

Aspetti Computazionali e Data Mining

- **Aspetti computazionali e gestione / mining dei dati:**
- **Data Mining,**
- **Web Mining,**
- **Spatial Data Mining,**
- **Text Mining .**

Introduzione al DM

- Il processo di *data mining*
 - Dai dati alle informazioni
 - Il *Data Warehouse*
 - Che cos'è il *Data Mining*

Dai dati alle informazioni

- **L'informazione è la chiave che permette ad un'azienda di essere sempre più competitiva e di guadagnare un mercato sempre più ampio.**
- **-le nuove strategie di mercato,**
- **-l'internazionalizzazione,**
- **-la crescita esponenziale delle informazioni connesse all'osservazione a monte dei fenomeni esterni ed interni all'impresa,**

Dai dati alle informazioni

hanno favorito una complicazione d'interpretazione di quei processi legati a eventi in grado di influenzare il buon andamento dell'impresa.

- Il mercato, di conseguenza, privilegia le imprese in grado di reagire in tempi brevi alle proprie richieste.**
- Le imprese si trovano quindi obbligate a migliorare i processi collegati**
 - -al time-to-market,**
 - -all'immagine aziendale,**
 - -alla soddisfazione delle esigenze dei clienti,**
 - - sostenibilità.**

Dai dati alle informazioni

- In questo contesto il ruolo dell'*Information Technology* è notevolmente cambiato
- -da un lato cerca ancora di fare le cose al minimo costo compatibile con la qualità richiesta,
- -dall'altro diventa strumento essenziale per offrire un migliore servizio alla clientela e promuovere lo sviluppo (proporre nuovi prodotti, nuovi servizi e più in generale nuove attività di business).

Dai dati alle informazioni

- **Prendendo in considerazione le modalità operative delle migliori organizzazioni imprenditoriali tra gli elementi che contribuiscono al successo vi è la maggiore sensibilità verso la conoscenza estesa e approfondita delle informazioni possedute**
- **L'obiettivo è "analizzare le informazioni per ricavare altre informazioni" che costituiscano una base di conoscenze utili per le decisioni strategiche ed operative.**

Dai dati alle informazioni

PROBLEMA:

**come gestire la grande quantità di dati “grezzi”
provenienti dalle fonti più svariate per renderle
accessibili e leggibili da chi in azienda deve
prendere decisioni.**

Il processo di *Knowledge Discovery*

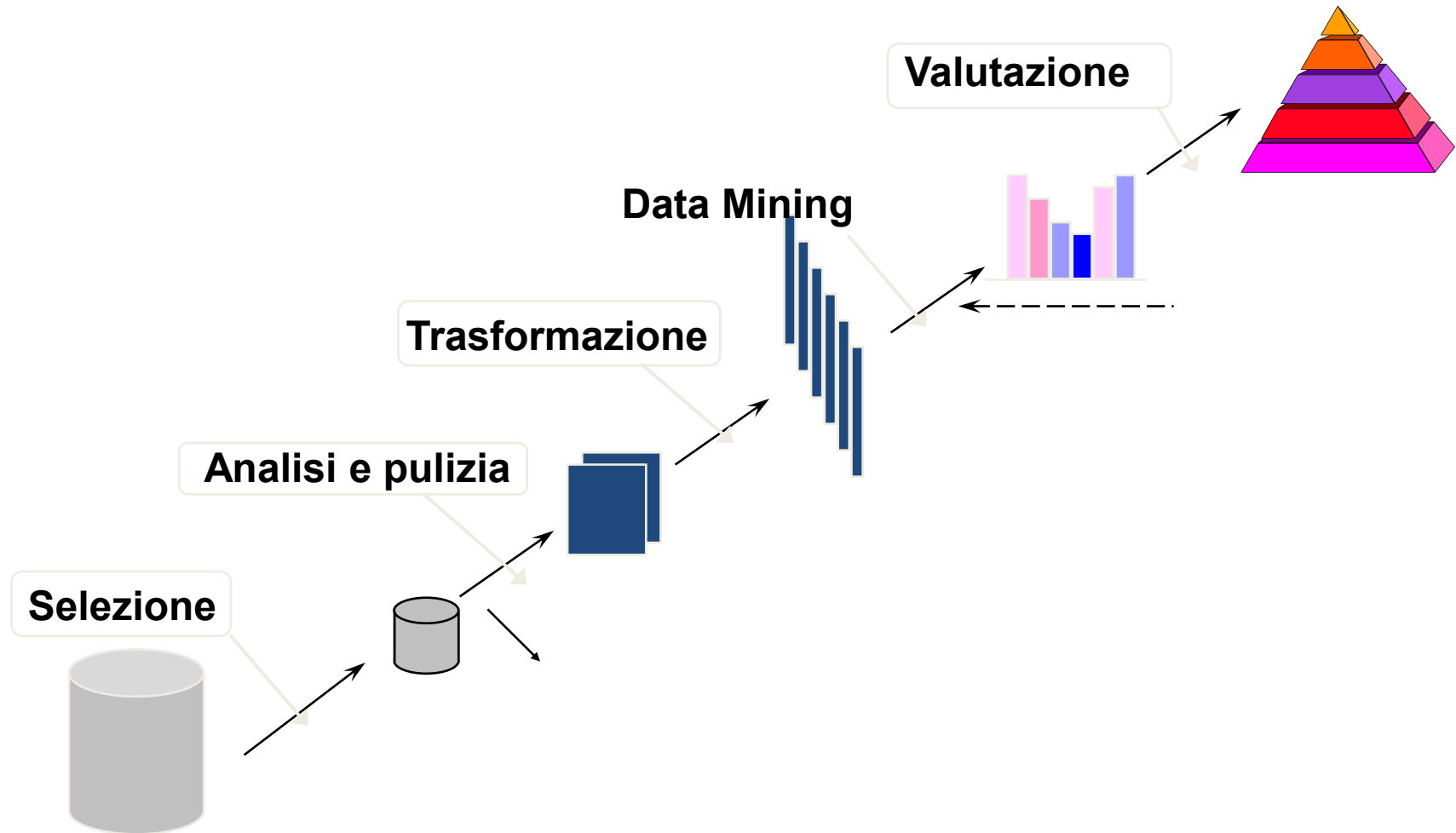
Il processo di *Knowledge Discovery* è una delle possibili soluzioni.

