

Data Mining2 – Advanced Aspects and Applications

Fosca Giannotti and Mirco Nanni
Pisa KDD Lab, ISTI-CNR & Univ. Pisa

<http://www-kdd.isti.cnr.it/>



DIPARTIMENTO DI INFORMATICA - Università di Pisa
anno accademico 2013/2014

- Churn analysis
 - Casi di studio

Contesto generale

- Attività e servizi caratterizzati da
 - un rapporto continuativo tra fornitore e fruitore
 - possibilità di tracciare il comportamento del fruitore
 - esistenza di fornitori concorrenti
- Principalmente attività commerciali
 - telefonia (fissa e mobile)
 - vendita al dettaglio

Definizione del problema

- Diversi sinonimi per lo stesso problema:
 - Churn
 - Abbandono
 - Defezione
- Nel nostro ambito: interruzione del rapporto fruitore-fornitore da parte del fruitore
 - Tipicamente non preannunciata
 - A volte esplicita (Es.: rescissione del contratto) a volte no (Es.: passaggio ad un supermercato concorrente)

Obiettivi generali

- Motivazione di base:
 - Preservare i clienti/fruitori fedeli è più conveniente del cercare nuovi clienti
 - Meno costoso
 - Da più garanzie di stabilità al servizio
- Obiettivo:
 - 1) Identificare i defezionanti
 - 2) Intraprendere iniziative per dissuaderli
 - Offerte, sconti, servizi aggiuntivi, trattamenti di favore

Identificazione tempestiva

- In alcuni casi, nel momento in cui la defezione si manifesta, è troppo tardi per intervenire
 - non è più possibile recuperare il defezionante, o
 - non è più conveniente recuperarlo
- Fondamentale identificare il defezionante *immediatamente*, o addirittura *in anticipo*
- Nuova formulazione del problema:
 - Churn Analysis = *Previsione dell'abbandono*

Interruzione del rapporto: modalità

- Interruzione esplicita
 - Tipica dei rapporti che richiedono un contratto o impegni da parte del fruitore
 - Es.: Telefonia, nei casi in cui è previsto un canone
 - Es.: Tesseramenti rinnovabili non gratuiti
- Interruzione implicita
 - Tipica dei rapporti non formalizzati o privi di costi per il fruitore
 - Es.: Tessere sconto e carte fedeltà

Interruzione implicita

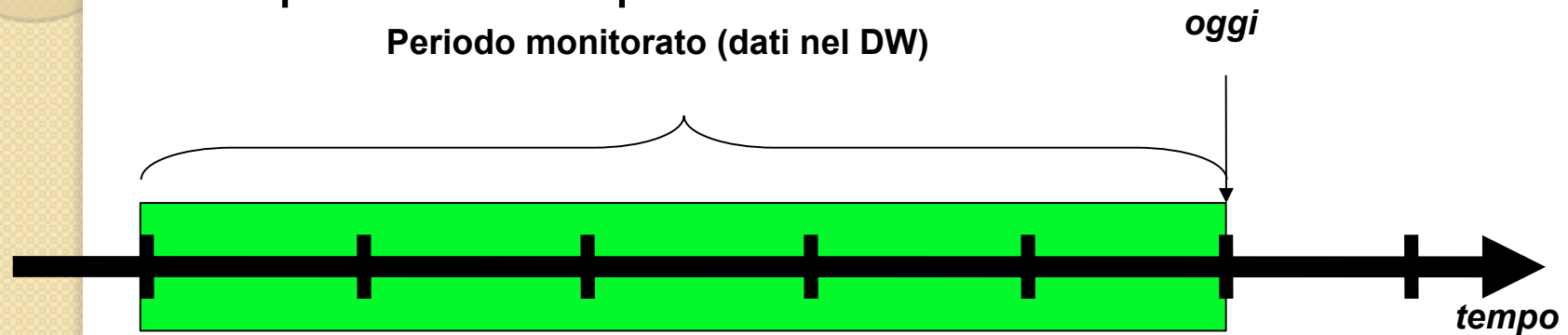
- E' la situazione più comune nel settore delle vendite al dettaglio
 - Carte fedeltà che non richiedono rinnovi né costi
 - Il defezionante semplicemente non la usa più
- Domanda: è sempre facile capire quando il cliente/fruttore ha abbandonato?
 - Non fa acquisti per un mese?
 - Non fa acquisti per un anno?
 - Visita il punto vendita meno di 2 volte al mese?
 - Spende meno del 50% di quanto faceva 3 mesi fa?

