), volume de *leads* e qualidade dos *leads*, escolhidos por refletirem os resultados relevantes para o negócio, como recomendam Järvinen e Karjaluoto (2015). Custo por aquisição e taxa de cliques são mantidos como métricas de diagnóstico, que informam a interpretação do gestor sem integrar as regras de classificação.

O ROAS constitui o indicador de resultado do artefato. É calculado a partir do valor do contrato informado pelo cliente, uma vez que o desfecho ocorre fora do ambiente digital. Cada contrato é atribuído ao período em que a mensagem do *lead* foi recebida, e não ao período da assinatura. Nos três contratos observados na série, as janelas de fechamento atravessaram a virada do mês, de modo que a atribuição pelo período de assinatura deslocaria a receita em um período em relação àquele que efetivamente originou o contato.

O volume de *leads* e a qualidade dos *leads* são tratados como indicadores distintos, e não como medida única de *leads* qualificados. O volume corresponde ao número de eventos de clique no botão de contato registrados pelo *pixel* instalado no site, medida objetiva e reproduzível. A qualidade decorre do *feedback* do cliente quanto à adequação dos contatos ao perfil do produto e é, por natureza, subjetiva. A separação é deliberada: como demonstrado na seção 4.6, os dois

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

indicadores variaram em sentidos opostos no período analisado, e sua fusão em métrica única suprimiria justamente a informação que orienta a decisão correta.

Para que a qualidade possa ser avaliada de forma reproduzível, o artefato especifica uma rubrica aplicada a cada contato, a partir da primeira conversa, conforme o Quadro 3. Considera-se qualificado o contato que atende aos critérios de pertinência, perfil e demanda, e a qualidade do período corresponde à proporção de contatos qualificados: boa a partir de 50%, regular entre 25% e 50% e ruim abaixo de 25%. Os cortes são preliminares e devem ser calibrados a partir dos primeiros períodos classificados. O critério de especificidade é registrado, mas não integra a classificação, para que se verifique posteriormente se antecipa o desfecho. A rubrica não considera características pessoais do contato, como idade, mas apenas o conteúdo da mensagem.

Quadro 3. Rubrica de classificação da qualidade dos leads

Critério	Pergunta ao avaliador	Evidência aceita
C1 Pertinência	A mensagem não é candidatura a emprego, oferta de fornecedor, engano, propaganda ou mensagem sem conteúdo?	Texto da mensagem
C2 Perfil	O contato representa empresa ou atividade que compra o tipo de produto oferecido?	Nome da empresa, cargo, atividade ou local de uso citados na conversa
C3 Demanda	A mensagem pede produto ou serviço do portfólio, ou informação para comprá-lo?	Produto, orçamento, preço ou visita citados
C4 Especificidade	A mensagem traz ao menos um dado concreto: quantidade, medida, prazo ou local?	Trecho com o dado

Fonte: elaborado pelos autores (2026).

A aplicação prevê dois avaliadores independentes, sem conhecimento do período nem da campanha de origem, o registro do trecho da conversa que justifica cada critério e o cálculo do kappa de Cohen (COHEN, 1960) para verificar a concordância entre eles.

A taxa de conversão, inicialmente prevista como métrica de suporte, foi excluída do artefato pela inadequação do evento de conversão discutida na seção 4.3. O retorno sobre investimento foi igualmente excluído, pois exige o lucro líquido da empresa, informação não disponibilizada para esta pesquisa, o que torna o ROAS a medida de retorno adequada ao contexto estudado.

4.2. Árvore de decisão da versão inicial

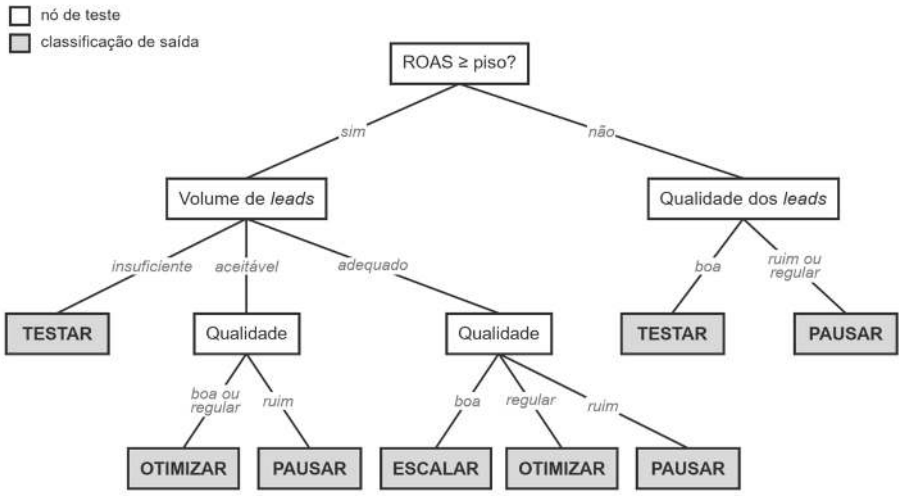
A versão inicial da árvore, apresentada na Figura 1, adota faixas preliminares definidas a partir da percepção do gestor antes da análise da série, confrontadas com a distribuição observada na seção 4.4. O ROAS é aceitável quando igual ou superior ao piso de 2, isto é,

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

receita atribuída equivalente ao dobro do investimento, retorno mínimo que, segundo o gestor, justifica manter o investimento, e crítico abaixo dele. O volume de *leads* é insuficiente abaixo de 8 contatos no mês, aceitável de 8 a 11 e adequado acima de 11. A qualidade é classificada como boa, regular ou ruim a partir do *feedback* do cliente. Complementa essas faixas a referência do negócio, segundo a qual três contratos por ano provenientes das campanhas já representam alto desempenho, dado o valor médio por transação.

