

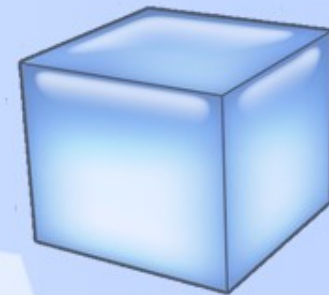
# BICOOP – Churn Analysis

Definizione del concetto di abbandono e creazione di modelli previsionali

*Mirco Nanni  
KDD Lab, ISTI-CNR, Pisa  
mirco.nanni@isti.cnr.it*

*DM 2013, 11 marzo*

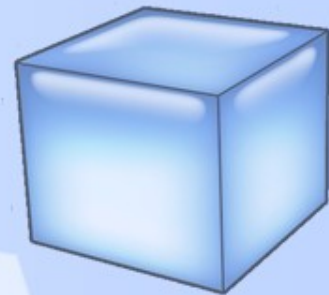
# Sommario



- Introduzione
- Preparazione dei dati
- Modelli
- Valutazione dei risultati
- Scenario d'uso
- Conclusioni



# Introduzione



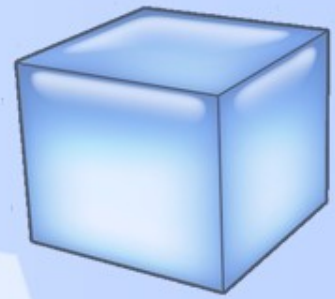
## Obiettivo:

- definire opportunamente il concetto di fedeltà del cliente
- fornire strumenti previsionali in grado di stimare la fedeltà dei clienti nel futuro prossimo:
- individuare:
  - clienti che diminuiranno la loro attività
  - clienti che cesseranno definitivamente il rapporto

## Finalità:

- arricchire la conoscenza dei propri clienti
- innescare meccanismi di *customer redemption*

# Abbandono



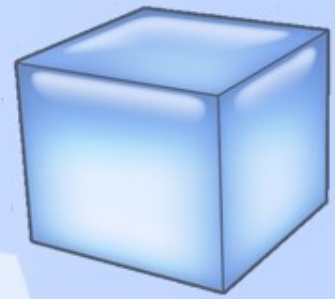
## Definizione del problema:

- Stimare la probabilità di abbandono dei clienti a partire dalle informazioni presenti nel data warehouse:
  - dati di vendita
  - dati anagrafici

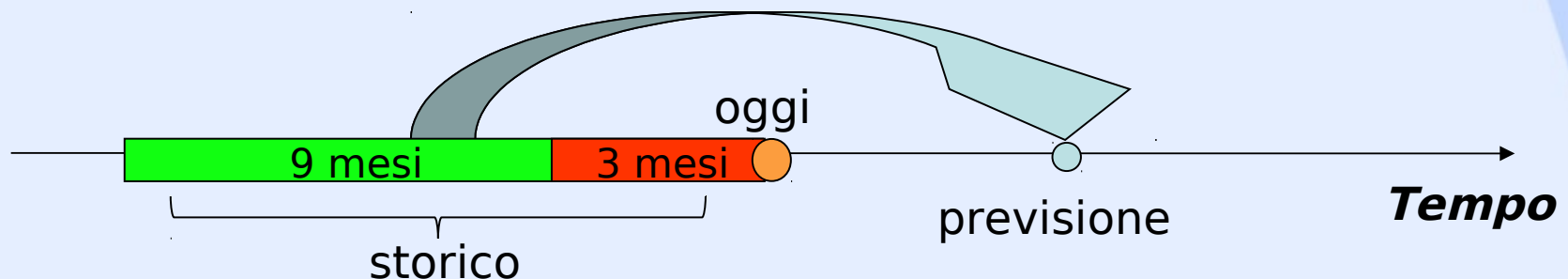
## Definizione di abbandono:

- Un cliente è a rischio di abbandono quando evidenzia un netto calo nelle sue misure di spesa:
  - visite
  - volumi di spesa
  - battute di cassa

