

Alberi Decisionali

**Per l'analisi del mancato rinnovo
all'abbonamento di una rivista**

Metodi Statistici per le Aziende 1

Il problema

L'anticipazione del fenomeno degli abbandoni da parte dei propri clienti, rappresenta un elemento fondamentale nel complesso delle azioni volte a stabilire un rapporto privilegiato e duraturo nel tempo con la clientela medesima.

Definizione

- n **“Abbandono”** è l’evento connesso alla decisione assunta da un cliente di non rinnovare l’abbonamento a una rivista.
- n Se è possibile che tale evento si verifichi all’interno di un intervallo di tempo compreso in quello che da ora in poi chiameremo “orizzonte previsivo”, esso costituirà oggetto della previsione in esame.

Definizione

L'evento abbandono può verificarsi secondo due modalità differenti:

- n *Disdetta esplicita* da parte dell'abbonato: può essere effettuata in corrispondenza di qualsiasi numero della rivista all'interno del suo ciclo di vita (n°1 – n°51)
- n *Mancato rinnovo* (tacito)

I diversi tipi di “abbandono”

La rescissione dei rapporti di abbonamento ad un periodico può essere dovuta a motivi molto diversi tra loro. A tal fine è utile distinguere due diversi tipi di abbandono:

- n **Abbandono inevitabile:** trasferimento di residenza all'estero, decesso, ..
- n **Abbandono volontario:** accade quando un cliente volontariamente interrompe il rapporto con l'azienda.

L'obiettivo primario di un modello di previsione è la previsione dell'abbandono volontario.

Obiettivo

- L'individuazione dei clienti a rischio di mancato rinnovo avviene tramite **l'assegnazione di una probabilità abbandono**, che rappresenta **una sintesi dei sintomi di insoddisfazione** che anticipano la decisione.
- Questo valore è **determinato da un modello statistico** che mette in relazione la scelta di rinnovare o meno un abbonamento con le variabili note, osservate in un periodo precedente la decisione.
- La possibilità di **selezionare una lista di clienti “a più alto rischio di abbandono”** consente di proporre iniziative di marketing mirate, massimizzandone l'effetto.

Metodo

- La tecnica statistica utilizzata per prevedere e classificare la probabilità di abbandono degli abbonati è:



L'Albero Decisionale

Cos'è un Albero Decisionale

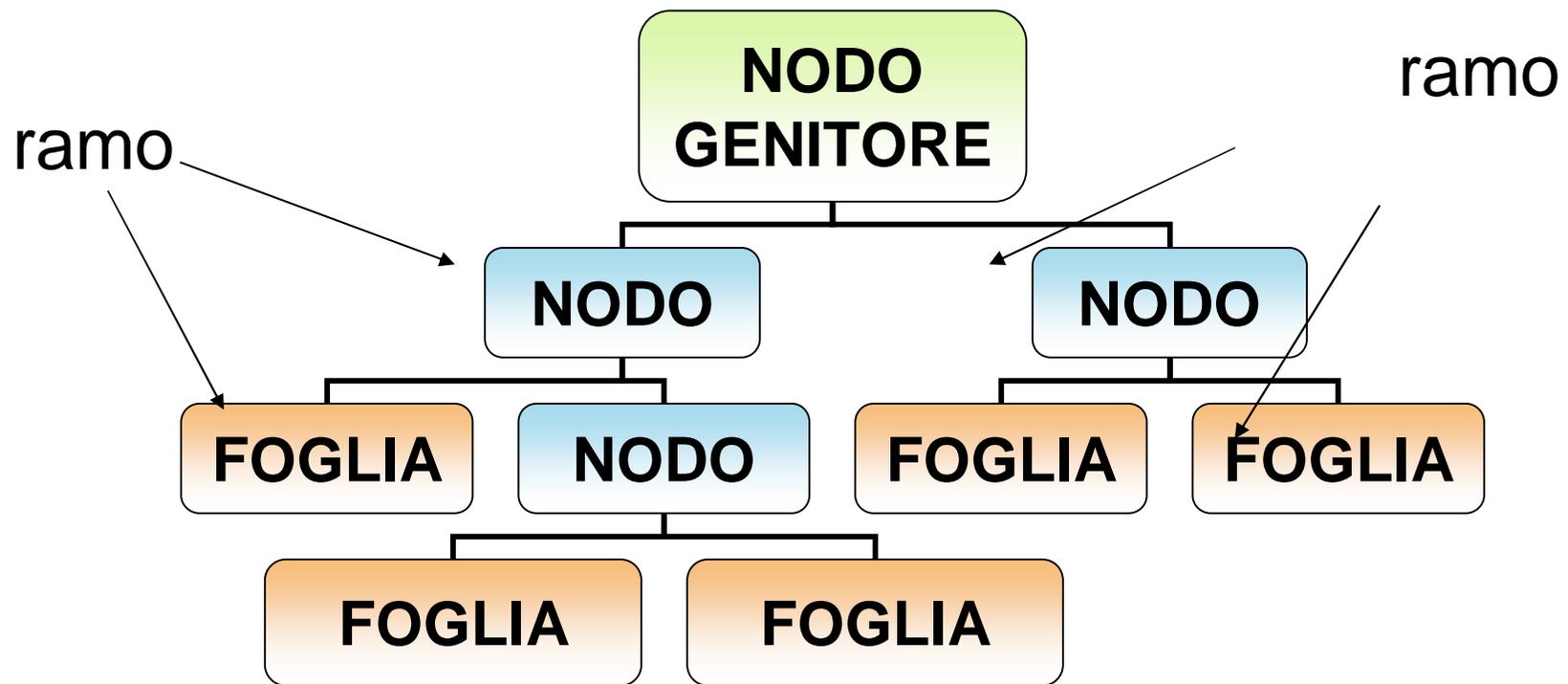
- n Gli alberi decisionali sono tecniche non parametriche che analizzano le **dipendenze di una variabile target Y da un insieme di variabili esplicative X** , adottando un **processo di partizione di tipo gerarchico**.
- n Ogni gruppo di unità viene cioè ricorsivamente diviso in sottogruppi per selezionare, **tra le variabili esplicative, quelle che maggiormente spiegano la variabilità della variabile target**.

