

# II Data Mining

---

**Susi Dulli**

***dulli@math.unipd.it***

---

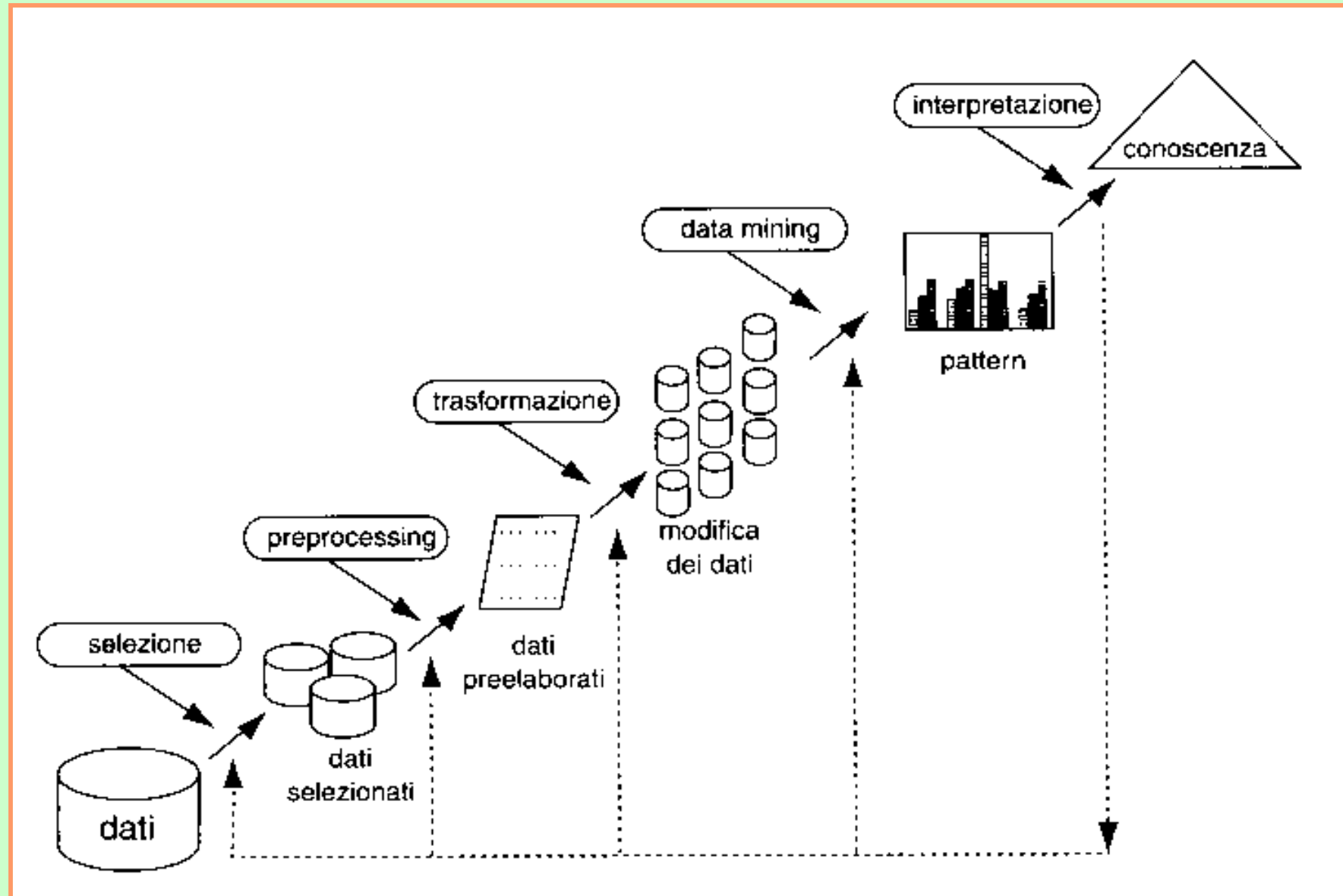
# Il Data Mining

Il Data Mining è il processo di scoperta di relazioni, **pattern**, ed informazioni precedentemente sconosciute e potenzialmente utili, all'interno di grandi basi di dati.

Un **pattern** indica una struttura, un modello, o, in generale una rappresentazione sintetica dei dati.

Le tecniche di Data Mining sono orientate alle esigenze aziendali.

# Knowledge Discovery in Database



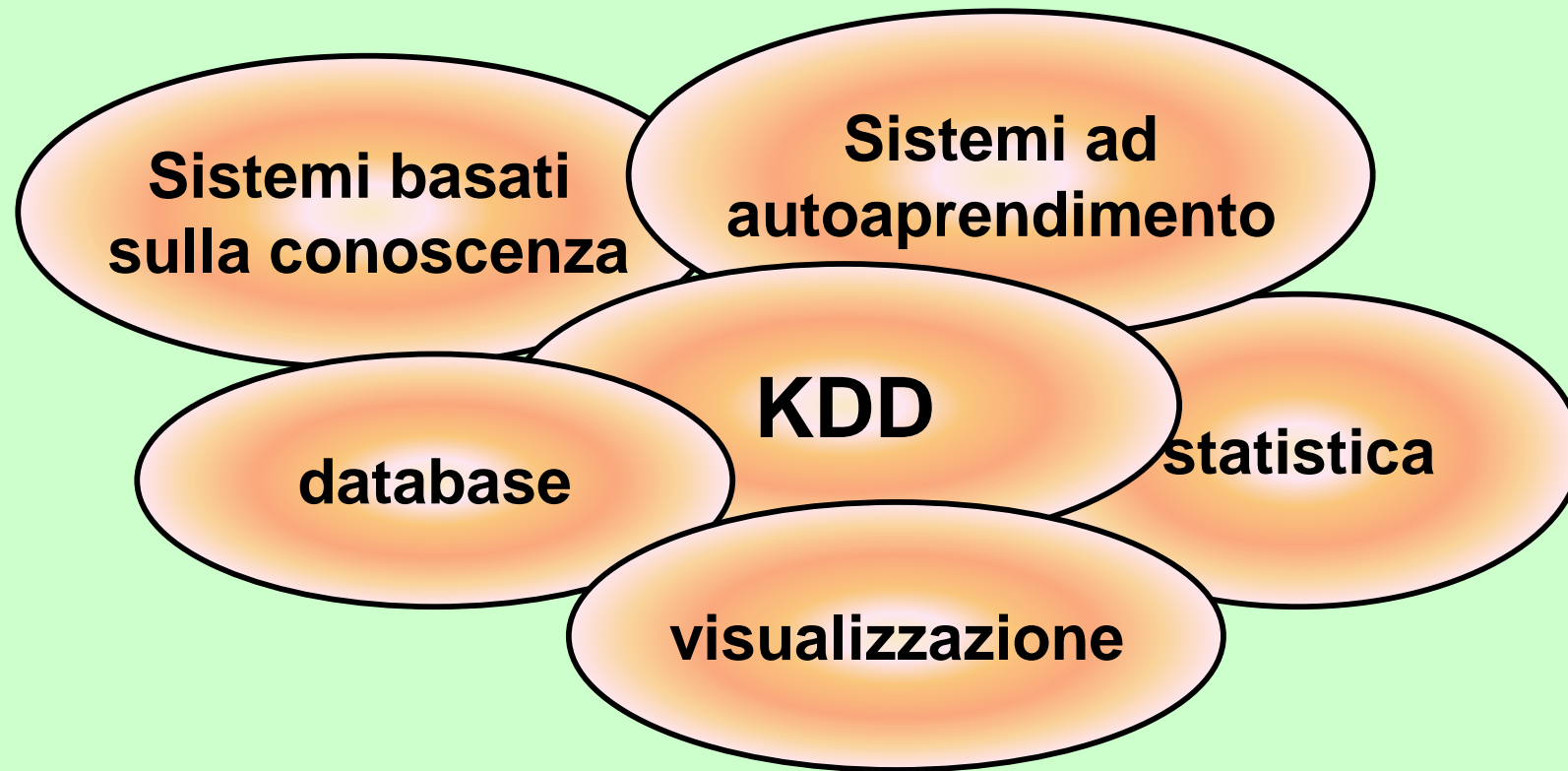
# Data mining e' un processo...