Abbandono “soft”

- Nozione alternativa di abbandono:
 - Passaggio da un tipo di rapporto ad uno diverso
 - Caso estremo: da “fedele” a “abbandono totale”
- Situazione naturale nella vendita al dettaglio
 - Il segmento “fedele” fornisce (parziali) garanzie su un indotto minimo dell'attività
 - Il degrado del cliente “fedele” a cliente “saltuario” ha effetti negativi sulla gestione aziendale
 - Valgono le stesse motivazioni dell'abbandono “hard”

Previsione dell'abbandono

- Il tracciamento del cliente ci consente di ricostruire la sua “storia” per un certo periodo

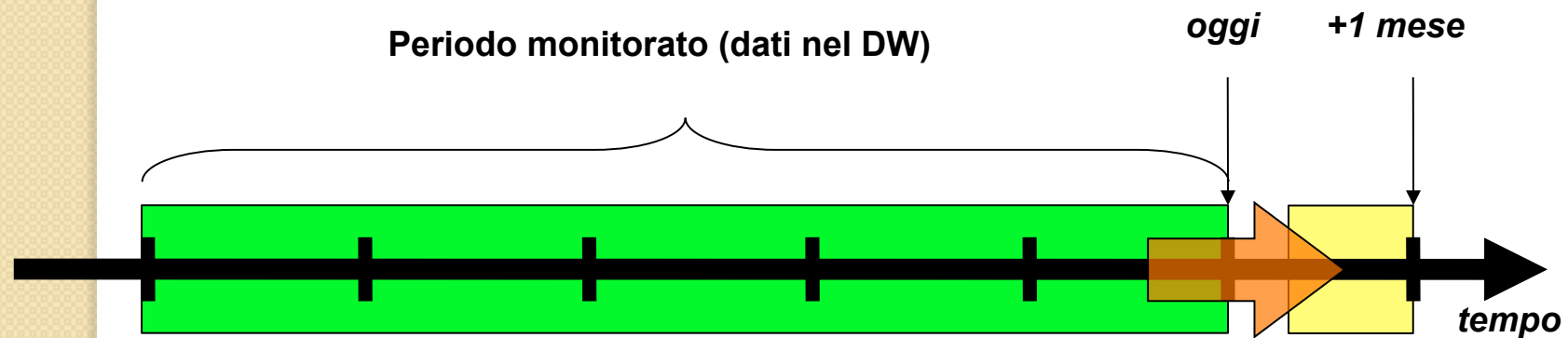


- La nozione di abbandono adottata sarà valutata su un segmento recente di tale storia



