

Conoscere il proprio target per impostare azioni commerciali mirate. Un progetto di data mining applicato ai risultati di una campagna promozionale condotta da **Findomestic**, società specializzata nell'erogazione di finanziamenti

La conquista della fedeltà dei propri clienti è uno dei principali obiettivi di qualsiasi impresa e, di conseguenza, l'orientamento alla customer satisfaction risulta essere il comune denominatore di tutte le attività aziendali. L'applicazione delle teorie del marketing nell'ambito delle decisioni gestionali ha favorito lo sviluppo di studi dedicati alla soddisfazione dei clienti. Ogni singolo consumatore possiede una serie di caratteristiche che lo rendono un soggetto unico, dotato di una propria personalità, di un sistema di valori e di un potere decisionale che gli consente di reagire agli stimoli provenienti dall'esterno. Ormai sempre più spesso si assiste a modelli organizzativi aziendali che, incentrati su una logica copernicana, pongono i consumatori al centro del proprio "universo" e rivolgono tutte le funzioni operative all'ottimizzazione dei processi di scambio.

Il progetto di Data Mining Strategico
L'importanza di approfondire la conoscenza del cliente al fine di realizzare efficaci azioni commerciali ha guidato questo studio reso possibile dall'accesso al database di Findomestic Banca S.p.A, società italiana leader nel settore del credito a consumo. Lo studio si è concretizzato in un progetto di data mining mirato a definire un modello statistico capace di analizzare le caratteristiche dei soggetti che hanno risposto positivamente ad una campagna promozionale lanciata da Findomestic. Il fine ultimo era quello di riuscire a valutare come, per un'azienda, fosse possibile mantenere la fedeltà dei propri clienti. La prima fase del progetto, durante la quale è stato possibile attingere

di **Eleonora Brichieri Colombi**,
Università di Firenze – Facoltà di Economia

Fedeli e contenti

alle informazioni disponibili nel customer database dell'impresa, ha portato alla realizzazione di un sistema di "response scoring" che ha permesso di individuare i clienti più ricettivi alle offerte promozionali e di evidenziarne le principali peculiarità. Tramite il sistema di scoring sono stati quindi identificati i segmenti di clientela con minore rischio di rifiuto in occasione di una nuova offerta (nel caso in questione uno specifico prodotto finanziario Findomestic). Il cuore della procedura è stato un modello statistico che ha permesso di collegare i caratteri di una popolazione (i clienti) ad una variabile obiettivo costituita dall'adesione o dalla non adesione di un soggetto a una proposta commerciale. Le variabili esplicative hanno riguardato invece le caratteristiche anagrafiche e demografiche dei soggetti e il loro comportamento nei confronti di passate iniziative commerciali simili a quelle in atto. Per realizzare l'analisi in questione è stato utilizzato il software SAS Enterprise Miner. I risultati ottenuti dovrebbero permettere di orientare un successivo rilancio del prodotto in modo più incisivo verso coloro che possiedono quelle specifiche caratteristiche emerse dal modello come significativamente influenti sulla probabilità di aderire alla promozione.

Lo scoring commerciale

Il compito principale dello scoring è quello di identificare segmenti di clientela che presentano minori rischi di rifiuto nell'ambito dell'offerta di nuovi prodotti. Allo scopo di massimizzare i risultati di azioni commerciali, vengono preliminarmente cercate le variabili discriminanti in grado

di separare nel migliore dei modi il gruppo dei "rispondenti" da quello dei "non rispondenti" a una passata strategia di Direct Marketing. Utilizzando tecniche di analisi statistica multivariata (in grado di sintetizzare i dati, di individuare talune variabili e di circoscrivere specifici segmenti di clientela), il sistema di scoring fornisce informazioni sul profilo del cliente. Tali informazioni devono essere gestite in modo da consentire l'indirizzabilità e la personalizzazione della comunicazione e da permettere la diagnosi degli atteggiamenti del cliente e la previsione dei suoi comportamenti. L'impiego delle informazioni dei futuri comportamenti del cliente sarà lo strumento che permetterà la massimizzazione del ritorno economico delle nuove iniziative promozionali.

