

Perché prendere decisioni in tempo reale è fondamentale

UN WHITE PAPER
DI PEGA



Introduzione

La customer experience (CX) è a un punto morto in molti settori. Le aziende stanno usando approcci e tecnologie obsoleti per cercare di risolvere problemi complessi relativi al coinvolgimento dei clienti, ma non ci stanno riuscendo. Tutto ciò sta generando un alto grado di frustrazione:

- **I consumatori** sono frustrati perché le loro interazioni con le aziende sono spesso difficili, noiose e non pertinenti. Un'esperienza positiva è seguita da dieci momenti da dimenticare. Si sentono come se fossero costantemente coinvolti in una presentazione di vendita.
- **I dirigenti aziendali** sono frustrati perché hanno investito in soluzioni che avrebbero dovuto risolvere il problema e distinguerli dalla concorrenza, e invece hanno ottenuto solo sporadici miglioramenti senza un significativo ritorno sull'investimento.
- **Gli investitori** sono frustrati perché le aziende si sono occupate solo a parole della customer experience, mentre le lamentele continuano ad aumentare. Le quote di mercato si assottigliano e i dividendi si riducono.

Il problema di fondo: le esigenze dell'azienda vs. problemi dei clienti

Il vero problema non sta nelle aspettative troppo alte dei clienti o nel fatto che le aziende sono restie a investire nella customer experience, perché non è vero. I consumatori vogliono semplicemente più esperienze positive, come quando le aziende anticipano le loro esigenze, rendono le informazioni facili da reperire e sembrano apprezzare davvero il rapporto instaurato. Allo stesso modo, le aziende hanno speso innumerevoli miliardi nella customer experience per cercare di soddisfare e superare le aspettative dei clienti. Ma dopo aver creduto alle promesse di fornitori, agenzie e partner hanno ottenuto ben poco. Tuttavia, le aziende continuano a spendere perché continuano a perdere terreno.

Il vero problema è che queste soluzioni sono state ideate per supportare le esigenze dell'azienda, non per risolvere i problemi dei clienti.

Le suite di automazione per il marketing, i desktop del servizio clienti, gli strumenti per sviluppare modelli e le piattaforme di Business Intelligence (BI) sono parte di un intero ecosistema di soluzioni progettate per facilitare il movimento di offerte e messaggi nel mercato da parte delle aziende. Queste soluzioni raccolgono dati da diversi sistemi in batch, consolidano storie, analizzano comportamenti, aggiungono informazioni di terzi, creano regole, costruiscono segmenti, ottimizzano il targeting e generano campagne. Permettono alle aziende di coinvolgere i clienti con estrema e crescente rapidità e con un costo minimo rispetto al passato. Tuttavia, ignorano un aspetto cruciale:

Le persone sono creature complesse che agiscono di impulso e cambiano idea rapidamente. Quando ricevono nuove informazioni, spesso nel giro di pochi secondi sono spinte in nuove direzioni.

Le persone sono fluide, dinamiche e flessibili. Questo è il motivo per cui un ecosistema costruito su dati creati in batch, database relazionali, decisioni basate sui segmenti e coinvolgimento basato su campagna non riesce a soddisfare i requisiti della customer experience moderna. Queste tecnologie sono state sviluppate per scegliere un messaggio e trasmetterlo a un ampio pubblico o a segmenti di mercato su larga scala, non per analizzare le esigenze individuali di un cliente in un determinato momento e gestirle rapidamente prima che cambino.

Il vantaggio competitivo del real-time decisioning

Il contesto del cliente (situazione, ambiente, emozioni, motivazioni e comportamenti) cambia continuamente mentre assimila nuove informazioni. Si presentano opportunità che possono svanire in pochi secondi. Ma quelle stesse persone sono esposte ogni giorno a un'enorme quantità di messaggi della concorrenza, ognuno dei quali assomiglia molto al vostro. Non vi distinguerete se non sarete in grado di leggere i segnali e ottimizzare il vostro messaggio in quella finestra ristretta. Ciò non significa necessariamente che perderete l'opportunità perché i vostri concorrenti sono migliori. Semplicemente, non vi siete mossi con la rapidità necessaria e i concorrenti erano lì, al posto giusto e al momento giusto. Ne consegue che le decisioni in tempo reale sono ora direttamente legate all'acquisizione del vantaggio competitivo.

Si avvantaggeranno i brand in grado di ridefinire una decisione 5, 10, 20 o anche 50 volte durante una singola esperienza. Ogni volta che queste aziende attivano nuovi dati e sfruttano modelli predittivi per reindirizzare una propensione del cliente ad agire, diventano sempre più rilevanti rispetto ai brand che non riescono a farlo, guadagnando nel mentre un **enorme vantaggio competitivo** durante il processo. Questo documento spiega come vengono prese quelle decisioni rapide in tempo reale, quali dati e strumenti analitici vengono utilizzati per alimentarle e quanto rapidamente deve muoversi un'azienda per avere un impatto significativo sulla customer experience.

I quattro pilastri del real-time decisioning

Rendere possibili interazioni rilevanti e contestuali implica qualcosa di più delle semplici decisioni in tempo reale. In tutto sono quattro i fattori che contribuiscono a un sistema decisionale in tempo reale ottimizzato: rilevamento, dati, decisioni ed erogazione. La comprensione di questi fattori vi consente di valutare in che misura siete pronti a prendere decisioni in tempo reale.



