

RAFFAELLA BETTELLA, ALBERTO SACCARDI
Nunatac

RENATO MIORELLI, PAOLO PESSI
Periodici San Paolo

MODELLI DI CHURN IN AMBITO EDITORIALE: IL CASO PERIODICI SAN PAOLO

Obiettivo del progetto

L'obiettivo del progetto è analizzare il fenomeno dell'abbandono degli **abbonati a Famiglia Cristiana**, settimanale edito da **Periodici San Paolo**.

I Clienti che hanno la maggiore probabilità di non rinnovare vengono individuati attraverso una regola che stima la **propensione all'abbandono**, che rappresenta una sintesi dei sintomi di insoddisfazione che ne anticipano la decisione.

Tale valore è determinato da un **modello statistico** che mette in relazione la scelta di rinnovare o meno un abbonamento con le variabili note – la storia del Cliente – osservate in un periodo precedente la decisione.

Le **fasi** attraverso le quali si articola l'analisi sono **quattro**:

1. Preparazione della base dati
2. Analisi dei contenuti
3. Modelling
4. Passaggio in produzione.

FASE 1. Preparazione della base dati

I **dati utilizzati** provengono da fonti interne:

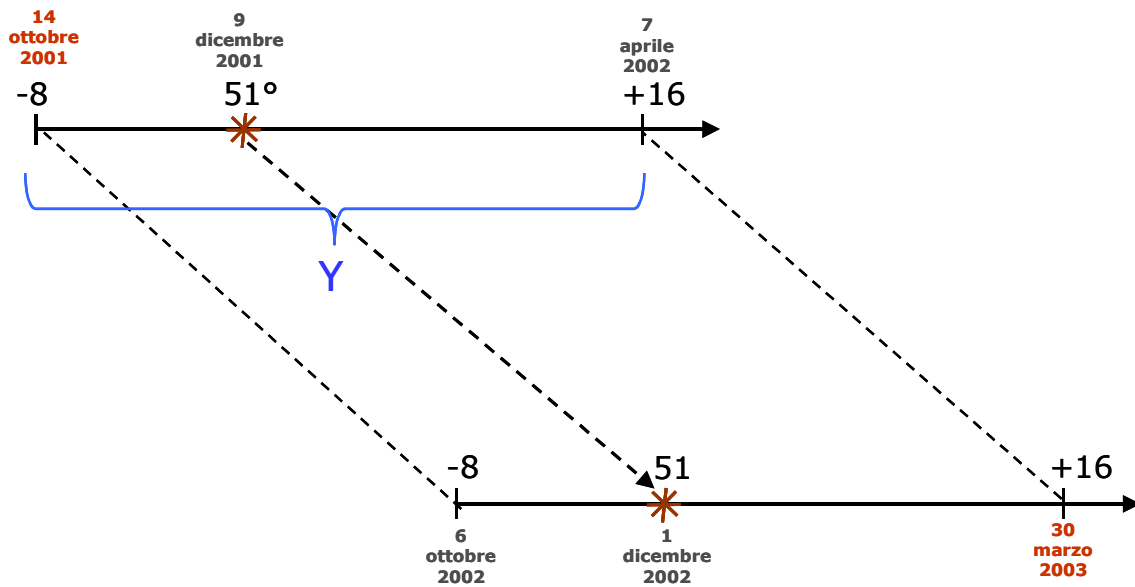
- Socio-demografici: sesso, dislocazione geografica, etc.;
- Abbonamento: anzianità dell'abbonamento, beneficiario/donatore, importi pagati, velocità di pagamento, etc.;
- Vendite per corrispondenza (VPC): prenotazioni, resi, pagamenti e importi relativi a vendite one shot;
- Inbound/Outbound: chiamate ricevute (disservizio, sofferenza, varie) e contatti effettuati (wrapper, print, mail, telemarketing);
- Pagamenti: velocità di pagamento per altri abbonamenti e VPC.

Il primo passo è stata la costruzione di un Data Mart integrato che collega le varietà delle informazioni disponibili e le storizza nel tempo, a livello di singolo cliente.

Successivamente, dall'universo degli abbonamenti a Famiglia Cristiana è stato selezionato il sottoinsieme oggetto di studio, costituito dai Privati, residenti in Italia, spontanei nell'anno corrente e non sospesi nell'anno di riferimento. Il periodo di ricevimento della prima copia dell'abbonamento deve essere compreso tra il 17 Dicembre 2000 e il 24 Novembre 2001.

L'universo di riferimento è di circa 270.000 abbonati.