Escalar exige ROAS aceitável, volume adequado e qualidade boa. Otimizar aplica-se quando ROAS e qualidade são satisfatórios, mas o volume está na faixa aceitável, e não na adequada, ou quando ROAS e volume são satisfatórios, mas a qualidade é regular, casos que pedem ajuste de palavras-chave ou segmentação. Testar, isto é, desenvolver nova estrutura em paralelo, aplica-se quando o volume é insuficiente para conclusões consistentes ou quando os *leads* chegam qualificados com retorno abaixo do piso, o que aponta problema na estrutura da campanha, e não na segmentação. Pausar aplica-se quando o retorno está abaixo do piso com qualidade regular ou ruim, ou quando predominam contatos desqualificados com volume aceitável ou elevado, sinal de segmentação muito aberta. A qualidade baixa é condição necessária em todos os ramos que levam a pausar: o volume elevado, por si só, não é indicador negativo, e o artefato não recomenda interrupção em período com qualidade boa, ainda que o retorno esteja abaixo do piso.

Figura 1. Árvore de decisão do artefato, versão inicial



Volume de leads: insuficiente abaixo de 8 contatos no mês; aceitável de 8 a 11; adequado acima de 11.

Fonte: elaborado pelos autores (2026).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

4.3. Qualidade do sinal de resultado

A instanciação exigiu, previamente, definir qual medida de conversão utilizar. A coluna de conversões da plataforma mostrou-se inadequada: em seis dos quinze meses, a taxa de conversão informada excedeu 100%, com máximo de 129,17% em fevereiro de 2026, valor logicamente impossível, pois indicaria mais conversões do que cliques. O evento de conversão havia sido configurado para contabilizar visualizações de página, e não a ação de contato pretendida. A plataforma registrou fielmente o evento tal como especificado, de modo que a inadequação decorre da especificação da medida, e não de falha da ferramenta. O caso ilustra a distinção de Wang e Strong (1996) entre acurácia do registro e adequação ao uso: o dado é acurado em relação ao que foi definido e inadequado em relação ao que se pretende decidir. A distorção, além disso, não é sinalizada na interface, que apresenta o valor sob o rótulo de conversões.

Um artefato de decisão construído sobre um sistema operacional herda, portanto, decisões de configuração que não são visíveis na saída. Desse achado decorre a verificação de plausibilidade incorporada à camada de tratamento: foi ela, e não a interface da plataforma, que tornou o problema detectável. Adotou-se em substituição o evento de *pixel* correspondente ao clique no botão de contato. Como cada relatório mensal informa também o valor do mês anterior, o confronto entre as duas leituras permitiu verificar a integridade dessa série, sem divergência nos quinze meses, de modo que o descarte recai sobre a coluna de conversões da plataforma, e não sobre parte do período.

4.4. Série analisada e calibração dos limiares

No período analisado, a conta registrou 173 eventos de contato e três contratos fechados provenientes das campanhas, volume compatível com a referência de desempenho do negócio apresentada na seção 4.2. O investimento total foi de R$ 16.877,29 e a receita atribuída às campanhas alcançou R$ 538.000,00, resultando em ROAS acumulado de 31,9. A Tabela 1 apresenta a série mensal do caso.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Tabela 1. Série mensal do caso analisado

Mês	Investimento (R$)	Volume de *leads*	Qualidade dos *leads*	Campanhas em operação
jun/2025	912,28	5	boa	uma
jul/2025	748,23	11	boa	uma
ago/2025	779,07	7	boa	uma
set/2025	910,39	5	boa	uma
out/2025	1.131,14	6	boa	duas (substituição)
nov/2025	1.300,35	10	boa	uma
dez/2025	1.202,36	11	boa	uma
jan/2026	1.304,69	16	ruim	duas (Performance Max)
fev/2026	1.273,50	14	ruim	duas (Performance Max)
mar/2026	1.122,26	26	ruim	duas (Performance Max)
abr/2026	1.231,64	18	boa	duas (substituição)
mai/2026	1.200,01	13	boa	uma
jun/2026	1.183,50	11	boa	uma
jul/2026	1.216,10	8	boa	uma
ago/2026	1.361,77	12	boa	uma
Total	16.877,29	173		

Fonte: elaborado pelos autores (2026).

O volume mensal de *leads* variou entre 5 e 26 eventos, com mediana de 11 e desvio padrão de 5,55. Os limiares preliminares de 8 e 11 eventos, definidos na seção 4.2 a partir da percepção do gestor, aproximam-se dos tercis da distribuição observada, de 9,3 e 12,3 eventos, e foram ajustados para 9 e 12, passando a decorrer da série observada, e não de estimativa. Aplicados à versão inicial, os limiares calibrados reduziriam sua convergência de quatro para três períodos, pois julho de 2026, com oito contatos, passaria da classificação otimizar para testar. Por essa razão, a avaliação da versão inicial mantém os limiares preliminares, definidos antes da análise, e os calibrados são empregados apenas na versão revisada (seção 4.7).

Registre-se uma condição de escopo relativa a esses limiares. O volume de contatos é função do orçamento aplicado, e a correlação observada entre investimento e volume foi de 0,39 no período completo e 0,49 quando excluídos os meses de operação da campanha Performance Max. Como o investimento mensal variou por fator de 1,82 entre o mínimo e o máximo, os limiares calibrados são válidos para a faixa de orçamento observada, não sendo extrapoláveis para patamares substancialmente distintos.