- analitico finalizzato ad selezionare, esplorare,modellare grosse quantita' di dati ...
- nella ricerca di relazioni e informazioni non note a priori...
- Informazioni tramutabili in azioni commerciali allo scopo di ottenere un vantaggio di business

Rilascia applicazioni integrate  
nei processi decisionali  
aziendali e NON studi su un  
fenomeno

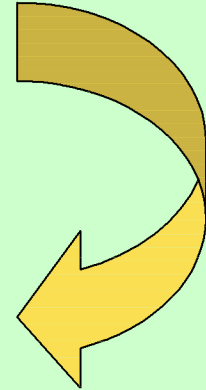
Grandi masse di dati si ma  
Integrazione fra analisi  
quantitative e conoscenza del  
business

# L'approccio multidisciplinare del Data Mining



# Knowledge Discovery in Database (KDD)

- ➔ Si applica a qualunque fonte di dati
- ➔ I campi di utilizzo degli strumenti di Data mining sono estremamente eterogenei
- ➔ l'elaborazione a dati aziendali e' recente



- **Comprensione del dominio (esperto di B.A., esperto di S.I.A., esperto di analisi dei dati)**
- **Preparazione del set di dati**
- **Data mining (descrizione dati attuali e previsione di andamenti futuri)**
- **Interpretazione e valutazione pattern**
- **Rappresentazione e utilizzo dei risultati**

# I vantaggi del Data Mining

- ➡ Trattamento di dati quantitativi, qualitativi, testuali, immagini e suoni;
- ➡ Non richiede ipotesi a priori da parte del ricercatore;
- ➡ Non richiede ipotesi sulla forma distributiva delle variabili;
- ➡ Possibilità di elaborare un numero elevato di osservazioni;
- ➡ Possibilità di elaborare un numero elevato di variabili;
- ➡ Algoritmi ottimizzati per minimizzare il tempo di elaborazione;
- ➡ Semplicità di interpretazione del risultato;
- ➡ Visualizzazione dei risultati.

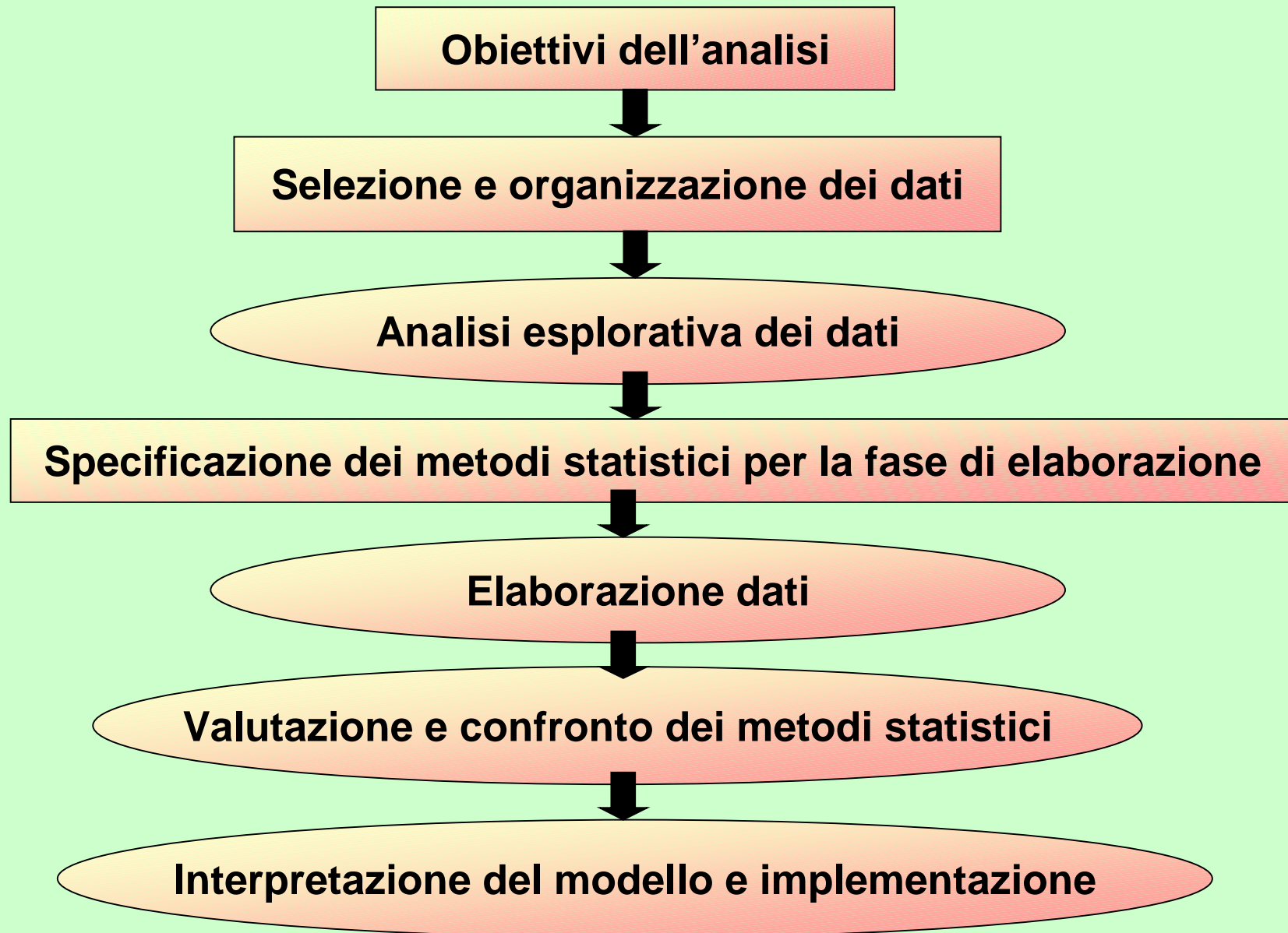


# Perché sono necessari strumenti di Data Mining

- Quantità dei dati
- Natura dei dati
- Rapida evoluzione del mercato
- Inadeguatezza degli strumenti tradizionali



# Il processo del Data Mining-kDD



# Fasi del processo di KDD

- 1. *definizione e comprensione del dominio applicativo:*** individuare le effettive problematiche di business e gli obiettivi da realizzare;
- 2. *creazione di un target data set:*** selezione di un sottoinsieme di variabili e di dati o di un campione dei dati;
- 3. *data cleaning e pre-processing:*** operazioni per attenuare il rumore nei dati, o degli outlier, selezione delle informazioni necessarie per generare il modello; decisioni sul trattamento dei campi mancanti o incompleti, dei dati rari (con un'eventuale sovra-campionatura) sulla definizione della storicità e sull'aggiornamento dei dati; aggiunta di variabili derivate e indicatori che hanno valori ricavabili da dati già esistenti.