- Esso consiste nella collezione dei dati da fonti diverse, cui segue la loro gestione ed infine la trasformazione da dati “grezzi” in informazione**

Il processo KDD

- **individuazione del dominio di conoscenza**
- **selezione ed estrazione dei dati dagli archivi**
- **pulizia dei dati**
- **trasferimento su “Data Warehouse”** 
- **trasformazioni e aggregazioni dei dati**
- **accesso ai dati per gli analisti**
- **Data Mining**
- **presentazione dei risultati**

Il processo di estrazione di conoscenza



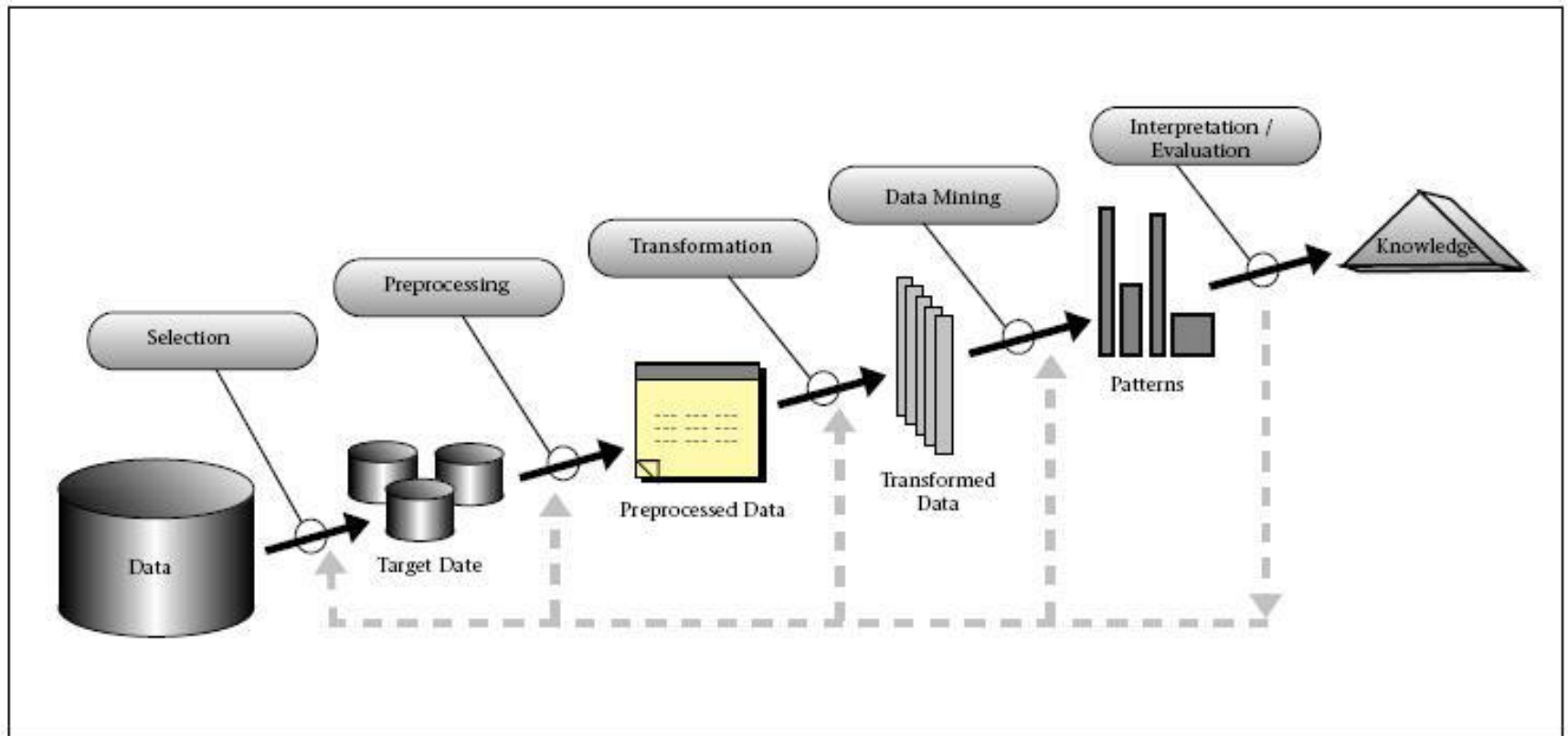


Figure 1. An Overview of the Steps That Compose the KDD Process.

Fayyad U., Piatetsky-Shapiro G., Padhraic S. (1996),
From Data Mining to Knowledge Discovery in Databases.
AI Magazine 17(3), 37-54.

<http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>

Fasi che portano dall'insieme dei dati grezzi all'estrazione della conoscenza

- ***Selezione***: selezione o segmentazione dei dati secondo alcuni criteri.
- ***Processing***: “pulizia” dei dati.
- ***Trasformazione***: i dati non sono semplicemente trasferiti da un archivio ad uno nuovo, ma sono trasformati in modo tale che sia possibile aggiungere informazione.

Fasi che portano dall'insieme dei dati grezzi all'estrazione della conoscenza

- ***Data Mining***: questo processo si occupa dell'estrazione dei modelli dai dati.
- ***Interpretazione e valutazione***: i modelli identificati dal sistema vengono interpretati in modo che la conoscenza acquisita possa essere di supporto alle decisioni.

Nascita e valore dell'informazione

- **Per informazione si intende un insieme di dati elementari, aggregati e integrati secondo specifiche regole che determinano un elemento di base per il patrimonio informativo aziendale.**
- **È rilevante per il manager allo scopo di attuare un modello di Business Intelligence.**
- **Business intelligence è il processo, e la tecnologia alla base, che permette la trasformazione dei dati in informazioni, informazioni in conoscenza. La conoscenza aiuta il processo decisionale ai vari livelli dell'organizzazione.**
- **B.I. combina business analytics, data mining, visualizzazione dei dati di modo che le diverse organizzazioni possano prendere decisioni basate sull'analisi dei dati.**

Differenza tra Business Intelligence e sistemi operazionali

- **Fare Business Intelligence in azienda vuol dire cercare, trovare, organizzare, interpretare informazioni che consentono di sfruttare, anticipare, controllare processi e fenomeni di mercato o interni all'azienda stessa, con l'obiettivo di accrescere il valore, ponendosi in una posizione di vantaggio nei confronti dei competitors.**
- **È impensabile fare BI affidandosi ai tradizionali database utilizzando sistemi operazionali, poiché in questo caso la struttura dei dati è orientata solamente ad ottimizzare le classiche operazioni necessarie alla gestione dell'azienda come l'inserimento di un ordine o di una fattura, il carico-scarico dei magazzini etc.**
- **Per fare tali operazioni ai giorni nostri vi è la necessità di utilizzare un Data Warehouse che a differenza di un sistema operativo rende disponibile i dati a supporto dei processi decisionali di più alto livello d'organizzazione.**

Data Warehouse

Il Data Warehouse (DW) è una piattaforma sulla quale vengono archiviati e gestiti dati provenienti dalle diverse aree dell'organizzazione.



