

Por que as decisões em tempo real são importantes

UM WHITE PAPER
DA PEGA



Introdução

A experiência do cliente está numa encruzilhada em diversos setores. As empresas estão usando abordagem e tecnologias defasadas para tentar resolver problemas complexos de engajamento, e os resultados têm sido ruins. Isso está gerando muita insatisfação:

- **Os consumidores** se irritam porque suas interações com as empresas são desagradáveis, tediosas e irrelevantes. A cada experiência boa, acontecem dez situações importunas. Eles sentem que estão sempre recebendo um discurso de vendas.
- **Os administradores** se irritam porque investiram em soluções que deveriam ter “acertado o rumo” e criado o diferencial em relação à concorrência, mas as melhorias foram esporádicas, sem retorno significativo sobre os investimentos.
- **Os investidores** se irritam porque as empresas falam o tempo todo da experiência do cliente, mas as reclamações viralizam o tempo todo. A fatia de mercado vai se evaporando e os dividendos estão indo embora.

O problema básico: os problemas da empresa contra os problemas do cliente

O grande desafio não é que a expectativa dos clientes seja muito grande, ou que as empresas não queiram investir na experiência deles. Nenhuma dessas afirmações é verdadeira. Os clientes só querem mais experiências “ideais”, aquelas em que a empresa prevê suas necessidades, facilita a busca de informações e valoriza a preferência deles. Por sua vez, as empresas já gastaram bilhões em experiência, na tentativa de cumprir e superar as expectativas dos clientes. Mas, apesar das promessas de fornecedores, agências e parceiros, o retorno tem sido pífio. E as empresas continuam gastando porque continuam perdendo espaço.

O verdadeiro problema é que essas soluções foram feitas para atender às necessidades da empresa, e não para resolver os problemas dos clientes.

Pacotes de automação de marketing, desktops de atendimento, ferramentas para criação de modelos e plataformas de inteligência de negócios (BI) fazem parte de um ecossistema completo de soluções feitas para que a empresa consiga enviar mais ofertas e mensagens para o mercado. Essas soluções coletam dados de sistemas diversos em lotes, consolidam históricos, analisam comportamentos, anexam dados de terceiros, criam regras, segmentam, otimizam o direcionamento e geram campanhas. Elas permitem que as empresas se conectem com os clientes cada vez mais rápido, a um custo muito menor do que antes. Mas elas ignoram um problema fundamental:

As pessoas são complexas, agem por impulso e mudam de ideia rapidamente. Quando consomem novas informações, às vezes elas decidem seguir outro caminho em segundos.

As pessoas são fluidas, dinâmicas e adaptáveis. É por isso que um ecossistema montado sobre dados em lote, bancos de dados relacionais e engajamento baseado em campanhas não pode satisfazer os requisitos dos clientes modernos. Essas tecnologias foram feitas para escolher uma mensagem e distribuí-la para grandes públicos ou segmentos em escala, não para analisar as necessidades de um cliente num momento específico e cuidar delas antes que mudem.

A vantagem competitiva das decisões em tempo real

O contexto dos clientes (situação, ambiente, emoções, motivações e comportamentos) está sempre mudando. As oportunidades aparecem e desaparecem em segundos. Mas os mesmos indivíduos ficam expostos a infinitas mensagens de concorrentes todos os dias, todas muito parecidas com a sua. Você não vai se destacar se não conseguir ler os sinais e otimizar sua mensagem durante esse rápido período. E a oportunidade não será perdida porque seus concorrentes são melhores, necessariamente. Você só não foi rápido o suficiente, e a concorrência estava lá: no lugar certo e na hora certa. Por isso, as decisões em tempo real agora estão diretamente ligadas à vantagem competitiva.

Estarão na frente dessa disputa as marcas que conseguirem refazer as decisões para o cliente cinco, 10, 20 ou até 50 vezes durante uma única experiência. Cada vez que ativarem novos dados e acionarem modelos preditivos para reavaliar a propensão do cliente, essas marcas ficarão muito mais relevantes do que as marcas que não conseguirem fazer isso e, assim, conquistarão **grande vantagem competitiva**. Este artigo explica como são tomadas essas decisões rápidas em tempo real, que dados e análises são usados e a com que rapidez a organização precisa agir para causar um impacto relevante na experiência do cliente.

As quatro capacidades fundamentais das decisões em tempo real

Proporcionar interações relevantes e contextualizadas requer mais do que a decisão em tempo real. São quatro fatores que formam uma estrutura otimizada de decisões em tempo real: detecção, dados, decisões e entrega. Se você compreendê-los, conseguirá avaliar com mais facilidade seu nível de preparo para o trabalho em tempo real.



Detecção

Identifique uma oportunidade com o cliente

Detectar uma oportunidade é a primeira etapa de um programa de engajamento em tempo real, e o primeiro aspecto a ser avaliado. Em resumo, nenhuma das capacidades ou processos seguintes terá importância se você não souber da oportunidade de vender, atender ou reter.

Use a detecção proativa

O aspecto mais importante da detecção é que ela não pode ser agendada. No mercado de hoje, ela precisa ser percebida pela análise de dados do contexto. Janelas de oportunidade digital são abertas e fechadas em segundos. É uma realidade muito diferente do passado, quando era comum agendar processos em lote para serem executados todas as noites ou toda semana.