Rilevamento

Identificare l'opportunità di un cliente

Il rilevamento di un'opportunità è il primo elemento portante del programma di coinvolgimento in tempo reale e il primo aspetto da valutare. In parole semplici, nessuna delle capacità o dei processi a valle è rilevante se non si è consapevoli dell'opportunità di vendita, assistenza e fidelizzazione.

Usare il rilevamento preventivo

L'aspetto più importante del rilevamento è che non può essere pianificato. Nel mercato di oggi deve essere intuito analizzando i dati contestuali. Le finestre di opportunità digitali si aprono e si chiudono in pochi secondi. Ciò è in netto contrasto con gli anni passati, quando era prassi comune pianificare processi in batch da eseguire ogni notte o settimanalmente.

Per riuscire a cogliere questi momenti di opportunità mentre si presentano, o anche prima, dovrete attivare un rilevamento degli eventi che utilizzi sia i dati slow-moving sia quelli più dinamici per la previsione di ciò che probabilmente accadrà. Molti casi d'uso richiederanno un rilevamento quasi istantaneo, come ad esempio l'estrazione di flussi di dati ad alta velocità per eventi rari, in particolare quelli con una finestra di opportunità piccola. A volte sono sufficienti gli attivatori di eventi semplici, ad esempio se accade questo fai quest'altro. Ma i casi d'uso emergenti più potenti sono costruiti su contesti solidi che contengono eventi complessi e interconnessi per cui sono richieste competenze più sofisticate.

L'aspetto più importante del rilevamento è che non può essere pianificato. Nel mercato di oggi deve essere intuito analizzando i dati contestuali.

Quali sono i diversi tipi di eventi?

Il rilevamento di un evento complesso va alla ricerca di modelli nelle origini dati e cerca di identificare quando potrebbe verificarsi un potenziale momento di opportunità.

- **Eventi semplici:** ascolto di una singola origine dati, ad esempio il clic su un link
- **Non eventi:** eventi attesi che non si sono verificati, ad esempio un deposito abituale, l'accesso online o da dispositivo mobile, una visita in filiale ecc.
- **Eventi complessi:** una combinazione di eventi semplici e non eventi che insieme rappresentano un momento potenzialmente importante

Rilevare la finestra di opportunità

Il rilevamento è un elemento portante fondamentale delle decisioni in tempo reale. Se la vostra capacità di rilevamento non è sufficientemente rapida, probabilmente non riuscirete a cogliere la vostra finestra di opportunità. Questo rende irrilevante qualsiasi altra cosa (combinazione dei dati, processi decisionali ed erogazione), perché avete già perso la vostra occasione.

Scenario 1: il rilevamento avviene troppo lentamente e non lascia tempo per reagire.



Scenario 2: il rilevamento avviene rapidamente, lasciando il tempo sufficiente per reagire.



Figura: latenza del rilevamento e finestra di opportunità

La progettazione di esperienze digitali spesso richiede tempi di risposta di 100 - 200 millisecondi (ms).

Quanto velocemente?

I requisiti aziendali per la velocità di rilevamento e l'esecuzione end-to-end possono essere misurati in microsecondi (un milionesimo di secondo) per alcuni casi d'uso, come il trading online e l'acquisto di pubblicità in tempo reale. Le esperienze digitali, come i clic sul sito web e le interazioni con app mobili, richiedono tempi di risposta inferiori a mezzo secondo. E la progettazione di esperienze digitali spesso richiede tempi di risposta di 100 - 200 millisecondi (ms). Gli intervalli di tempo superiori a un secondo possono portare progressivamente a risultati non ottimali o anche negativi.

Il settore delle telecomunicazioni è relativamente avanzato nell'approccio alle decisioni in tempo reale. Ad esempio, un consumatore che ha raggiunto il limite massimo per il credito telefonico prepagato può trasformarsi in un'opportunità di presentare un'offerta contestualizzata correlata al raggiungimento di quell'obiettivo. Questo è un semplice evento scatenante, ma è necessario agire immediatamente.

Analogamente, nel settore dei servizi finanziari, una banca potrebbe voler rilevare se un cliente ha lasciato il denaro o la carta in uno sportello bancomat durante un'interazione, con l'obiettivo di avvisarlo dell'errore al più presto. In un esempio più complesso, una società di

gestione patrimoniale può combinare i dati provenienti dall'attività di un conto, l'utilizzo dei prodotti, la navigazione in rete e l'utilizzo delle applicazioni mobili per dedurre la probabile intenzione di una persona in un momento specifico. L'azienda potrebbe poi usare questi dati per presentare ai clienti suggerimenti, messaggi e offerte contestualmente più pertinenti.

Dati

Accedere e combinare le informazioni in tempo reale

Per una vera soluzione in tempo reale è necessario combinare i dati in tempo reale e utilizzarli per comprendere il contesto specifico del cliente. Le aziende interagiscono con le persone tramite più canali e acquisiscono dati da più fonti. Il contesto particolare di una persona (le cose che contano in questo preciso momento) è in genere molto più ampio rispetto a ciò che cogliamo con una singola interazione ed è in costante evoluzione. La maggior parte dei sistemi e degli approcci fallisce perché non è in grado di considerare il quadro generale.