Data Warehouse

Il rapido sviluppo dei sistemi informativi ha reso possibile l'immagazzinamento di dati di notevoli dimensioni (gigabyte, terabyte, etc.) in modo sempre più efficiente,

cosicché un obiettivo aziendale fondamentale è la costruzione di un *Data Warehouse* anello di congiunzione tra i sistemi informazionali gestionali e le applicazioni cosiddette *business-critical*, cioè funzionali al perseguimento della missione aziendale.

Caratteristiche del Data Warehouse

Integrato	Raccoglie i dati di tutte le fonti utili interne/esterne
Orientata al soggetto	Soggetti: clienti, prodotti, servizi
Variabile nel tempo	Orizzonte dei dati storico, più anni del gestionale
Non volatile	Dati non modificabili, prevista solo consultazione

Caratteristiche del Data Warehouse

Il *Data Warehouse* è un insieme di dati estratto dai database aziendali e gestito secondo alcuni criteri:

- ***Orientamento decisionale:*** cambiamento dalla prospettiva di dati per le applicazioni (dati per supportare il processo produttivo) a quella di dati per il management (dati per supportare il processo decisionale). Si focalizza sulla natura di base del contesto di business, indipendentemente dai processi informativi interni.

Caratteristiche del Data Warehouse

- ***Integrazione:*** il database consolida i dati provenienti dai differenti sistemi presenti in azienda che utilizzano diversi sistemi di codifica, unità di misura o altro, ed elimina le inconsistenze tra i dati.
- ***Gestione storica dei dati:*** mentre i dati operazionali danno una visione sincronica e sono validi nel momento in cui si leggono, i dati in un *Data Warehouse* tengono conto della variabile tempo, e possono essere utilizzati per analizzare la "storia" degli eventi.

Caratteristiche del Data Warehouse

- ***Non volatilità:*** i dati nuovi vengono aggiunti e non vanno a rimpiazzare quelli vecchi, preservando la temporalità e accrescendo il contenuto informativo.

Data Warehouse dell'ISTAT: I.Stat

I.Stat: il data warehouse dell'Istat

Offre un patrimonio informativo statistico completo e omogeneo. I dati sono organizzati per temi, presentati in tavole multidimensionali e corredati di un ampio apparato di metadati

Data Warehouse dell'ISTAT: I.Stat

Benvenuto in I.Stat

Cosa è I.Stat

I.Stat è la banca dati, sempre aggiornata, delle statistiche correntemente prodotte dall'Istituto nazionale di statistica.

Le statistiche sono ricercabili per tema. Il sistema è interrogabile anche per parola chiave.

I dati sono presentati sotto forma di tavole multidimensionali che gli utenti possono esportare in formato xls, csv. È inoltre possibile creare tabelle personalizzate agendo sulle variabili, il periodo di riferimento e la disposizione di testate e fiancate.

Attraverso un web service che consente l'interrogazione diretta machine-to-machine, enti e organizzazioni ma anche privati cittadini, possono formulare specifiche query sui dati, effettuare il download dei risultati. Per usufruire del servizio:

<http://www.istat.it/it/strumenti/web-service>

<https://www.istat.it/it/metodi-e-strumenti/web-service-sdmx>

Data Warehouse dell'ISTAT: I.Stat

Altre banche dati derivanti da I.Stat

Per alcuni argomenti di particolare interesse, l'Istat ha riorganizzato le informazioni già presenti nella banca dati I.Stat all'interno dei seguenti sistemi tematici:

- [Giovani.Stat](#) Vedi anche il Quadro informativo [#Giovani](#)
- [Anziani.Stat](#) Vedi anche il Quadro informativo [#Anziani](#)
- [Immigrati.Stat](#) Vedi anche il Quadro informativo [#Immigrati e nuovi cittadini](#)
- [Congiuntura.Stat](#) Vedi anche il Quadro informativo [#Congiuntura economica](#). Per approfondimenti sulle politiche di revisione accedi a [#Revisioni](#)
- [PubblicaAmministrazione.Stat](#)
- [CapitaleUmano.Stat](#): statistiche e indicatori su istruzione, formazione, mercato del lavoro e reddito, integrate con informazioni derivanti da altri enti (Invalsi, Unioncamere, Eurostat, Ocse).

Banche dati sui censimenti (a partire dalla tornata 2010-11)

- [Agricoltura](#)
- [Popolazione e abitazioni](#)
- [Industria, istituzioni pubbliche e non profit](#)

Cited (5/3/25)

← → ↻

Non sicuro dati.istat.it

☆

9

SAS OnDemand for...

Posta - SCHIOIER GA...

Avvia riunione - Zo...

Gestionali | Universi...

U-Web Cedolini

U-Web Missioni

BiblioUniTS | Accedi...

SAS Documentation...

>>

Tutti i

 **I.Stat** | il tuo accesso diretto alla statistica italiana

[Clicca qui per il login](#) | [FAQs e Contatti](#) | [Manuale utenti](#)
English

Per

Esplora Temi

>> **Annulla**

Tutti i temi

Censimento agricoltura 2010

Censimenti permanenti della popolazione e delle abitazioni, imprese, istituzioni pubbliche e nonprofit

Censimento popolazione e abitazioni 2011

Ambiente ed energia

Caratteristiche del territorio

Popolazione e famiglie

Condizioni economiche delle famiglie e disuguaglianze

Salute e sanità

Assistenza e previdenza

Istruzione e formazione

Cultura, comunicazione, viaggi

Giustizia e sicurezza

Vita quotidiana e opinione dei cittadini

Conti nazionali

Imprese

Agricoltura

Benvenuto in I.Stat, la banca dati completa per gli esperti

La nuova banca dati IstatData

Il nuovo sistema di accesso ai dati che sostituirà progressivamente I.Stat e sarà arricchito di grafici, mappe e dashboard interattive. I dati presenti potranno essere interrogati direttamente machine-to-machine tramite web service

Per accedere direttamente ai dati più richiesti

Si può consultare la banca dati [StatBase](#)

Come si usa I.Stat

I dati sono presentati sotto forma di tabelle. Gli utenti possono:

- modificarle
- esportarle
- interrogare direttamente i dati machine-to-machine tramite [web service](#)

Per approfondimenti sui termini statistici utilizzati si può consultare il [Glossario](#)

Altre banche dati derivanti da I.Stat

- [Giovani.Stat](#)
- [Anziani.Stat](#)
- [Immigrati.Stat](#)
- [Congiuntura.Stat](#)
- [PubblicaAmministrazione.Stat](#)
- [CapitaleUmano.Stat](#)
- [ViolenzaSulleDonne.Stat](#)

Cited (6/3/25)

Statistiche Smart da Big Data

L'intensa e continua generazione di Big Data su tutti gli aspetti della vita, attraverso l'uso costante di dispositivi intelligenti (Smart), strumenti digitali e reti elettroniche/social network, sta caratterizzando la realtà sociale, ambientale ed economica e l'interazione tra persone, organizzazioni e oggetti.