Ripartizione delle unità

- n Le unità statistiche vengono quindi suddivise, secondo una regola divisiva, in gruppi sempre più omogenei al loro interno (**varianza minima internamente al gruppo**) e disomogenei al loro esterno (**varianza massima fra i gruppi**) rispetto alla variabile target.

Costruzione del grafo (albero)

- una segmentazione binaria
- una segmentazione ternaria
- una segmentazione multipla o a K vie



Costruzione del grafo (albero)



Costruire un grafico detto albero di decisione o dendrogramma, dove:

- n **i nodi** rappresentano i gruppi ai diversi livelli di segmentazione (i **nodi genitori** sono quelli la cui partizione determina la nascita di due o più sottogruppi detti figli)
- n **i rami** indicano le condizioni che hanno determinato le suddivisioni
- n **le foglie** si riferiscono ai nodi finali, per i quali non è ritenuta utile un'ulteriore segmentazione

Metodo

Il metodo degli Alberi Decisionali, **particolarmente adatto nel trattamento di dati qualitativi**, permette di:

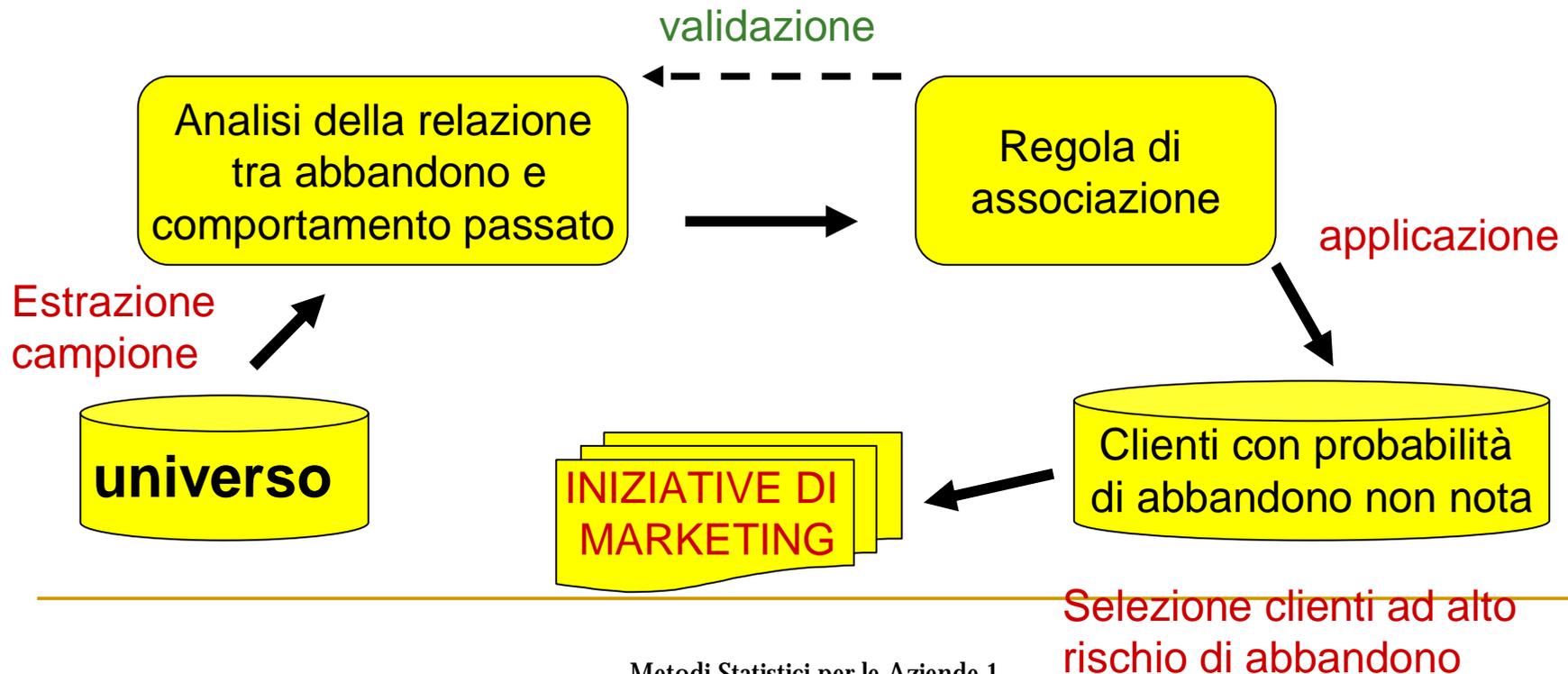
- n selezionare le variabili più correlate con il fenomeno dell'abbandono, considerando anche l'effetto congiunto (analisi multivariata);
- n sulla base delle modalità assunte da queste variabili "spaccare" la popolazione in sottoinsiemi il più possibile diversi tra loro e il più possibile omogenei al loro interno a cui assegnare uno stesso tasso di abbandono.