Sua capacidade de detectar esses momentos de oportunidade imediatamente, ou até antes que aconteçam, exige a detecção de eventos a partir de dados rápidos e lentos para prever o que pode acontecer. Muitos casos de uso exigem detecção quase imediata, como a mineração de eventos especiais em fluxos de dados de alta velocidade, principalmente aqueles com janelas de oportunidade curtas. Às vezes, um simples gatilho (ex.: se acontecer x, fazer y) pode ser suficiente. Mas os casos de uso mais relevantes são construídos sobre cenários tortuosos, com eventos complexos e interconectados, que exigem funções mais sofisticadas.

O aspecto mais importante da detecção é que ela não pode ser agendada. No mercado de hoje, ela precisa ser percebida pela análise de dados do contexto.

Quais são os diferentes tipos de eventos?

Uma detecção de eventos complexos busca padrões em diferentes fontes de dados e tenta identificar quando existe um momento de oportunidade.

- **Eventos simples:** escutar uma única fonte de dados (ex.: clique num link)
- **Não eventos:** evento que era esperado mas não aconteceu (ex.: depósito bancário regular, login em site ou aplicativo, visita à agência, etc.)
- **Eventos complexos:** uma combinação de diversos eventos simples e “não eventos” que, juntos, representam um momento importante

Detecte sua janela de oportunidade

A detecção é uma etapa fundamental das decisões em tempo real. Se sua capacidade de detecção não for rápida o bastante, é provável que a janela de oportunidade seja perdida, tornando irrelevantes todas as outras etapas (combinação de dados, tomada de decisão e entrega).

Cenário 1: A detecção é muito lenta, e não há tempo para reagir.



Cenário 2: A detecção é rápida, e há tempo suficiente para reagir.



Ilustração: latência de detecção e a janela de oportunidade

O design das experiências digitais costuma exigir resposta em 100 a 200 milissegundos.

Qual é a velocidade necessária?

A velocidade de detecção (e execução de ponta a ponta) necessária para sua empresa pode ser medida em microssegundos (milionésimos de segundo) em alguns casos de uso, como negociações online ou compras de anúncios online. Experiências digitais, como cliques na web e interações em aplicativos móveis, exigem tempo de resposta de menos de meio segundo. E o design das experiências digitais costuma exigir resposta em 100 a 200 milissegundos (ms). Um intervalo de mais de um segundo pode trazer resultados insatisfatórios ou até prejudiciais.

O setor de telecomunicações está relativamente avançado nas decisões em tempo real. Por exemplo: um cliente que coloca crédito em um telefone pré-pago pode ser uma oportunidade para apresentar uma oferta contextualizada. É um gatilho simples, mas precisa ser acionado imediatamente.

Outro exemplo: no setor de finanças, um banco pode querer detectar se o cliente deixou o dinheiro ou o cartão no caixa eletrônico durante a interação, para avisá-lo o mais rápido possível. Um caso mais complexo seria uma empresa de gestão de fortunas que pode combinar dados de atividade da conta, utilização de produtos, navegação na web e uso

de aplicativos móveis para prever a intenção do cliente em determinado momento. Essa inferência pode ser usada para mostrar recomendações, mensagens e ofertas mais relevantes.

Dados

Acesse e combine informações em tempo real

Para funcionar verdadeiramente em tempo real, uma solução precisa receber dados em tempo real para entender o contexto específico do cliente. As empresas interagem com pessoas em diversos canais, recebendo dados de diversas fontes. O contexto relevante de uma pessoa (aquilo que importa no momento) geralmente é muito mais amplo do que o que observamos em uma única interação, além de estar sempre mudando. A maioria dos sistemas e abordagens falha porque não analisa o quadro completo de cada pessoa.

Por exemplo: você pode detectar que um cliente está passando por uma área georreferenciada que pode ser importante (entrando em sua loja ou na loja de um concorrente). Mas esse evento, em si, é uma pequena parte de um quadro muito mais amplo para o cliente. O gatilho é importante, mas você precisa de mais dados para saber se deve agir agora e, se for o caso, qual deve ser a sua ação.

Se o cliente estiver com dificuldades para usar seu produto, por exemplo, ele pode ter entrado na loja para pedir ajuda, e não para fazer outra compra. Você não vai saber disso se não estiver analisando os dados certos (motivos de chamadas, probabilidade de atendimento, histórico de navegação). Sem analisar o quadro completo, você provavelmente tentaria vender algo, e seria o pior momento possível. Esse tipo de interação pode prejudicar seriamente o relacionamento de longo prazo com esse cliente, principalmente se ele se repetir.

Em muitos casos, não existem dados significativos ou eles não são considerados no processo decisório.

A capacidade de integrar esses dados adicionais é crucial para as decisões em tempo real inteligentes. Infelizmente, os dados não podem ficar parados num banco de dados relacional criado para processamento em lote. Também é preciso acomodar dados “federados” para que possam ser analisados e aplicados em tempo real.

