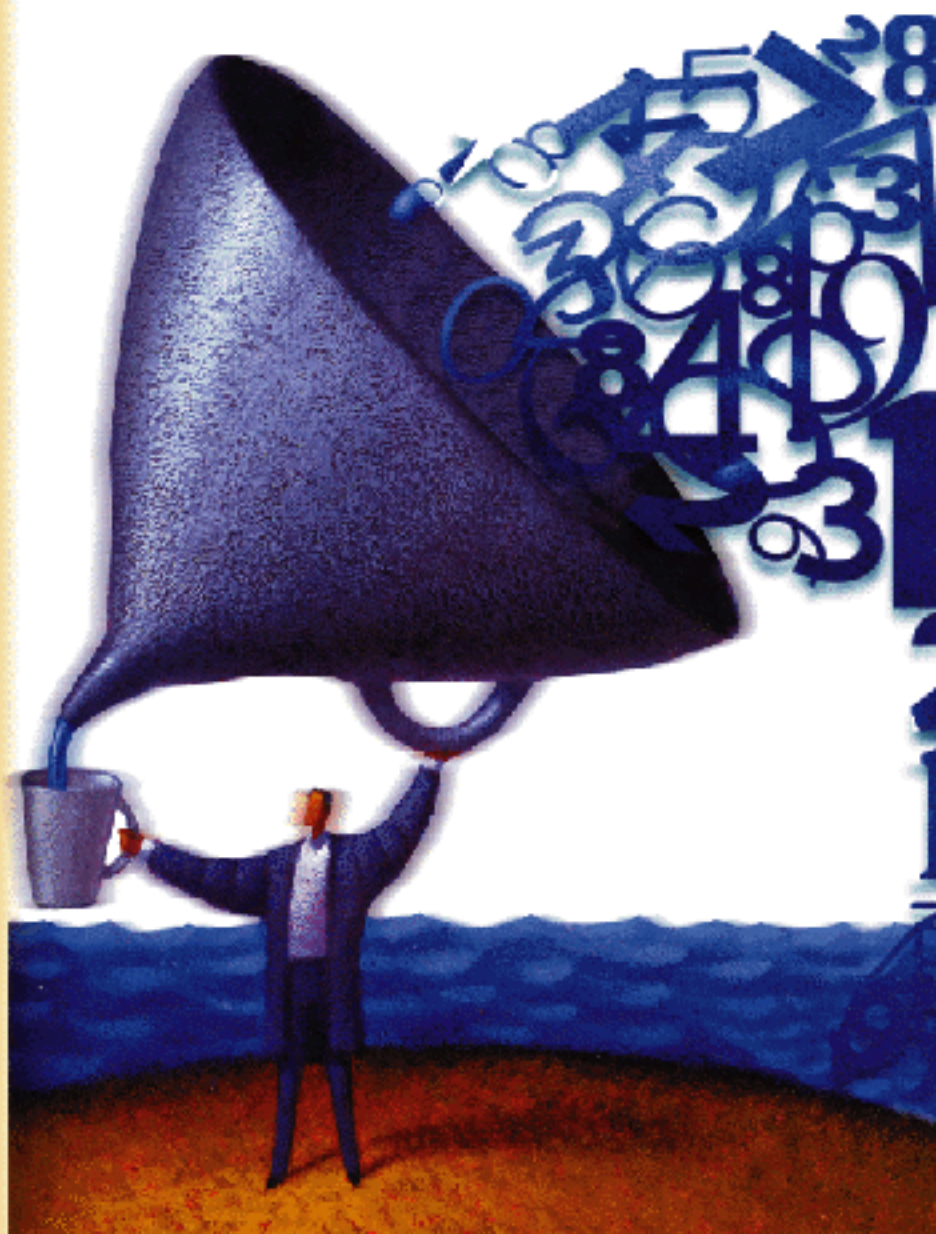


Jill Dyché

**Prefazione
di Martha Rogers**

e-Data

Trasformare
i dati in
informazioni
con tecniche
di Data
Warehousing
e di Database
marketing



