



La fidelizzazione dei Clienti attraverso la Customer Retention

Alfredo Roccato
alfredo.roccato@fastwebnet.it

1° KUDI – KNIME User Day Italia
Milano, 9 ottobre 2012

Agenda

- **Proporre una metodologia per la Customer Retention basata su tre passi:**
 - ✓ Predire i clienti importanti che sono più a **rischio d'abbandono** (Risk of Leaving)
 - ✓ Definire un **indice di valore** per ogni cliente (Customer Value Index)
 - ✓ Costruire una **Matrice Rischio/Valore** in base alle classificazioni di questi indici per segmentare i clienti e decidere chi mantenere come **priorità**

Perché la Customer Retention

L'obiettivo della Customer Retention è quello di evitare di riacquisire clienti

E' dimostrato che

- acquisire nuovi clienti, può costare 5 volte di più che soddisfare e mantenere i clienti attuali¹
- un aumento del 5% nella fidelizzazione dei clienti può aumentare i profitti per cliente dal 25% al 100%²

¹ C. Hart, J. Heskett e W. Sasser (1990), *The Profitable Art of Service Recovery*

² F. Reichheld (1996-2001), *The Loyalty Effect*

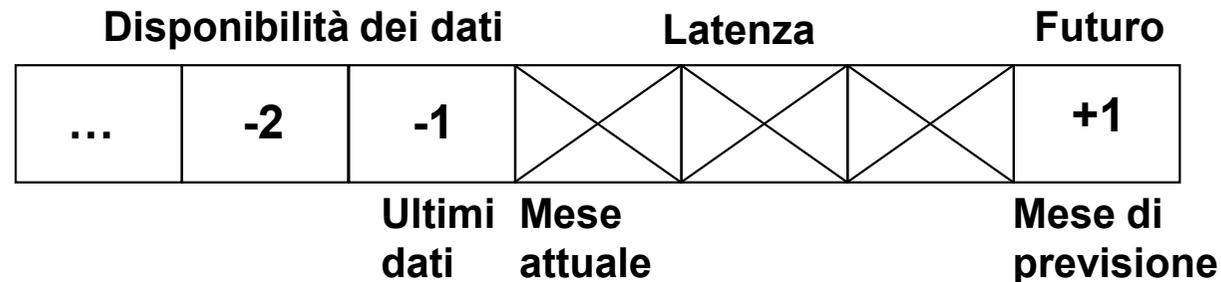
La metodologia proposta

- **Strategia**

- ✓ Si utilizzano i dati storici per costruire un modello predittivo
- ✓ Le regole fornite dal modello verranno applicate ai dati attuali per predire la probabilità di abbandono di ogni cliente in base al periodo di latenza prescelto
- ✓ I risultati verranno incrociati con valutazioni di profittabilità dei clienti per definire le priorità in base alle quali saranno attivati i contatti delle iniziative commerciali

La struttura

- **La finestra temporale**



+1 indica il mese nel quale si verificherà l'abbandono (il futuro)

X indica i mesi di "latenza"¹ (il presente)

-1 indica il mese degli ultimi dati disponibili (il passato recente)

¹ Il periodo di latenza è il tempo che intercorre tra il mese relativo agli ultimi dati di riferimento e il mese previsto di abbandono; questo periodo tiene conto sia del tempo necessario per la disponibilità dei nuovi dati e il tempo necessario per contattare il cliente

I dati

- **I dati storici**

18 mesi (ultimo mese: settembre 2012)

- **Clienti selezionati (~54.000)**

- ✓ Segmento "UPPER MASS" (50k-100k €)
- ✓ Anzianità di rapporto (TENURE) > 16 mesi
- ✓ Senza informazioni negative
- ✓ Non dipendenti

- **Variabili di calcolo relative a settembre 2012 (“month- 1”)**

- Anagrafiche**

- ✓ Age
 - ✓ Tenure
 - ✓ ...