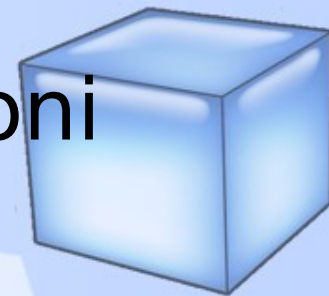
# Analisi previsionale



- Raccolta dei dati storici per l'estrazione di:
  - Variabili di vendita e anagrafiche, i predittori (periodo verde)
  - Variabili obiettivo (periodo rosso)
- Costruzione di un modello predittivo
  - Addestrato in modo opportuno su dati storici
  - Utilizzabile per ottenere informazioni previsionali



# Preparazione dati - osservazioni (periodo verde)

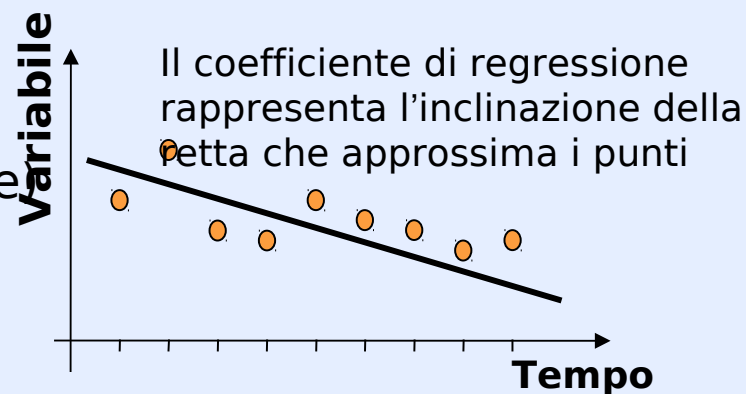


Si sono estratte dal data warehouse, per il periodo di 9 mesi (Dicembre 2006 - Agosto 2007) le seguenti informazioni:

- Dati anagrafici (sesso, età, professione etc.)
- Dati di spesa
  - Globale
  - Settori specifici: fresco, carni, pesce, ortofrutta
  - Pesata (abbattimento no-food )

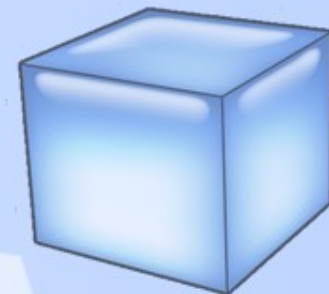
## • Trend di spesa:

- Tipologia cliente (per ogni mese)
- Regressione spese
- Regressione spesa
- Regressione battute



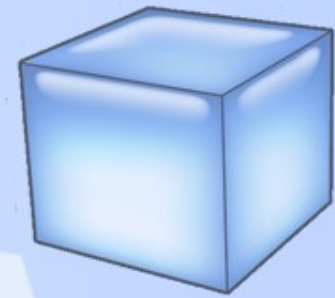


# Preparazione dati - target (periodo rosso)



- Si sono estratte dal data warehouse, per il periodo di 3 mesi (Settembre 2007 - Novembre 2007) le seguenti informazioni:
  - Numero di spese
  - Variazione di spesa rispetto al periodo verde
    - Volume di spesa
    - Battute di cassa
    - Numero di visite