L'analisi simulativa dei dati con SAS Enterprise Miner

L'analisi, supportata dal software SAS, si dipana attraverso un percorso a fasi. La prima è la creazione di un archivio dei dati e la definizione di alcuni dettagli riguardanti le variabili. Dopo aver indicato il data set di riferimento, Enterprise Miner assegna automaticamente alle variabili i ruoli (input, target, rejected...) e i livelli di misurazione (binary, nominal, ordinal interval). Prima di iniziare la costruzione dei modelli, i dati possono essere suddivisi in sottoinsiemi attraverso il nodo Data Partition. Esso consente di dividere i dati input in tre distinti data set: Training (per la stima dei modelli), Validation (per la valutazione dei modelli nel nodo Assessment), Test (un data set opzionale che può essere utilizzato per la valutazione finale del modello). Superate le fasi Input Data Source e Data Partition, il passo successivo consiste nello stimare, dall'archivio training, il modello prescelto (nel caso in questione, il modello di regressione logistica). Successivamente il software permette una valutazione del modello, resa possibile dall'inserimento del nodo Assessment. Questo nodo, oltre a produrre una tabella che mostra per ciascun modello le principali statistiche sintetiche per valutarne l'adeguatezza, crea diversi grafici



in grado di mostrare simultaneamente i risultati dei modelli stimati. L'ultima fase è quella che prevede l'inserimento del nodo Reporting. Grazie alla sua visualizzazione è possibile infatti avere un riassunto dettagliato di tutta l'analisi effettuata, direttamente in formato Html.

Il caso Findomestic

Nel caso analizzato si è fatto riferimento a un data set comprendente 103.471 osservazioni relative a tutti i clienti Findomestic che, attraverso la spedizione di una lettera, sono stati stimolati ad aderire ad un determinata iniziativa commerciale connessa a un finanziamento. Sono state considerate 21 variabili, ovvero una variabile risposta binaria Y (dove $Y=1$ stava a significare che il cliente aveva aderito all'iniziativa richiedendo il finanziamento propostogli nei tre mesi successivi alla data di spedizione della lettera e dove $Y=0$ stava a significare una non adesione) e 20 covariate relative alle possibili motivazioni per cui il cliente

aveva deciso di richiedere il finanziamento. In particolare si sono considerate variabili relative alla situazione socio-anagrafica, finanziaria, personale accanto a variabili relative al prestito (impegno mensile di pagamento ed esposizione totale al credito). Per semplificare l'analisi si è deciso di modificare il data set originario in modo da avere per lo più variabili dicotomiche. La dicotomizzazione effettuata ha permesso così di condurre un interessante studio sugli odds ratio che ha fornito chiare indicazioni circa l'identità delle variabili effettivamente in grado di spiegare il comportamento del cliente. La prima fase dell'analisi, di tipo esplorativo-descrittivo, è stata finalizzata a sintetizzare l'insieme delle informazioni utilizzate. Dopodiché si è passati all'approccio multivariato attraverso il quale sono stati combinati insieme tutti i "segnali" che arrivavano dalle diverse variabili esplicative per poi ottenere un "segnale complessivo" che individuasse in misura sintetica il grado e le caratteristiche

di risposta alla proposta commerciale effettuata. Per interpretare il segno e la forza dei legami tra il finanziamento richiesto in seguito alla lettera e le variabili esplicative ad esso associate, si è fatto ricorso ad un modello di regressione logistica. Il modello finale è stato costruito dopo aver ponderato l'importanza delle variabili sia tramite metodologie statistiche sia in base a considerazioni di carattere logico-commerciale.

Risultati ottenuti

Per interpretare correttamente il modello si è fatto riferimento alle tre curve di assestment: la Roc Curve, la Lift Chart e la Threshold Chart. L'analisi ha dimostrato che per riuscire a prevedere correttamente i clienti fedeli è necessario scegliere un valore di soglia del 20%. Ciò significa che tutti i clienti che mostrano una probabilità stimata di richiedere un finanziamento in seguito alla lettera promozionale superiore al 20% devono essere considerati "fedeli", viceversa tutti coloro con probabilità stimata inferiore devono essere considerati "occasionalisti". Lo studio ha evidenziato che solo lo 0,6% dei 103.471 clienti ai quali era stata inoltrata la lettera ha risposto richiedendo un finanziamento entro tre mesi: anche se limitata, tale percentuale è da ritenersi economicamente significativa. È stato anche possibile tracciare le caratteristiche che accomunano coloro che hanno aderito alla campagna proposta da Findomestic: solitamente rientrano in una delle due fasce di età 18-30 e 50-70 anni, sono genitori e hanno un basso impegno di pagamento mensile nei confronti dell'impresa e una bassa esposizione totale al credito. Tutto ciò può rivelarsi estremamente utile: una successiva campagna similare potrebbe essere indirizzata alle sole categorie di clienti più sensibili. Naturalmente, questa è solo una delle possibili analisi effettuabili sul data set clienti, sia perché i raggruppamenti ideati all'inizio dell'analisi potrebbero essere realizzati in più modi nel tempo e secondo discrezionalità diverse, sia in quanto l'obiettivo di future campagne potrebbe essere diverso da quello considerato. ■