- Comportamentali**

- ✓ Transactions and Amounts for Withdrawal and Purchases (Credit/Debit cards)
 - ✓ Average Balances
 - ✓ Transaction and Amounts for Regular Income
 - ✓ Transactions and Amounts for Standing Order
 - ✓ Transactions and Amounts for Direct Debits
 - ✓ Transactions and Amounts for Outgoing and Incoming Bank Transfers
 - ✓ Transactions and Amounts for Withdrawals and Deposits
 - ✓ Transactions and Amounts for Withdrawals and Deposits at the Branch Cash
 - ✓ ...

I dati

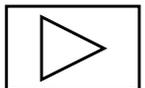
- **Variabili di calcolo**

Per ciascuna variabile comportamentale sono state calcolate le **differenze percentuali** tra l'ultimo mese (-1) e i rispettivi 3, 6, 9 mesi prima (-4, -7, -10):

$$\text{Nome_Variabile_P3} = \frac{(\text{Nome_Variabile_1} - \text{Nome_Variabile_4})}{\text{ass}(\text{Nome_Variabile_4})}$$

$$\text{Nome_Variabile_P6} = \frac{(\text{Nome_Variabile_1} - \text{Nome_Variabile_7})}{\text{ass}(\text{Nome_Variabile_7})}$$

$$\text{Nome_Variabile_P9} = \frac{(\text{Nome_Variabile_1} - \text{Nome_Variabile_10})}{\text{ass}(\text{Nome_Variabile_10})}$$



I dati

- **Customer Value Index**

Non essendo disponibile una misura di profittabilità, è stato calcolato un indice ponderato basato sui valori assoluti dei Saldi, Volumi e prodotti detenuti (Cross Selling)

$$\Sigma_i \text{ SaldoMedio} = \frac{\Sigma_m \text{ ass (Saldi}_m * \text{giorni}_m)}{\text{Media del Segmento}} \quad \text{dove } m=-1,-2,\dots,-12 \text{ e } i=C,D,P$$

$$\Sigma_i \text{ Volumi} = \frac{\Sigma_m (\text{Tot. Importi Movimentati}_m)}{\text{Media del Segmento}} \quad \text{dove } m=-1,-2,\dots,-12 \text{ e } i=C,D,P$$

$$\text{Cross Selling} = \text{CONTI} * \text{peso} + \text{DEPOSITI} * \text{peso} + \text{PRESTITI} * \text{peso}$$

$$\text{Customer Value Index} = \text{SaldoMedio} * \text{peso} + \text{Volumi} * \text{peso} + \text{Cross Selling} * \text{peso}$$

Prestazioni del modello

- I risultati sono basati sul Validation set (40%)

Matrice di confusione

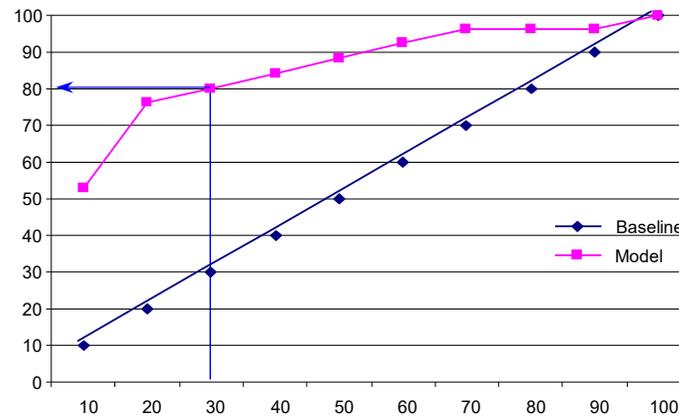
	Predicted Loyal	Predicted Lost
Truly Loyal	21.077	58
Truly Lost	106	331

N. of observations
on the Validation
set:

21.572

Correct Classified 21.408 Accuracy 99,2%
Wrong Classified 164 Error 0,8%

Risposta catturata



I risultati finali

- **La Matrice Rischio/Valore**

- ✓ Il modello finale è stato applicato a uno Score set per classificare i clienti in base alla Matrice Rischio/Valore e definire la lista delle priorità