Ad esempio, potreste essere in grado di rilevare un cliente che attraversa un recinto virtuale, che potrebbe essere potenzialmente un evento importante (ad esempio quando entra in un negozio o nella filiale di un concorrente). Ma quel singolo evento è solo una piccola parte di un quadro più ampio per quel cliente. Anche se l'attivatore è importante, avete bisogno di più dati per decidere se agire subito e, in tal caso, quale azione intraprendere.

Ad esempio, se il cliente sta cercando di utilizzare un prodotto, potrebbe entrare nel negozio per chiedere assistenza, non per un nuovo acquisto. Questo non potete saperlo se non state estraendo i dati giusti (il motivo della chiamata, la probabilità di richiesta di assistenza e la cronologia della navigazione). Senza una visione d'insieme, probabilmente cerchereste di vendere al cliente qualcosa... nel peggior momento possibile. Questo tipo di interazione può danneggiare notevolmente la relazione a lungo termine con quel cliente, specialmente se si ripete nel tempo.

In molti casi i dati significativi sono assenti o non vengono presi in considerazione nel processo decisionale.

Per prendere una decisione intelligente in tempo reale, è fondamentale essere in grado di integrare quei dati aggiuntivi. Purtroppo i dati non possono rimanere invariati in un database relazionale progettato per processi in batch. I dati devono spesso adattarsi a un sistema federato per essere analizzati e applicati in tempo reale.

Le cinque dimensioni del contesto

Per consentire un'interazione effettivamente rilevante ed efficace, è necessario combinare le origini dati in grado di mostrare ogni aspetto del contesto relativo al cliente nel momento specifico in cui si verifica l'interazione.

Esistono cinque principali categorie di contesto:

Contesto	Descrizione	Esempi di origini dati
Ambientale	Posizione del cliente: luogo fisico o digitale dove si trova la persona	Posizioni segnalate da dispositivi mobili, domini, ID container
Comportamentale	Azioni del cliente: in che modo ha interagito o reagito	Transazioni, cronologia delle interazioni, navigazione sul Web
Emotivo	Sentiment del cliente: come si sente adesso	Tassi di abbandono, probabilità predefinita, esito delle chiamate
Motivazionale	Intenzione del cliente: cosa vuole ottenere	Propensione alle chiamate, motivi delle chiamate, moduli inviati
Situazionale	Lo stato del cliente: progressi verso il raggiungimento dell'obiettivo	Fasi del percorso, stato del caso, punteggi del rischio di credito

Molte delle cosiddette soluzioni in tempo reale tralasciano erroneamente questo aspetto di costruzione del contesto del tutto, saltando direttamente da un attivatore dell'evento a un'azione predeterminata.

Prendiamo in considerazione un esempio semplice. Il decoder della TV smette di funzionare quindi l'utente va sul sito web del fornitore in cerca di una soluzione rapida. Tuttavia, questo motivo specifico per visitare il sito web attiva un'invasiva offerta di marketing o una finestra popup che si intromette tra l'utente e il risultato che cerca di ottenere (riparare il decoder), aumentando ancora di più la sua frustrazione. Questo tipo di situazione è comune e avviene perché il brand non ha la comprensione contestuale di quel cliente specifico, anche se i riferimenti al suo contesto erano facilmente accessibili tramite i dati del decoder.

Il contesto fornisce una prospettiva aggiuntiva con cui prendere una decisione. I dati contestuali di maggior valore sono, ad esempio, i dati CRM, le attività di ricerca sul Web, i clic sull'applicazione mobile, le transazioni, i post sui social, i dati dei dispositivi e dell'IoT, le opinioni espresse in forma scritta (chat, social, e-mail, trascrizione da voce a testo) e i dati di terzi, come le previsioni meteo o la prossimità.

Nell'esempio del decoder, il contesto del dispositivo non funzionante rappresenta un indizio importante nel contesto situazionale e motivazionale in cui si trova la persona e determina cosa potrebbe essere giusto fare (o non fare) successivamente.

Quando si valutano i dati contestuali, è necessario tenere in considerazione due aspetti:

- **Stagnazione:** in quanto tempo i dati diventano obsoleti e perdono valore? Le persone sono complesse e cambiano continuamente, e di conseguenza i loro dati. La latenza dei dati può avere un notevole impatto sulla pertinenza, in particolare laddove sono

coinvolti processi di elaborazione in batch. L'utilizzo di dati stagnanti vi porterà ad affrontare problemi contestuali che non esistono più. Ad esempio, potreste indirizzare la vostra offerta tramite una notifica push su app mobile a una persona che si trova in un determinato luogo dopo che ha già lasciato il punto vendita.

- **Portata:** i dati contestuali che combinate tra loro coprono ogni aspetto del contesto? In molti casi i dati significativi sono assenti o non vengono presi in considerazione nel processo decisionale. Spesso quelle informazioni avrebbero potuto fornire una visione più dettagliata del contesto del cliente in quel momento e avrebbero cambiato completamente la decisione. Ad esempio, quando un brand ignora il contesto motivazionale (come nel caso del decoder non funzionante) ed erroneamente fa un'offerta commerciale anziché cercare di risolvere un problema effettivo.

La velocità di combinazione dei dati è fondamentale

Non appena viene rilevata l'esigenza o l'opportunità di un cliente, come spiegato successivamente, la rapidità nel combinare i dati contestuali assume improvvisamente un'importanza enorme. Il tempo inizia a stringere e dovete combinare le informazioni con la minor latenza possibile per prendere una decisione intelligente e ben ponderata.