Essendo il periodo di potenziale rinnovo "continuo" (ogni settimana si assiste all'apertura di nuovi abbonamenti, al rinnovo di abbonamenti in essere, alla disdetta esplicita o tacita) è stato necessario procedere ad un **riallineamento** delle diverse unità rispetto ad un comune evento: **la scadenza dell'abbonamento in corso**.



L'**orizzonte previsivo** del modello è fissato in **24 settimane** a partire dal momento in cui è possibile effettuare il rinnovo dell'abbonamento, ovvero inizia con lo stato -8 dell'abbonamento e si conclude con lo stato +16.

Il modello sarà in grado di prevedere al tempo t_0 (momento della previsione), sulla base delle informazioni comportamentali disponibili ("Storia" disponibile fino all'istante t_1), gli **abbandoni che si verificheranno nelle 24 settimane successive**.

FASE 2. Analisi dei contenuti

La tabulazione e l'analisi di tutte le variabili a disposizione è un passo importante e indispensabile per individuare le caratteristiche dei clienti e per eliminare variabili non attendibili o poco informative.

FASE 3. Modelling

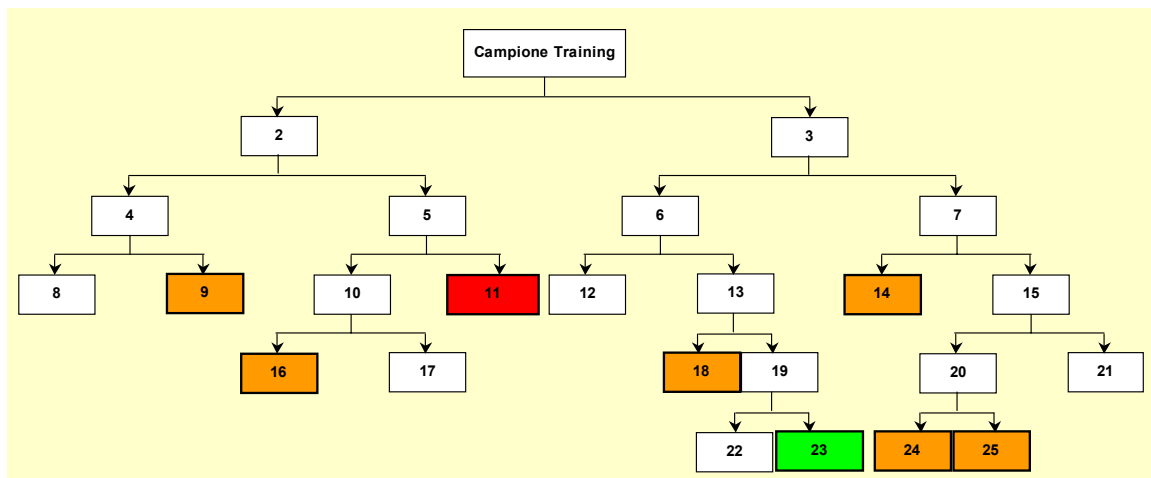
Sulla base dei risultati dello studio di cui al punto 2 si è deciso di concentrare l'analisi sui **clienti con almeno tre anni di anzianità** di abbonamento, che costituiscono il 93% del parco clienti e sono caratterizzati da un tasso di abbandono inferiore alla media: ponendo a 100 la media, la loro probabilità è pari a 83.

La **tecnica statistica utilizzata** per prevedere e classificare la propensione all'abbandono dei Clienti abbonati a Famiglia Cristiana è nota come “**Albero decisionale**”.

Tale metodo seleziona le variabili più connesse con il fenomeno dell'abbandono (target) e sulla base delle modalità assunte da queste variabili “spacca” la popolazione in sottoinsiemi il più possibile diversi tra loro e il più possibile omogenei al loro interno, cui assegna uno stesso tasso di abbandono.

Come risultato finale si ottiene una struttura, ad albero appunto, le cui **foglie finali** rappresentano insiemi di Clienti sottoposti allo stesso rischio di abbandono.

I **rami** invece rappresentano le condizioni che hanno determinato una suddivisione delle unità in più gruppi.



La **regola** di assegnazione dei Clienti alle foglie è lo strumento che ci permette di attribuire all'intero parco Clienti una stima della propensione all'abbandono.

Il progetto è stato realizzato utilizzando **SAS Enterprise Miner™**, Versione 4.1.

Gli step di realizzazione e verifica del processo statistico sono:

1. selezione di un campione di popolazione, di cui sia nota la variabile esplicativa (target);
2. suddivisione di tale campione in un gruppo di training (70%) e in uno di validation (30%);
3. stima della regola sul campione di training con variabile esplicativa nota;
4. validazione della bontà del modello utilizzando il campione di validation;
5. applicazione della regola alla popolazione in un periodo successivo T+1, con variabile target nota.