# Fasi del processo di KDD

- 4. *data reduction e projection*:** definizione della modalità di rappresentazione dei dati secondo gli obiettivi posti, utilizzo di metodi per ridurre il numero delle variabili;
- 5. *scelta del ruolo dei sistemi di data mining per l'analisi*:** utilizzo dei sistemi di data mining per classificazione, regressione, clusterizzazione, etc.
- 6. *scelta del o degli algoritmi di data mining*:** selezione dei metodi per la ricerca di pattern, decidendo quali modelli o parametri possono essere più appropriati, integrazione dei metodi di data mining scelti con l'intero processo di scoperta della conoscenza;

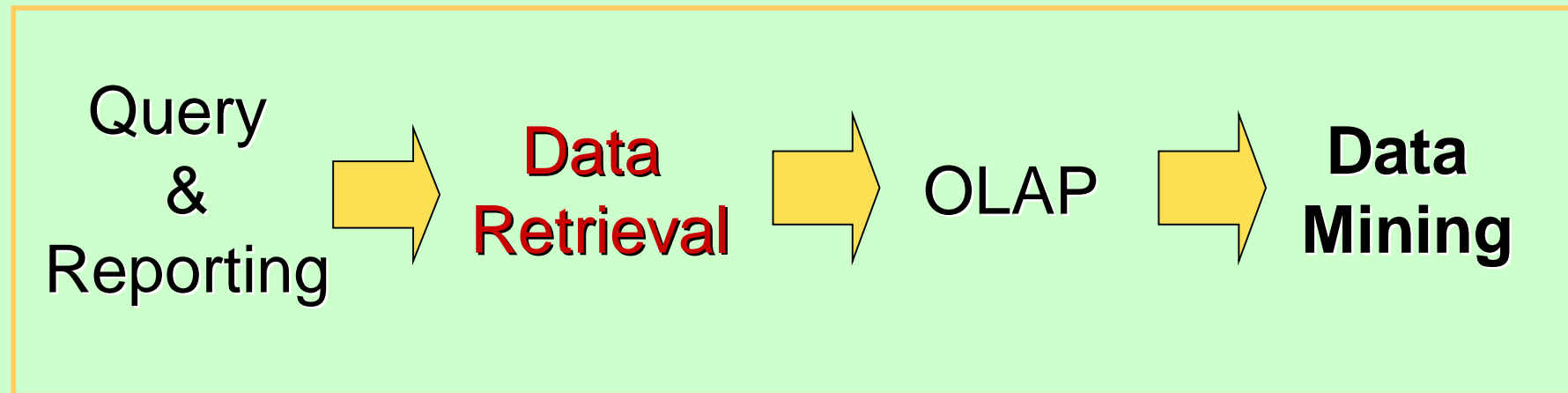
# Fasi del processo di KDD

- 7. *data mining*:** ricerca di modelli di interesse per l'utente, con raffinamenti successivi, presentati secondo definite modalità di rappresentazione (classificazione, alberi di decisione, regressione, cluster analysis...)
- 8. *interpretazione dei modelli identificati*:** analisi e verifica dei risultati con possibile retroazione ai punti precedenti per ulteriori iterazioni al fine di migliorare l'efficacia dei modelli trovati;
- 9. *consolidamento della conoscenza scoperta*:** integrazione della conoscenza e valutazione delle performance del sistema, mettendo a confronto i risultati con l'effettivo andamento nella realtà dei fatti e produzione della documentazione agli utenti finali o a terze parti interessate.

# Evoluzione degli strumenti di B.I. nell'ottica KDD

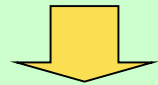
## OBIETTIVO:

Estrazione della conoscenza da un DB



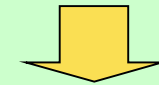
# Differenze tra Data Retrieval e Data Mining

## Data Retrieval



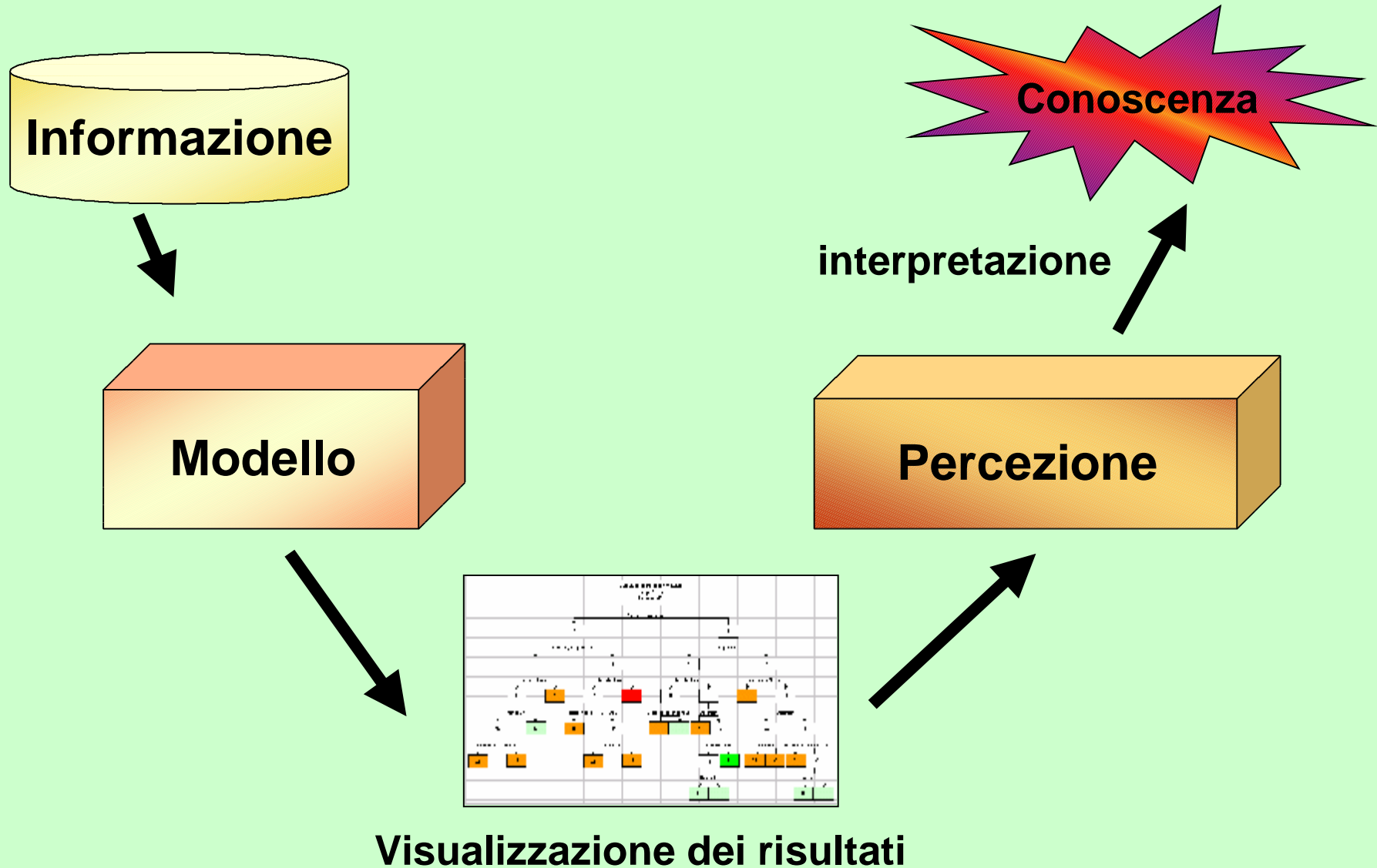
- Quanti sono i clienti che hanno età tra 30 e 50 anni e comprano Diet Coke
- Quali documenti contengono la parola "Sanità"
- Quanti brevetti ha depositato la società Colgate nel 1999