Scenario 1: il rilevamento avviene rapidamente, ma i dati vengono combinati tra loro troppo lentamente senza lasciare il tempo di reagire entro la finestra di tempo.



Scenario 2: la combinazione dei dati avviene rapidamente, fornisce una conoscenza sufficiente per consentire di prendere decisioni intelligenti e lascia abbastanza tempo per la fase successiva.



Figura: latenza della combinazione dei dati e finestra di opportunità

La latenza non è una grande preoccupazione per certi tipi di dati come indirizzi postali, numeri di telefono, date di nascita o possesso dei prodotti. Semplicemente, sono dati che non cambiano molto spesso. Invece i dati di localizzazione e provenienti dalle applicazioni mobili e altre tracce digitali cambiano in modo estremamente veloce. Il cliente ha appena cliccato sulla nuova offerta della carta di credito Gold? Oppure stava passando il cursore del mouse sopra i

termini per la chiusura dell'account? È ancora sulla stessa pagina web o nello stesso negozio? Informazioni come queste sono fondamentali per identificare il contesto.

Il contesto non è solo importante per l'iper personalizzazione nel canale stesso, ma anche per l'interazione tra più canali. Senza una comprensione del contesto nel complesso, i brand possono essere colti alla sprovvista in un "punto cieco" del comportamento. Un cliente potrebbe non passare semplicemente da un canale all'altro, ma usarli contemporaneamente, ad esempio per consultare i termini e le condizioni sul sito web mentre parla al telefono con un operatore oppure per visualizzare i dettagli di un prodotto sull'applicazione mobile mentre fa acquisti in negozio. Le attività su un canale devono fornire informazioni sulla next best action su qualsiasi altro canale per ottimizzare quella specifica interazione in quel momento specifico. Ciò dimostra perché le funzionalità in tempo "veramente" reale contano davvero.

La vostra capacità decisionale dipende interamente dai dati che utilizzate per alimentarla. In molti casi, se la vostra azienda non è in grado di rendere operativi i dati contestuali che precedono l'interazione, non sarete in grado di prendere la decisione giusta per quel cliente. Quando si verifica questa situazione, il coinvolgimento, la soddisfazione e i tassi di conversione ne risentono di conseguenza. Nella cultura dell'opt-out di oggi, ogni messaggio irrilevante ricevuto dal cliente lo avvicina sempre di più al disimpegno.

Decisione

In che modo le regole possono ostacolare la modellazione adattiva in tempo reale

Il rilevamento in tempo reale e la combinazione dei dati in tempo reale forniscono solide basi su cui prendere una decisione intelligente. Tuttavia, non tutte le decisioni in tempo reale vengono prese in modo uguale. La semplice logica deterministica del "se si verifica questa situazione, allora accade questo" frustra i clienti e ostacola anziché aiutare. Quei tipi di regole non rappresentano le esigenze, le emozioni o le motivazioni singolari del cliente in quel momento specifico. I fattori contestuali come questi rappresentano la fase successiva dell'evoluzione della customer experience per le aziende, che deve prendere in considerazione un livello di decisione più sofisticato.

Livelli di decisione in tempo reale

I livelli di maturità del processo decisionale in tempo reale sono quattro:



Figura: maturità del processo decisionale per la next best action

Livello 1: regole deterministiche

Le regole aziendali deterministiche vengono spesso utilizzate per configurare la logica decisionale. Tali regole devono essere relativamente semplici. In caso contrario, diventano ingestibili e non sono in grado di determinare risultati contestuali e probabilistici a causa di tutti i diversi fattori e permutazioni da tenere in considerazione. Esempi di regole deterministiche includono le istruzioni if-then-else, le scorecard lineari e le strutture decisionali.

Livello 2: analisi con punteggio precalcolato (offline) e regole deterministiche

Questo livello aggiunge i modelli analitici (propensione) per prevedere il comportamento del cliente. A questo livello le conclusioni analitiche sono state predeterminate e il punteggio è stato precalcolato prima che si verifichi l'interazione o l'opportunità. In questi casi i modelli vengono elaborati in batch programmati (ad esempio, ogni notte/settimana/mese) in base ai dati disponibili in quel momento. I risultati vengono poi caricati nell'ambiente decisionale per essere utilizzati quando si verifica l'interazione. Quando arriva il momento, i dati con punteggio precalcolato vengono utilizzati nell'arbitraggio della decisione in tempo reale.

Indubbiamente, questa capacità è meglio che non avere alcuna previsione. Ma ci sono svantaggi significativi che spesso conducono a una decisione sbagliata. Supponiamo che un cliente rifiuti l'offerta di una carta di credito Gold. Questo cambiamento del contesto dovrebbe condizionare immediatamente la probabilità che il cliente accetti altre carte di credito, prodotti di finanziamento o anche il rischio di abbandono. Tuttavia, il nuovo contesto non verrà preso in considerazione se si utilizzano le previsioni con punteggio precalcolato. I punteggi non cambieranno fino al seguente aggiornamento programmato, portando a previsioni imprecise e suggerimenti potenzialmente irrilevanti (se non deleteri).

Consideriamo l'esempio precedente di interazione su più canali, in cui un cliente ha appena iniziato l'interazione sul sito web o sul sistema di risposta vocale interattiva (IVR) con un operatore. I modelli analitici saranno in grado di considerare quelle nuove e preziose tracce digitali nei loro calcoli e ridefinire il punteggio di tutte le conclusioni in tempo reale? Il processo decisionale basato su previsioni con punteggio precalcolato non tiene in considerazione questo contesto critico. Anche i modelli offline sono incredibilmente inefficienti e sprecano potenza di calcolo nel calcolare probabilità che non verranno mai usate e che sono destinate a stagnare rapidamente.