O mesmo procedimento não se mostrou aplicável ao ROAS. Com três contratos em quinze meses, o ROAS mensal foi nulo em doze períodos e assumiu valores de 80,7, 30,9 e 331,4 em dezembro de 2025, abril de 2026 e julho de 2026, de modo que o primeiro quartil, a mediana e o terceiro quartil da distribuição são todos iguais a zero. A distribuição é, portanto, degenerada, e qualquer limiar situado entre 1 e 30 classifica os quinze meses de maneira

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

idêntica. Essa constatação não decorre de insuficiência da coleta, mas da natureza do negócio, no qual o desfecho é evento raro por definição.

4.5. Validação retroativa da versão inicial

A versão inicial convergiu com as decisões históricas em quatro dos quinze períodos (Tabela 2, seção 4.7), o que corresponde a 27%, com kappa de 0,05, valor próximo ao esperado por acaso. A divergência concentra-se em causa única e estrutural. Nos doze meses sem contrato atribuído à coorte, o ROAS mensal é nulo e o artefato o interpreta como retorno insatisfatório, classificando o período como testar ou pausar, enquanto o gestor tratou o mês sem contrato como situação esperada, dada a referência de três contratos anuais, e manteve o refinamento de palavras-chave e segmentação, correspondente a otimizar. O ROAS, tomado na granularidade mensal, não discrimina desempenho quando o desfecho constitui evento raro. Sob o enquadramento adotado, uma avaliação que identifica limitação de projeto e orienta a revisão do artefato integra o próprio ciclo de pesquisa (PEFFERS *et al.*, 2007), e não configura invalidação da proposta.

4.6. O caso da campanha Performance Max

Entre janeiro e março de 2026 operou, em paralelo à campanha principal, uma campanha do tipo Performance Max destinada à captação de *leads*, único período da série com duas campanhas de naturezas distintas em operação simultânea. Ela gerou 14.358 cliques, contra 2.563 da campanha de pesquisa, com investimento de R$ 933,92. Os três meses concentraram 56 dos 173 eventos de contato da série (32%). O *feedback* do cliente quanto à qualidade desses contatos foi classificado como ruim, incluindo pessoas fora do perfil do produto e mensagens sem relação com a oferta, e nenhum contrato foi atribuído às coortes desses três meses.

O contraste com o restante da série constitui o achado central desta avaliação. Março de 2026 registrou o maior volume mensal de contatos de todo o período, 26 eventos, e não gerou contrato algum. Julho de 2026 registrou 8 eventos, valor abaixo do primeiro quartil da distribuição, e originou o contrato de maior valor da série, correspondente a 75% da receita (Tabela 1). Uma regra orientada exclusivamente pelo volume recomendaria ampliar o investimento em março e trataria julho como período de desempenho insuficiente, equivocando-se em ambos os casos. O resultado sustenta empiricamente a decisão de projeto de tratar volume e qualidade como indicadores distintos (seção 4.1).

A instanciação não pôde aplicar a rubrica descrita na seção 4.1, porque o histórico de conversas não estava disponível. A qualidade foi informada pelo cliente em julgamento retrospectivo, por período, e classificada como ruim exatamente nos três meses de operação da

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

campanha Performance Max, o que impede tratar a recomendação do artefato nesses meses como evidência independente: com qualidade boa em todos os períodos, a versão revisada recomendaria escalar de janeiro a março de 2026. Segundo o cliente, a inadequação dos contatos foi percebida por volta do final de fevereiro de 2026. Com a informação disponível em cada momento, a versão revisada teria recomendado escalar em janeiro, e a inicial, testar, e ambas teriam recomendado pausar ao final de fevereiro, antecipando a interrupção em cerca de um mês. O investimento da campanha em março somou R$ 292,57, correspondente a 31% do total aplicado nela, sem gerar contrato. A antecipação depende da data exata da percepção: tivesse ela ocorrido já em março, o artefato não teria antecipado a decisão do gestor.

Indicadores registrados pela própria plataforma, independentes do julgamento do cliente e disponíveis no mês da decisão, são coerentes com essa avaliação: a campanha Performance Max recebeu 25% do investimento do trimestre e gerou 85% dos cliques, com custo por clique entre R$ 0,05 e R$ 0,11, contra valores entre R$ 1,05 e R$ 1,12 na campanha de pesquisa, e taxa de cliques de 19,1% em fevereiro de 2026. O padrão indica tráfego de baixo custo e grande volume, compatível com o relato de contatos fora do perfil, mas não mede a qualidade dos contatos, uma vez que o evento de contato não é atribuído por campanha. Esses indicadores não foram incorporados às regras, pois um limiar definido a partir deste episódio configuraria sobreajuste.

Registra-se aqui a principal limitação de atribuição. O evento de *pixel* é registrado no âmbito do site, e não por campanha de origem, e inclui contatos de tráfego orgânico e direto, de modo que não é possível determinar quantos dos 26 contatos de março de 2026 decorreram da campanha Performance Max. A distinção entre as campanhas apoia-se no *feedback* do cliente, e não em atribuição mensurada, o que reforça o argumento em favor de regras explícitas: quando a atribuição não é recuperável nem retrospectivamente, a decisão depende de critérios declarados e auditáveis.