## Data Mining



- Quali sono le caratteristiche dei miei clienti
- Quali sono gli argomenti trattati da un insieme di documenti
- Quali sono i miei concorrenti e come evolvere la loro attività

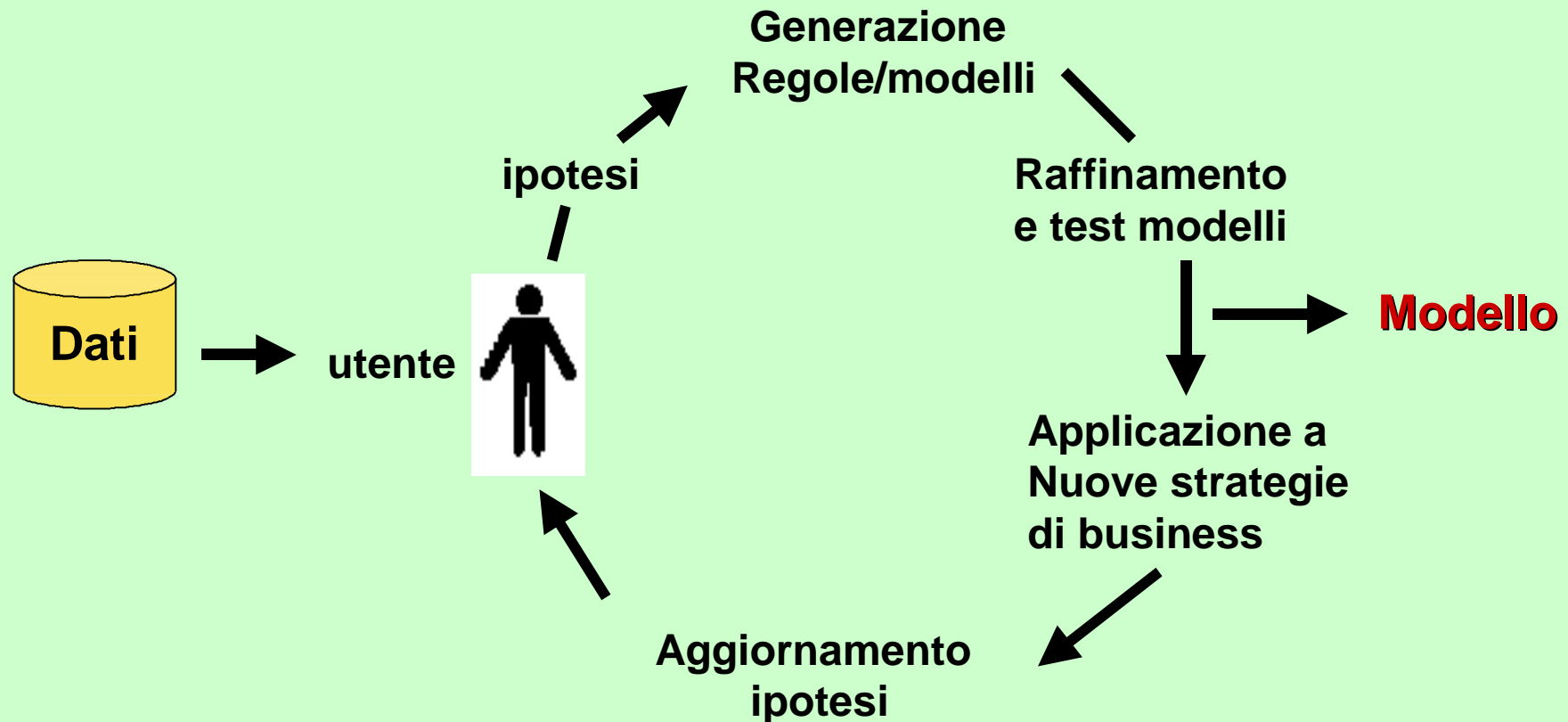
# Tecniche di Knowledge Discovery



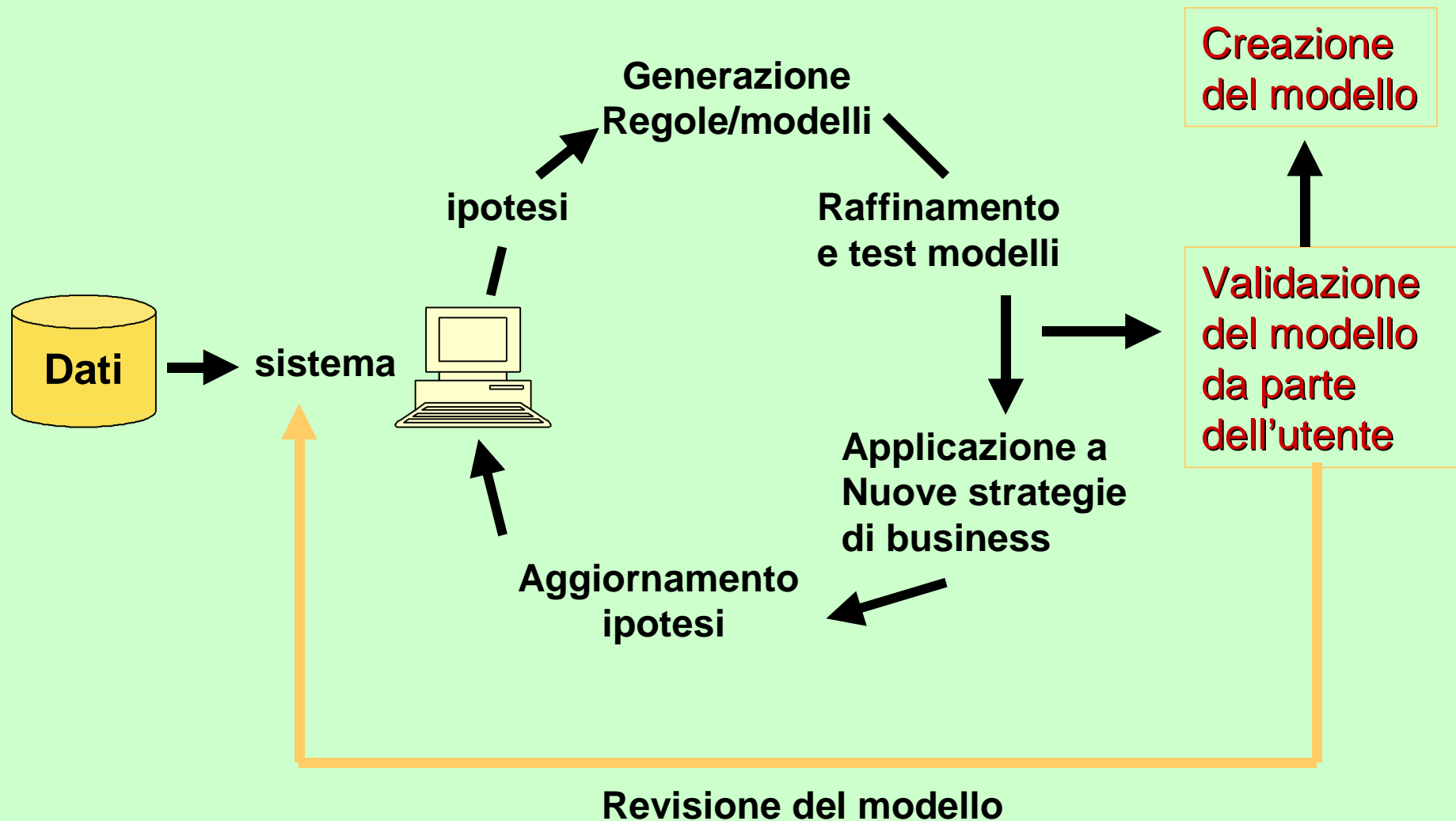


Utilizzo del DM per la definizione di modelli

# Interazione tra **utente** e sistema di Data Mining






























# Generazione del modello a iniziativa del **sistema**



# Strumenti di Data Mining

- **Clementine** di SPSS Inc.
- **Intelligent Miner** dell'IBM S.p.a.
- **Enterprise Miner** di SAS Institute Inc.
- **SuperQuery** di Azmy Thinkware Inc.
- **Tlearn** freeware
- **R System** freeware
- **Weka** freeware