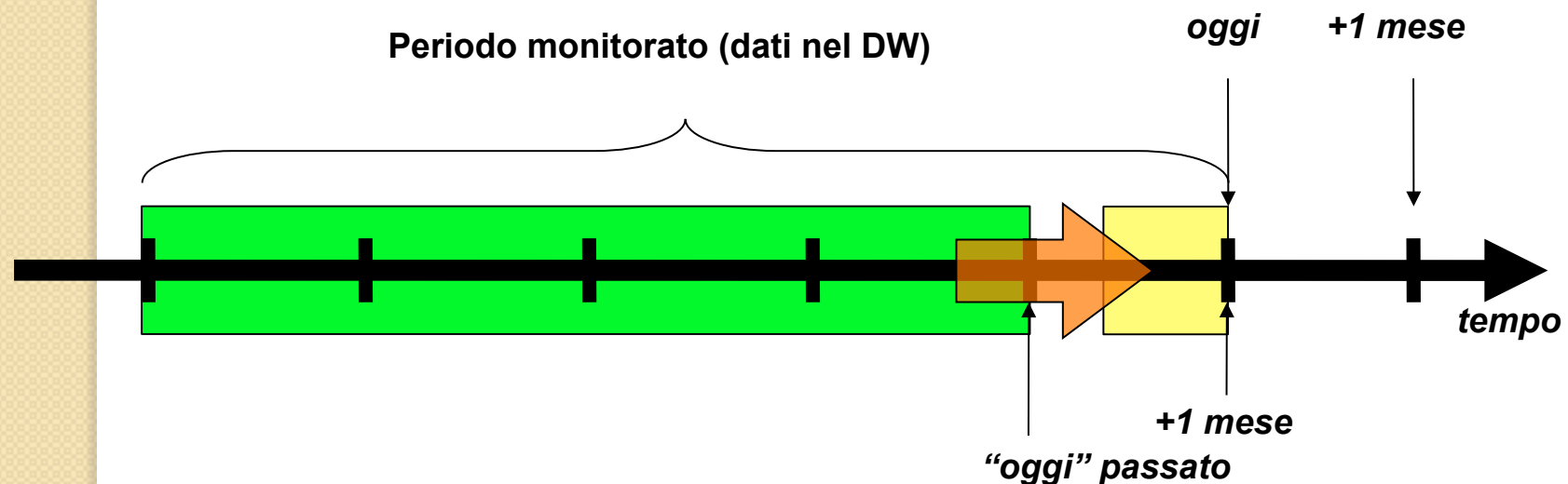
Previsione dell'abbandono

- Obiettivo: previsione dello stato di abbandono *futuro*, conoscendo la storia recente del cliente



- **La storia recente fornisce indizi sul comportamento che il cliente si presta a tenere**
 - alcuni indizi permettono di discriminare i futuri defezionanti, altri no
 - alcuni indizi sono espliciti nei dati a disposizione, altri vanno derivati da essi

- Come determinare oggi le correlazioni tra situazione attuale e stato futuro?
 - Cerchiamo queste correlazioni nel *passato*
 - Le relazioni “passato → oggi” verranno sfruttate per predire il futuro dall'oggi



Schema dell'applicazione

- Definizione/estrazione delle variabili di lavoro
 - Variabili predittive: gli *indizi* disponibili *oggi/passato*
 - Variabile target: lo stato di abbandono *futuro/oggi*
- Estrazione del modello predittivo
 - Ricerca di correlazioni tra variabili predittive e variabile target, da sfruttare in fase di predizione
- Applicazione del modello predittivo
 - Le relazioni variabile predittiva → target vengono applicate alla situazione odierna (in termini di variabili predittive) per stimare il valore più probabile della variabile target

Casi di studio

- Due esperienze molto simili in ambito retail:



- **COOP: predizione dell'abbandono su diverse nozioni base di defezione**



- **Carrefour: enfasi sulla sintesi di un indicatore dello stato di abbandono del cliente**

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*

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* focalizzati.

Abbandono

Definizione del problema:

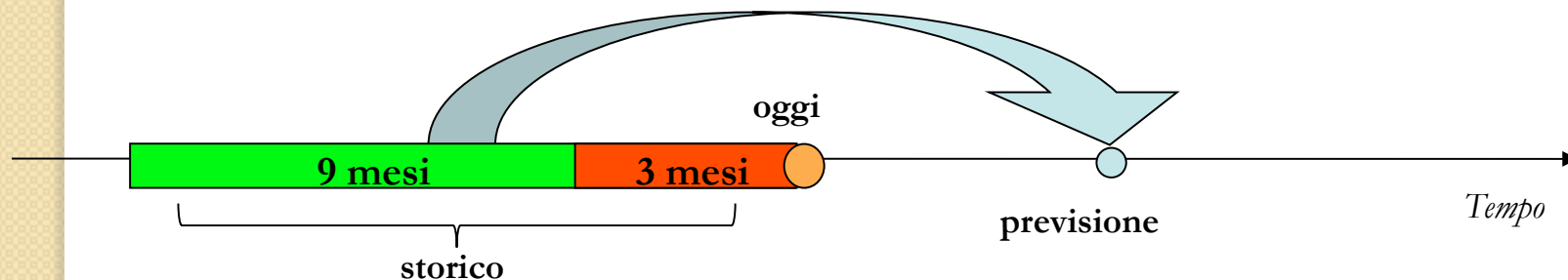
- Stimare la probabilità di abbandono dei clienti a partire dalle informazioni presenti nel data warehouse:
 - dati di vendita
 - dati anagrafici

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 – target (periodo rosso)

- Si sono estratte dal data warehouse, per il periodo rosso di 3 mesi (Settembre 2007 – Novembre 2007) le seguenti informazioni:
 - Numero di spese
 - Variazione di spesa rispetto al periodo verde
 - Volume di spesa (Valore degli acquisti)
 - Battute di cassa (Varietà di prodotti acquistati)
 - Numero di visite (Frequenza di visita)

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:

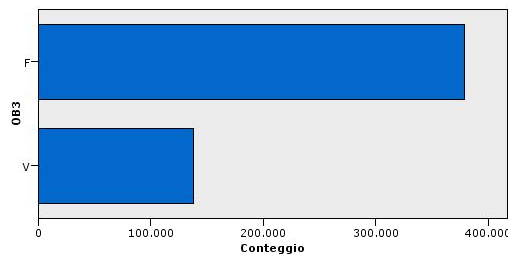
- 517.000 righe
- 47 attributi

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	



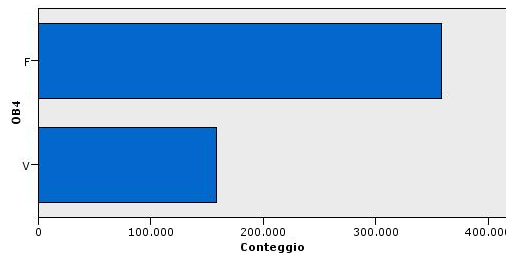
F. Obiettivo – Soglie

- 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:



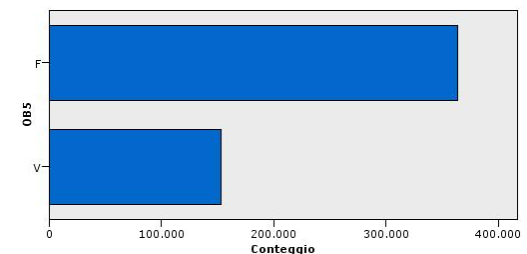
**OB1:
RAPP_SPESE**

Frequenza di visita



OB2: RAPP_SPESA

Volume di spesa



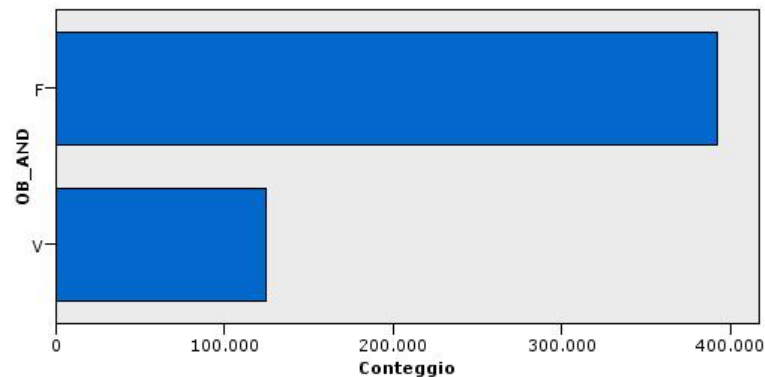
OB3: RAPP_BATTUTE

Varietà di items

F. Obiettivo – Sintesi

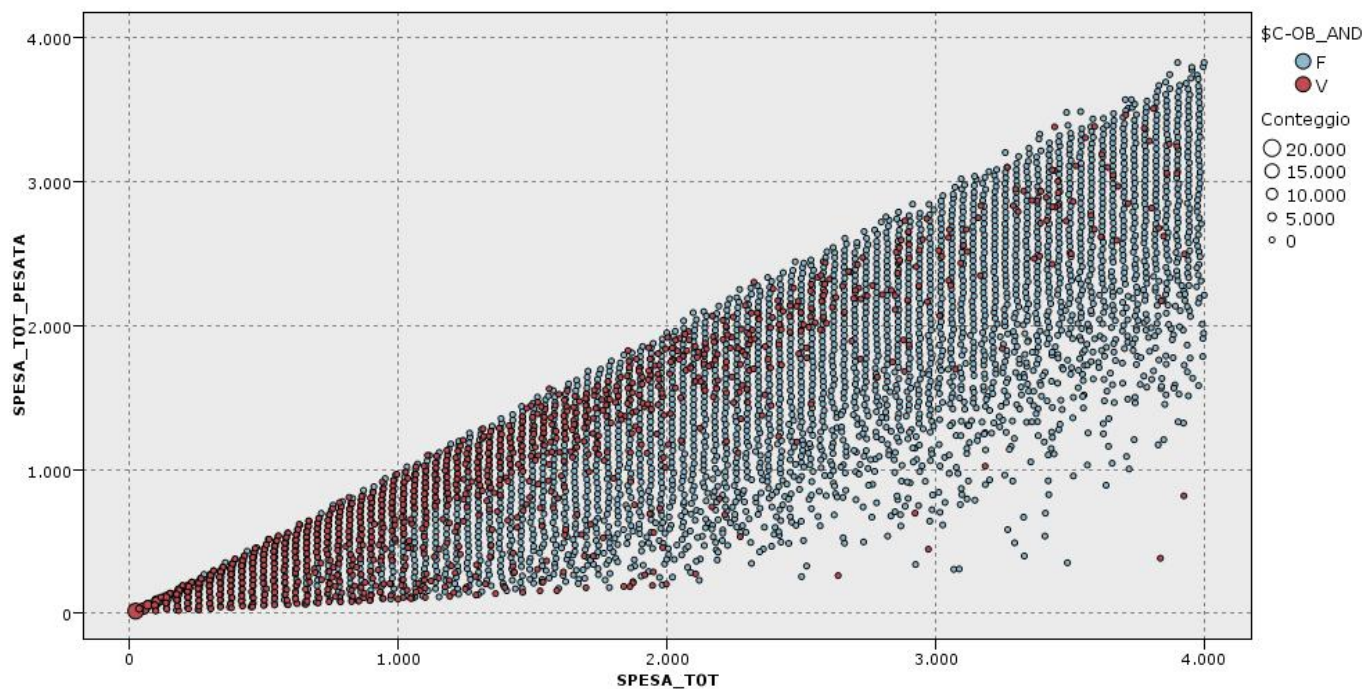
Per la funzione obiettivo finale si è deciso di considerare come potenziali defezionari tutti i clienti che superato la soglia di allarme, in ognuno dei tre indicatori OB1, OB2, OB3:

OB_AND: OB1 and OB2 and OB3



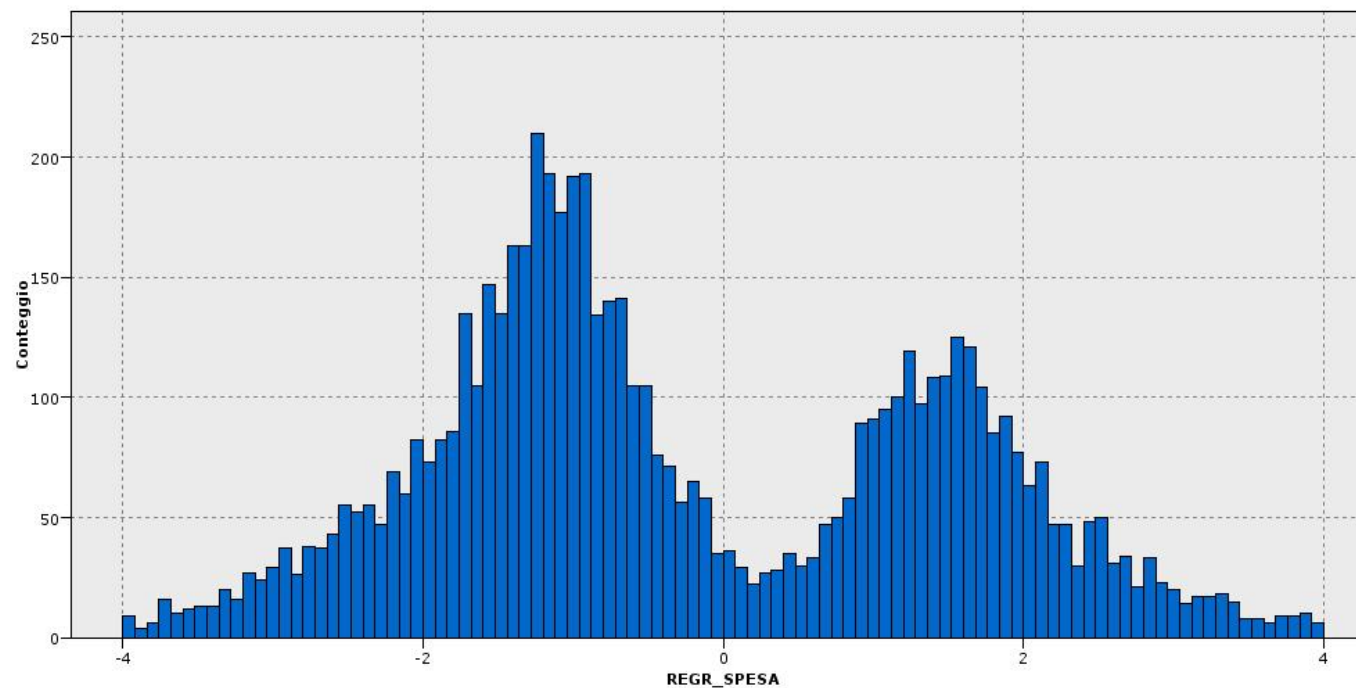
Modello previsionale e Risultati

- 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 <= 128

e TIPOLOGIA_01 = 7

e TIPOLOGIA_09 = 0

e TIPOLOGIA_ZERI > 2

e REGR_BATTUTE <= -0,98

allora V (confidenza 82,8%)