Data mining in supporto alle decisioni

Testo tratto dal volume: [e-Data](#)

Autore: Dyché

Copyright © 2001 – APOGEO srl

Via Natale Battaglia 12 – 20127 Milano (Italy)

Telefono: 02-02-28970277 (5 linee r.a.) – Telefax: 02-26116334

Email apogeo@apogeoonline.com

U.R.L. <http://www.apogeoonline.com>

Responsabile editoria digitale Alberto Mari

Copertina e progetto grafico di Enrico Marcandalli

Tutti i diritti sono riservati a norma di legge e a norma delle convenzioni internazionali. Nessuna parte di questo e-book può essere riprodotta con sistemi elettronici, meccanici o altri, senza l'autorizzazione scritta dell'Editore.

Nomi e marchi citati nel testo sono generalmente depositati o registrati dalle rispettive case produttrici.

E-Data

Autore: Dyché

ISBN: 88-7303-712-7

Pagine: 400

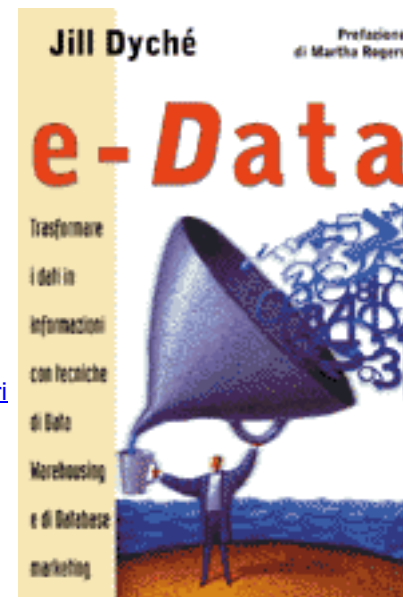
Formato: 14,5 x 21

Prezzo: € 25,21

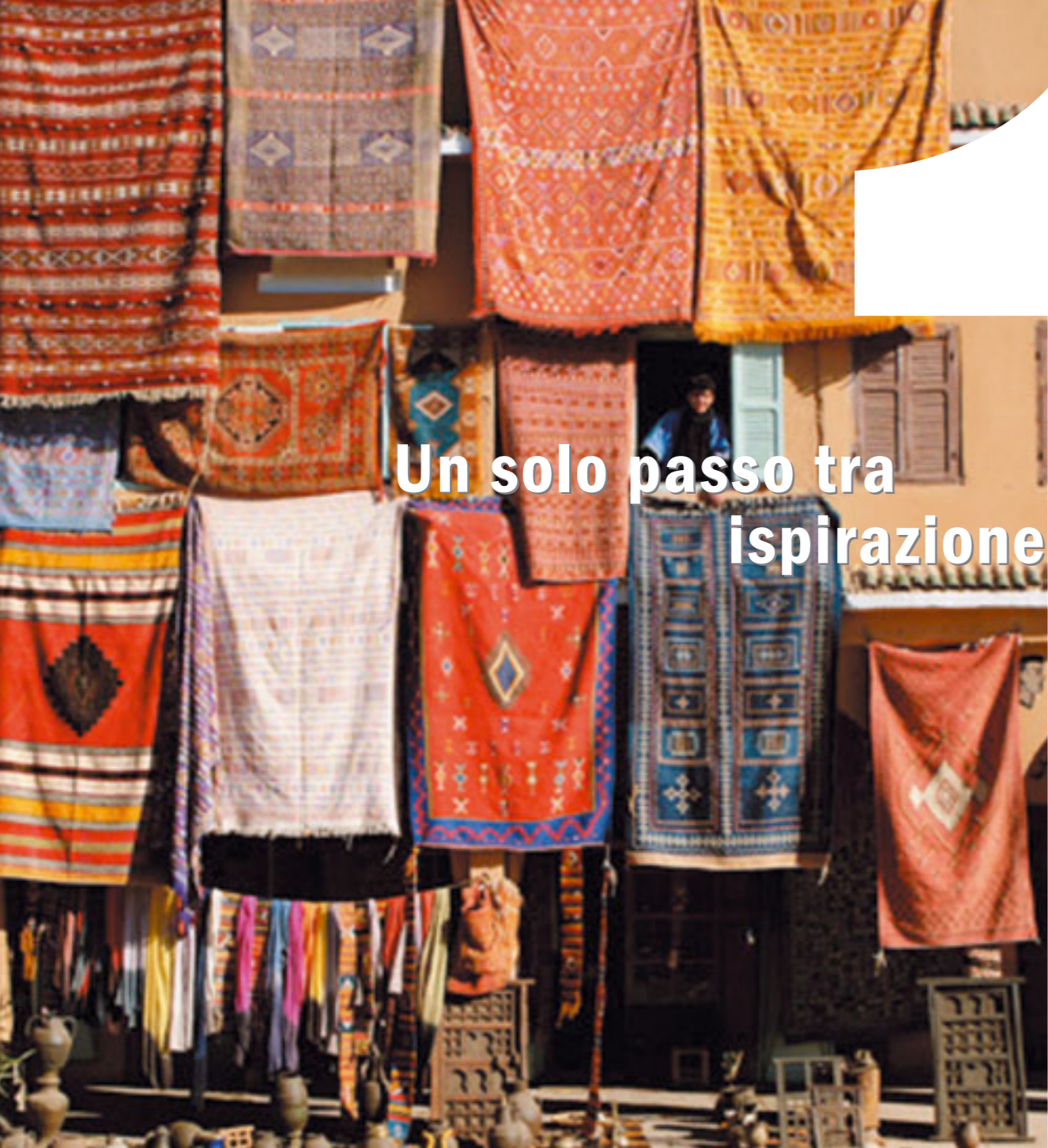
Collana: Cultura digitale

Il testo del presente e-book è tratto dal volume e-Data. Maggiori informazioni online all'indirizzo:

<http://www.apogeoonline.com/libri/00712/scheda>



e-Data fornisce informazioni inestimabili riguardo al Data Warehousing in generale, al suo sviluppo e al suo valore strategico, alle tecnologie che supporta, alla sua influenza sulle procedure di Decision Making; nozioni queste che vi consentiranno di trasformare la miniera d'oro dei dati di cui disponete in informazioni altrettanto preziose, predisponendo la vostra azienda alla leadership del mercato in cui operate e migliorando i livelli di soddisfazione della vostra clientela.