# Confronti tra strumenti di DM: **Problemi**

<b>PROBLEMI</b>	Clementine	Intelligent Miner	Enterprise Miner	Super Query	Tlearn
Analisi Cluster					
Analisi testuale					
Raggruppamenti demografici		+ 			
Classificazione					
Associazioni					
Sequenze					
Previsioni					
Web mining		+ 			






























Tecnica presente



Punto di forza

# Confronti tra strumenti di DM: **Tecniche**

<b>TECNICHE</b>	Clementine	Intelligent Miner	Enterprise Miner	Super Query	Tlearn
Analisi Fattoriale					
Altre tecniche di preparazione dati	+ 		+ 		
Visualizzazione scientifica	+ 				
Reti Neurali					
Alberi Decisione					
Regole induzione				+ 	
Previsioni					
Click-stream					



Tecnica presente

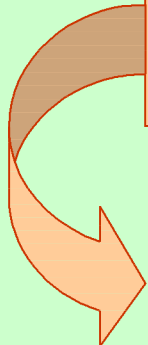


Punto di forza

# Data Mining: integrazione dell'informazione

La novità offerta dagli strumenti di Data Mining non sta nel rinnegare il tipo tradizionale di conoscenza, **ma nell'integrare** (laddove la cultura aziendale lo permette) **i processi decisionali con regole costruite sintetizzando estesi patrimoni informativi.**

**I risultati ottenuti con gli strumenti di Data Mining devono perciò essere presentati, comunicati e condivisi con le aree aziendali che ne trarranno i maggiori benefici.**



Modelli organizzativi di tipo reticolare

# Metodi di Data Mining

- ➡ **Esplorazione mediante visualizzazione multidimensionale** (scaling multidimensionale, analisi di regressione logistica, stepwise, analisi delle corrispondenze)
- ➡ **Associazione e sequenze** (usate nella market basket analysis per misurare l'affinità dei prodotti)
- ➡ **Clustering** (segmentazione della clientela in gruppi omogenei)
- ➡ **Analisi Fattoriale** (per determinare il numero di fattori da estrarre e per componenti principali)
- ➡ **Modelli previsivi**
  - di Classificazione (**Alberi di Decisione**)
  - **Reti Neurali**
- ➡ **Mappe di Kohonen** (Reti Neurali non supervisionate)
- ➡ **Algoritmi Genetici**

# Esempi di applicazioni

<b><i>Esempio</i></b>	<b><i>Tipo di problema</i></b>	<b><i>Tecnica adottabile</i></b>
Quali sono i tre principali motivi che hanno indotto il mio cliente a passare alla concorrenza?	Classificazione	Reti Neurali Decision Tree
Quali sono le fasce di clienti a cui posso offrire nuovi prodotti?	Clustering	Reti Neurali *kohon. Cluster Analysis
Quali sono le probabilità che un cliente ha aperto un c/c acquisterà anche il prodotto x in breve tempo?	Sequencing	Tecniche statistiche Rule induction
Quali sono le probabilità che un cliente acquisti due prodotti completamente differenti?	Associazione	Tecniche statistiche Rule induction
Quale sarà il prezzo del titolo tra un giorno/mese ecc?	Previsione	Reti neurali Tecniche statistiche



<b><i>Tecnica</i></b>	<b><i>Vantaggi</i></b>	<b><i>Limiti</i></b>
<b>Visualizzazione</b>	L'utente è in grado di visualizzare grandi moli di dati, scoprire relazioni e di testarle	Richiede un utente esperto in statistica e in grado di utilizzare altre tecniche di Data Mining
<b>Reti Neurali</b>	Elevata capacità elaborativa con dati in cui si nascondono relazioni non lineari. Anche con dati incompleti e <i>rumorosi</i>	Incapacità di spiegare i risultati sebbene sia possibile utilizzare altri sistemi per interpretazioni. Converta dati qualit. in quantit.
<b>Algoritmi Genetici</b>	Buona capacità previsionale usando dati in cui si nascondono relazioni non lineari	Incapacità di spiegare i risultati sebbene sia possibile utilizzare altri sistemi per interpretazioni. Converta dati qualit. in quantit.
<b>Fuzzy Logic</b>	Può classificare variabili e risultati sulla base di <i>vicinanza</i> alla soluzione	Numero limitato di fornitori e applicazioni disponibili sul mercato
<b>Decision Tree e Rule induction</b>	Creano regole e modelli sulla base di dati storici. Le regole e i modelli sono trasparenti all'utente e facilmente interpretabili.	Richiedono un <i>tuning</i> ottimale per evitare la produzione di elevati numeri di regole difficilmente interpretabili e gestibili.

# Problemi di Data Mining

- **L'Analisi per gruppi** suddivide una popolazione in sottoinsiemi disgiunti secondo definiti criteri.
- **La Classificazione** cataloga un fenomeno in una certa classe secondo un insieme di regole predeterminate.
- **Le Regole d'associazione** sono legami di causalità validi tra gli attributi delle osservazioni di un *data set*.

# APPLICAZIONI

## Score System

- Modello che mette in relazione le caratteristiche di un sistema di individui (lista potenziali clienti ad es.) con una variabile obiettivo.
- Si assegna un punteggio numerico(score) ad ogni individuo
- *Score e' collegato alla probabilita' che ogni individuo risponda all'azione commerciale*

# Applicazioni

- Credit Scoring
- Metodologia che sulla base di variabili che meglio esprimono il comportamento del pagamento del cliente ( griglia) calcola lo score che rappresenta la sua dignita' di credito
- Si utilizza per decidere o meno la concessione di un prestito o di un fido in base alla classe di rischio a cui appartiene il cliente

# Applicazioni

- *Liquidazione sinistri*: un'assicurazione analizza i sinistri denunciati per decidere quali fattori possono ridurre il tempo necessario per liquidare un sinistro
- *Rilevazione frodi*: una società che gestisce carte di credito può rilevare quali transazioni d'acquisto possono essere state effettuate con carte rubate o falsificate e bloccare quelle carte

# Applicazioni

- *ANALISI TESTUALE*: individua gruppi omogenei di documenti in termini di argomento trattato.

Consente di classificare nuovi documenti

- *Market basket analysis*: individua le associazioni per es. su dati di vendita.

quali prodotti sono acquistati congiuntamente

# Conclusioni:

- ➡ E' importante osservare che non esiste una tecnica 'superiore' alle altre, ma **ogni tecnica è riferita a determinati obiettivi e tipologie di dati da analizzare.**
- ➡ Spesso i migliori risultati per trasformare i dati in informazioni si ottengono attraverso la **combinazione di diverse tecniche di analisi.**

# DATA MINING PER CRM ANALITICO



**AZIENDA**



Sistema aperto e dinamico,  
fortemente influenzato dal  
mercato esterno e ricco di  
variabili

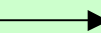
opera



**MERCATO**



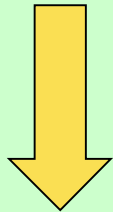
incerto



difficile da prevedere



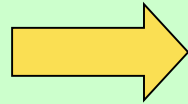
competitivo



La **SOPRAVVIVENZA** delle aziende è garantita dalla ricerca di un **VANTAGGIO COMPETITIVO** strategico in termini di **INFORMAZIONI** che permettano di attuare scelte strategiche in tempi brevi.