Os cinco elementos do contexto

Além de tudo isso, é necessário reunir fontes de dados que mostrem cada aspecto do contexto do cliente no momento específico em que a interação está acontecendo, criando uma interação verdadeiramente relevante e interessante.

Existem cinco categorias de contexto principais:

Contexto	Descrição	Exemplos de fontes
Ambiental	Localização do cliente; onde a pessoa está, física ou digitalmente	Localização de celular, domínios, IDs de contêiner
Comportamental	Ações do cliente; como a pessoa interagiu ou respondeu	Transações, histórico de interações, navegação
Emocional	Emoção do cliente; como a pessoa está se sentindo	Pontuação de churn, probabilidade de inadimplência, atitude na chamada
Motivacional	Intenção do cliente; o que ele deseja obter	Probabilidade de contato telefônico, motivo do contato, envio de formulários
Situacional	Estado do cliente; progresso da pessoa em relação a suas metas	Estágios da jornada, status do caso, avaliação de crédito

Muitas das soluções ditas “em tempo real” cometem o erro de ignorar totalmente esse aspecto da contextualização, passando direto de um gatilho para uma ação predeterminada.

Considere este exemplo simples: seu receptor de TV a cabo não está funcionando, então você acessa o site da empresa para ver se existe uma solução rápida. Você está acessando o site por um motivo bem específico, mas aparece um pop-up ou oferta de marketing insistente, atrapalhando seu acesso ao resultado desejado (consertar o receptor quebrado), o que só traz mais irritação. Essa desconexão é comum e acontece porque a marca não entende o contexto desse cliente, ainda que haja indícios desse contexto facilmente acessíveis, através dos dados do próprio receptor da empresa.

O contexto traz novas perspectivas para as decisões. Exemplos de dados contextuais valiosos: dados de CRM, atividades de navegação, cliques em aplicativos móveis, transações, publicações em redes sociais, dados de dispositivos e IdC, estado emocional indicado por texto (bate-papos, redes sociais, e-mail, conversor de voz em texto) e dados de terceiros, como clima ou localização.

No exemplo do receptor de TV a cabo, o contexto do dispositivo inoperante seria uma dica importante do contexto situacional e motivacional dessa pessoa, que poderia determinar ações corretas ou incorretas para a interação.

Dois coisas devem ser consideradas em dados de contexto.

- **Estagnação:** com que rapidez os dados ficam estagnados e perdem valor? As pessoas são complexas e mudam sempre, assim como seus dados. A latência dos dados pode afetar muito a relevância, principalmente quando eles são processados em lote. O uso de

dados estagnados pode fazer com que você considere contextos que deixaram de existir, como interagir com uma pessoa que esteve em sua loja, enviando uma oferta para o celular quando ela já saiu do local.

- **Escopo:** os dados contextualizados que você está reunindo abordam todos os elementos do contexto? Em muitos casos, não existem dados significativos ou eles não são considerados no processo decisório. Muitas vezes, essas informações poderiam ter ilustrado melhor o contexto do cliente naquele momento, mudando totalmente a decisão. Por exemplo: quando a marca ignora o contexto motivacional, como o receptor de TV a cabo estragado, e faz uma oferta de venda em vez de tentar resolver o problema.

A velocidade da combinação dos dados é crucial

Assim que a necessidade ou oportunidade do cliente for detectada (abordada na próxima seção), a velocidade de sua combinação de dados de contexto torna-se muito importante. O relógio começa a rodar imediatamente, e é preciso que a latência seja a menor possível para que a decisão seja inteligente e bem informada.

Cenário 1: A detecção é rápida, mas os dados são combinados muito devagar e não há tempo para reagir na janela de oportunidade.



Cenário 2: Os dados são combinados com rapidez e oferecem informações suficientes para decisões bem informadas, com tempo para os processos seguintes.



Ilustração: latência da combinação dos dados e a janela de oportunidade

A latência não é uma questão relevante para certos tipos de dados, como endereços de correspondência, telefones, datas de nascimento e às vezes até produtos comprados. São dados que não mudam com frequência. Mas os dados de aplicativos móveis e localização, bem como outras pegadas digitais, são muito dinâmicos. Seu cliente acabou de clicar naquela oferta de cartão de crédito? Ou estava lendo os termos do fechamento da conta?

O cliente ainda está na página, ou ainda está na loja? Informações assim são fundamentais para identificar o contexto.

O contexto é importante não só para a hiperpersonalização do próprio canal, mas também para interações intercanais. Sem compreender o contexto completo, a marca pode ser pega desprevenida por um “ponto cego” comportamental. O cliente pode não estar passando de um canal para outro, talvez esteja consumindo os dois ao mesmo tempo, por exemplo, lendo os termos e condições no seu site enquanto está ao telefone com um atendente, ou vendo os detalhes do produto no aplicativo enquanto compra na loja. As atividades de um canal precisam informar a próxima melhor ação em qualquer outro canal, para otimizar aquela interação e momento específicos. Isso mostra por que as funções em tempo real “de verdade” são tão importantes.