Un solo passo tra
ispirazione

e creazione.



Soluzioni software che vi permettono di trasformare velocemente le nuove idee in nuovi prodotti. Questo è .NET di Microsoft. Essere sicuri che l'informazione giusta arrivi a destinazione al momento giusto: solo così avrete a disposizione gli elementi per concretizzare le vostre idee. Con .NET di Microsoft le persone possono collaborare in modo più efficace, diventare più

produttive e realizzare più rapidamente nuovi prodotti. Questo significa un solo passo tra ispirazione e creazione. Questo è il business con .NET di Microsoft. Per ulteriori informazioni chiamate il Servizio Clienti allo 02.70.398.398 o visitate il sito microsoft.com/italy/business/

Costa Crociere ospita sulle sue navi migliaia di passeggeri ogni giorno, ognuno con desideri e preferenze diversi. Costa Crociere impara a conoscerli uno per uno: quando scelgono sul Web la crociera che hanno sempre sognato, quando sono in navigazione, quando arrivano in porto. Anche a casa continuano a far parte del club Costa e possono raggiungerlo su Internet, perché Costa ha una soluzione per essere sempre in contatto con i propri clienti e con tutte le agenzie di viaggio che prenotano crociere con la sua flotta. Una soluzione integrata di Customer Relationship Management basata su software .NET di Microsoft.