Artefice e protagonista della quarta rivoluzione industriale, Internet delle Cose (Internet of Things – IoT) applicato alla vita quotidiana o ai processi delle imprese sta trasformando le operazioni di produzione di dati e sta facendo crescere l'integrazione tra i sistemi intelligenti.

E' quindi prevedibile che nel prossimo futuro, la maggior parte dei dati scaturirà direttamente dalle persone, con lo svolgimento quotidiano di attività, l'uso di sistemi e di oggetti personali, tra cui gli innumerevoli dispositivi intelligenti in grado di connettersi a Internet.

In tale contesto, l'Istat è chiamato a svolgere un ruolo chiave nella trasformazione in statistiche ufficiali di questa mole di dati, producendo le cosiddette statistiche smart affidabili (Trusted Smart Statistics). Per questa ragione l'Istat è impegnato nell'innovazione attraverso gli sviluppi metodologici e l'uso di nuove fonti di dati per una più efficiente e tempestiva produzione di statistiche ufficiali, per rispondere meglio alle esigenze degli utenti e ridurre l'onere statistico su famiglie, individui e imprese. Tali statistiche sono prodotte in conformità agli standard adottati a livello nazionale e internazionale, con particolare riguardo agli aspetti relativi alla protezione della privacy delle persone a cui i dati si riferiscono e agli aspetti di qualità.

+ Cosa sono e a che servono

+ Come cambiano il processo di produzione

+ Aspetti relativi alla privacy

.

Data Mart

- È definito come un sistema che raccoglie i dati necessari ad un certo settore di utenti e di applicazioni. Si specializza su una singola area funzionale aziendale e contiene dati di sintesi necessari per eseguire delle specifiche funzioni.
- Si indentifica in un sotto-insieme di dati aziendali che possono essere estratti da un database operativo o da un DW.
- Contenendo un numero più limitato di dati rispetto al DW è molto più veloce ed economico, per questo motivo viene usato molto dalle piccole imprese.



Definizioni OLTP - OLAP

On line transaction processing (OLTP) è l'insieme dei sistemi transazionali pensati e ottimizzati per garantire la massima sicurezza nella gestione delle transazioni.

On line Analytical processing (OLAP) è l'insieme dei sistemi di analisi dei dati pensati e ottimizzati per garantire la massima Performance sui tempi di esecuzione di *queries* su database relazionali e la massima “estensione” delle interrogazioni

On-line Transactional processing (OLTP)

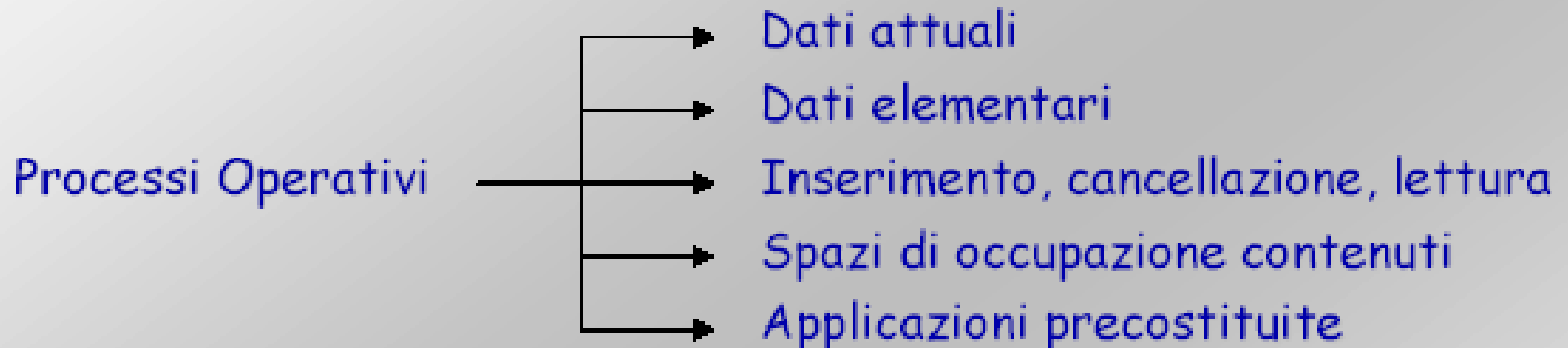
- È un complesso di sistemi caratterizzato da transazioni on-line (insert, update, delete) che forniscono i dati di origine al DW e garantiscono la massima sicurezza nella gestione delle transazioni giornaliere, mantenendo un'efficacia misurata dal numero di transazioni al secondo.
- Tali sistemi sono elaborati da query molto veloci e spesso utilizzate per registrare, modificare e visualizzare singoli record in tempo reale.

On-line Analytical processing (OLAP)

- È un complesso di tecniche e strumenti di analisi di un grosso quantitativo di dati che permettono di effettuare avanzate indagini on-line sui dati del DW per il supporto alle decisioni di tipo manageriale.
- Questo modello è studiato tenendo conto di due priorità fondamentali:
 - La semplicità di analisi da parte dell'utente
 - La velocità nell'estrazione dei dati