Sua capacidade de tomar decisões depende totalmente dos dados que você usa para alimentá-las. Em muitos casos, se sua empresa não conseguir operacionalizar os dados de contexto anteriores à transação, a decisão não será a melhor para esse cliente. E isso prejudica o engajamento, a satisfação e as taxas de conversão. Muitas pessoas evitam receber contatos e interações hoje em dia, por isso toda mensagem irrelevante faz com que o cliente se afaste mais e tenda ao desengajamento total.

Decisão

Como as regras podem prejudicar os modelos adaptativos em tempo real

A detecção e a combinação de dados em tempo real criam uma base sólida para a tomada de decisões inteligentes. Mas nem todas as decisões em tempo real são iguais. Uma lógica determinista de “se acontecer isso, fazer aquilo” irrita os clientes e mais atrapalha do que ajuda. Esse tipo de regra não leva em conta as necessidades, emoções ou motivações exclusivas do cliente naquele momento. Fatores contextuais como esses são a próxima etapa da evolução da experiência do cliente para as empresas, onde um nível decisório mais sofisticado deve ser considerado.

Níveis de decisões em tempo real

Existem quatro níveis básicos de maturidade em tomadas de decisão em tempo real:



Ilustração: maturidade da tomada de decisões de próxima melhor ação

Nível 1: Regras deterministas

Regras de negócios deterministas costumam ser usadas para configurar a lógica decisória. São regras que precisam ser relativamente simples e diretas, caso contrário ficam impossíveis de manejar e não conseguem determinar contexto e probabilidades porque existem muitos fatores e permutas a serem considerados. Exemplos de regras deterministas: se isso, então, aquilo; pontuações lineares e árvores de decisão.

Nível 2: Análises pré-classificadas (offline) e regras deterministas

Este nível acrescenta modelos de análise (de propensão) para prever o comportamento do cliente. Aqui, as conclusões da análise foram predeterminadas e pré-classificadas antes da própria interação ou oportunidade. Nesses casos, os modelos são processados em lotes agendados (toda noite, toda semana ou todo mês) conforme os dados então disponíveis. As classificações obtidas são carregadas para o ambiente decisório e usadas quando ocorre uma interação. No momento propício, os dados pré-classificados são usados para arbitrar a decisão em tempo real.

Pode-se dizer que essa função é melhor do que não ter nenhuma previsão. Mas as desvantagens são significativas e podem resultar em decisões incorretas. Imagine que um cliente recusa uma oferta de cartão de crédito. Essa mudança de contexto deve influenciar imediatamente a probabilidade de aceitação de outros cartões de crédito, empréstimos ou mesmo o risco de perda do cliente. Mas o novo contexto não será considerado com o uso de previsões pré-classificadas. As classificações não serão alteradas até a próxima atualização agendada, o que gera previsões imprecisas e recomendações possivelmente irrelevantes (ou até prejudiciais).

Considere o exemplo anterior, da interação intercanais em que um cliente acaba de iniciar uma interação no site ou por IVR com um atendente humano. Os modelos analíticos vão conseguir considerar essa valiosa nova pegada digital nos seus cálculos e reavaliar as conclusões em tempo real? As previsões pré-classificadas deixam pontos cegos de contexto nas suas decisões. Os modelos offline também são muito ineficientes, gastando um poder de processamento considerável no cálculo de probabilidades que provavelmente nunca serão usadas e se estagnarão rapidamente.

Nível 3: Modelos de análises em tempo real (ao vivo) e regras deterministas

Uma arquitetura superior usa previsões sob demanda em tempo real com o contexto mais recente. Neste nível, toda a decisão, de cima a baixo, é tomada em tempo real, apoiando uma abordagem de próxima melhor ação durante o engajamento. Os próprios modelos analíticos são executados verdadeiramente em tempo real (não são só classificações trazidas por upload), usando todo o arsenal de dados em tempo real. Aqui, toda a decisão da próxima melhor ação é feita com o contexto completo do cliente. A decisão ideal para esse momento é selecionada.

No exemplo do defeito no receptor de TV a cabo, dados de Internet das Coisas (IdC) do dispositivo, combinados com o comportamento em tempo real de cliques na web, podem ser considerados na decisão em tempo real, inclusive as previsões que apoiam essa conclusão.

Nível 4: Modelos adaptativos em tempo real e regras deterministas

Os exemplos mais sofisticados combinam regras deterministas e modelos em tempo real e adaptativos (autoaprendizado) para uso nas decisões em tempo real. Aqui, os próprios modelos estão sempre mudando conforme o comportamento do cliente. Nesse processo de aprendizado de máquina, os modelos aprendem com cada previsão (correta ou incorreta) que é feita. Combinados com regras deterministas, eles trazem mais eficiência, previsões mais precisas, decisões aprimoradas e resultados de negócios e experiência do cliente muito melhores (que, no fim das contas, eram os objetivos iniciais).

Velocidade das decisões

Desconsiderando a precisão, a velocidade da decisão também é crucial, como a ilustração abaixo comprova. Por exemplo: uma empresa detecta que um cliente em potencial está procurando um empréstimo imobiliário e reúne todos os dados necessários para essa decisão. Mas se a empresa não conseguir fazer a oferta com a velocidade necessária, a janela já pode ter se fechado e o cliente pode estar no concorrente.

Cenário 1: O evento é detectado e os dados são reunidos rapidamente, mas a decisão em si demora demais e não é possível reagir no tempo disponível.