Livello 3: modelli analitici in tempo reale (live) e regole deterministiche

Un'architettura superiore utilizza le previsioni in tempo reale on-demand con il contesto più recente. A questo livello tutta la decisione, dall'inizio alla fine, viene presa in tempo reale per supportare un approccio al coinvolgimento basato sulla next best action. I modelli analitici, anziché i semplici punteggi caricati, vengono eseguiti davvero in tempo reale sfruttando l'intero spettro di dati in tempo reale. In questo caso la decisione per la next best action nella sua interezza viene presa utilizzando il contesto completo del cliente. Viene poi selezionata la decisione ottimale per quel momento.

Nell'esempio del decoder, ad esempio, i dati dell'IoT provenienti da quel dispositivo, combinati con il comportamento dei clic sul Web in tempo reale, possono essere inclusi nella stessa decisione in tempo reale, insieme alle previsioni che supportano tale conclusione.

Livello 4: modelli adattivi in tempo reale e regole deterministiche

Gli esempi più sofisticati combineranno le regole deterministiche con i modelli adattivi in tempo reale (self-learning) da utilizzare nelle decisioni in tempo reale. In questo caso i modelli stessi cambiano continuamente in base al comportamento del cliente. Grazie al machine learning (ML), i modelli apprendono da ogni previsione (corretta o sbagliata) che fanno. Se vengono combinati con le regole deterministiche, il risultato è una maggiore efficienza, previsioni più accurate, decisioni migliori, risultati aziendali e customer experience decisamente migliori (che è ovviamente l'obiettivo).

Rapidità di decisione

Precisione a parte, anche la rapidità con cui si prendono le decisioni è fondamentale, come mostrato nella figura. Ad esempio, un'azienda rileva che un potenziale cliente si trova "nel mercato" per un mutuo e raccoglie tutti i dati necessari per quella decisione. Ma se l'azienda non riesce a presentare l'offerta con la rapidità necessaria, la finestra potrebbe chiudersi e il cliente potrebbe rivolgersi a un concorrente diretto.

Scenario 1: viene rilevato un evento e i dati vengono combinati rapidamente, ma la decisione viene presa troppo lentamente e non consente di reagire entro la finestra di tempo.



Scenario 2: viene rilevato un evento, i dati vengono combinati rapidamente, la decisione viene presa in modo tempestivo lasciando tempo per le attività a valle.



Figura: latenza del processo decisionale e finestra di opportunità

Anche in questa fase le esigenze e gli obiettivi dell'azienda vengono ponderati per decidere la next best action, inclusi aspetti quali la fidelizzazione, il lifetime value del cliente, le priorità del servizio clienti, il margine, l'inventario, la propensione all'opt-out e molto altro.

Decisioni su larga scala

Alla luce di quanto sopra, è necessario prendere in considerazione anche la portata e il volume delle decisioni. Per qualsiasi interazione, possono esistere centinaia o migliaia di azioni possibili per le quali è necessario valutare la rilevanza, utilizzando il contesto completo di quel cliente. In genere il volume delle decisioni è di 500 - 1000 decisioni al secondo nelle grandi aziende (circa 86 milioni di decisioni al giorno). Molte aziende hanno bisogno di una scalabilità maggiore. Alcune società di servizi finanziari raggiungono volumi di picco di oltre 200 milioni di decisioni al giorno.

L'altra considerazione chiave nel processo decisionale riguarda la scalabilità della modellazione stessa. Le tecniche come la modellazione adattiva (basata sul machine learning) permettono alle aziende di creare e gestire automaticamente modelli più predittivi di quelli che potrebbero realizzare con le sole risorse della scienza dei dati.

Erogazione

Rispondere entro i tempi della finestra di opportunità

Per erogazione si intende l'azione intrapresa dal brand per rispondere all'opportunità nel canale, coinvolgendo direttamente il cliente con l'azione consigliata.

Valutare il punto di inflessione

È importante capire che i primi tre pilastri (rilevazione, dati, decisione) devono realizzarsi in tempo il più possibile reale. Di conseguenza l'azienda può sapere se deve intraprendere una qualche azione e se è in grado di farlo entro i tempi della finestra di opportunità. Ciò serve come importante "punto di inflessione", poiché la decisione può tener conto della potenziale latenza/irrelevanza nella fase di erogazione. Ad esempio, se un'azienda riesce a determinare di non essere in grado di reagire con l'azione desiderata nella finestra temporale oppure si rende conto che esiste un rischio di interazione irrilevante o negativa, la decisione migliore potrebbe essere semplicemente quella di non agire e annullare l'erogazione di un'azione potenzialmente tardiva.

Man mano che aumenta il tempo necessario per elaborare ed erogare una decisione, generalmente diminuiscono la qualità del coinvolgimento del cliente e la probabilità di conversione dell'azienda.

Ad esempio, una società di servizi finanziari del Nord America rileva gli eventi, combina i dati e prende decisioni in tempo reale su larga scala, ma considera anche l'opzione di non erogazione in un punto di contatto del cliente. Quando si verifica questa situazione, la decisione viene comunque presa per registrare il contesto del cliente al fine di migliorare le decisioni future.