## RUOLO del DATA WAREHOUSE:

DATI  
AZIENDALI



D.W.H.



SISTEMI  
INFORMATIVI  
STRATEGICI

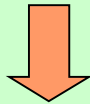
PERMETTE DI **TRASFORMARE** DATI ETEROGENEI  
IN INFORMAZIONI UTILI E PERTINENTI AL  
MODELLO DI ATTIVITA' SCELTO, NEL CONTESTO  
DI UTILIZZO DEI DATI STESSI

**CARATTERISTICHE DEI DATI**

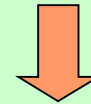
- integrati
- riutilizzabili
- non volatili

# ALGORITMI DI RICERCA e ANALISI

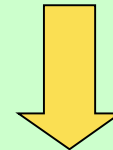
sviluppati dai Data Warehouse



OLAP



DATA MINING



ALBERI DECISIONALI  
RETI NEURALI  
MARKET BASKET ANALYSIS  
CLUSTER ANALYSIS

- 
- 
-

## DATABASE MARKETING



APPLICAZIONE DELLE TECNICHE DI DATA MINING ALLA **SEGMENTAZIONE** DELLA **CLIENTELA**, PER INDIVIDUARE GRUPPI OMOGENEI DAL PUNTO DI VISTA SOCIO – DEMOGRAFICO E DEL COMPORTAMENTO D'ACQUISTO



**SEGMENTAZIONE della CLIENTELA:**  
PROCESSO MEDIANTE IL QUALE VENGONO RAGGRUPPATI IN INSIEMI OMOGENEI, ELEMENTI CONTENUTI IN UN INSIEME ETEROGENEO

**CLIENTE (definizione MARKETING ORIENTED):**  
COMPRENDE TUTTI COLORO CON I QUALI SI FA BUSINESS, SIANO ESSI INTERNI OD ESTERNI ALL'AZIENDA

# SEGMENTAZIONE della CLIENTELA

(PARAMETRI da CONSIDERARE)

**PARAMETRI  
GEOGRAFICI**  
(stato, regione,  
città,...)

**PARAMETRI  
PSICOLOGICI**  
(classe sociale, stile di  
vita, personalità,...)

**PARAMETRI  
DEMOGRAFICI**  
(età, reddito, sesso,...)

**PARAMETRI  
COMPORAMENTALI**  
(occasioni di acquisto,  
vantaggi attesi, situazio-  
ne dell'utente, abitudini  
d'acquisto)

# CRM (customer relationship management)

## EVOLUZIONE TEMPORALE della STRUTTURA AZIENDALE

### AZIENDA ORIENTATA ALLA PRODUZIONE

Prodotte in quantità elevata per poter far fronte alla crescita rapida e diversificata della domanda;

### AZIENDA ORIENTATA VERSO LE VENDITE:

- cerca di garantire e attuare politiche tali da permettere uno sbocco sicuro ai prodotti;
- attua politiche di contenimento dei prezzi per offrire un prodotto che costi meno per battere la concorrenza;

### AZIENDA ORIENTATA AL MERCATO, SOSTENUTA da POLITICHE di MARKETING

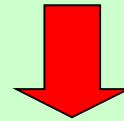
- rivolge un'attenzione particolare al rapporto **prezzo/qualità** del prodotto;
- attenzione anche nei confronti degli aspetti "qualitativi" della clientela;

**CRM** = ESTENSIONE NATURALE E PREVEDIBILE DEL DATABASE MARKETING VERSO UNA MAGGIORE ATTENZIONE NEI CONFRONTI DEL CLIENTE

# STRUTTURA del CRM

1<sup>^</sup> FASE →

Analisi approfondita degli attributi e dei comportamenti del consumatore per raggiungere una CONOSCENZA completa di lui, dei suoi desideri e delle sue necessità