# Dataset



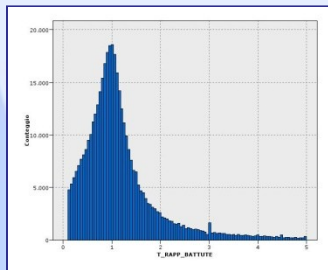
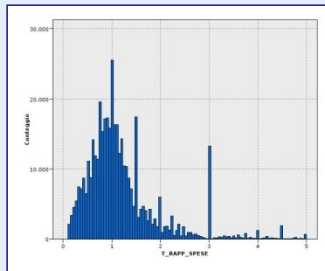
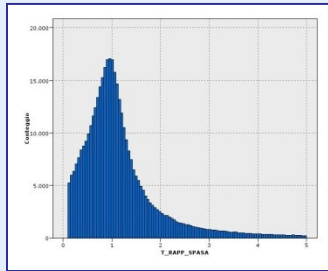
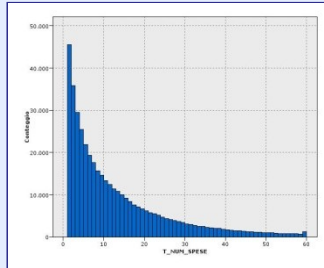
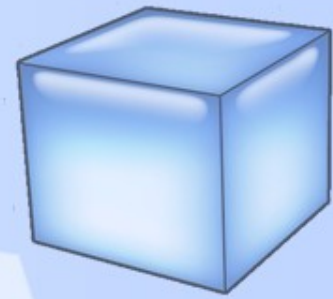
- Il dataset così ottenuto presenta una riga per ogni cliente che ha effettuato almeno una spesa nei nove mesi di osservazione. In tutto abbiamo ottenuto:

- 517.000 righe
- 47 attributi

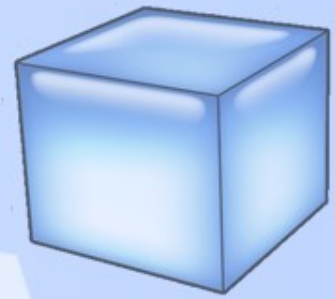
Predittori Anagrafici	Predittori di spesa	Predittori di trend	Variabili target
CLIENTE_ID	DATA_ULTIMA_SPESA	TIPOLOGIA_01	T_NUM_SPESE
SESSO	NUM_SPESE	TIPOLOGIA_02	T_RAPP_SPESE
STATO_CIVILE	SPESA_TOT	TIPOLOGIA_03	T_RAPP_SPESA
PROFESSIONE	SPESA_TOT_PESATA	TIPOLOGIA_04	T_RAPP_BATTUTE
TITOLO_STUDIO	SPESA_MEDIA	TIPOLOGIA_05	
PROVINCIA	SPESA_MEDIA_PESATA	TIPOLOGIA_06	
REGIONE	BATTUTE	TIPOLOGIA_07	
ANNO_SOCIO	FRESCHI_TOT	TIPOLOGIA_08	
FASCIA_ANNO_SOCIO	FRESCHI_SPESE	TIPOLOGIA_09	
FL_INVIO_RIVISTA	CARNI_TOT	TIPOLOGIA_MEDIA	
COD_NEGOZIO	CARNI_SPESE	TIPOLOGIA_ZERI	
ETA	PESCE_TOT	REGR_NUM_SPESE	
ETA_FASCIA	PESCE_SPESE	REGR_SPESA	
	ORTOFRUTTA_TOT	REGR_SPESA_PESATA	
	ORTOFRUTTA_SPESE	REGR_BATTUTE	