Erogare rapidamente

Quando una risposta è ritenuta appropriata, l'ultimo miglio è rappresentato dalla capacità dell'azienda di reagire in tempo reale. La figura seguente illustra l'impatto dell'erogazione latente sui risultati, nel contesto di una decisione in tempo reale.



Figura: latenza dell'erogazione e finestra di opportunità

La rapidità dell'erogazione può essere ostacolata dalla natura dell'azienda, dall'approccio nei confronti del coinvolgimento e dai canali usati. Per massimizzare il potenziale delle decisioni in tempo reale, potrebbe essere necessario cambiare il modo in cui pensate ai clienti e ai canali in generale. Ad esempio, il vostro motore decisionale potrebbe identificare un "lead", ossia un cliente con un'elevata propensione ad agire ora. Il vostro team potrebbe chiamare quel cliente, perché è molto bravo a convertire i lead in vendite. Quindi caricate il lead nell'elenco delle chiamate in uscita affinché un operatore possa chiamarlo non appena possibile, cioè il mattino seguente.

Avete combinato i dati, individuato l'opportunità, preso una decisione in tempo reale e l'avete fornita sul canale, tutto in poche centinaia di millisecondi. Ma potreste comunque fallire nell'impresa. Avete scelto il canale migliore per voi, non per il cliente. Una chiamata può subire un notevole ritardo prima che il canale riesca a coinvolgere effettivamente il cliente. La mancata attenzione verso i tempi del cliente può tradursi in un'opportunità mancata. Molte azioni possono trarre vantaggio dall'uso di un mezzo più rapido, come l'invio di un SMS, una notifica sull'applicazione o un'e-mail. Tutto ciò deve far parte del vostro processo decisionale.

Considerate questa regola generale: man mano che aumenta il tempo necessario per elaborare ed erogare una decisione, generalmente diminuiscono la qualità del coinvolgimento del cliente e la probabilità di conversione dell'opportunità da parte dell'azienda. Ecco qualche

Legenda: Buona (verde), Sufficiente (grigio), Insufficiente (arancione)

Rapidità delle decisioni in tempo reale	Sito web	App mobile	Assistente virtuale	Contact center	Punto vendita	Notifica push su app	E-mail e SMS	Pubblicità a pagamento	Pubblicità diretta
< 200 millisecondi	✓	✓	✓	✓	✓	✓	✓	✓	✓
< 500 millisecondi	—	—	—	✓	✓	✓	✓	✓	✓
< 1 secondo	✗	✗	✗	—	—	✓	✓	✓	✓
< 10 secondi	✗	✗	✗	✗	✗	✓	✓	✓	✓
< 1 minuto	✗	✗	✗	✗	✗	—	—	✓	✓
< 1 ora	✗	✗	✗	✗	✗	✗	✗	—	✓
< 1 giorno	✗	✗	✗	✗	✗	✗	✗	✗	—
< 1 settimana	✗	✗	✗	✗	✗	✗	✗	✗	✗

Qualità del coinvolgimento dei clienti

Figura: relazione tra rapidità della decisione e qualità del coinvolgimento

esempio a livello di canale:

- Sui **canali digitali in entrata** (Web, app mobili ecc.), la necessità di decisioni in tempo reale è molto elevata. I clienti consumano continuamente contenuti creando tracce digitali che possono attivare modelli e aumentare la pertinenza di un ordine di grandezza, ma stagneranno quasi all'istante. Se non riuscite a prendere una decisione in meno di 200 ms, dovrete tornare a precalcolare il punteggio delle esigenze del cliente, il che comporterebbe una drastica riduzione della probabilità di conversione del cliente.
- Sui **canali in entrata gestiti dagli operatori**, come un call center o un punto vendita, la necessità di prendere decisioni rapide in tempo reale è ancora molto elevata, ma gli accordi sul livello di servizio (SLA) delle prestazioni diminuiscono leggermente. Le conversazioni dal vivo hanno un ritmo più lento e comprendono delle pause naturali. In questo tipo di canali è necessario restituire le decisioni in un tempo compreso tra 500 ms e 1 secondo affinché l'operatore possa elaborarle e utilizzarle nel corso naturale della conversazione. Se l'attesa si prolunga ulteriormente, queste pause diventano imbarazzanti, l'operatore smette di usare i vostri suggerimenti e torna a fare affidamento solo sulla propria esperienza in materia, ignorando quindi molti aspetti del contesto del cliente.
- Sui **canali digitali in uscita**, come e-mail, SMS o notifiche push su app mobili, la rapidità è importante ma non si misura più in millisecondi. Di solito si misura in secondi o minuti. Ad

esempio, i dati di geolocalizzazione intorno a un aeroporto, a una stazione ferroviaria o a un punto vendita possono identificare una grande finestra di opportunità per coinvolgere un cliente e individuare un bisogno specifico. Tuttavia, non è necessario prendere quella decisione in meno di 200 ms. Di solito quella finestra dura per un tempo che va da qualche secondo a qualche minuto circa. Ma se l'attesa è più lunga, il cliente passa ad altro. Quando la finestra si chiude, il vostro raggio di azione diventa molto meno efficace e potrebbe diventare sospetto o invadente.