Software per il business agile.

Microsoft®

Informazioni sul libro

Sommario

Data mining in supporto alle decisioni

L'evoluzione del supporto alle decisioni

Query standard: il motore del DSS

Analisi multidimensionali: il potere delle sezioni e dei dati

Modellazione segmentazione: analisi per gli "utenti esperti"

Knowledge Discovery: il potere dell'ignoto

Alcuni esempi tratti dalla vita reale

Query standard

Analisi multidimensionali

Modellazione e segmentazione

Knowledge Discovery

Perché si ricorre al Data Mining?

Il Data Warehousing nell'effettiva realtà aziendale

Cosa è necessario fare per raggiungere la cima

Data mining in supporto alle decisioni

La seguente sceneggiatura è tratta da un incontro che ho avuto con un potenziale cliente che mi aveva invitata a passare per discutere dell'analisi dei dati. Sebbene qualsiasi somiglianza con persone e fatti realmente accaduti sia puramente casuale, potreste riconoscervi in qualcuno dei personaggi descritti.

APERTURA CON DISSOLVENZA:

SALA CONFERENZE DELLA SEDE CENTRALE DI UNO DEI MAGGIORI ISTITUTI DI CREDITO. GIORNO.

Vestiti di tutto punto e seduti attorno a un ampio tavolo di mogano ci sono MARIANNE, di professione Product Manager, SHANE, analista marketing, CHARLIE, amministratore di database che sta bevendo un analcolico tutto d'un fiato e JILL consulente Data Warehouse. A grandi passi arriva il SIG. B., vicepresidente Marketing, il quale stringe frettolosamente la mano a Jill e si siede a un capo del tavolo sporgendosi in avanti.

SIG. B.: Jill la ringraziamo per essere venuta, come è andato il viaggio?

JILL: Oh, è stato ...

SIG. B.: Bene, l'abbiamo convocata perché abbiamo bisogno di qualche dritta. Recentemente ho letto un articolo che mi ha tenuto sveglio per due notti. Uno dei nostri concorrenti ha iniziato a fare un qualche tipo di analisi dei dati e la cosa mi ha preoccupato.

JILL: Che tipo di analisi?

SIG. B.: Che cos'è che hanno fatto, Shane?

SHANE: Del Data Mining.

SIG. B.: Proprio così del Data Mining! Vogliamo fare del Data Mining e vogliamo che lei ci spieghi come procedere.

JILL: Benissimo. Che tipo di Data Mining hanno fatto i vostri concorrenti?

SIG. B.: Che tipo di Data Mining, Marianne?

MARIANNE: Sono stati in grado di stabilire quali loro clienti avrebbero acquistato un nuovo prodotto, inoltre avrebbero elaborato delle liste di nomi che avrebbero utilizzato come base per azioni di mailing.

JILL: Un intelligente strategia di marketing mirato! Benissimo! Ditemi, che tipo di marketing mirato state facendo attualmente?

Si guardano gli uni con gli altri.

JILL: OK, cerchiamo di vederla in altri termini, dove tenete i dati relativi ai clienti?

Tutti guardano Charlie.

CHARLIE: Um, quali dati relativi ai clienti?

JILL: Sa di cosa parlo, dei dati relativi ai clienti della vostra società. Come accedete alle informazioni sul loro conto?

CHARLIE: Beh, alcuni di questi dati si trovano nel sistema di fatturazione ed altri sono memorizzati nel nostro sistema per le analisi di marketing. Per quanto riguarda i dati dei clienti che abbiamo acquisito da meno di un anno si possono trovare nel nostro sistema di gestione degli ordini e dell'approvvigionamento fino a quando non riusciremo a raccoglierci ... abbiamo anche i dati di alcuni ex-clienti in un database di Access nel portatile di Craig.