Cenário 2: O evento é detectado, os dados são reunidos rapidamente, a decisão em si é tomada no tempo necessário e há tempo para as atividades seguintes.



Ilustração: latência do processo decisório e a janela de oportunidade

É nesta fase também que as necessidades e objetivos do negócio são avaliadas durante a decisão da próxima melhor ação, inclusive aspectos como retenção, valor do ciclo de vida do cliente, prioridades de atendimento, margem, inventário, propensão à recusa, etc.

Decisões em escala

Considerando os dados acima, a escala e o volume das decisões também devem ser considerados. Para qualquer interação, pode haver centenas (ou milhares) de ações possíveis, cuja relevância deve ser avaliada em tempo real a partir do contexto completo do cliente. Em empresas grandes, é comum que o volume atinja 500 a 1000 decisões por segundo (por dia, seriam 86 milhões de decisões). E diversas empresas precisam de escala muito maior do que isso. Algumas empresas de serviços financeiros têm picos de mais de 200 milhões de decisões ao dia.

Outra consideração importante para as decisões é a escalabilidade dos próprios modelos. Técnicas como modelos adaptáveis (aprendizado de máquina) permitem que as empresas criem e mantenham automaticamente mais modelos preditivos do que seria possível somente com recursos de processamento de dados.

Entrega

Responda dentro da janela de oportunidade

A entrega é a ação da marca em resposta à oportunidade no canal, interagindo diretamente com o cliente e usando a ação recomendada.

Avalie o ponto de inflexão

É importante compreender que os primeiros três pilares (detecção, dados e decisão) precisam acontecer em tempo real, sempre que possível. Assim, a empresa sabe se é necessário fazer alguma coisa e se é possível fazê-la no tempo disponível. Esse é um “ponto de inflexão” importante, porque a decisão pode provocar latência ou irrelevância durante a etapa da entrega. Por exemplo: se uma empresa concluir que não será possível reagir dentro dessa janela com a ação desejada, ou que existe o risco de a interação ser irrelevante ou negativa, a melhor decisão pode ser não fazer nada e suprimir a entrega dessa ação, que talvez já esteja atrasada.

Conforme aumenta o tempo necessário para tomar e entregar a decisão, a qualidade do engajamento com o cliente e a probabilidade de conversão geralmente diminui.

Uma empresa de serviços financeiros norte-americana, por exemplo, detecta eventos, reúne dados e toma decisões em tempo real em escala, mas também considera a opção de não entregar nada durante um ponto de contato com o cliente. Quando isso ocorre, toma-se a decisão de registrar o contexto do cliente para melhorar decisões futuras.

Entregue com rapidez

Quando a resposta é considerada adequada, a capacidade da empresa de reagir em tempo real é a reta de chegada. A ilustração abaixo mostra o impacto de uma entrega atrasada no resultado, no contexto de uma decisão em tempo real.