		<i>PROBABILITY OF LEAVING</i>					
		LOW	MEDIUM	HIGH			
<i>CUST. VALUE</i>	HIGH	3.673	1.741	3.620	Segment A	3.620	24%
	MEDIUM	741	731	1.738	Segment B	4.210	28%
	LOW	383	286	1.952	Segment C	7.035	47%
					<hr/>	<hr/>	
						14.865	100%

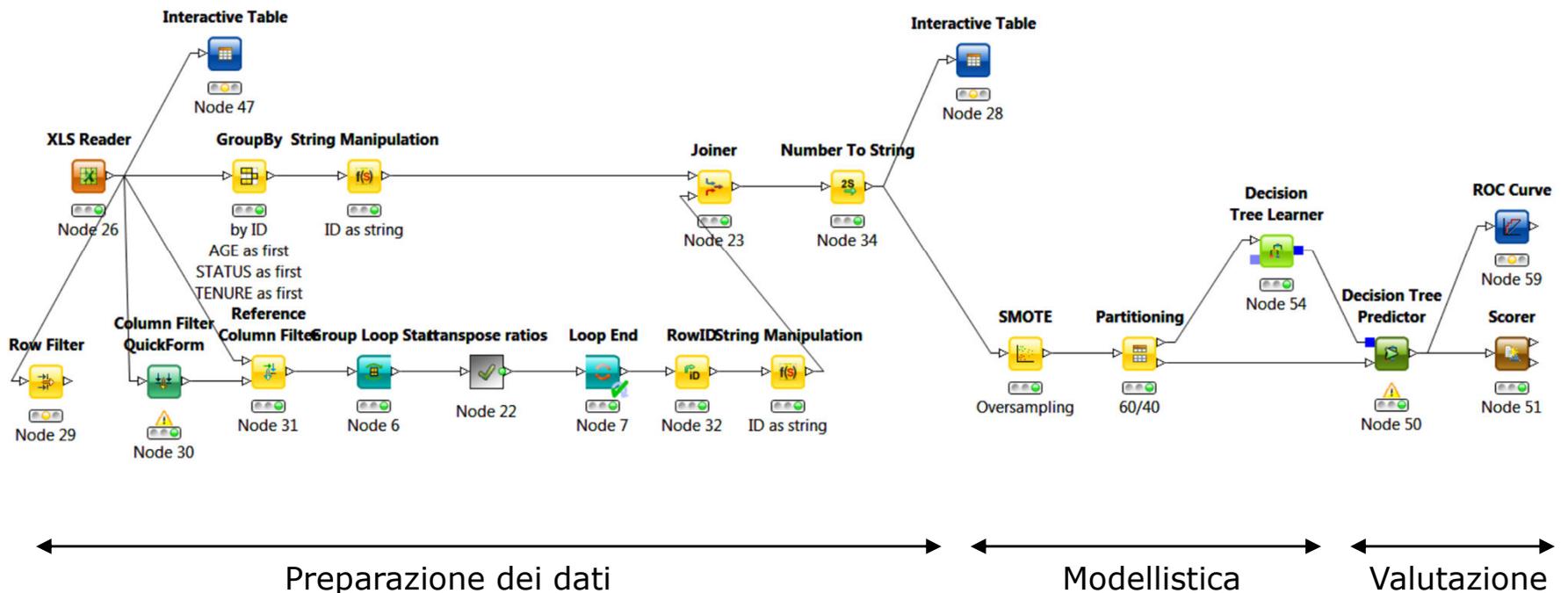
Software

- Perché la scelta 

- ✓ Preparazione dei dati
- ✓ Suite completa di metodi per la data analysis, il data mining e la visualizzazione grafica
- ✓ Facilità d'uso
- ✓ Open-Source → *Il software è distribuito gratuitamente*