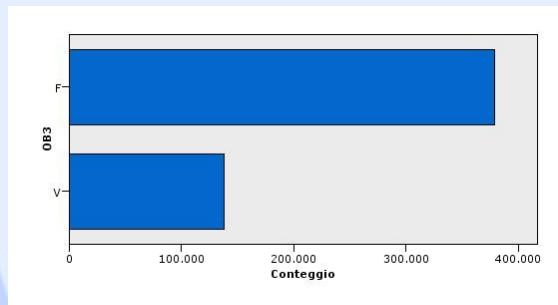
# Funzioni Obiettivo



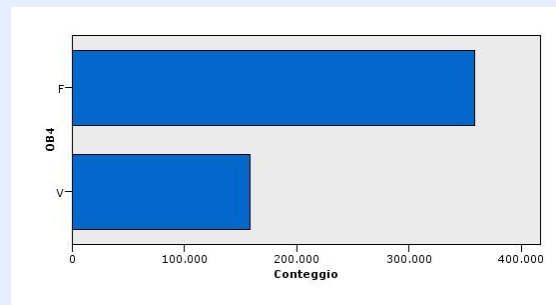
# F. Obiettivo - Soglie



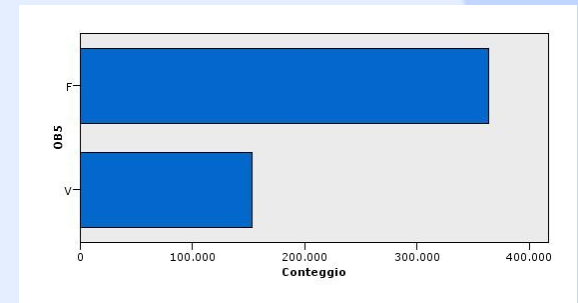
- Scelta una soglia di allarme per indicare un possibile cliente defezionario i rapporti si trasformano in tre indicatori di abbandono
- Abbiamo scelto come soglia una diminuzione sulle 3 misure del 50%
- Otteniamo le seguenti distribuzioni:



OB1: RAPP\_SPESE

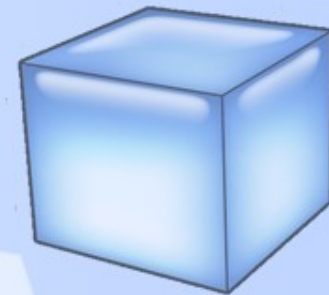


OB2: RAPP\_SPESA



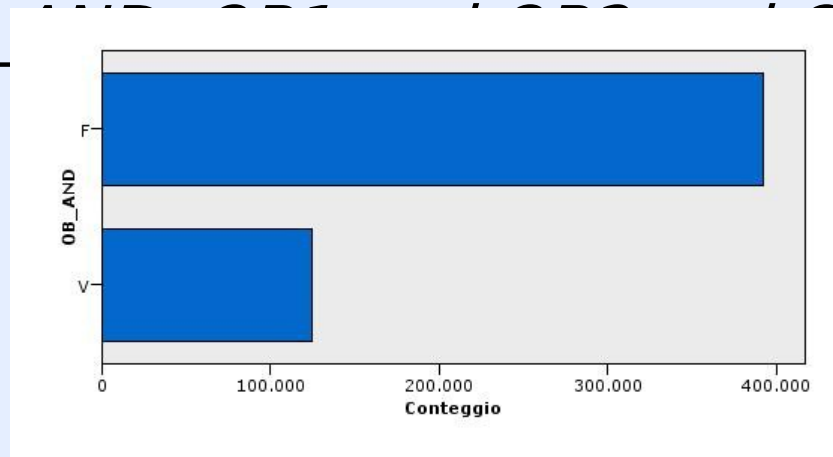
OB3: RAPP\_BATTUTE

# F. Obiettivo - Sintesi

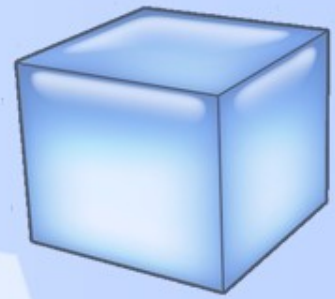


Per la funzione obiettivo finale si è deciso di considerare come potenziali defezionari tutti i clienti che superato la soglia di allarme, in ognuno dei tre indicatori OB1, OB2, OB3:

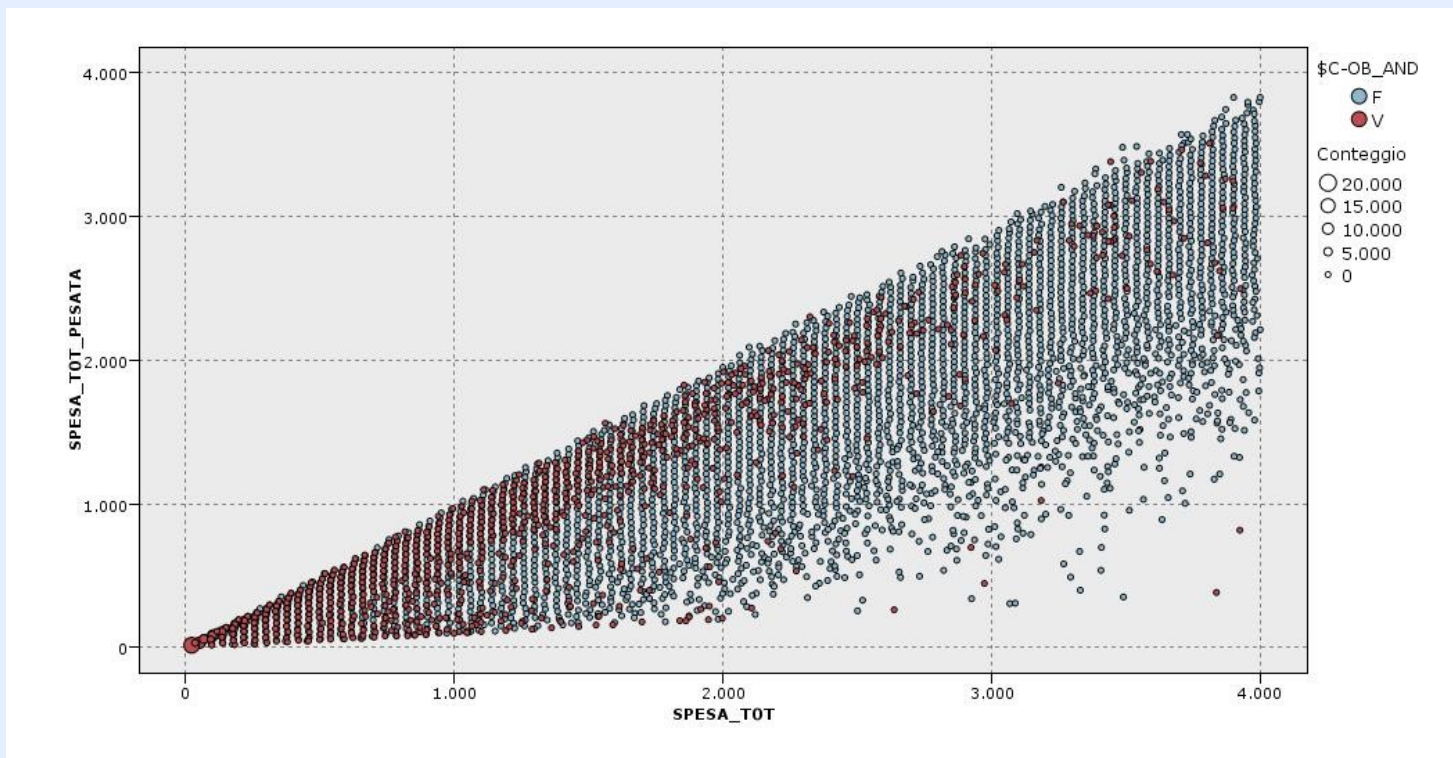
*OB* \_\_\_\_\_ *B3*



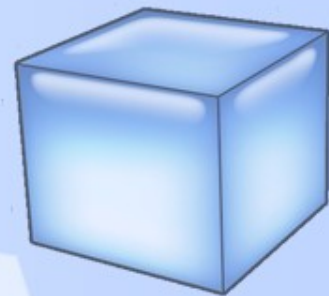