E' stato estratto un campione sbilanciato, così da analizzare tutti i casi positivi (abbandono); i risultati sono stati riespansi in modo da rispettare la distribuzione reale della popolazione.

Si sono studiate le relazioni tra le variabili esplicative e con la variabile target e, in funzione dei risultati, è stata fatta la prima selezione delle variabili da dare in pasto al modello.

Questo passaggio ha consentito anche la costruzione di nuove variabili con maggior livello informativo, ottenute aggregando modalità e/o variabili elementari.

Una selezione più raffinata delle variabili è stata ottenuta durante la fase di costruzione dell'albero ricorrendo all'utilizzo del test Chi-square.

In definitiva, le **variabili selezionate** sono:

1. velocità di pagamento dell'abbonamento,
2. acquisti VPC,
3. periodo di scadenza dell'abbonamento,
4. anzianità abbonamento,
5. regolarità dell'abbonamento (51 numeri),
6. adesioni ad iniziative proposte tramite mail,
7. pagamenti di opere one shot,
8. importo totale pagato dal cliente.

La combinazione delle modalità di queste variabili determina l'appartenenza di un cliente ad una foglia (regola).

Tale regola consente di "riordinare" la popolazione sulla base della propensione all'abbandono, che è costante per tutti i clienti che appartengono ad una stessa foglia.

Prima di accettare i risultati del modello si è proceduto alla sua **validazione**, attuata dapprima utilizzando il canonico campione di validation a parità di tempo T, che ha portato a lift chart pressoché sovrapposte.

Successivamente, essendo entrati in possesso di informazioni relative all'abbandono in un periodo successivo, si è verificata la bontà delle stime della popolazione al tempo T+1. L'esito positivo di queste prove ha consentito di passare allo step successivo.

Il focus si è spostato a questo punto su due sottoinsiemi della popolazione d'analisi:

- **CLIENTI A RISCHIO**: il 10% circa della popolazione a più alto rischio di abbandono, e
- **CLIENTI FEDELI**: il 15% circa a più basso rischio di mancato rinnovo.

L'analisi dei comportamenti delle variabili nei diversi gruppi ha permesso di cogliere le peculiarità dei clienti in funzione della loro propensione all'abbandono.

I **clienti a rischio** (propensione all'abbandono = 236 rispetto alla media 100) hanno un'anzianità di abbonamento inferiore alla media, regolarizzano i pagamenti in ritardo e più frequentemente sono stati beneficiari dell'abbonamento almeno un anno.

Ricorrono poco all'acquisto per corrispondenza (per cui i loro investimenti totali sono bassi) e anche in questo caso il pagamento è ritardato rispetto alla media.

I **clienti fedeli** (propensione all'abbandono = 39 rispetto alla media 100) risultano avere l'abbonamento da molti anni, con regolarità di copie ricevute annualmente. La scadenza del loro abbonamento, che pagano senza ritardi, non coincide col periodo estivo (periodo di debolezza per i rinnovi).

Fanno acquisti one shot con frequenza decisamente al di sopra della media.

FASE 4. Passaggio in produzione

Il progetto si è concluso con la messa in produzione dei risultati acquisiti, fase attraverso la quale i risultati sono stati integrati in un **processo** in grado di generare indicatori aggiornabili nel tempo.

Il valore aggiunto del modello sviluppato sta proprio nella possibilità di **monitorare costantemente nel tempo l'intera popolazione degli abbonati**, dalla quale è possibile estrarre liste selezionate di clienti.

Possiamo dividere il processo informatico da quello analitico: il primo prevede la disponibilità di una Customer Table sull'intera popolazione degli abbonati a Famiglia Cristiana, contenente dati comportamentali elementari aggiornabili **settimanalmente**; il secondo è in grado di fornire ogni settimana una lista nominativa dei clienti a più alto (o a più basso) rischio di abbandono, cui sottoporre iniziative mirate di marketing.

Approfondimenti

Come abbiamo visto, l'Albero decisionale fornisce in output gruppi di abbonati caratterizzati da una uguale propensione all'abbandono (appartenenza alla stessa foglia).

Potrebbe però essere necessario, o semplicemente utile, **graduare tale rischio** anche all'interno di una stessa foglia, ad esempio per poter decidere la numerosità esatta di clienti da sottoporre ad una iniziativa.

Per ottenere questo risultato si sono integrati i risultati già ottenuti con quelli derivanti dall'applicazione di un modello di regressione logistica, che ordina i clienti rispetto alla probabilità di abbandono assegnata dal modello.

Questa applicazione, oltre ad affinare il processo descritto, ha fornito una conferma dei risultati.