Il workflow

- Preparazione dei dati, Modellistica e Valutazione dei risultati



Il workflow

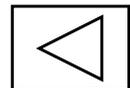
- **Preparazione dei dati**

Il passo principale è declinare ogni variabile in 3 nuove variabili identificate dal valore del campo TIME_WINDOW per definire le differenze percentuali **P3**, **P6** e **P9**

The image shows two screenshots of a data table interface. The top screenshot shows a table with columns: Row ID, TIME_WINDOW, ID, STATUS, AGE, TENURE, D AMT_CHEQUES, D AMT_..., and D AMT. The bottom screenshot shows the same data transformed into three new columns: D AMT_CHEQUES_P3, D AMT_CHEQUES_P6, and D AMT_CHEQUES_P9, representing percentage differences for time windows of 3, 6, and 9 units respectively. Arrows indicate the mapping from the original 'D AMT_CHEQUES' column to the new columns.

Row ID	TIME_WINDOW	ID	STATUS	AGE	TENURE	D AMT_CHEQUES	D AMT_...	D AMT
Row45	+1	75480...	0	88	320	4,550	16,991.2	415.71
Row46	-1	75480...	0	87	311	2,840	14,592.67	1,675.45
Row47	-4	75480...	0	87	308	5,290	14,490.1	1,589.95
Row48	-7	75480...	0	88	305	5,000	14,421.72	1,763.95
Row49	-10	75480...	0	86	302	6,201.81	10,816.29	2,463.85
Row50	+1	79558...	0	73	313	3,476.84	57,408.31	62,256.1
Row51	-1	79558...	0	72	304	7,219.03	141,039.26	106,540

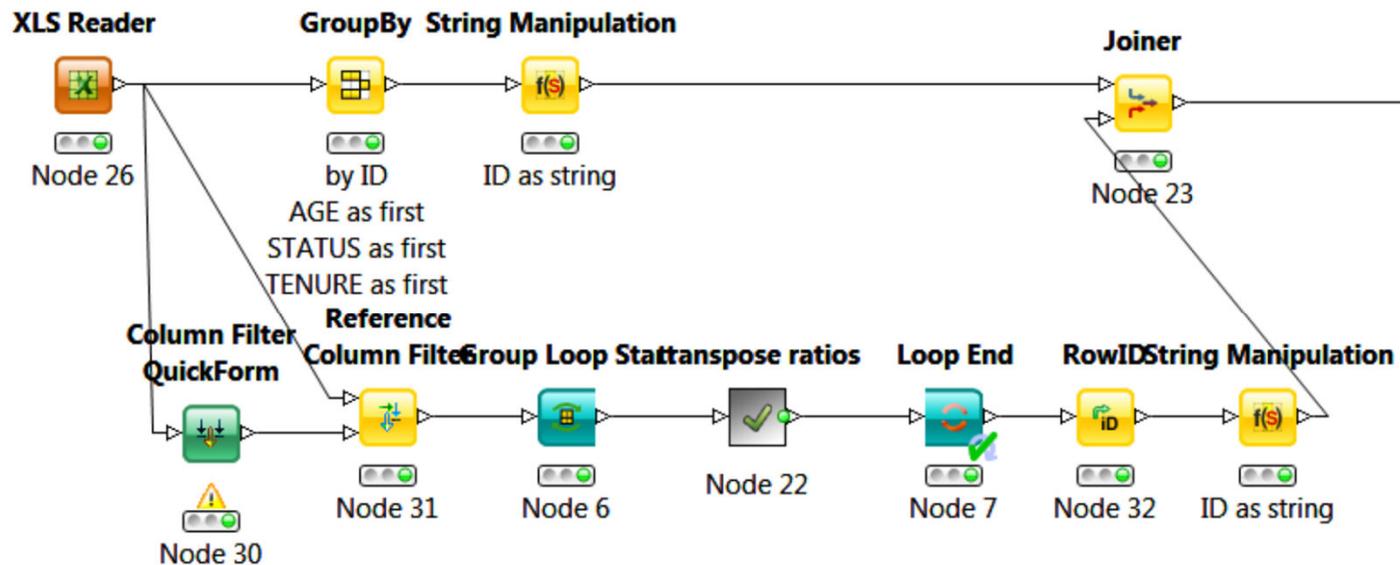
Row ID	D AMT_CHEQUES_P3	D AMT_CHEQUES_P6	D AMT_CHEQUES_P9	D AMT_AT...
Row9_7548...	-0.432	-0.542	0	
Row10_795...	-0.139	-0.498	0	
Row11_862...	-0.059	0.016	0.023	-0.027
Row12_904...	0	0	0	0
Row13_957...	0	0	0	0
Row14_1.0...	-0.002	-0.002	0.005	0.241