# Modello previsionale e Risultati



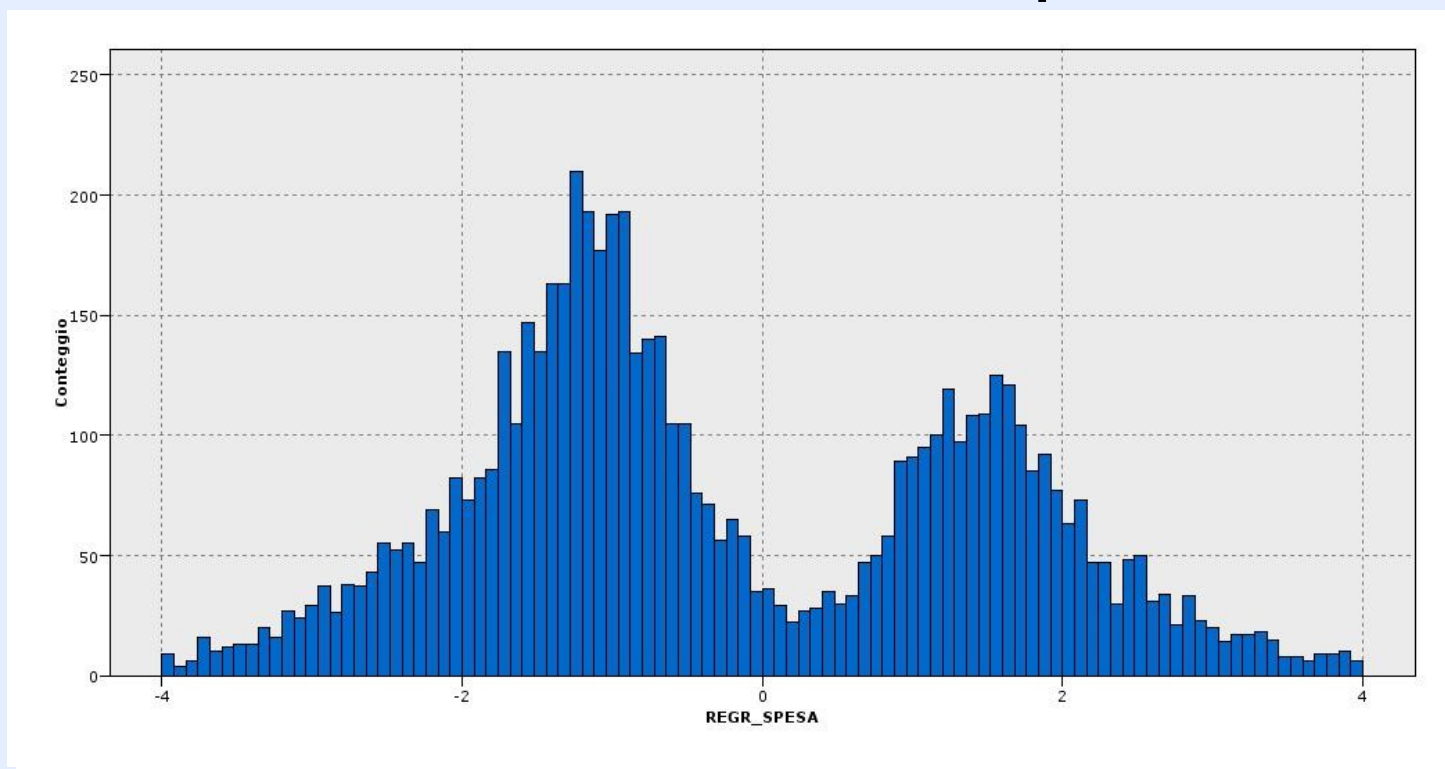
- Distribuzione spesa totale vs. spesa pesata