4.7. Revisão do artefato

A avaliação orientou a formulação de uma segunda versão do artefato, com três alterações na camada de decisão. A primeira substitui o ROAS do período corrente pelo ROAS acumulado móvel, definido na Equação 1, para que o retorno reflita o desempenho agregado, e não a ocorrência pontual de desfecho. A segunda introduz o estado de retorno não observado, no qual a decisão é orientada pelos indicadores antecedentes de volume e qualidade e a classificação escalar não é atribuída, pois a ampliação de investimento pressupõe confirmação de retorno. A terceira antecipa a qualidade dos *leads* para o primeiro nível da árvore, de modo que qualidade ruim conduz a pausar independentemente do retorno. Ela decorre da primeira:

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

como o retorno acumulado transporta para os períodos seguintes o efeito de contratos anteriores, sem ela um período com qualidade ruim não seria classificado como pausar enquanto houvesse contrato recente na janela. As regras completas constam do Quadro 4 e da Figura 2.

Quadro 4. Regras de decisão da versão revisada do artefato, avaliadas em ordem

Regra	Condição	Classificação
R1	Qualidade ruim	Pausar, qualquer que seja o retorno ou o volume
R2	Retorno não observado e volume insuficiente	Testar
R2	Retorno não observado e volume aceitável ou adequado	Otimizar (escalar bloqueado)
R3	ROAS acumulado abaixo do piso e qualidade boa	Testar
R3	ROAS acumulado abaixo do piso e qualidade regular	Pausar
R4	ROAS acumulado igual ou acima do piso e volume insuficiente	Testar
R4	ROAS acumulado igual ou acima do piso, volume adequado e qualidade boa	Escalar
R4	ROAS acumulado igual ou acima do piso, demais casos	Otimizar

Fonte: elaborado pelos autores (2026).

O ROAS acumulado do período t corresponde à razão entre a receita atribuída às coortes da janela e o investimento aplicado nos mesmos períodos, conforme a Equação 1, em que R é a receita atribuída à coorte do período k, I é o investimento do período k, J é a extensão da janela e d é a defasagem de maturação.

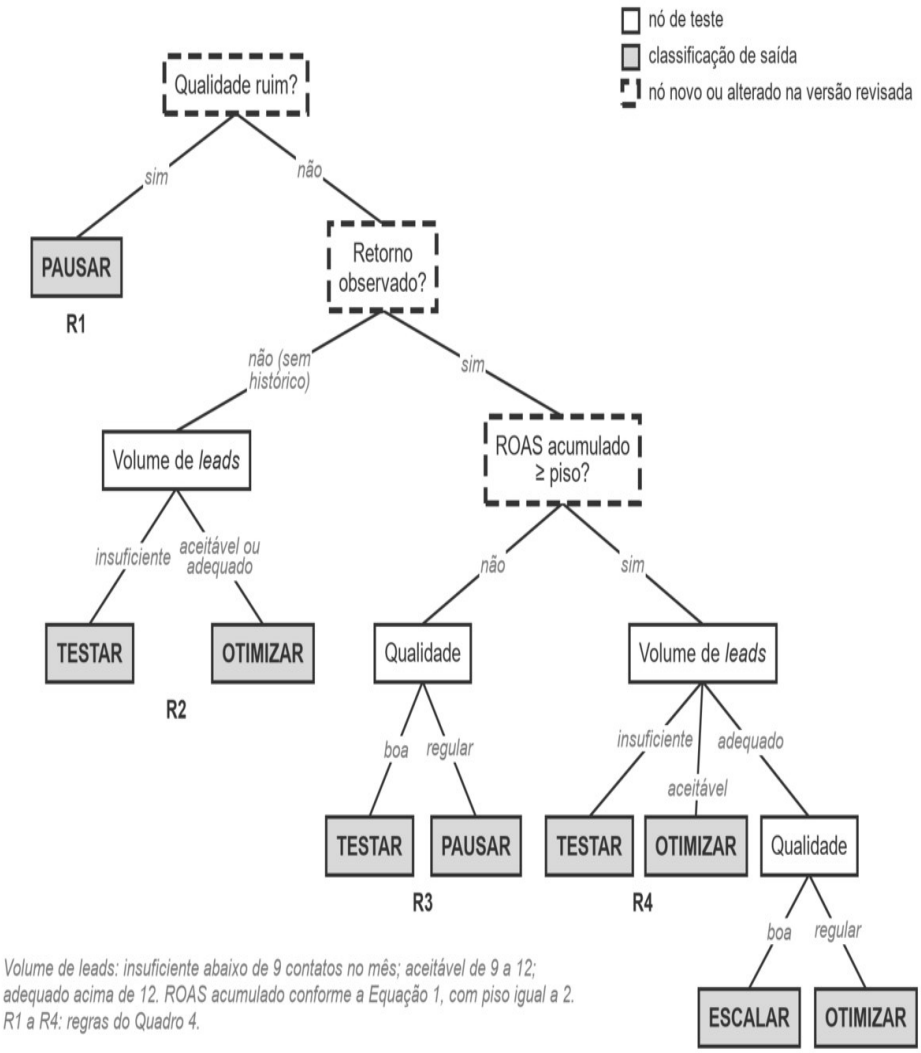
$$\mathrm{ROAS}_t^{\mathrm{acum}} = \frac{\sum_{k=\max(1,t-d-J+1)}^{t-d} R_k}{\sum_{k=\max(1,t-d-J+1)}^{t-d} I_k} \quad (1)$$

Adotou-se janela de doze meses, compatível com a referência de desempenho do negócio, expressa em contratos por ano. Nos primeiros períodos da série, a janela é truncada em junho de 2025, primeiro mês disponível. Mantiveram-se o piso de ROAS igual a 2 e os limiares de volume de 9 e 12 contatos calibrados na seção 4.4.