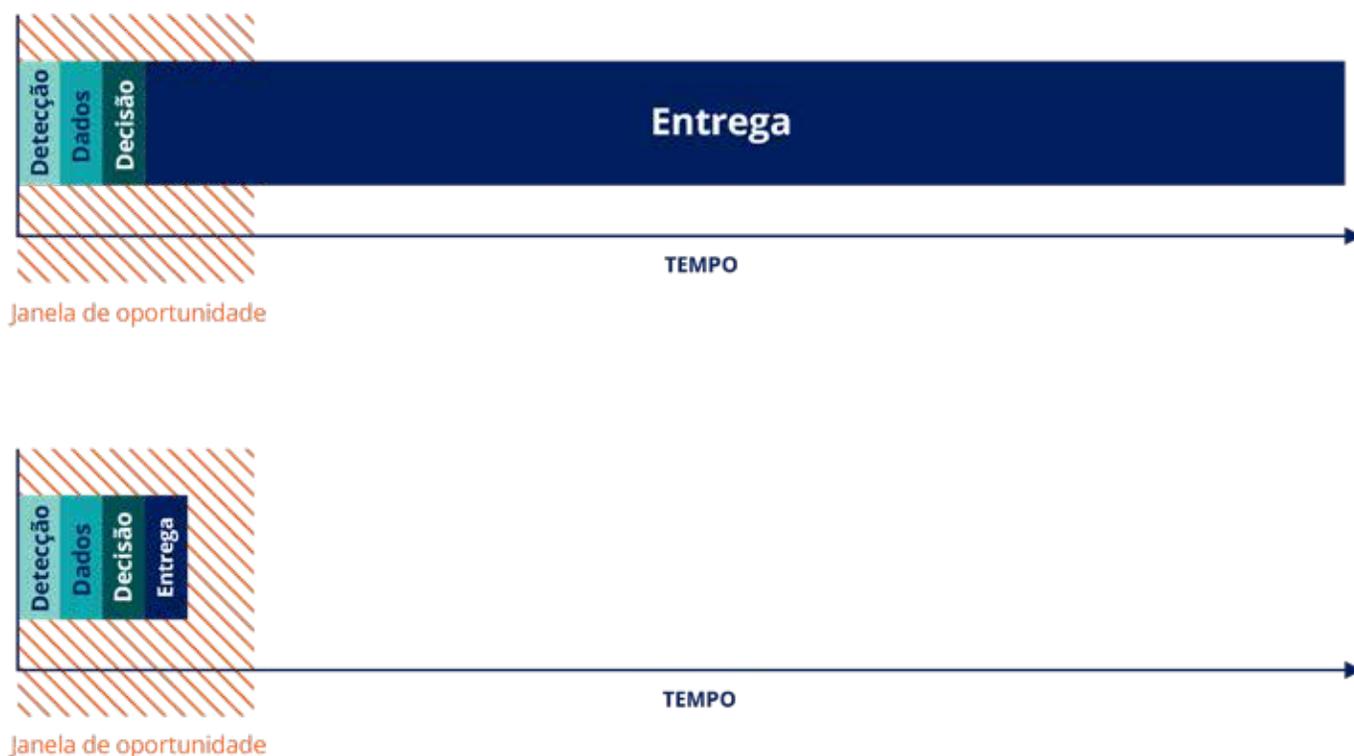


Ilustração: latência da entrega e a janela de oportunidade

A rapidez da entrega pode ficar em um gargalo devido à natureza de seus negócios, à sua abordagem de engajamento e aos canais adotados. Pode ser necessário mudar sua forma de pensar em clientes e canais em geral para maximizar seu potencial em tempo real. Por exemplo: seu mecanismo decisório pode identificar um “lead”, um cliente com alta propensão a agir imediatamente. Você pode querer que a equipe ligue para o cliente, porque sua equipe é muito boa na conversão em vendas. Então você carrega o lead para uma fila de chamadas para que seu representante ligue na próxima oportunidade: a manhã seguinte.

Você reuniu os dados, detectou a oportunidade, tomou uma decisão em tempo real e entregou-a ao canal, tudo em milissegundos. Mas, mesmo assim, pode estar se preparando para um fracasso. Você escolheu o canal que era melhor para suas necessidades, não as do cliente. Pode demorar consideravelmente para que o telefonema e o engajamento aconteçam. Ignorar o tempo do cliente pode resultar na perda da oportunidade. Muitas ações podem se beneficiar de um contato mais rápido, como um SMS, notificação no aplicativo ou e-mail em tempo real. E isso precisa fazer parte de seu processo decisório.

Considere esta regra geral: conforme aumenta o tempo necessário para tomar e entregar a decisão, a qualidade do engajamento com o cliente e a probabilidade de conversão geralmente diminui. Alguns exemplos simples, por canal:

Legenda Bom Razoável Ruim

Velocidade da decisão em tempo real	Site	Aplicativos móveis	Assistente virtual	Central de atendimento ao cliente	Loja física	Push no celular	E-mail e SMS	Anúncios pagos	Mala direta
< 200 milissegundos	✓	✓	✓	✓	✓	✓	✓	✓	✓
< 500 milissegundos	—	—	—	✓	✓	✓	✓	✓	✓
< 1 segundo	✗	✗	✗	—	—	✓	✓	✓	✓
< 10 segundos	✗	✗	✗	✗	✗	✓	✓	✓	✓
< 1 minuto	✗	✗	✗	✗	✗	—	—	✓	✓
< 1 hora	✗	✗	✗	✗	✗	✗	✗	—	✓
< 1 dia	✗	✗	✗	✗	✗	✗	✗	✗	—
< 1 semana	✗	✗	✗	✗	✗	✗	✗	✗	✗

Qualidade do engajamento do cliente

Ilustração: relação entre a velocidade de decisão e a qualidade do engajamento

- Em **canais digitais de entrada** (web, aplicativos móveis, etc.), a necessidade de decisões em tempo real é muito alta. Os clientes estão sempre consumindo conteúdo e criando pegadas digitais que podem alimentar os modelos e aumentar a relevância em ordem de magnitude, mas isso pode se estagnar quase imediatamente. Você é forçado a voltar às necessidades do cliente anteriores à avaliação se não tomar decisões em menos de 200 ms, o que diminuirá astronomicamente a probabilidade de conversão.
- Em **canais de entrada com atendentes**, como centrais de atendimento ou lojas físicas, a necessidade de reação em tempo real ainda é alta, mas o desempenho dos acordos de nível de serviço (SLA) é levemente reduzido. Conversas ao vivo acontecem mais devagar, com pausas naturais. Nesse tipo de canal, as decisões normalmente precisam acontecer entre 500 ms e 1 segundo para que o agente possa processá-las e usá-las durante uma conversa natural. Se demorar mais do que isso, as pausas ficam desconfortáveis e o agente para de usar as recomendações e volta a depender exclusivamente de seu próprio conhecimento do assunto, o que exclui quase todos os aspectos do contexto do cliente.
- Em **canais digitais de saída**, como e-mail, SMS ou notificação por push, a velocidade é importante, mas já não é medida em milissegundos. Geralmente, é medida em segundos ou minutos. Por exemplo: dados de geolocalização perto de um aeroporto, estação de

trem ou loja física podem identificar uma ótima janela de oportunidade para engajar o cliente e descobrir uma necessidade importante. Mas essa decisão não precisa ser feita em menos de 200 ms. Geralmente, a janela dura de segundos a minutos. Mais do que isso, o cliente vai embora. Quando essa janela se fecha, seu alcance é reduzido e pode ser inclusive desagradável ou intrusivo.