Processi operativi e Processi decisionali

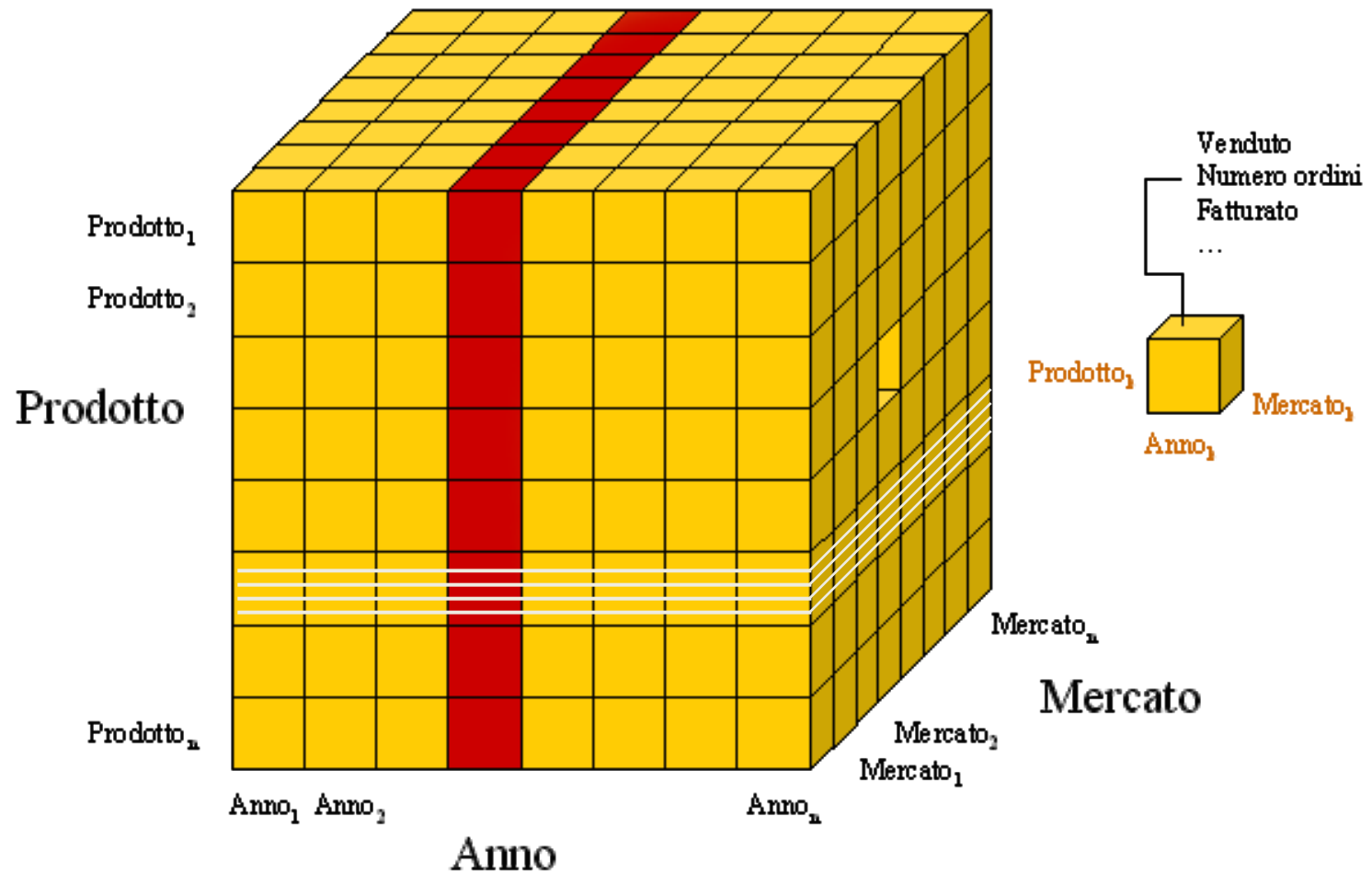
OLTP: On Line Transactional Processing



OLAP: On Line Analytical Processing



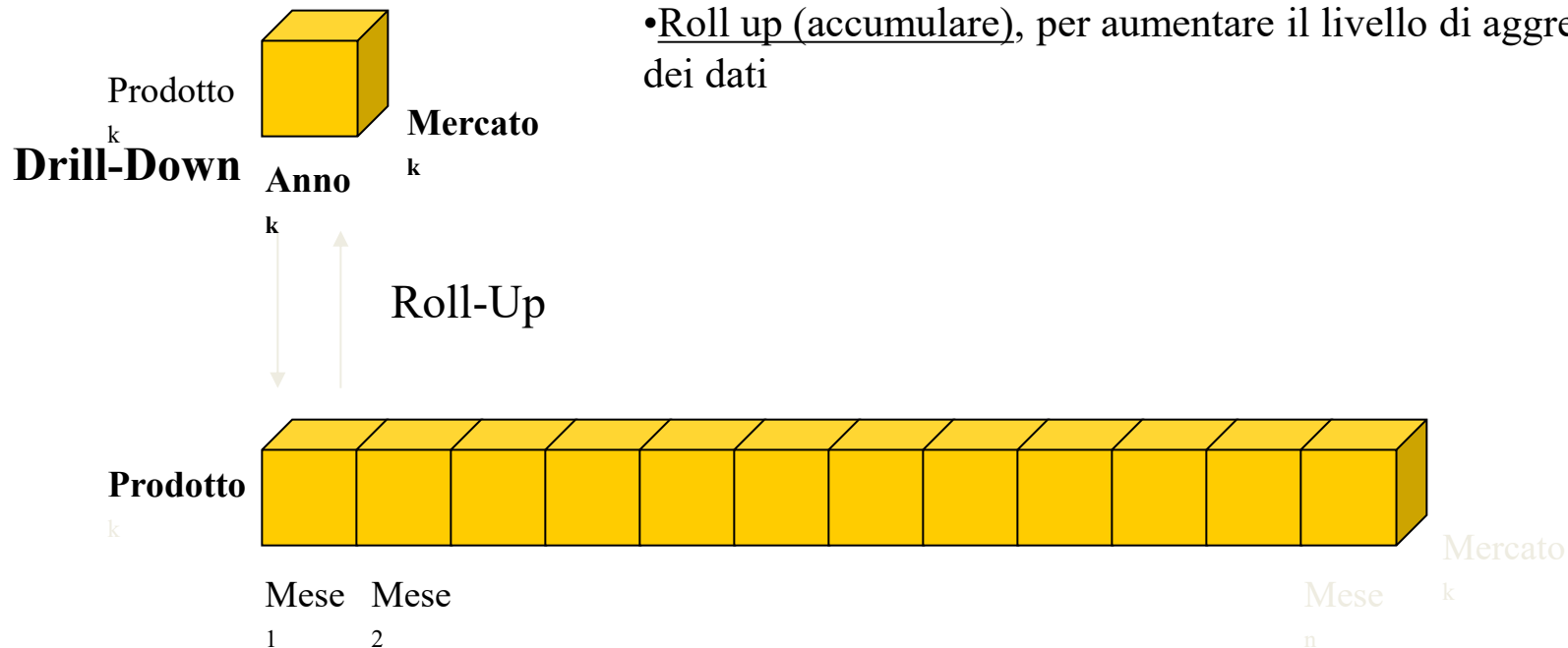
Il cubo OLAP

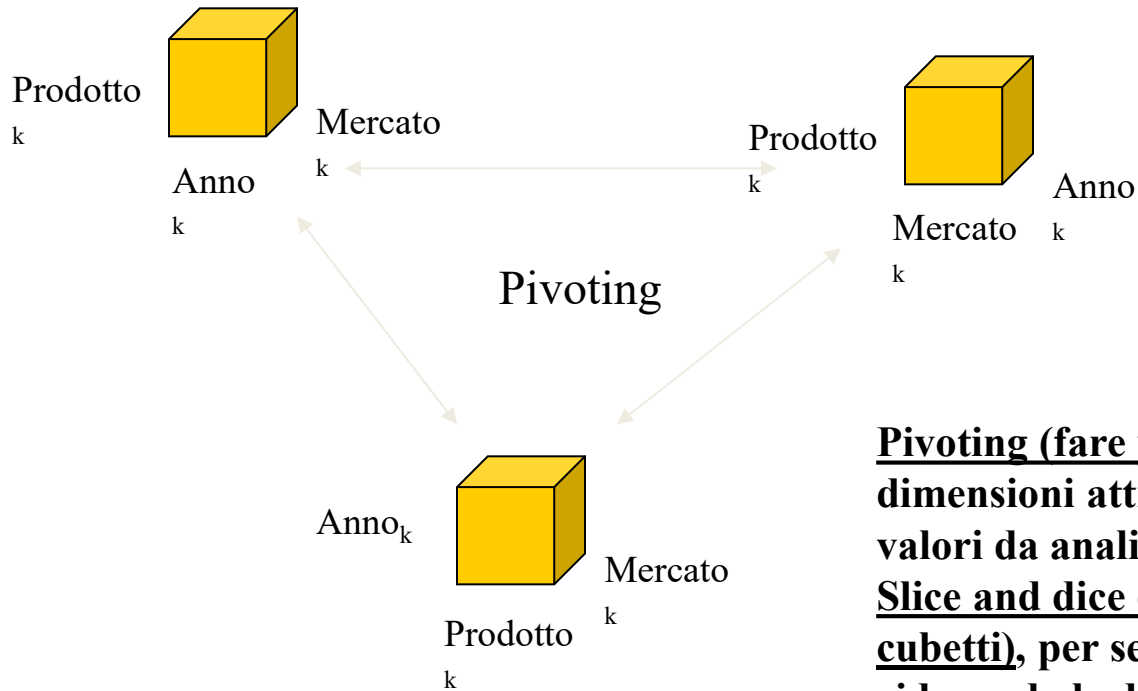


Operazioni sul cubo OLAP

Le tipiche operazioni (OLAP) che si eseguono sul Data Cube sono:

- Drill down (perforare), per aumentare il livello di dettaglio dei dati
- Roll up (accumulare), per aumentare il livello di aggregazione dei dati



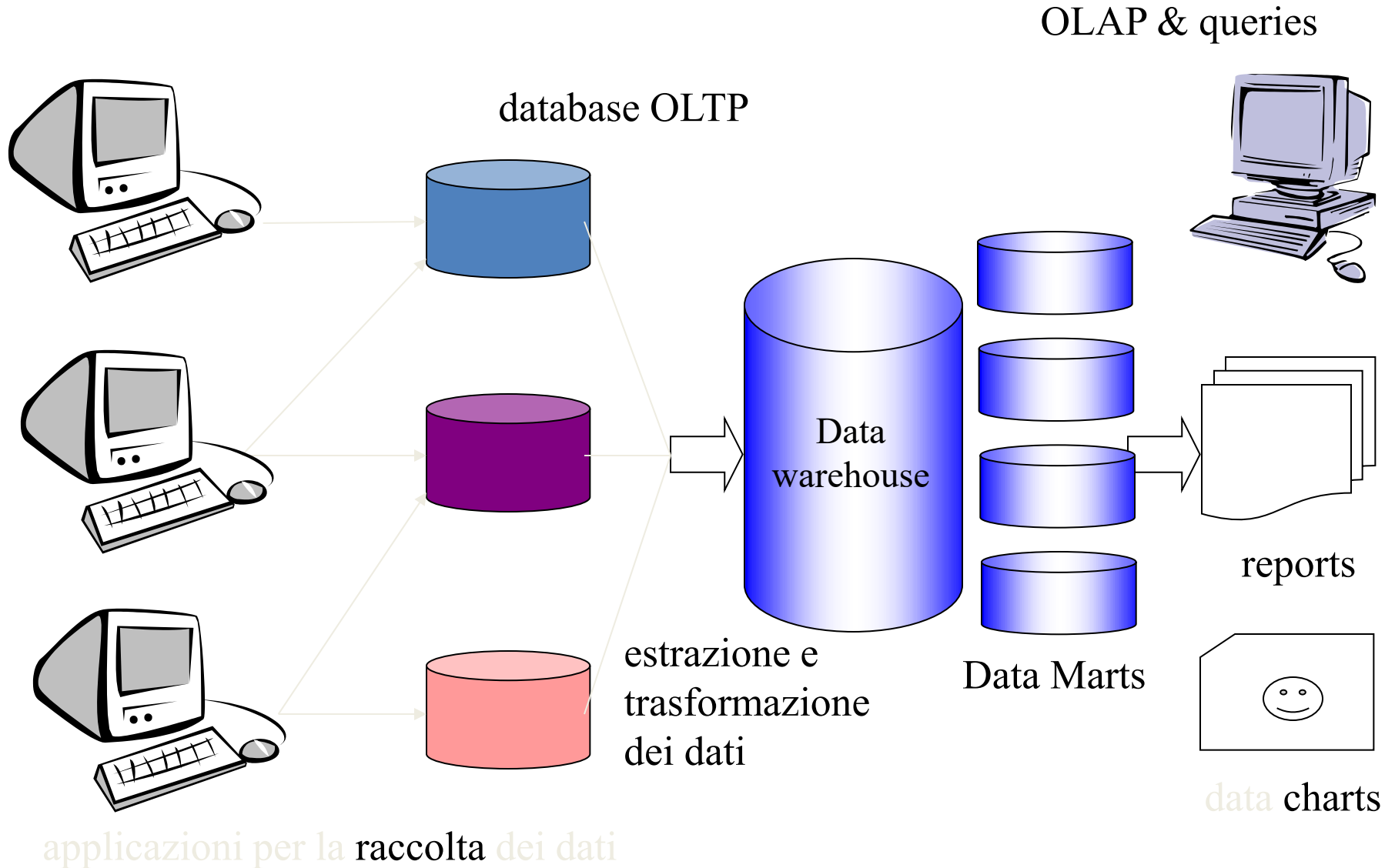


Pivoting (fare perno), per selezionare due dimensioni attraverso le quali aggregare i valori da analizzare

Slice and dice (affettare e ritagliare a cubetti), per selezionare e proiettare i dati riducendo le dimensioni

Ranking (attribuire una classe di merito), per ordinare i dati secondo diversi criteri

Data Warehouse



Che cos'è il Data Mining

Data Mining

è il processo di estrazione di conoscenza da banche dati di grosse dimensioni tramite l'applicazione di algoritmi che individuano le “relazioni nascoste” tra le informazioni e le rendono visibili.