# Modello previsionale e Risultati



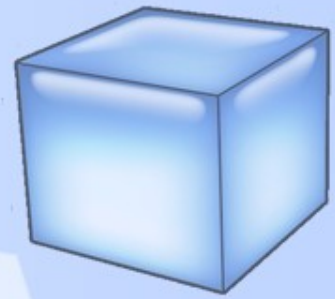
- Distribuzione trend di spesa



Trend dei clienti con spesa totale > 400€



# Modello previsionale e Risultati



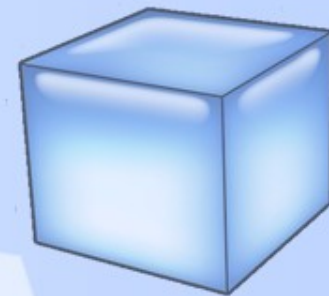
- Esempio di regole associative:

se REGIONE = TOSCANA  
e NUM\_SPESE  $\leq$  128  
e TIPOLOGIA\_01 = 7  
e TIPOLOGIA\_09 = 0  
e TIPOLOGIA\_ZERI  $>$  2  
e REGR\_BATTUTE  $\leq$  -0,98  
allora V (confidenza 82,8%)

se DATA\_ULTIMA\_SPESA  $>$  183  
e NUM\_SPESE  $\leq$  21  
e TIPOLOGIA\_ZERI  $>$  1  
e REGR\_NUM\_SPESE  $\leq$  -0,02  
e REGR\_BATTUTE  $\leq$  -0,98  
allora V (confidenza 92%)

# Modello previsionale

## Risultati Globali



- Correttezza generale del modello:
  - 81.06% sul training set (70% del dataset, 360.000 righe)
  - 80.94% sul test set (30% del dataset, 155.000 righe)

Valori Predetti

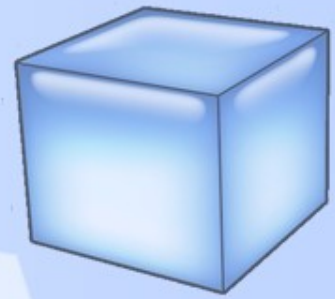
• ~~Matrici di confusione!~~

Valori Reali	Training Set		Test Set	
	F	V	F	V
F	256.608	17.920	110.029	7.767
V	50.540	36.466	21.855	15.734