Il workflow

- **Preparazione dei dati**

Ciò è stato facilmente ottenuto utilizzando gli strumenti "Data Manipulation" di Knime



Il workflow

- **Modellazione e Valutazione dei risultati**

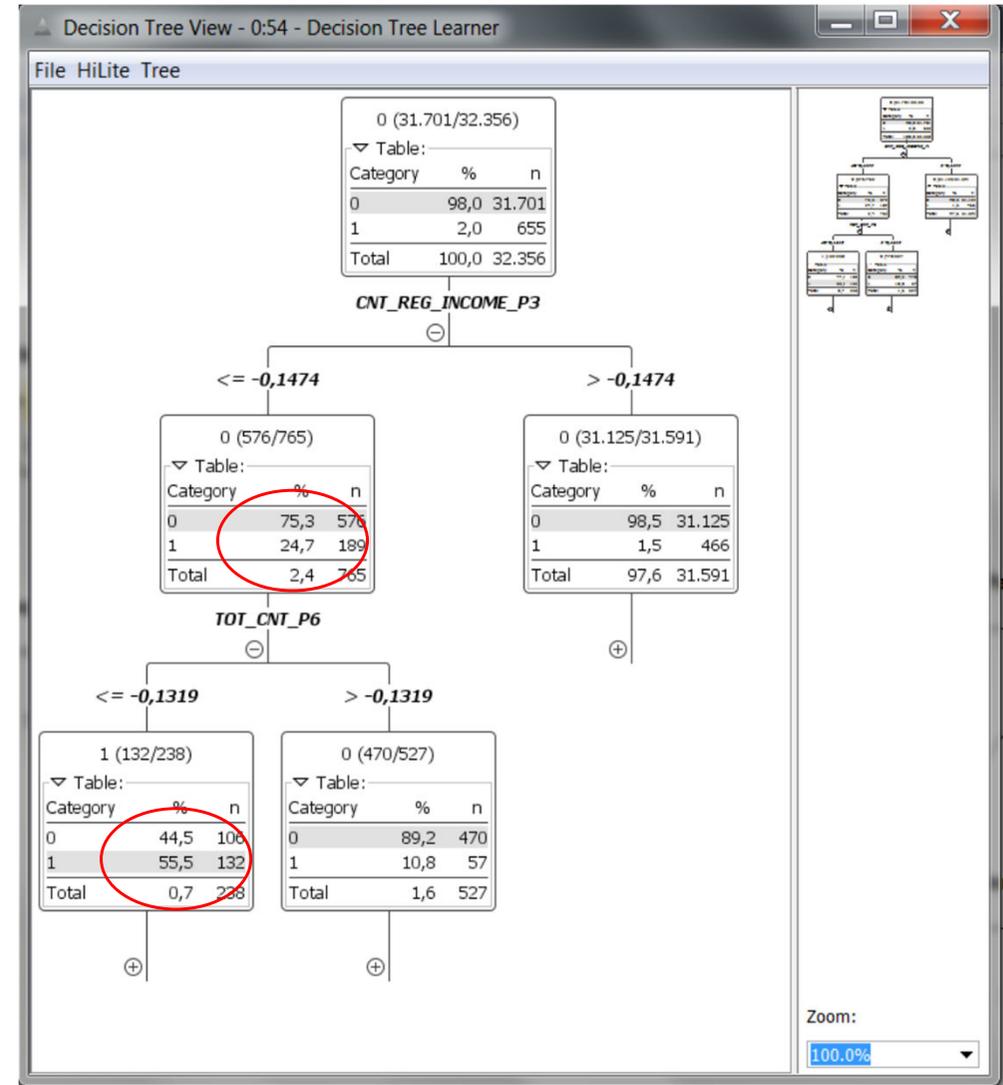
- ✓ Sono stati considerati diversi metodi utilizzando lo strumento "Data Mining", il metodo Decision Trees è stata la scelta finale
- ✓ Poiché l'evento dell'abbandono è poco frequente, per migliorare le capacità predittive del modello, la sua proporzione è stata aumentata con un sovracampionamento portandola dal 2% al 30%
- ✓ Allo scopo di fornire stime corrette su dati sconosciuti la prestazione del modello si è basata su dati non utilizzati per la sua costruzione (Validation set)