Che cos'è il Data Mining

Esistono diverse definizioni di *Data Mining*:

- ***Data Mining* è l'estrazione non banale di informazioni (potenzialmente utili, implicite e sconosciute in precedenza) dai dati. Questo comporta un certo numero di approcci quali il raggruppamento (*clustering*), l'aggregazione dei dati, la conoscenza di regole di classificazione, la ricerca di reti di dipendenza, l'analisi dei cambiamenti, la scoperta di anomalie". (Frawley, Piatetsky-Shapiro e Matheus 1996).**

Che cos'è il Data Mining

- ***“Data Mining è la ricerca di relazioni e modelli globali che esistono nei grossi database ma sono nascosti nella vastità del numero dei dati. Queste relazioni rappresentano una preziosa conoscenza sul database e dunque sugli oggetti che esso contiene”. (Holshemier&Siebes, 1994).***

Il processo di *Data Mining*

- Il processo di *Data Mining* permette di scoprire delle relazioni nei dati che forniscono al *decision-maker* le informazioni necessarie a fronteggiare le difficoltà proprie dei mercati competitivi.

Il processo di *Data Mining*

- **I fattori critici di successo in un progetto di *Data Mining* sono:**
 - **la conoscenza del business,**
 - **l'esperienza maturata nel tempo.**
- **I software di analisi relativi al processo di *Data Mining* gestiscono autonomamente i dettagli tecnici in modo tale che ci si possa concentrare sulle decisioni.**

Il processo di *Data Mining*

- **La novità offerta dalla nuova tecnologia e dal *Data Mining***
- **- non sta nel rinnegare il tipo tradizionale di conoscenza, che rimane fondamentale,**
- **- ma nell'integrare i processi decisionali con regole costruite sintetizzando complessi ed estesi patrimoni informativi. Tali regole sono il risultato operativo di progetti finalizzati ad attività di business.**

Cosa distingue il processo di *Data Mining* da una tradizionale analisi statistica

- **non tanto la quantità dei dati che vengono analizzati o le particolari tecniche che vengono impiegate,**
- **quanto la necessità di operare in un ambiente in cui la metodologia dell'analisi quantitativa e le conoscenze di business devono essere integrate.**

Fare *Data Mining* significa

- **seguire una metodologia che va dalla traduzione della problematica di business in una problematica di analisi quantitativa, all'implementazione di regole decisionali economicamente misurabili.**

Il processo di *Data Mining*

Realizzare processi di *Data Mining* significa anche organizzare un efficace team di lavoro tenendo conto delle diverse risorse ed esigenze aziendali:

**quelle dei sistemi informativi,
del marketing centrale ,
della rete commerciale.**

Il processo di *Data Mining*

Un'efficace attività di *Data Mining* coinvolge tre diverse figure in grado di interagire tra loro:

- un esperto di business aziendale;**
- un esperto di sistemi informativi aziendali;**
- un esperto di metodi quantitativi per l'analisi dei dati.**

Esempio di processo di *Data Mining*

- **Obiettivo:** campagna di promozione di un nuovo prodotto.
- **Necessità di business:** individuare la tipologia del cliente
- **Area di applicazione:** gestione del mercato
- **Tipo di operazione:** segmentazione del mercato
- **Tecnica utilizzata:** cluster analysis



esempi di processo di Data Mining



cited 03/03/2122

DATAMINING

- <https://www.egon.com/it/blog/664-tecniche-esempi-data-mining-marketing>

tecniche ed esempi pratici di data mining nel marketing

- 2) REGRESSION ANALYSIS PER FARE PREVISIONI DI MARKETING
- 3) CLASSIFICATION ANALYSIS PER INDIVIDUARE SPAM E NON SOLO
- 4) ANOMALY DETECTION PER RICONOSCERE LE INCONGRUENZE
- 5) INTRUSION DETECTION PER UNA MAGGIORE SICUREZZA DEL SISTEMA
- 6) ASSOCIATION RULE LEARNING PER SCOPRIRE LE RELAZIONI FRA DATI
- 7) DECISION TREES PER OTTIMIZZARE IL PROJECT RISK MANAGEMENT
- 8) NEURAL NETWORKS PER AUTOMATIZZARE L'APPRENDIMENTO
- 9) RULE INDUCTION PER ANALISI PREDITTIVE BASATE SUI DATI
- 10) DATA WAREHOUSING PER IL TRATTAMENTO DI BIG DATA

<https://docs.microsoft.com/en-us/azure/architecture/data-science-process/overview>



esempi di processo di Data Mining



cited 03/03/22

DATAMINING

<https://www.dataskills.it/data-mining-esempi/#gref>

esempi di Data Mining

10 tecniche ed esempi pratici di data mining nel marketing

I segreti del data mining per la tua strategia di marketing



Valorizzare le informazioni aziendali residenti in grandi depositi dati è uno degli obiettivi più noti del data mining. Ma le potenzialità delle tecniche, delle metodologie e degli esempi che rientrano nella definizione di data mining vanno ben oltre la semplice valorizzazione di dati. In questo articolo ci concentreremo sul marketing e su ciò che puoi fare tu stesso per promuovere la tua azienda o il tuo business, anche online, grazie al data mining. Nella lista delle 10 tecniche ed esempi pratici selezionati abbiamo dato voce a

risorse e link di approfondimento, in modo che tutti possano conoscere meglio questo ambito vasto e in continuo divenire. Sfruttando le tecniche e prendendo spunto dagli esempi che ti mostreremo fra poco, sarai in grado di potenziare e raffinare la tua strategia di marketing e distinguerti dalla concorrenza. Cominciamo!



ESEMPI DI DATA MINING

Home > Predictive Analytics > Esempi di Data Mining

Con il termine Data Mining, conosciuto negli ultimi anni come Analisi Predittiva, si intende l'applicazione di algoritmi di Machine Learning a basi di dati, con il fine di **estrarne informazioni di valore per il business**. Differentemente dalle analisi statistiche tradizionali, che si limitano a riorganizzare lo storico di dati raccolti nel tempo, l'**analisi di dati tramite Data Mining** sfrutta i dati storici per fornire previsioni sullo stato dei fatti nel futuro.