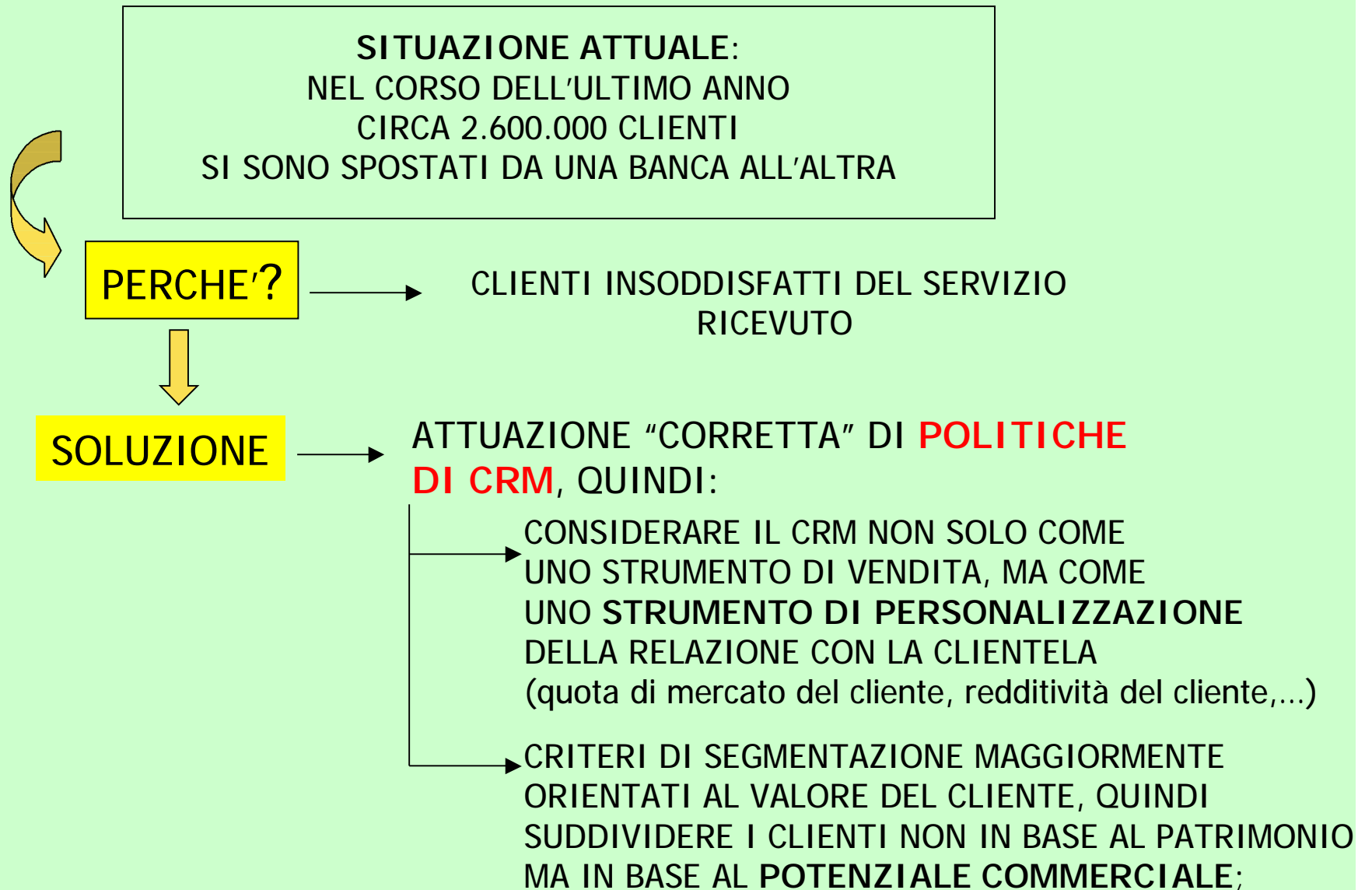
MEZZI

SISTEMI  
AUTOMATIZZATI  
DI VENDITA

CALL  
CENTER

WEB

# BANCHE, QUALE FUTURO?





**INIZIATIVE  
del  
DATABASE  
MARKETING**

**CARTE FEDELTA'** a CARTE UTILIZZATE DAI CLIENTI  
PER OTTENERE VANTAGGI NELLO  
ACQUISTO IN ALCUNI NEGOZI

**CUSTOMER SATISFACTION**

a PROMUOVERE INDAGINI SUCCESSIVE PER CAPIRE  
IL LIVELLO DI SODDISFAZIONE DEI CLIENTI NEI  
CONFRONTI DEI SERVIZI E DEI PRODOTTI CHE SONO  
STATI VENDUTI LORO

**CROSS SELLING** a VENDITA ALLO STESSO CLIENTE DI  
PRODOTTI APPARTENENTI A DIVERSE  
FAMIGLIE/PRODUCT GROUP

# CROSS-SELLING

## → APPROCCIO TRADIZIONALE "INGANNEVOLE" O "PRODOTTO-CENTRICO"

Es: LA X-STORE DECISE DI ANALIZZARE LA REDDITIVITA' DEL PORTAFOGLIO PRODOTTI E DI ELIMINARE TUTTI QUELLI NON PROFITTEVOLI.  
DOPO QUALCHE TEMPO ALTRI NEGOZI NACQUERO VICINO ALL X-STORE E GUADAGNORONO QUOTE DI MERCATO.  
DI CONSEGUENZA, ALTRI PRODOTTI DIVENNERO NON PROFITTEVOLI PER X-STORE CHE LI HA ELIMINATI.  
X-STORE, PROCEDENDO IN QUESTO MODO, DOPO QUALCHE TEMPO E' FALLITA!



SI EVIDENZIA, QUINDI, COME I CLIENTI NON ACQUISTINO SOLO UN PRODOTTO ALLA VOLTA, MA COME ACQUISTINO PIU' PRODOTTI NELLO STESSO PROCESSO D'ACQUISTO. L'ELIMINAZIONE DI ALCUNI PRODOTTI PUO' COMPORTARE L'INSODDISFAZIONE DEL CLIENTE E OPPORTUNITA' DI ULTERIORE BUSINESS PER I CONCORRENTI.

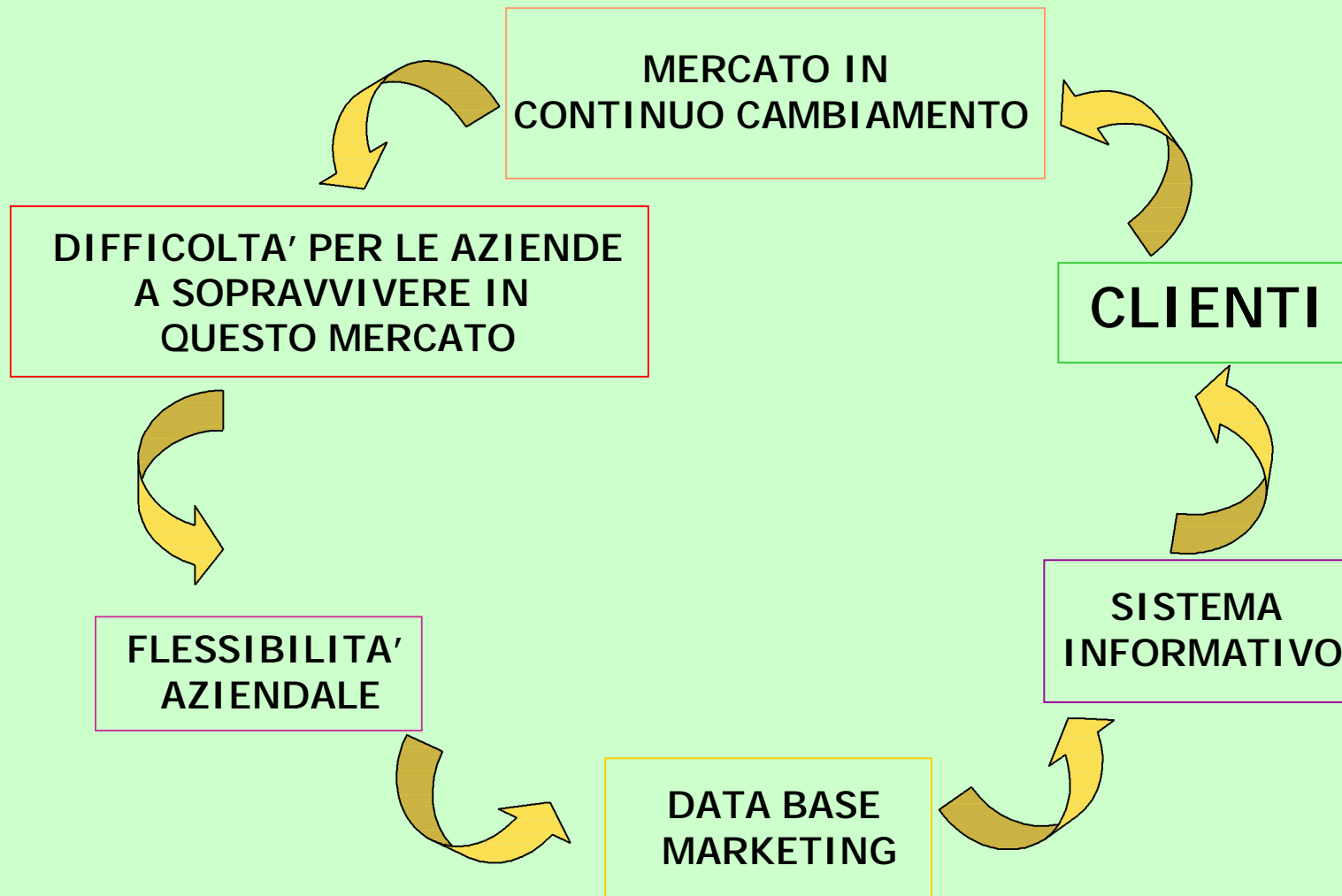
## → CROSS SELLING COME ANALISI

PREVEDE:

- DI CODIFICARE IN MODO OPPORTUNO CLIENTI E PRODOTTI;
- DI CURARE LA RACCOLTA E LA GESTIONE DEI DATI RELATIVI ALLE SINGOLE TRANSAZIONI;
- DI ANALIZZARE REGOLARMENTE I RISULTATI CONSEGUITI E LE CAUSE;
- DI PIANIFICARE LE ATTIVITA' SULLA BASE DELLE ANALISI.

## → CROSS-SELLING COME STRATEGIA

CARATTERIZZATO DA UNA VISIONE DEL BUSINESS ORIENTATA PREVALENTEMENTE AL CLIENTE. CONSISTE NEL CONSIGLIARE AL CLIENTE CHE HA ACQUISTATO UN PARTICOLARE PRODOTTO, L'ACQUISTO ANCHE DI PRODOTTI COMPLEMENTARI.