JILL: A quali di questi clienti volete vendere i vostri nuovi prodotti?

Si guardano gli uni con gli altri.

MARIANNE: Beh, idealmente a tutti loro.

SIG. B.: Assolutamente, a tutti loro!

JILL: Così come fate quando avete bisogno di informazioni sugli acquisti di un determinato cliente oppure quando volete sapere come sta andando un prodotto in una determinata area?

SHANE: (timidamente) Chiamiamo Charlie.

Tutti guardano Charlie di nuovo. Charlie beve un sorso della sua bevanda e la versa sul tavolo.

JILL: (togliendosi la giacca) Quanto tempo abbiamo

CHIUSURA IN DISSOLVENZA ...

Questa società aveva un'effettiva necessità aziendale: individuare il prodotto giusto per il cliente giusto. Tuttavia, la tecnologia in uso non era in grado di fornire una facile soluzione al problema, i dati relativi ai clienti erano decentralizzati e l'azienda si affidava a un'unica risorsa umana, Charlie, invece di implementare un Data Warehouse che avrebbe potuto consolidare e interpretare intelligentemente l'informazione relativa alla clientela.

Mettendola in altri termini, l'istituto di credito voleva correre prima di sapere almeno camminare.

La carenza di tecnologie appropriate non era l'unico problema, le stesse procedure aziendali erano estremamente contorte: il personale del reparto marketing doveva chiamare Charlie o inviargli un'e-mail tutte le volte che aveva bisogno di informazioni (in entrambi i casi sperando che lui avrebbe risposto), spiegare il problema, aspettare una risposta, analizzare il risultato e, nella maggior parte dei casi, accorgersi che era necessario ridefinire la domanda e riproporla a Charlie.

Infine, anche se questi utenti fossero stati in grado di accedere a un Data Warehouse e avessero potuto richiamare

i dati relativi ai clienti, la maggior parte di loro non avrebbe saputo cosa farsene dei risultati della ricerca.

L'evoluzione del supporto alle decisioni

In questa sezione ci proponiamo di analizzare le varie categorie di analisi per il supporto alle decisioni partendo da semplici query, per giungere ad alcune delle analisi più avanzate esistenti. Queste categorie sono pensate in modo da suggerire un approccio evolutivo nella realizzazione e nell'utilizzo di un Data Warehouse, soprattutto se questo è del tipo aziendale, ovvero consente l'esecuzione di diverse funzioni aziendali nell'ambito dell'impresa, o se si tratta di Data Mart con funzione specifica.

Query standard: il motore del DSS

Nonostante sul mercato siano reperibili i più recenti e sofisticati algoritmi, nonché innovativi strumenti di visualizzazione, le query (domande poste ad un database) rappresentano sicuramente il metodo di analisi più diffuso. In effetti, i dirigenti di un'azienda preferiscono proporre delle query ai Data Warehouse su scala giornaliera piuttosto che ricorrere agli altri tre tipi di analisi combinate.

L'ironia in questo caso vuole che molte aziende nuove al Data Warehousing pretendano di eludere il ricorso alle query per passare subito ad analisi avanzate; questo è l'atteggiamento descritto dalla scenetta riportata all'inizio nel presente capitolo, un atteggiamento tipico di coloro che vogliono utilizzare il supporto alle decisioni per ragioni "strategiche".

Ma che cosa c'è di più strategico che rendere accessibile ai PC di tutti gli utenti aziendali l'informazione dettagliata e significativa relativa ai clienti, ai prodotti e al comportamento dell'impresa? La possibilità di poter effettuare da soli semplici query consente a molte imprese di disporre di informazioni

nuove e significative che possono coadiuvare qualsiasi tipo di decisione strategica. Inoltre, è più probabile che gli utenti che sono in grado di analizzare i dati richiamati da query semplici siano più propensi a sostenere l'adozione di tipologie più avanzate di supporto alle decisioni, avendo già potuto sperimentare i vantaggi offerti da queste tecnologie.