A defasagem de maturação d exclui da soma as coortes mais recentes, cujo desfecho ainda pode ocorrer. Adotou-se d igual a um mês: como as três negociações concluídas na série duraram de duas a três semanas e atravessaram a virada do mês, o contrato de uma coorte só é conhecido no período seguinte, e incluí-lo no período da própria coorte utilizaria informação indisponível no momento da decisão. Esse prazo descreve os negócios que fecharam, e não a população de contatos, razão pela qual a sensibilidade do resultado a d foi verificada, conforme descrito adiante.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 2. Árvore de decisão do artefato, versão revisada



Fonte: elaborado pelos autores (2026).

O retorno é classificado como não observado quando nenhuma coorte maturada está disponível, ou quando nenhuma coorte da janela registrou desfecho e a série ainda não completou uma janela inteira. O estado representa ausência de histórico, e não retorno insatisfatório. Decorrida uma janela completa sem desfecho, o retorno nulo passa a ser observado e o período é avaliado pela regra R3. A Tabela 2 apresenta a classificação das duas versões em cada período.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Tabela 2. Classificação das duas versões do artefato e decisão real

Mês	Volume	Qualidade	ROAS acumulado (12 meses)	Regra (v2)	Versão inicial	Versão revisada	Decisão real	Origem da decisão	Converge (v1)	Converge (v2)
jun/2025	5	boa	não observado	R2	testar	testar	otimizar	relato	não	não
jul/2025	11	boa	não observado	R2	testar	otimizar	otimizar	relato	não	sim
ago/2025	7	boa	não observado	R2	testar	testar	otimizar	relato	não	não
set/2025	5	boa	não observado	R2	testar	testar	otimizar	relato	não	não
out/2025	6	boa	não observado	R2	testar	testar	testar	documento	sim	sim
nov/2025	10	boa	não observado	R2	testar	otimizar	otimizar	relato	não	sim
dez/2025	11	boa	não observado	R2	otimizar	otimizar	otimizar	relato	sim	sim
jan/2026	16	ruim	13,89	R1	pausar	pausar	testar	documento	não	não
fev/2026	14	ruim	11,70	R1	pausar	pausar	otimizar	relato	não	não
mar/2026	26	ruim	10,14	R1	pausar	pausar	pausar	documento	sim	sim
abr/2026	18	boa	9,08	R4	escalar	escalar	testar	documento	não	não
mai/2026	13	boa	11,33	R4	testar	escalar	otimizar	relato	não	não
jun/2026	11	boa	10,29	R4	testar	otimizar	otimizar	relato	não	sim
jul/2026	8	boa	10,08	R4	otimizar	testar	otimizar	relato	sim	não
ago/2026	12	boa	38,83	R4	testar	otimizar	otimizar	relato	não	sim
Total	173								4 de 15	7 de 15

Fonte: elaborado pelos autores (2026).

A aplicação da segunda versão à mesma série elevou a convergência de quatro para sete dos quinze períodos, conforme a Tabela 2. As oito divergências remanescentes distribuem-se em quatro grupos. No primeiro, com junho, agosto e setembro de 2025, o volume é insuficiente e o retorno ainda não é observável, e o artefato recomenda testar, enquanto o gestor manteve o refinamento incremental. No segundo, com janeiro e fevereiro de 2026, o artefato recomenda interromper a campanha Performance Max antes da decisão efetivamente tomada, mas apenas a divergência de fevereiro antecipa a decisão do gestor com a informação disponível na época, conforme a seção 4.6. No terceiro, o artefato recomenda escalar, enquanto o gestor optou por testar nova estrutura em abril de 2026 e por manter o refinamento em maio de 2026.

O quarto grupo compreende um único período, julho de 2026, e é desfavorável ao artefato: com oito contatos, abaixo do limiar inferior, a versão revisada recomendou testar nova estrutura no período que originou o contrato de maior valor da série, correspondente a 75% da receita, reproduzindo o equívoco atribuído na seção 4.6 à regra orientada exclusivamente pelo volume. O limiar inferior não distingue um período de baixo volume com contatos qualificados de

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

um período de baixo volume por insuficiência de demanda, distinção que dependeria de medida de qualidade mais fina do que a classificação mensal adotada.

Não foram realizados ajustes adicionais nos limiares com o propósito de elevar a convergência. Com quinze períodos e instanciação única, tal procedimento configuraria sobreajuste ao caso analisado, comprometendo a validade da avaliação. Ressalte-se ainda que a segunda versão foi formulada após o exame desta série, de modo que sua convergência constitui demonstração dentro da amostra, e não avaliação independente.

A sensibilidade da segunda versão aos parâmetros foi verificada variando-se um parâmetro por vez. Com defasagem de maturação de zero a três meses, a convergência permaneceu em sete períodos, com as mesmas recomendações, e também se manteve em sete com janelas de sete a doze meses e com qualquer piso de ROAS entre 1 e 9, valor próximo ao menor retorno acumulado observado fora dos períodos de qualidade ruim, de 9,08. Janelas de três a seis meses reduziram a convergência para cinco ou seis períodos, e pisos superiores a 9 a fizeram oscilar entre seis e oito. Com limiares de volume de 8 e 11 contatos, a convergência manteve-se em sete, mas julho de 2026 passa a convergir e agosto de 2026, a divergir: a classificação de períodos próximos aos limiares depende de sua escolha.