Risultato atteso

Creazione di una struttura ad albero le cui foglie finali rappresentano insiemi di clienti sottoposti allo stesso rischio di abbandono.

La regola di assegnazione

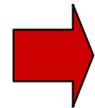
La regola di assegnazione dei clienti alle foglie è lo strumento che ci permette di attribuire all'intero parco clienti una stima della probabilità di abbandono



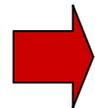
Come costruire l'universo

I dati a disposizione provengono da fonti diverse:

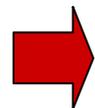
- q Fonti interne
- q Fonti esterne



Analisi dell'anzianità di abbonamento



Analisi della distribuzione geografica dei clienti



Analisi della distribuzione per sesso

Come “funziona” un albero decisionale

- n Quando un modello ad albero viene applicato ai dati, **ciascun record scorre lungo la struttura dell'albero, lungo un percorso determinato da una serie di test**, fino a giungere ad un nodo terminale.
- n E' auspicabile ottenere un **albero “bilanciato”**, dove cioè la suddivisione dei nodi procede in parallelo.

Obiettivo dell'analisi

Ricavare la regola di classificazione delle unità statistiche, al fine di eseguire previsioni sul valore assunto dalla variabile dipendente:

- Si crea, sulla base del campione osservato, una regola di predizione del valore della variabile target di un nuovo soggetto per il quale siano state osservate le variabili esplicative.
- Ciascuno dei nodi terminali indica una chiara *regola decisionale* di allocazione delle unità statistiche, che si legge collegando il nodo iniziale a ciascuno di essi.

Variabile target

Target	Descrizione	Frequenza	%
99	Abbonati che pagano in anticipo	9.000	9%
0	Clienti che rinnovano l'abbonamento	83.000	83%
1	Clienti che non rinnovano l'abbonamento	8.000	8%

100.000

Metodi Statistici per le Aziende 1

Selezione del campione

Target	Freq.	%	Freq. Cum.
99	910	5,7%	910
0	7.090	44,3%	8.000
1	8.000	50,0%	16.000

La scelta del campione:

- n tutti gli abbandoni (target1)
- n un campione di non abbandoni (target0)
- n un campione di coloro che pagano in anticipo (target99)

Modelling

Per la costruzione del modello sono stati costruiti due sottocampioni:

- n **Training:** estratto casualmente, ampio il 70% del campione ed utilizzato in sede di **stima** del modello;
- n **Validation:** sottocampione ampio il 30% del campione composto dalle rimanenti unità statistiche e utilizzato per verificare la **robustezza** del modello.