Il seguente esempio di query può fornire una serie di dati illuminanti per gli utenti finali che prima non aveva accesso a nessun tipo di dati:

Visualizza tutti i clienti che lo scorso anno hanno acquistato il prodotto X

Spesso le richieste dati sono "preconfezionate", ovvero sono state predefinite, proprio come avviene per il rapporto aziendale standard che siete soliti vedere sulla vostra scrivania alla fine di tutti i mesi. Le query preconfezionate vengono rappresentate sotto forma di icona o immagine sul monitor del vostro PC così, mentre i dati sottostanti possono cambiare di giorno in giorno o persino di ora in ora, le query preconfezionate non devono essere riscritte tutte le volte.

In una azienda con la quale ho collaborato, gli analisti finanziari erano costretti ad attendere il montaggio di un nastro (si tratta di un processo mediante il quale un nastro viene inserito in un apposito lettore collocato nel centro elaborazione dati al fine di leggerne i contenuti) per poter accedere alle informazioni relative alle entrate di un determinato prodotto, informazioni queste che avevano già un anno di vita; a seconda delle richieste accumulate, questa procedura poteva richiedere settimane. Quando furono messe a disposizione degli analisti delle query preconfezionate questi furono in grado di accedere allo storico delle vendite con un semplice clic del mouse. Questi utenti non erano esperti analisti di dati, ma impiegati che necessitavano di quei dati per svolgere il loro lavoro.

Un altro tipo di query può essere effettuato senza preparazione, spesso nello stesso istante in cui l'utente la pensa, si tratta delle query "ad hoc", solitamente specifiche di un determinato utente o di un determinato problema (si tratta di query relative a un determinato prodotto, canale di distribuzione o conto clienti). A differenza delle query preconfezionate spesso utilizzate in modo più generalizzato, le query ad hoc sono sufficientemente originali e dettagliate da rendere poco pratico il ricorso alla predefinizione, per questa ragione questo tipo di query viene creato ex novo ogni volta che deve essere eseguito. La Figura 1 mette a confronto queste due tipologie di ricerche dati.

Oltre a consentire tempi di risposta più ridotti le query permettono agli impiegati di pensare e di agire autonomamente. Dato che virtualmente le risposte ai problemi sono immediate, l'azienda può adottare rapidamente azioni atte a controbattere le improvvise strategie di marketing dei concorrenti, a modificare l'indirizzo di un cliente, a fornire risposte in tempo reale o ad incoraggiare un cliente scontento a restare.

Secondo Tony Marshall della Hallmark Cards Inc., il 20 per cento del vantaggio competitivo dell'impresa in cui lavora deriva da report preconfezionati, il restante 80 per cento è dato dall'abilità degli utenti finali di rispondere alle nuove problematiche effettuando query ad hoc; egli afferma inoltre che è proprio grazie a questo tipo di analisi che i loro partner riescono a identificare i trend di mercato più recenti supportando processi decisionali innovativi.

Le query ad hoc infatti offrono su richiesta l'informazione necessaria per consentire all'azienda di diventare più vivace e reattiva. Sottovalutare le potenzialità offerte dalle query semplici o attendere di poter usufruire di livelli di analisi più avanzati può essere in effetti un grave errore in termini di concorrenzialità.



Figura 1. Query preconfezionate e query ad hoc.

Analisi multidimensionali: il potere delle sezioni e dei dati

L'analisi multidimensionale rappresenta il gradino successivo nella scala delle tipologie di analisi. Mentre alcuni utenti si accontentano di effettuare semplici query standard, altri (in particolar modo coloro che devono effettuare ricerche più approfondite nell'ambito di dati specifici) necessitano di tecnologie di analisi più potenti che gli consentano di studiare e confrontare l'informazione.