Para verificar o efeito de calibrar os limiares sobre a própria série avaliada, adotaram-se dois procedimentos. Na divisão temporal, os limiares foram derivados apenas dos oito primeiros meses, de junho de 2025 a janeiro de 2026, resultando em 6 e 11 contatos, e aplicados aos sete meses seguintes. Na calibração expansível, os limiares de cada período foram derivados exclusivamente dos períodos anteriores, a partir de outubro de 2025. Os tercis deslocaram-se de 6,3 e 10,7 contatos na primeira metade da série para 12,0 e 14,0 na segunda, acompanhando o aumento do investimento, como prevê a condição de escopo da seção 4.4, mas o resultado pouco se alterou: na segunda metade, a versão revisada convergiu em três dos sete períodos tanto com os limiares da primeira metade quanto com os da série inteira, e a calibração expansível convergiu em cinco dos onze períodos avaliáveis, contra seis com os limiares fixos. Ambos os procedimentos mantêm fora da amostra apenas os limiares, uma vez que as regras da versão revisada foram formuladas após o exame da série completa.

4.8. Regras de referência e leitura das divergências

Para situar a convergência obtida, as duas versões do artefato foram comparadas a três regras de referência: manter a otimização em todos os períodos; pausar quando a qualidade é ruim e otimizar nos demais; e classificar apenas pelo volume de contatos, com os limiares da versão revisada. A Tabela 3 apresenta os resultados.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Tabela 3. Convergência com decisões históricas: artefato e regras de referência

Regra	Convergência (15 períodos)	Períodos documentados	Kappa de Cohen	Intervalo de 95% do kappa
Artefato, versão revisada	7 (47%)	2 de 4	0,21	-0,03 a 0,50
Artefato, versão inicial	4 (27%)	2 de 4	0,05	-0,15 a 0,29
Sempre otimizar	11 (73%)	0 de 4	0,00	0,00 a 0,00
Só qualidade (ruim pausa, demais otimizam)	11 (73%)	1 de 4	0,33	-0,06 a 0,76
Só volume (limiares 9 e 12)	6 (40%)	1 de 4	0,13	-0,05 a 0,38

Fonte: elaborado pelos autores (2026).

A regra de manter a otimização em todos os períodos alcançou 73% de convergência, acima das duas versões do artefato, com kappa nulo. O resultado decorre da composição da referência: em onze dos quinze períodos, a decisão real foi otimizar, registrada por relato da prática de rotina, de modo que qualquer regra que responda quase sempre otimizar obtém convergência elevada sem discriminar situação alguma. Nos quatro períodos com decisão documentada, nos quais a operação alterou a estrutura de campanhas, essa regra não convergiu em nenhum, e as duas versões do artefato convergiram em dois.

A regra baseada apenas na qualidade alcançou a mesma convergência de 73%, com kappa de 0,33, superior ao da versão revisada. O resultado reforça o papel da qualidade como indicador antecedente, mas exige cautela: a qualidade foi classificada como ruim exatamente nos três períodos de operação da campanha Performance Max, de modo que essa regra reproduz, na prática, a identificação daquele episódio. Pela escala de Landis e Koch (1977), a concordância da versão revisada classifica-se como razoável e a da versão inicial, como leve. Com quinze períodos, contudo, os intervalos de confiança de 95% do kappa, estimados por reamostragem *bootstrap* com 5.000 repetições (EFRON; TIBSHIRANI, 1993), incluem o valor zero para todas as regras comparadas, e a série não permite afirmar que qualquer delas concorde com as decisões históricas acima do esperado por acaso.

Dado esse limite, a avaliação de mérito recai sobre as divergências, examinadas à luz do desfecho posterior no Quadro 5. Das oito divergências da versão revisada, uma foi favorável ao artefato, em fevereiro de 2026, quando recomendou interromper a estrutura que não gerou contrato, já com a inadequação dos contatos percebida pelo cliente, e uma foi desfavorável, em julho de 2026, quando recomendou testar nova estrutura no período que originou 75% da receita. A de janeiro de 2026 depende da informação considerada; em quatro, o desfecho indica uma direção sem permitir conclusão, por ausência de contrafactual; e em uma não há desfecho que discrimine as alternativas.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Quadro 5. Divergências da versão revisada avaliadas pelo desfecho posterior

Período	Versão revisada	Decisão real	O que o desfecho mostrou	Leitura
jun, ago e set/2025	testar	otimizar	Nenhum contrato nas coortes de junho a novembro de 2025. O teste realizado em outubro de 2025 criou a campanha de pesquisa em operação exclusiva em dezembro de 2025, coorte do primeiro contrato (R$ 97.000,00).	Indício favorável ao artefato
jan/2026	pausar	testar	Coorte sem contrato e com qualidade ruim; a inadequação só foi percebida pelo cliente no final de fevereiro, e com a informação da época a recomendação seria escalar.	Depende da informação considerada
fev/2026	pausar	otimizar	Inadequação percebida pelo cliente no final do mês; coortes de fevereiro e março sem contrato; a campanha seguiu em março (R$ 292,57) e foi pausada nesse mês.	Favorável ao artefato
abr/2026	escalar	testar	O teste criou a campanha em operação exclusiva em julho de 2026, coorte do contrato de R$ 403.000,00.	Indício desfavorável ao artefato
mai/2026	escalar	otimizar	Coortes de maio e junho de 2026 sem contrato.	Não avaliável
jul/2026	testar	otimizar	Coorte do contrato de R$ 403.000,00, correspondente a 75% da receita.	Desfavorável ao artefato

Fonte: elaborado pelos autores (2026).

Dois aspectos do Quadro 5 merecem registro. O artefato opera no nível da conta e recomenda uma ação para o período, cabendo ao gestor identificar qual estrutura interromper, de modo que a leitura de que a recomendação de fevereiro de 2026 se refere à campanha Performance Max é do analista, e não da regra. Além disso, as duas decisões de testar efetivamente tomadas, em outubro de 2025 e abril de 2026, criaram as estruturas em operação nos períodos que originaram 93% da receita: o artefato convergiu com o gestor no primeiro caso e, no segundo, recomendou escalar, divergindo da decisão que se mostrou mais produtiva.