Sommario

Data mining: esempi pratici

Churn Analysis

Fraud Detection

Time series Forecast

Clustering

Market Basket Analysis



Sviluppo di un processo di Data Mining

L'attività di *Data Mining* si sviluppa attraverso le seguenti fasi (Saccardi, Cuzzocrea):

- predisposizione dei dati provenienti da fonti esterne;**
- definizione di un campione;**
- attuazione di un'analisi preliminare;**
- costruzione di un modello;**
- scelta del modello finale;**
- implementazione delle regole nei processi decisionali.**

Sviluppo di un processo di Data Mining:

Predisposizione dei dati provenienti da fonti esterne

- **consiste nella definizione degli obiettivi aziendali. In funzione di questi viene sviluppata l'attività di analisi.**

Un esempio tipico può essere la selezione del target di clientela per la promozione di un nuovo prodotto.

Sviluppo di un processo di Data Mining:

Identificazione delle fonti dei dati

- **consiste nel reperire i dati necessari al raggiungimento degli scopi prima definiti.**
- **Le fonti dei dati possono essere interne, esterne oppure una combinazione delle due, dettata dall'esigenza di arricchire i dati con nuove dimensioni descrittive o con nuovi nominativi non presenti nel sistema informativo aziendale.**

Sviluppo di un processo di Data Mining:

Definizione del campione

- ***Data Mart delle Analisi (DMA)***, ovvero la matrice dei dati disegnata sulla base dalle esigenze di analisi e dagli obiettivi preposti.
- Il rilascio del DMA è subordinato ad un'attività di controllo di qualità:
- controllo formale per l'individuazione delle variabili non utilizzabili ;
- controllo sostanziale per la verifica del contenuto delle variabili implementate, presenza di modalità non previste.

Sviluppo di un processo di Data Mining:

Analisi mediante un campione

- **Ridurre i tempi di elaborazione;**
- **Mantenere delle garanzie sui risultati;**
- **Disporre di più basi di confronto per la costruzione e per la validazione e successiva la scelta del modello di sintesi;**

Sviluppo di un processo di Data Mining:

Attuazione di analisi preliminari

- **Prima di procedere con l'analisi vera e propria, si effettua uno studio preliminare del fenomeno oggetto d'indagine.**
- **Tale analisi può portare ad una selezione o ad una trasformazione delle variabili originarie. (Ad esempio se le variabili non spiegano in modo chiaro il fenomeno, oppure se vi sono esigenze di sintesi in quanto le variabili originarie conducono ad un'analisi troppo dettagliata e dispersiva rispetto alle esigenze).**

Sviluppo di un processo di Data Mining:

Costruzione di un modello

- **A partire dai dati si procede alla determinazione di regole generali a seconda dell'obiettivo di partenza.**
- **Se ad esempio l' obiettivo riguarda lo studio dell'*interdipendenza* tra variabili,**
 - **per variabili di tipo qualitativo si possono utilizzare tecniche quali l'analisi delle corrispondenze o i modelli log-lineari ecc.;**
 - **per variabili di tipo quantitativo si possono utilizzare tecniche quali l'analisi dei gruppi e l'analisi delle componenti principali ecc..**

Sviluppo di un processo di Data Mining:

Costruzione di un modello

- **A partire dai dati si procede alla determinazione di regole generali a seconda dell'obiettivo di partenza. Se ad esempio l'obiettivo riguarda lo studio della *dipendenza* vale a dire lo studio di una variabile rispetto ad altre variabili considerate esplicative,**
- **nel caso di variabili categoriali si utilizzano, ad esempio, modelli logit, probit;**
- **nel caso di variabili quantitative, si utilizzano modelli di regressione lineare o logistica, a seconda dalla natura della variabile dipendente.**

Sviluppo di un processo di Data Mining:

Costruzione di un modello

OSSERVAZIONE:

In molti casi considerazioni sullo spazio delle variabili, sulla dimensione del problema, sulla complessità del modello esplicativo sottostante portano alla necessità di utilizzare un altro tipo di modellistica:

reti neurali, alberi decisionali, algoritmi genetici, random forest, boosting, support vector machines, nearest-neighbors, graphical models, ecc..

Sviluppo di un processo di Data Mining:

Scelta del modello finale

La scelta del modello finale viene effettuata tenendo conto di vari fattori, quali:

- i risultati ottenuti con le analisi preliminari;**
- i risultati ottenuti con la costruzione delle regole generali da seguire;**
- gli obiettivi da raggiungere ed i vincoli di business, sia in termini di budget che di tempo.**

Sviluppo di un processo di Data Mining:

Scelta del modello finale

La scelta del modello finale viene effettuata tenendo conto di vari fattori, quali:

- La redazione di un modello ottimale potrebbe richiedere un periodo di elaborazione superiore a quello disponibile. In tal caso si preferisce giungere a soluzioni più semplici ma comunque in grado di fornire indicazioni più efficaci rispetto ai criteri tradizionalmente usati in azienda.**

Sviluppo di un processo di Data Mining:

Scelta del modello finale

Elementi necessari per la scelta ottimale del modello finale:

- **produrre una grande quantità di modelli in modo semplice e rapido, confrontare i risultati da essi prodotti,**
- **dare una quantificazione economica della regola costruita,**

Sviluppo di un processo di Data Mining:

Implementazione delle regole nei processi decisionali

Elementi necessari per la scelta ottimale del modello finale:

- **Elemento distintivo dell'attività di *Data Mining*, rispetto a quella di semplice analisi dei dati, è l'integrazione dei risultati nei processi decisionali aziendali.**
- **La conoscenza del business, l'estrazione delle regole e il loro inserimento nelle procedure gestionali, permettono di passare dalla fase di analisi al rilascio in produzione di un “motore” decisionale.**

Sviluppo di un processo di Data Mining:

Implementazione delle regole nei processi decisionali

Esempi:

- **la selezione del target di riferimento per un gestore di campagne di marketing,**
- **lo *score* associato al rilascio di un prestito,**
- **la probabilità di abbandono di un contratto telefonico,**
- **la probabilità di transazioni fraudolente provenienti dall'utilizzo di carte di credito.**

Sviluppo di un processo di Data Mining:

Esempi di utenti del *Data Mining*

Esempi:

- **gli istituti di credito per conoscere i reali rischi di insolvenza;**
- **le imprese per conoscere meglio la propria clientela e offrire così prodotti adeguati alle esigenze di quest'ultima;**
- **le aziende turistiche per predire con irrisori margini d'errore il target dei loro futuri clienti;**

Sviluppo di un processo di Data Mining:

Esempi di utenti del *Data Mining*

Esempi:

- **i meteorologi per fare previsioni a medio e lungo termine, “leggendo” le correlazioni tra milioni di fotogrammi registrati dai satelliti;**
- **le Borse Valori per predire le oscillazioni di un titolo;**
- **i ricercatori scientifici, ad esempio nel campo della genetica.**