se DATA_ULTIMA_SPESA > 183

e NUM_SPESE <= 21

e TIPOLOGIA_ZERI > 1

e REGR_NUM_SPESE <= -0,02

e REGR_BATTUTE <= -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)
- Matrici di confusione:

Valori Predetti

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

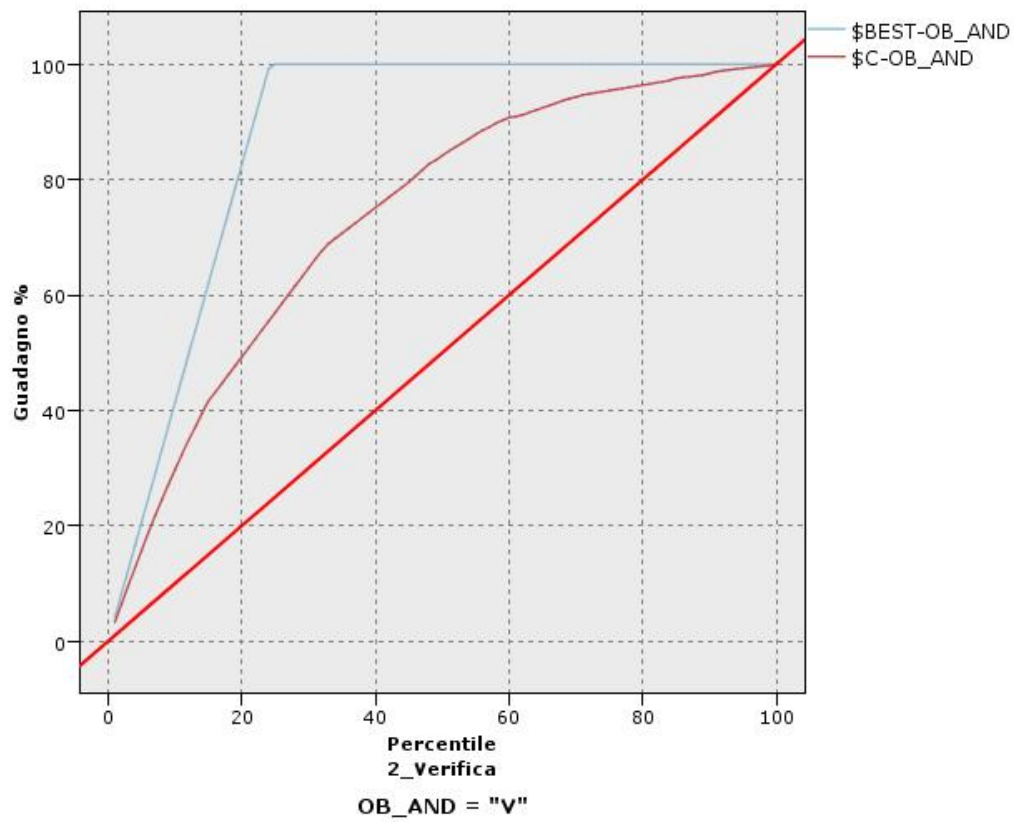
66.9%

Con un guadagno netto del 42.8%

Modello previsionale

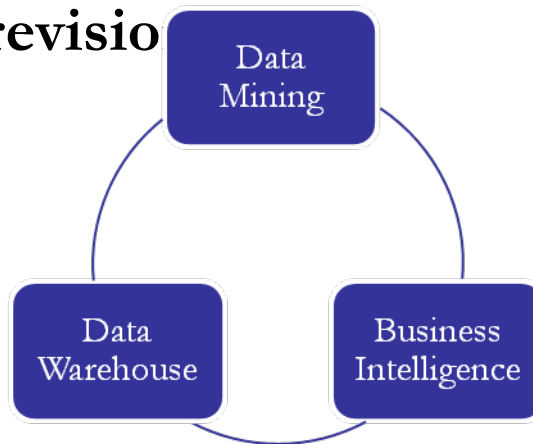
Risultati Globali

- 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 previsiono**



- **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