Modelling

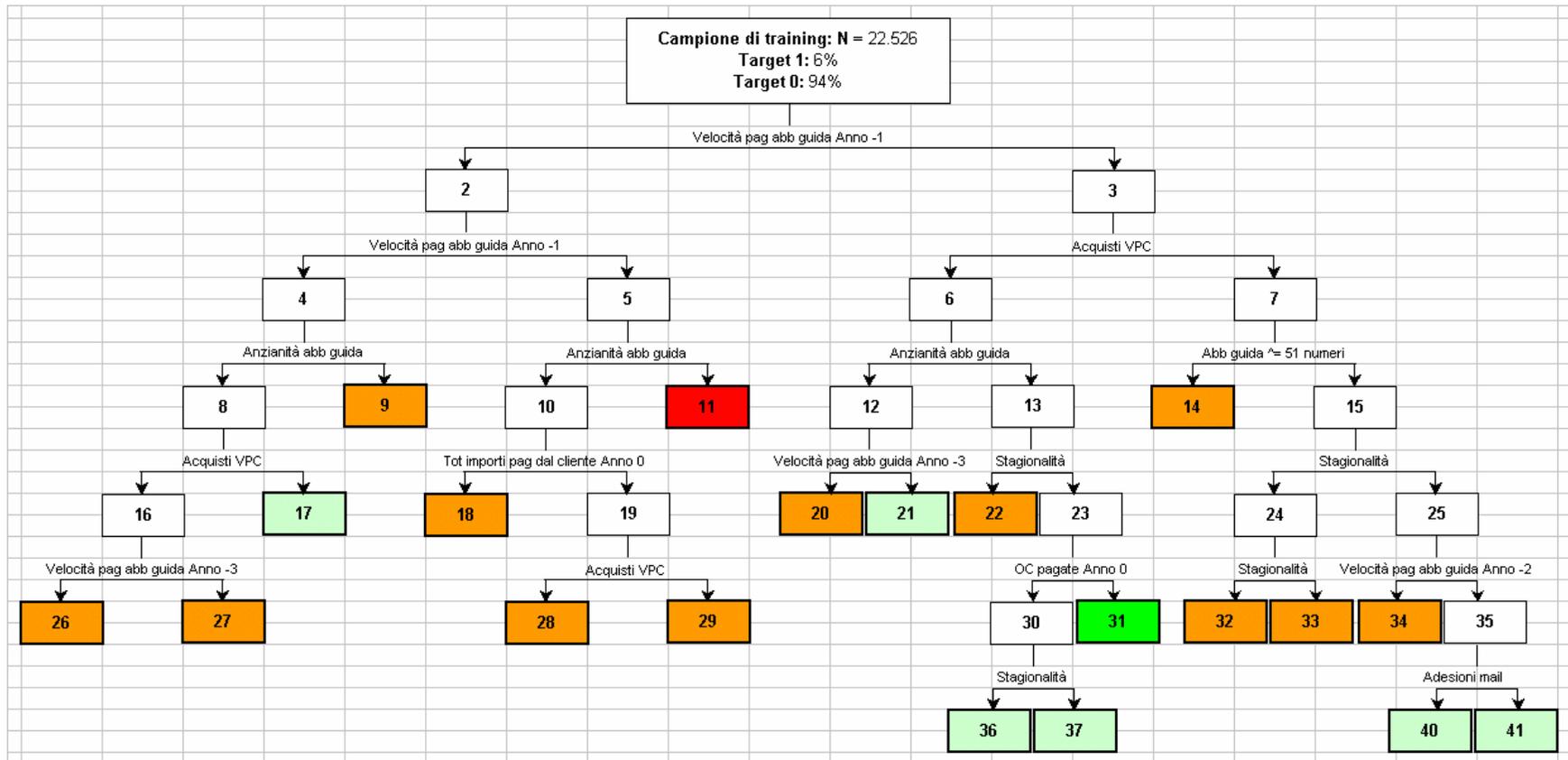
- n La **tecnica di validazione** consiste nell'applicare la regola costruita a dati non impiegati nella fase di stima del modello, per valutare a posteriori le probabilità stimate dal modello.
- n La valutazione sui dati di training non sarebbe sufficiente, poiché il modello potrebbe essere ben stimato su tali dati, ma non robusto su dati nuovi (*over fitting*).

Costruzione dell'albero

Le variabili discriminatorie utilizzate per costruire l'albero derivano dalle fonti a disposizione e valutate negli anni in esame:

- n Variabile target
- n Tipo di abbandono
- n Velocità di pagamento
- n Anzianità di abbonamento
- n Modalità di pagamento
- n Mancati pagamenti
- n Acquisti
- n Totale importi pagati dal cliente
- n Abbonamenti ad altre riviste
- n Sesso
- n Ecc.

Interpretazione dei risultati



Applicazione ed Interpretazione

- n Il colore della foglia è data dall'incidenza del target nella foglia stessa. La foglia di colore rosso avrà un'incidenza massima di $\text{target}=1$ (21,8%), mentre le foglie di colore verde “brillante” sono quelle con minore presenza (3% circa).