Il concetto di "rottura e ricompensa"

Quando valutiamo le funzionalità in tempo reale, ci può essere una grande differenza tra "rottura" (l'intervallo di tempo che impedisce alla vostra azienda di partecipare) e "ricompensa" (l'opportunità di vincere davvero il gioco). La rottura e la ricompensa sono cumulative. Ogni ritardo nelle quattro categorie si combina e incide sulle fasi successive. Se mancate la finestra di opportunità, perdete l'opportunità stessa.



Figura: comprensione dell'impatto delle larghezze di banda in tempo reale

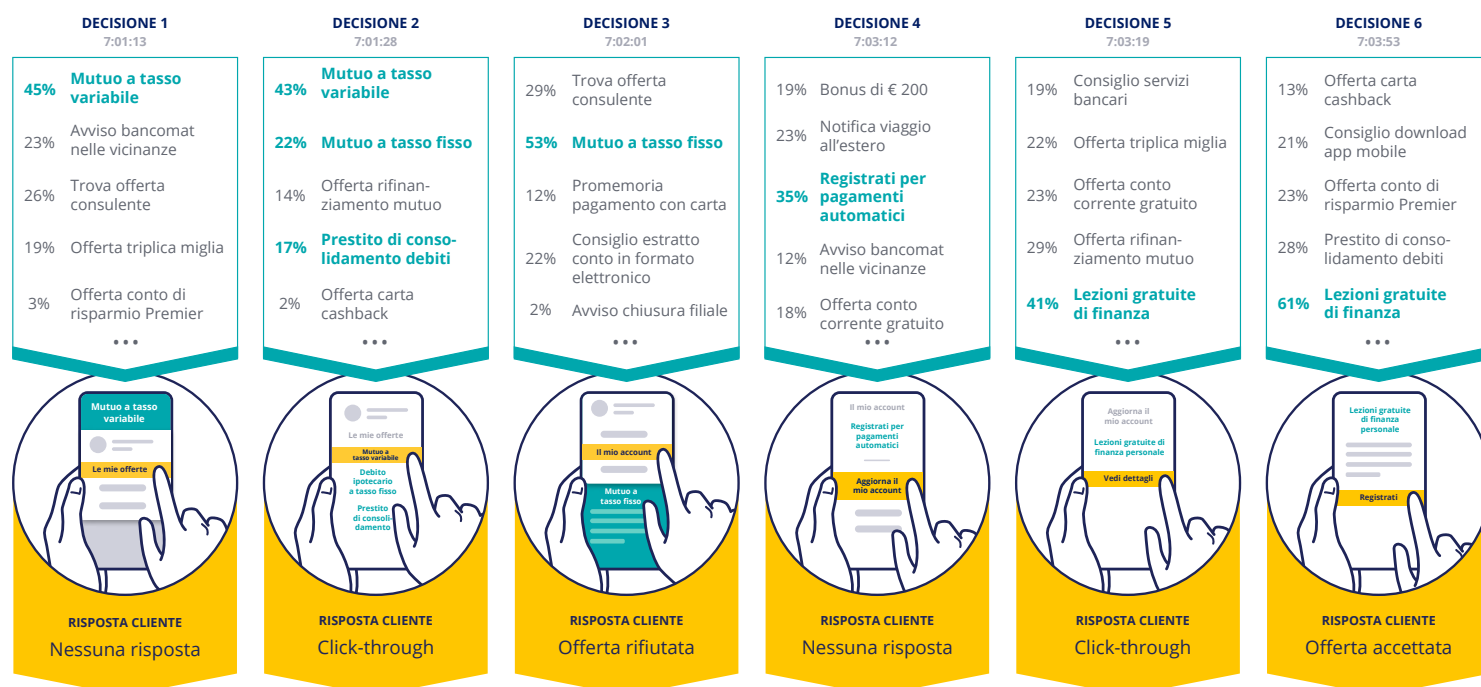
Quanto velocemente? Se non riuscite ad adattarvi al cliente e al processo decisionale, peggio per voi. State sprecando l'opportunità di usare quel momento per creare valore e state mettendo a rischio la customer experience.

Aggregatori e nuovi intermediari si intromettono continuamente tra voi e il vostro (potenziale) cliente. Non siete solo in competizione con il tempo, ma anche con ogni vostro concorrente. E non è solo questione di rispondere per primi, ma rispondere per primi con il messaggio più accattivante. All'interno della finestra di opportunità, lo spazio tra la vostra reattività e il concorrente successivo più veloce è esattamente la vostra opportunità di vantaggio competitivo e il potenziale di "ricompensa in tempo reale".

La "rottura in tempo reale" è il divario tra ottimizzare un'opportunità e il non averla affatto. Le soluzioni più sofisticate di next best action ridefiniranno più volte l'intero processo decisionale in tempo reale durante una singola interazione con il cliente dal vivo. Questo può significare prendere da 10 a 50 decisioni diverse in pochi minuti, tutte durante una singola esperienza del cliente. I punteggi di propensione vengono ricalcolati ogni volta per ogni potenziale azione, tenendo conto dei nuovi dati e del contesto. Uno dei principali operatori telefonici degli Stati Uniti applica questo approccio agli scenari in tempo reale per la fidelizzazione dei clienti: se un cliente chiama per disdire l'abbonamento, si raccomanda di fare domande e risposte per capire meglio il contesto tramite i dati. Le risposte determinano l'offerta di fidelizzazione e ogni calcolo aggiuntivo a supporto. In questi casi, i nuovi dati, che vengono continuamente elaborati, determinano il passo successivo nel corso dell'interazione.