4.9. Discussão

Os resultados permitem precisar a contribuição central deste trabalho. Se o resultado somente se observa após a maturação da coorte, enquanto volume e qualidade são observáveis no próprio período de decisão, então a decisão sob essas condições depende necessariamente de indicadores antecedentes, cabendo ao indicador de resultado a função de confirmação posterior. A composição do artefato por três indicadores não constitui escolha de conveniência de projeto, mas imposição da estrutura temporal dos dados disponíveis.

Essa formulação dialoga com a desconexão estrutural entre métricas digitais e resultados de negócio apontada por Saura, Ribeiro-Soriano e Palacios-Marqués (2021). O

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

presente caso demonstra que a desconexão não se limita à ausência do dado de desfecho no sistema operacional, mas alcança a temporalidade do indicador de resultado, que somente se torna informativo depois de decorrido o período em que a decisão precisava ser tomada.

Dialoga igualmente com Wang e Strong (1996). O caso do evento mal especificado mostra que um artefato decisório herda decisões de configuração do sistema operacional que não são visíveis na saída apresentada ao gestor. Disso decorre um requisito de projeto de caráter geral: a camada de tratamento de um artefato decisório deve incorporar verificação de plausibilidade das entradas, uma vez que a inadequação do indicador não se manifesta como erro, mas como valor plausível sob rótulo incorreto.

5. CONSIDERAÇÕES FINAIS

Este trabalho projetou e avaliou um artefato de apoio à decisão baseado em regras explícitas para contextos em que o indicador de resultado é incompleto e tardio. A resposta à questão de pesquisa é parcial. Demonstrou-se ser possível formalizar o raciocínio decisório em regras que operam sobre três indicadores disponíveis em contexto operacional de recursos limitados e confrontá-las sistematicamente com as decisões tomadas em quinze meses de operação real. A avaliação, contudo, não permite afirmar que o artefato decida melhor do que a prática que pretende apoiar.

Quando o desfecho constitui evento raro, o retorno calculado na granularidade do período de decisão assume valor nulo na maior parte dos meses por razões estruturais do negócio e não discrimina desempenho. A versão revisada, com retorno acumulado, estado de resultado não observado e prioridade à qualidade dos *leads*, elevou a convergência de 27% para 47%, abaixo dos 73% da regra trivial de manter a otimização em todos os períodos, e os intervalos de confiança do kappa incluem o valor zero para todas as regras comparadas: a convergência mede aderência à prática, e não eficácia. Esses resultados sustentam as duas contribuições discutidas na seção 4.9: a decisão sob resultado incompleto e tardio depende de indicadores antecedentes, cabendo ao indicador de resultado a função de confirmação posterior, e a camada de tratamento de um artefato decisório deve verificar a plausibilidade das entradas.

A hipótese de que um artefato baseado em regras reduziria a subjetividade e aumentaria a consistência das avaliações sustenta-se parcialmente. A formalização tornou explícitos e auditáveis critérios antes implícitos, mas a qualidade decorreu, nesta instância, de julgamento retrospectivo do cliente, coincidente com a operação da campanha Performance Max, o que limita o valor probatório desse episódio. Ao considerar a qualidade, o artefato teria antecipado em cerca de um mês a interrupção da estrutura improdutiva, a partir do momento em que o

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

cliente percebeu a inadequação dos contatos, e não em dois meses, como sugere a classificação retrospectiva. Em sua versão revisada, entretanto, reproduziu o equívoco da regra de volume no período de maior receita, o que aponta a definição operacional da qualidade como condição para que o volume seja interpretado corretamente.

Como limitações, registram-se a instanciação em contexto único, sem pretensão de generalização estatística; o registro parcial das decisões reais, documentadas em quatro dos quinze períodos e reconstituídas por relato do gestor nos demais; a classificação da qualidade por julgamento retrospectivo do cliente, sem aplicação da rubrica especificada; a impossibilidade de atribuir os eventos de contato por campanha de origem, uma vez que o *pixel* opera no âmbito do site; a permanência da etapa de coleta como procedimento manual; a referência de avaliação, que em onze dos quinze períodos provém de relato da prática de rotina; a coincidência entre o primeiro autor e o gestor da conta, responsável tanto pelas decisões de referência quanto pelo relato da prática de rotina; o tamanho da série, insuficiente para distinguir a concordância obtida da esperada por acaso; a formulação da versão revisada após o exame da própria série; e a validade dos limiares calibrados restrita à faixa de orçamento observada.

Como trabalhos futuros, indicam-se a integração de outras plataformas de anúncios; a substituição da exportação manual de relatórios pela conexão direta via interface de programação de aplicações, com ferramentas de *Business Intelligence* na camada de visualização; e a definição operacional do indicador de qualidade a partir da classificação automatizada das conversas de atendimento por modelos de linguagem de grande escala, que converteria o único indicador subjetivo do artefato em medida reproduzível e auditável. A rubrica especificada na seção 4.1 fornece os critérios para essa classificação, e a concordância entre avaliadores humanos, a referência para validá-la.

DECLARAÇÃO DE USO DE INTELIGÊNCIA ARTIFICIAL

Este trabalho utilizou a ferramenta de inteligência artificial generativa Claude (Anthropic) para revisão textual, para a tradução do resumo para o inglês e o espanhol e para a elaboração das Figuras 1 e 2, geradas por script produzido com a ferramenta a partir das regras de decisão definidas pelos autores. A concepção da pesquisa, a coleta e a análise dos dados e a interpretação dos resultados são dos autores, que revisaram todo o conteúdo e se responsabilizam integralmente por ele.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

CONFLITO DE INTERESSES E FINANCIAMENTO

O primeiro autor atuou como gestor remunerado da conta de anúncios analisada durante todo o período da pesquisa, condição cujas implicações para a avaliação são discutidas na seção 3.3. Não há outros conflitos de interesses a declarar. A pesquisa não recebeu financiamento.