O que é “atraso X recompensa”

Para avaliar sua capacidade de trabalhar em tempo real, pode haver uma grande diferença entre o seu “atraso” (o tempo que ainda impede sua empresa de entrar no jogo) e a “recompensa” (a oportunidade de ganhar o jogo). O atraso e a recompensa são cumulativos. Cada atraso nas quatro categorias se multiplica e afeta mais etapas adiante. Se você passar da janela de oportunidade, perderá a própria oportunidade.



Ilustração: qual é o impacto da capacidade em tempo real

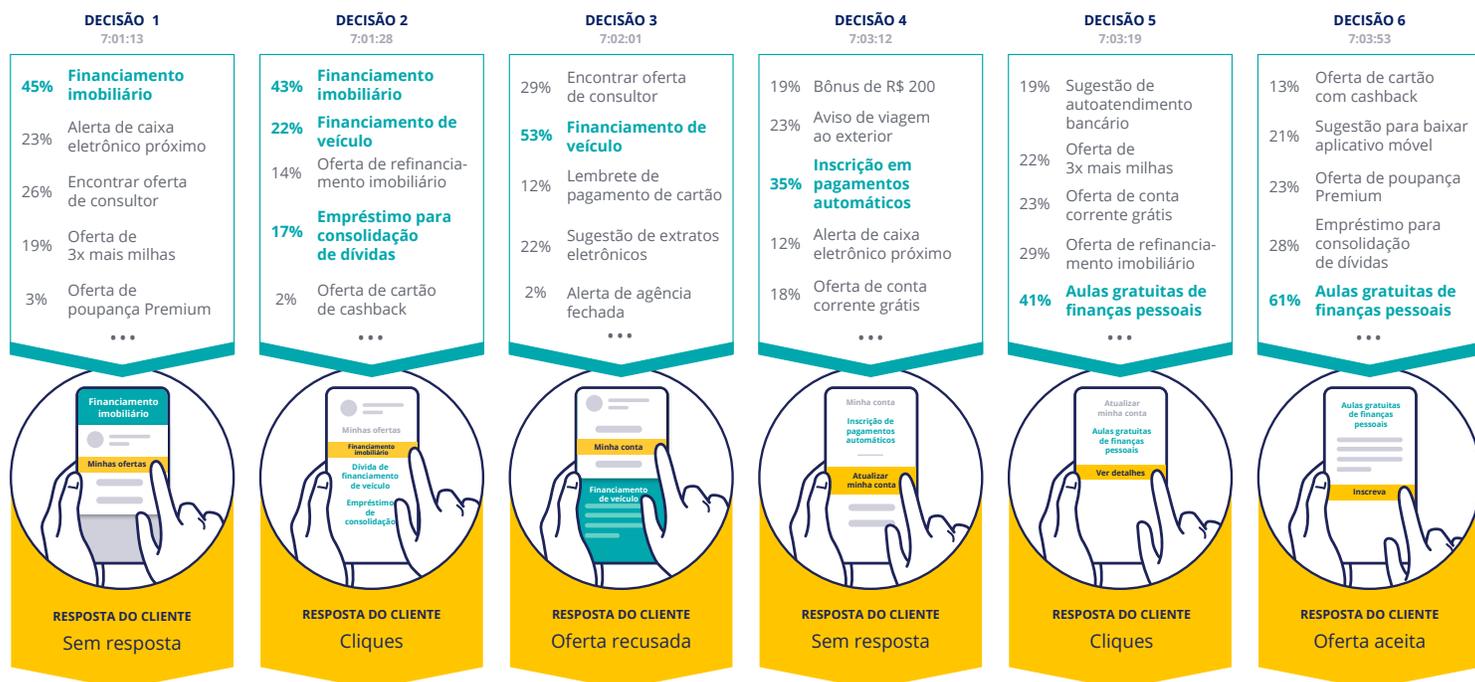
Qual é a velocidade necessária? Se você não se adaptar ao cliente e ao processo decisório dele, não vai dar certo. Você perde a oportunidade de usar o momento para gerar valor e coloca em risco a experiência do cliente.

Agregadores e novos intermediários estão sempre se colocando entre você e seu (possível) cliente. Você não concorre só com o relógio, mas também com todos os seus rivais. Não é só uma questão de quem vai responder, mas quem vai responder primeiro com a mensagem mais adequada. Dentro dessa janela de oportunidade, o espaço entre sua resposta e o próximo concorrente mais rápido é exatamente a sua oportunidade de obter vantagem competitiva e o potencial da “recompensa em tempo real”.

O “atraso em tempo real” é o espaço entre a otimização da oportunidade e não ter oportunidade nenhuma. As soluções mais sofisticadas de próxima melhor ação “refazem a decisão” várias vezes durante todo o processo em tempo real, em uma única interação com o cliente. Podem ser feitas de 10 a 50 decisões distintas em poucos minutos, em uma mesma interação. A cada momento, a pontuações de propensão são recalculadas para cada ação possível, considerando os novos dados e contextos. Uma das maiores empresas de telecomunicações dos EUA aplica essa abordagem a cenários de retenção de clientes em tempo real. Se um cliente liga para cancelar, recomenda-se um questionário para entender melhor o contexto pelos dados. As respostas orientam a oferta de retenção e cada cálculo adicional feito para apoiá-la. Nesses casos, os novos dados configuram a próxima etapa da interação, processando novos dados o tempo todo.