# **Dall'inductive learning al Data Mining**



# L'inductive Learning

---

- ➡ **L'inductive learning** è un processo che permette di costruire modelli di dati a partire dai dati ricavati da un database.
- ➡ Oggetti con caratteristiche simili vengono raggruppati in classi e regole attraverso le quali è possibile **prevedere a quale classe apparterrà un nuovo oggetto**.
- ➡ Il database sul quale viene applicato questo processo è un ambiente dinamico e di conseguenza il modello d'induzione deve essere adattativo, deve cioè essere in grado di imparare.

# Tipologie di apprendimento

---

Per costruire i modelli sulla base dei dati si possono differenziare due tipologie di apprendimento che differenziano le tecniche di Data Mining.

Ø **Data Mining *supervisionato***, fondato su algoritmi predittivi.

Ø **Data Mining *non supervisionato***, che sfrutta algoritmi di tipo descrittivo

# Data Mining *supervisionato*

---

E' un approccio **top down**, applicabile quando è chiaro l'obiettivo da prevedere, che genera previsioni, stime, caratterizzazioni rispetto al comportamento di alcune variabili target, individuate in funzione di variabili di input.

Nei modelli previsionali l'obiettivo è quello di apprendere in modo che la conoscenza acquisita sia applicabile anche in futuro, quindi **il modello migliore non è solo quello che presenta migliore efficacia (lift) ma quello meglio performante con i dati futuri.**

# Data Mining *supervisionato*

---

Confrontare i risultati del modello con il reale andamento della realtà risulta un passo necessario soprattutto per fare previsioni sui comportamenti futuri.

Le fasi della costruzione e dell'applicazione di un modello previsionale sono:

- ➡ La costruzione di un ***set di apprendimento (training)***
- ➡ La costruzione di un ***set di valutazione (evaluation)***
- ➡ La costruzione di un ***set di test***
- ➡ Applicazione del modello al ***set di calcolo (score)***



**Obiettivo:** Classificare alcuni clienti di un'azienda secondo due classi A e B

**Costruzione del Training set:**

(Ottenuto utilizzando la tavola dei numeri casuali e criteri di buon senso)

- **Variabile obiettivo** del Training set (DIP):  
DIP= 0 se il cliente  $\in$  A  
DIP= 1 se il cliente  $\in$  B
- **Variabili esplicative** del Training set ( $X_1, X_2, \dots, X_{14}$ ):  
 $X_1$ = SESSO (Maschile=0, Femminile=1)  
 $X_2$ = ETA' (< 10 anni=0, 10-20 anni=0.1, 20-30 anni=0.2, ..., >100 anni=1)

**Costruzione del Valid set:**

$X_1, X_2, X_3, \dots, X_{14} \Rightarrow Y$

Si validano i risultati ottenuti nel Training

# Costruzione di un set di apprendimento: *set di training*

---

- ➡ Il modello viene sperimentato usando dati preclassificati **da un sottoinsieme del set di costruzione** (campione dell'universo di riferimento), dove gli algoritmi trovano pattern di valore previsionale.
- ➡ In pratica il modello quando esercita su tale set **impara a riconoscere le strutture di questo set.**

# Costruzione di un set di valutazione: *set di validation*

---

- ➡ Per verificare il rendimento del modello o per paragonare i risultati di modelli diversi, viene usato un set distinto dagli altri due.
- ➡ Il set di prova consente al modello di generalizzare meglio, **evitando il sovradattamento ai dati**, e di stimarne le prestazioni.

# Costruzione di un set di test: *testing*

---

- ➡ Viene rifinito il modello attraverso un altro sottoinsieme del set di costruzione, **per evitare che il modello memorizzi il set di training** garantendo che sia più generale e che funzioni meglio con i dati sconosciuti.

# Applicazione del modello al *set di calcolo: set di score*

---

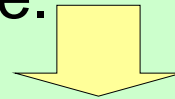
- ➡ Set **non preclassificato** e non appartenente al set di costruzione.
- ➡ Non si conoscono i risultati corretti per tali dati e quindi **verranno usati i calcoli previsionali per prendere decisioni su di essi.**
- ➡ La fase di scoring è importante per **ottimizzare i risultati** di un'operazione commerciale.

# Applicazione del modello al *set di calcolo: set di score*

---

Lo scoring system è un insieme di procedure per costruire un modello predittivo in cui l'**obiettivo si basa sull'individuazione di una relazione tra una serie di variabili predittive ed una variabile target.**

E' un approccio di analisi incentrato **sull'assegnazione ai clienti della probabilità di adesione ad una campagna di marketing**, in modo da classificarli per applicare azioni diversificate.

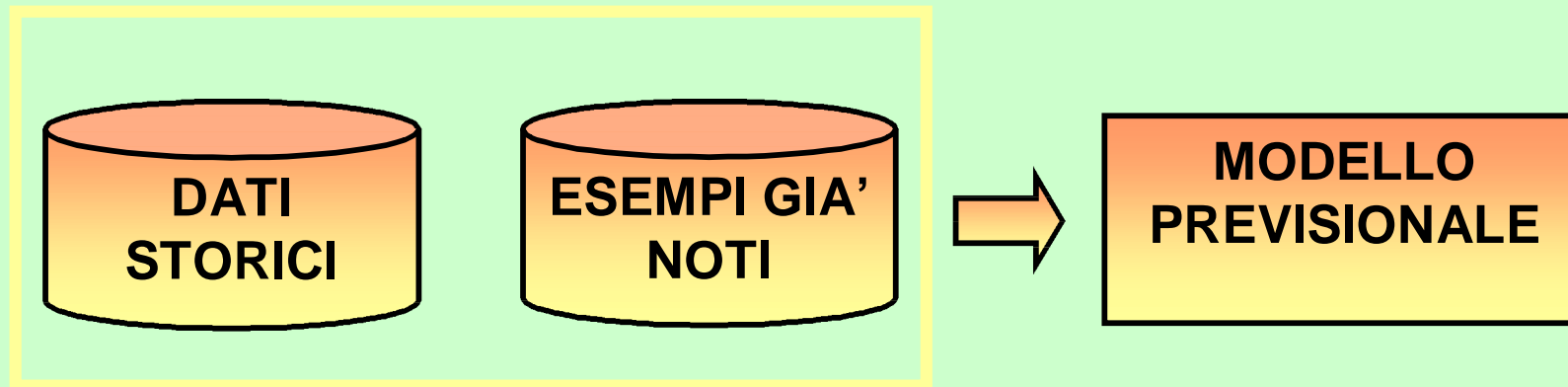


I modelli previsionali devono quindi essere addestrati prima di venir usati e creati prima di esser valutati.

# Costruzione dei modelli previsionali

---

L'addestramento di un modello previsionale consiste in pratica in **un processo di creazione di un modello con dati storici ed esempi già noti** di quello che si intende prevedere.



# Costruzione dei modelli previsionali

---

- ➡ L'utilizzo di modelli previsionali presuppone che il **passato sia un buon predittore per il futuro.**
- ➡ Viene applicato il modello migliore ad un set di calcolo per effettuare previsioni.
- ➡ ***Viene quindi inserita una sorta di probabilità di risposta all'azione commerciale, una probabilità condizionata ai dati raccolti ed elaborati che costituisce la confidence.***





# Data Mining *non supervisionato*

---

E' un approccio **bottom up** in cui si lascia che i dati stessi indichino un risultato, dove **non esiste una variabile target** usata per la descrizione e l'individuazione di segmenti.

Tale approccio viene spesso **applicato nella fase esplorativa** per cogliere nelle strutture decisionali un pattern interessante.