REFERÊNCIAS

CHAPELLE, Olivier. Modeling delayed feedback in display advertising. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 20., 2014, New York. **Proceedings [...]**. New York: ACM, 2014. p. 1097-1105. DOI: https://doi.org/10.1145/2623330.2623634.

CHEN, Hsinchun; CHIANG, Roger H. L.; STOREY, Veda C. Business intelligence and analytics: from big data to big impact. **MIS Quarterly**, Minneapolis, v. 36, n. 4, p. 1165-1188, 2012. DOI: https://doi.org/10.2307/41703503.

COHEN, Jacob. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, Durham, v. 20, n. 1, p. 37-46, 1960. DOI: https://doi.org/10.1177/001316446002000104.

DUMAS, Marlon; LA ROSA, Marcello; MENDLING, Jan; REIJERS, Hajo A. **Fundamentals of business process management**. 2. ed. Berlin: Springer, 2018. DOI: https://doi.org/10.1007/978-3-662-56509-4.

EFRON, Bradley; TIBSHIRANI, Robert J. **An introduction to the bootstrap**. New York: Chapman & Hall, 1993.

GRANDHI, Balakrishna; PATWA, Nitin; SALEEM, Kashif. Data-driven marketing for growth and profitability. **EuroMed Journal of Business**, Bingley, v. 16, n. 4, p. 381-398, 2021. DOI: https://doi.org/10.1108/EMJB-09-2018-0054.

GREGOR, Shirley; HEVNER, Alan R. Positioning and presenting design science research for maximum impact. **MIS Quarterly**, Minneapolis, v. 37, n. 2, p. 337-355, 2013. DOI: https://doi.org/10.25300/MISQ/2013/37.2.01.

HEVNER, Alan R.; MARCH, Salvatore T.; PARK, Jinsoo; RAM, Sudha. Design science in information systems research. **MIS Quarterly**, Minneapolis, v. 28, n. 1, p. 75-105, 2004. DOI: https://doi.org/10.2307/25148625.

JÄRVINEN, Joel; KARJALUOTO, Heikki. The use of web analytics for digital marketing performance measurement. **Industrial Marketing Management**, New York, v. 50, p. 117-127, 2015. DOI: https://doi.org/10.1016/j.indmarman.2015.04.009.

KARCZMARCZYK, Artur; JANKOWSKI, Jarosław; WĄTRÓBSKI, Jarosław. Multi-criteria decision support for planning and evaluation of performance of viral marketing campaigns in social networks. **PLOS ONE**, San Francisco, v. 13, n. 12, e0209372, 2018. DOI: https://doi.org/10.1371/journal.pone.0209372.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

LANDIS, J. Richard; KOCH, Gary G. The measurement of observer agreement for categorical data. **Biometrics**, Washington, v. 33, n. 1, p. 159-174, 1977. DOI: https://doi.org/10.2307/2529310.

LAUDON, Kenneth C.; LAUDON, Jane P. **Gerenciamento de sistemas de informação**. 3. ed. Rio de Janeiro: LTC, 2001.

MCAFEE, Andrew; BRYNJOLFSSON, Erik. Big data: the management revolution. **Harvard Business Review**, Boston, v. 90, n. 10, p. 60-68, 2012.

NYGÅRD, Robert; MEZEI, József. Automating lead scoring with machine learning: an experimental study. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 53., 2020, Maui. **Proceedings [...]**. Honolulu: University of Hawai'i at Mānoa, 2020. p. 1439-1448. DOI: https://doi.org/10.24251/HICSS.2020.177.

PEFFERS, Ken; TUUNANEN, Tuure; ROTHENBERGER, Marcus A.; CHATTERJEE, Samir. A design science research methodology for information systems research. **Journal of Management Information Systems**, Armonk, v. 24, n. 3, p. 45-77, 2007. DOI: https://doi.org/10.2753/MIS0742-1222240302.

PROVOST, Foster; FAWCETT, Tom. **Data science for business**: what you need to know about data mining and data-analytic thinking. Sebastopol: O'Reilly Media, 2013.

SAURA, José Ramón; RIBEIRO-SORIANO, Domingo; PALACIOS-MARQUÉS, Daniel. Setting B2B digital marketing in artificial intelligence-based CRMs: a review and directions for future research. **Industrial Marketing Management**, New York, v. 98, p. 161-178, 2021. DOI: https://doi.org/10.1016/j.indmarman.2021.08.006.

SHAH, Denish; MURTHI, B. P. S. Marketing in a data-driven digital world: implications for the role and scope of marketing. **Journal of Business Research**, New York, v. 125, p. 772-779, 2021. DOI: https://doi.org/10.1016/j.jbusres.2020.06.062.

TROISI, Orlando; MAIONE, Gennaro; GRIMALDI, Marco; LOIA, Francesca. Growth hacking: insights on data-driven decision-making from three firms. **Industrial Marketing Management**, New York, v. 90, p. 538-557, 2020. DOI: https://doi.org/10.1016/j.indmarman.2019.08.005.

WANG, Richard Y.; STRONG, Diane M. Beyond accuracy: what data quality means to data consumers. **Journal of Management Information Systems**, Armonk, v. 12, n. 4, p. 5-33, 1996. DOI: https://doi.org/10.1080/07421222.1996.11518099.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.