Exemplo de decisão refeita

A figura abaixo mostra o caminho de um cliente durante uma interação típica por celular, com a marca usando decisões em tempo real. Durante apenas 2 minutos e 40 segundos, a marca “refaz a decisão” para esse cliente seis vezes, sempre contabilizando novos dados de cliques em dispositivo móvel e outros comportamentos demonstrados pelo cliente durante a experiência. Ao final, o indivíduo recebeu seis próximas melhores ações muito diferentes, mas altamente relevantes, enquanto navegava pelo aplicativo móvel e mudava seu contexto.



Para entender verdadeiramente o poder e a complexidade das decisões deste exemplo, é importante saber o seguinte:

- Uma marca que consegue “refazer decisões” para seus clientes precisa ter uma **biblioteca de ações possíveis** como recurso, incluindo ofertas de vendas, recomendações de retenção, “sugestões” proativas de serviço e mensagens de relacionamento.
- É comum que as empresas tenham **mais de 1000 ações** possíveis para um só cliente, em qualquer ponto da decisão. Cada uma dessas ações também tem diversos tratamentos criativos associados, representando diversas combinações de imagens, texto, cor, aplicação de canal etc.
- Cada ação e tratamento também está associada a uma série de **modelos de propensão**, que são reavaliados com novos dados sempre que uma decisão é tomada. A pontuação desses modelos é usada para arbitrar ações e definir a mais adequada para o cliente naquele exato momento, considerando seu perfil completo e o contexto atual (dos últimos segundos).
- Cada clique, espera, recusa de oferta, etc. representa uma **mudança no contexto do cliente**. Quando esses eventos acontecem (o cliente clica em “não, obrigado” ou vai para outra página), esse dado é imediatamente aplicado nas decisões em tempo real, mapeadas para o modelo de dados mantido no aplicativo, e o perfil é reavaliado para o cálculo de uma nova “próxima melhor ação”.
- No exemplo acima, conforme o cliente recebe e seleciona/recusa/ignora as ações, e conforme o comportamento de cliques é agregado e organizado, a marca está **sempre mudando sua abordagem**.
- O mecanismo decisório se adapta: nas primeiras três interações, foram recomendadas ofertas de vendas. Depois, somente recomendações de serviços, retenção e relacionamento durante o resto da experiência, porque o comportamento da pessoa mudou a pontuação de propensão e mostrou que essas ações “menos agressivas” eram mais **relevantes para seu contexto e intenção**.
- As marcas sempre tentaram se alinhar ao contexto do cliente, usando somente segmentação e regras de negócios, mas essas abordagens antigas de “um para muitos” **não se adaptam ao contexto em tempo real** nem atingem o nível de relevância desejado. Por isso, elas costumam ser ineficazes.
- Sem uma abordagem baseada em propensão e constante reavaliação das decisões para cada perfil de cliente com base em novos dados, é impossível atingir o **aumento de 5 vezes na relevância** que é associado ao investimento em tempo real.

As soluções mais sofisticadas de próxima melhor ação “refazem a decisão” várias vezes durante todo o processo em tempo real, em uma única interação com o cliente.

Conclusão

Nem todo tempo real é igual. Selecionar uma solução de próxima melhor ação que não aceite todas as quatro etapas do “tempo real verdadeiro” arrisca a experiência de seu cliente e seu potencial de ganhos. Analistas independentes de setores como telecomunicações, serviços financeiros e seguros constataram que, em média, uma implementação bem-sucedida de próxima melhor ação em tempo real traz mais de 225 milhões de dólares em vendas e valor de retenção a cada 10 milhões de clientes, por ano.

Para cada mês que sua empresa deixa de usar uma solução de ponta de próxima melhor ação com funções em tempo real verdadeiro, você pode estar perdendo cerca de 20 milhões de dólares para cada 10 milhões de clientes.

É importante começar a estratégia com simplicidade, mas não cometer o erro de usar um software simples. Sua base para próximas melhores ações em tempo real precisa administrar toda essa complexidade. A escolha correta vai fazer você voar à frente da concorrência. Do contrário, você não vai ter espaço para decolar.

É hora de levar seu futuro em tempo real a sério.





A Pega é líder em software em nuvem para interação com clientes e excelência operacional. As maiores marcas do mundo usam softwares controlados por IA da Pega para otimizar todas as interações com clientes em qualquer canal, sem deixar de cumprir as promessas de marca. A plataforma de desenvolvimento de aplicativos low-code da Pega permite que as empresas criem e aprimorem aplicativos para ajudar clientes e funcionários, além de estimular a transformação digital em escala global. Há mais de 35 anos, a Pega ajuda a promover satisfação do cliente, custos mais baixos e maior valor vitalício do cliente.

Saiba mais. Acesse www.pegacom