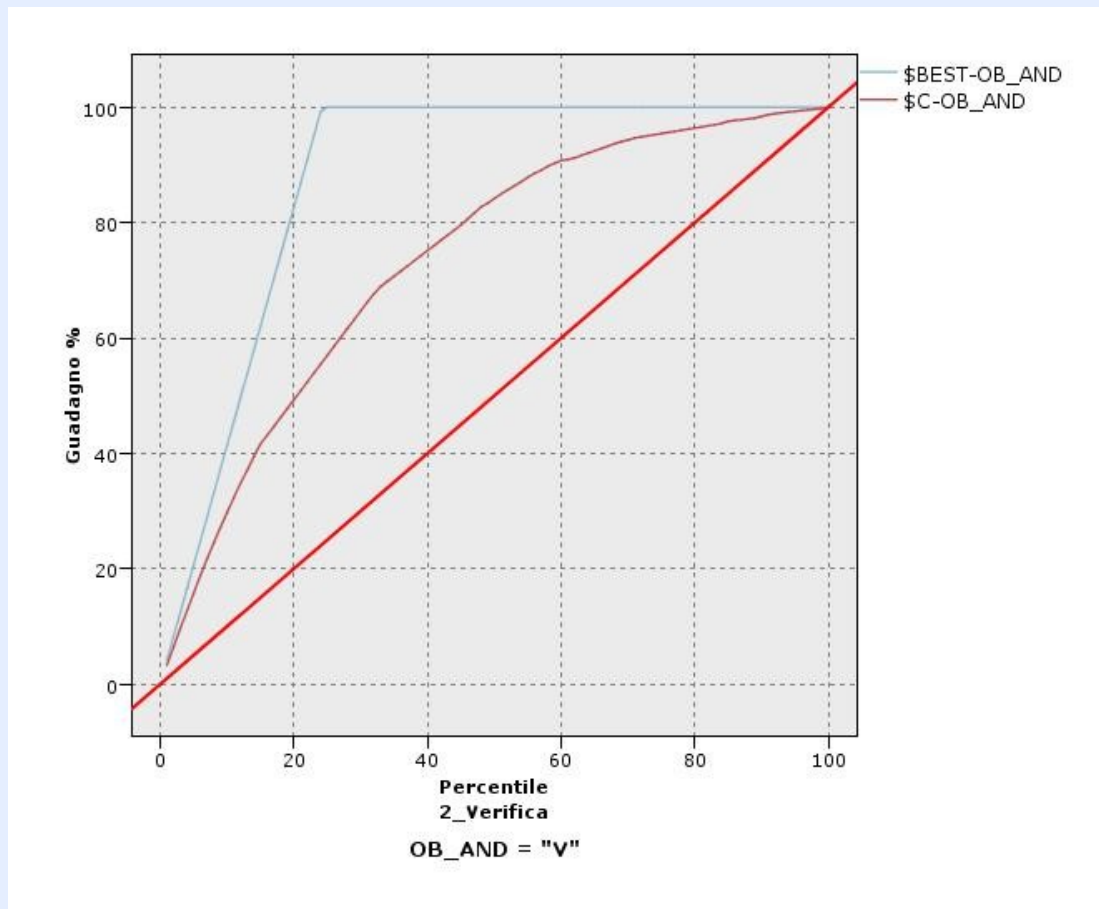
**66.9%**

Con un guadagno netto del **42.8%**

# Modello previsionale Risultati Globali

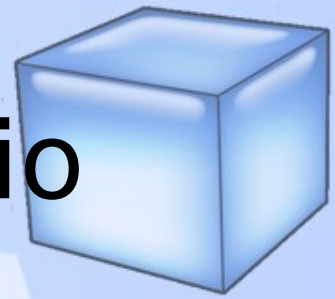


- Lift chart

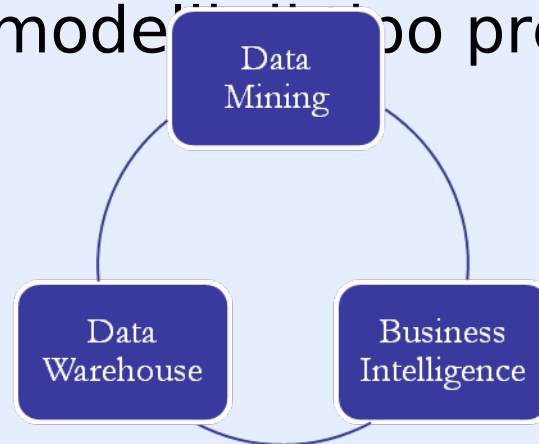




# Scenario d'uso - Esempio

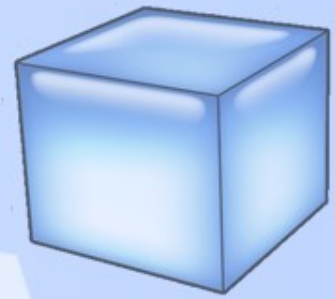


- Creare un ambiente aperto e dinamico nel quale i dati forniti dal data warehouse vengono elaborati e trasformati in modelli di tipo previsionale.



- I modelli previsionali possono essere usati per arricchire il data warehouse, innestando un circolo virtuoso di informazioni utilizzabili anche direttamente in ambienti di Business Intelligence.

# Conclusioni



- Per concludere:
  - Sono stati utilizzati dati provenienti dal data warehouse, risparmiando tempo e ottenendo dati di buona qualità
  - Abbiamo usato tecniche di mining avanzate per generare modelli predittivi, principalmente regole associative e alberi di decisione.
  - I risultati ottenuti sono soddisfacenti e si intravedono buone prospettive di miglioramento
- Possibili sviluppi futuri
  - Sperimentazione di altri tipi di analisi: sub group analysis, market segmentation, clustering ect.
  - Consolidamento e validazione dei risultati ottenuti
  - Incrementare la collaborazione con gli esperti del dominio per una migliore taratura del problema, delle definizioni usate e delle funzioni obiettivo
  - Integrazione dei dati previsionali forniti dai modelli predittivi all'interno della struttura di business intelligence aziendale