Esempio di ridefinizione della decisione

La figura seguente mostra il percorso di un cliente mentre avanza durante una tipica interazione su dispositivo mobile con un brand che sfrutta le decisioni in tempo reale. Nel corso di soli 2 minuti e 40 secondi, il brand ridefinisce per ben sei volte la decisione in merito a quel profilo del cliente, tenendo conto ogni volta dei nuovi dati relativi ai click-through su dispositivo mobile e altri comportamenti mostrati dal cliente durante l'esperienza. Il risultato finale è che alla persona vengono presentate sei next best action diverse, ma tutte estremamente pertinenti, mentre naviga nell'applicazione mobile e cambia il suo contesto.



Per comprendere veramente la potenza e la complessità delle decisioni in tempo reale in questo esempio, è importante sapere quanto segue:

- Un brand in grado di ridefinire le decisioni per i propri clienti deve avere una **libreria di azioni potenziali** a cui attingere con una combinazione di offerte di vendita, consigli per la fidelizzazione, suggerimenti di assistenza proattivi e comunicazioni per consolidare il coinvolgimento.
- Le aziende hanno spesso **più di 1000 azioni** che possono mettere in gioco per un singolo cliente, in qualsiasi momento della decisione. Ad ognuna di quelle azioni sono inoltre associate più modalità di trattamento creative, con varie combinazioni di immagine, testo, colore, canale ecc.
- A ogni azione e a ogni trattamento a sua volta è associata una serie di **modelli di propensione** il cui punteggio viene ricalcolato con nuovi dati ogni volta che viene presa una decisione. I punteggi di quei modelli vengono usati per mediare tra le azioni al fine di determinare quella più appropriata per il cliente in quel preciso momento, considerando il profilo del cliente completo e il suo contesto attuale "al secondo".
- Ogni click-through, passaggio di mouse, rifiuto di offerta ecc. rappresenta un **cambiamento del contesto del cliente**. Quando avvengono tali eventi (ad esempio quando un cliente seleziona "no grazie" o semplicemente accede a una nuova pagina), quei dati vengono immediatamente attivati per l'uso in tempo reale e associati al modello di dati gestito dall'applicazione. Poi viene ricalcolato il punteggio del profilo per proporre una nuova "next best action".
- Nell'esempio sopra citato, man mano che a un cliente vengono presentate azioni da selezionare/rifiutare/ignorare e il comportamento dei suoi click-through viene aggregato ed elaborato, il brand **cambia costantemente il suo approccio**.
- Il motore decisionale si adatta dal suggerimento di offerte di vendita nelle prime tre interazioni, ai suggerimenti di assistenza, fidelizzazione e azioni di consolidamento del resto dell'esperienza, perché il comportamento della persona ha cambiato i punteggi di propensione e ha mostrato che le azioni più "morbide" sono più **pertinenti rispetto al loro contesto e intento**.
- Tradizionalmente i brand hanno cercato di allinearsi al contesto del cliente utilizzando solo la segmentazione o le regole aziendali. Questi sono i vecchi approcci "uno a molti" che **non riescono ad adattarsi al contesto in tempo reale** o a raggiungere il livello di pertinenza desiderato. Di conseguenza falliscono.
- Senza un approccio basato sulla propensione, che ridefinisce continuamente le decisioni per ogni profilo del cliente in base a nuovi dati, è impossibile raggiungere **un aumento della pertinenza di cinque volte**, che è associata a un investimento in tempo reale.

Le soluzioni più sofisticate di next best action ridefiniranno più volte l'intero processo decisionale in tempo reale durante una singola interazione con il cliente dal vivo.

Conclusione

Non tutto ciò che viene creato in tempo reale è uguale. Selezionare una soluzione per la next best action non in grado di supportare tutti le quattro fasi "in tempo reale" mette a rischio l'esperienza del cliente e i ricavi potenziali. Gli analisti finanziari indipendenti in settori quali telecomunicazioni, servizi finanziari e assicurazioni, hanno scoperto che applicare con successo la next best action in tempo reale porta a un guadagno medio di 225 milioni di dollari in termini di maggiori vendite e valore di fidelizzazione per ogni 10 milioni di clienti, ogni anno.

Ogni mese che non sfruttate la soluzione più adatta per la next best action con funzionalità in tempo reale è probabile che costi alla vostra azienda circa 20 milioni di dollari per ogni 10 milioni di clienti.

È importante cominciare per gradi con la strategia, ma non fate l'errore di cominciare con troppa semplicità con il software. Le basi della next best action in tempo reale devono essere in grado di gestire complessità. Scegliete saggiamente e vi lascerete alle spalle la concorrenza. Se non lo fate, vi troverete presto fuori dai giochi.

È ora di pensare seriamente al vostro futuro in tempo reale.





Pega è un'azienda leader nel software cloud per il customer engagement e l'eccellenza operativa. I brand più noti al mondo si affidano al software di Pega basato su IA per ottimizzare ogni interazione con i clienti a prescindere dal canale e mantenere le promesse del brand. La piattaforma di sviluppo di applicazioni low-code consente alle imprese di implementare e sviluppare le app rapidamente per rispondere alle esigenze di clienti e dipendenti e consentire la trasformazione digitale su scala globale. Da oltre 35 anni Pega assicura una maggiore soddisfazione dei clienti, la riduzione dei costi e l'aumento del valore del customer lifetime value.

Per ulteriori informazioni, visita la pagina www.pegasystems.com/it