Il workflow

• Modellazione e Valutazione

Il Decision Tree View mostra le regole che definiscono il modello, le prime evidenziano che

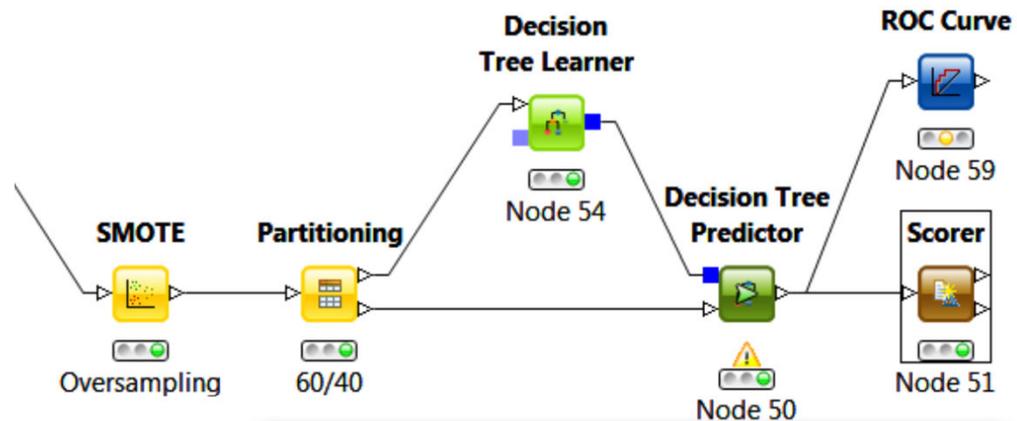
- ✓ un decremento delle operazioni di accrediti periodici dal trimestre precedente inferiore a -14,7% porta il numero di defezioni al 24,7%
- ✓ un decremento delle operazioni totali dal semestre precedente inferiori al -13% porta, insieme alla regola precedente, un ulteriore aumento del numero di defezioni al 55,5%



Il workflow

- **Modellazione e Valutazione**

- ✓ Il nodo Scorer, oltre a calcolare le probabilità sul Validation set, illustra le prestazioni del modello (Precisione 99,2%)



Confusion Matrix - 0:51 - Scorer

File Hilite		
STATUS ...	0	1
0	21077	58
1	106	331

Correct classified: 21.408 Wrong classified: 164
Accuracy: 99,24 % Error: 0,76 %

Conclusioni

- La **fidelizzazione** dei clienti passa attraverso un'attività di Customer Retention
- E' di fondamentale importanza stabilire un **processo** che porti a **individuare i clienti profittevoli più a rischio di defezione** nel prossimo futuro
- Questo processo richiede la possibilità di **strutturare i dati** in modo appropriato e di **utilizzare modelli** di analisi che predicono la defezione dei clienti prima che questa accada.
- **Knime** ha le funzionalità che soddisfare queste esigenze in modo **facile e veloce, risparmiando** tempo e, soprattutto, ... ***senza costi per il software!***