L'analisi multidimensionale offre differenti prospettive, "dimensioni" o "punti di vista", attraverso i quali osservare i dati: "periodo di tempo", "luogo" e "prodotto" sono esempi di dimensioni comunemente utilizzate.

Le ragioni che spingono a effettuare analisi multidimensionali risiedono nell'esigenza di vedere

l'informazione in una particolare prospettiva così, per esempio, si possono suddividere i clienti per aree geografiche, le vendite per città o le chiamate per periodo della giornata. Un'altra giustificazione a questo tipo di analisi è da ricondurre alla necessità di raccogliere un insieme di risposte da utilizzare più volte secondo prospettive diverse, ma facciamo un esempio:

Visualizza tutti i nuovi clienti dell'azienda suddividendoli per zone di vendita.

La dimensione utilizzata per classificare le vendite legate ai nuovi clienti in questa query ("zone di vendita") è di tipo geografico. Una volta che l'utente ha ottenuto questa informazione può approfondire l'indagine richiedendo un livello di dettaglio maggiore:

Visualizza tutti i nuovi clienti dell'azienda suddividendoli per aree di vendita.

Questo report multidimensionale applicherà la tecnologia drill-down per estrarre i dati relativi alle singole province di vendita a partire dai dati regionali aggregati.

Gli strumenti utilizzati per le analisi multidimensionali sono simili a quelli utilizzati per effettuare query standard, spesso si tratta proprio degli stessi. In effetti sia le query standard che le analisi multidimensionali possono essere preconfezionate per consentirne un uso ripetitivo e vengono eseguite ricorrendo a strumenti software user-friendly. La differenza tra le due tipologie di indagine risiede nel fatto che mentre le query standard richiamano un'ampia sezione trasversale di dati differenti, le analisi multidimensionali vengono utilizzate per vedere gli stessi dati sotto prospettive diverse.

Anche se spesso gli utenti che richiedono analisi multidimensionali sono gli stessi che si affidano anche alle query standard, i primi godono sicuramente di una maggiore consapevolezza degli strumenti che stanno impiegando. Va

sottolineato inoltre che l'analisi multidimensionale si è dimostrata una manna per quegli utenti aziendali che volevano estrarre il meglio dai dati aziendali.

Modellazione segmentazione: analisi per gli "utenti esperti"

All'espansione della comunità di utenti che si affida ad analisi di DSS corrisponde un incremento della specializzazione degli utenti. L'utente aziendale medio, probabilmente, si accontenterà di report preconfezionati e di qualche analisi multidimensionale per il lungo termine, tuttavia quando i dati diventano più voluminosi e dettagliati e vengono a crearsi nuove prospettive aziendali, gli "utenti esperti" ovvero gli analisti marketing e gli statistici richiedono maggiori capacità analitiche altamente sofisticate per poter raccogliere informazioni nuove.

Utilizzare i dati esistenti nel Data Warehouse, principalmente dati storici, può aiutare gli analisti a prevedere eventi futuri. Il lavoro di previsione a fronte di ampi volumi di dati dettagliati può essere effettuato utilizzando strumenti di analisi specifici che estraggono i dati dal Data Warehouse e li analizzano per realizzare una serie di modelli. Un modello è semplicemente una raccolta di costanti proprie di una determinata caratteristica che come tale può essere rappresentata graficamente o mediante una serie di regole o notazioni. Sebbene molti degli algoritmi di modellazione (tecniche di analisi logiche) esistano da una dozzina di anni o più, solo con l'avvento del supporto alle decisioni essi hanno potuto sfruttare la ricchezza di dettaglio e di informazione storica (e-data) che viene centralizzata nell'ambito del Data Warehousing.

Anche la segmentazione non è un'invenzione recente. In effetti, questa disciplina del marketing viene utilizzata fin dall'introduzione degli studi demografici degli anni Sessanta e Settanta.

La segmentazione suddivide i clienti o altre aree di dati in determinati gruppi o segmenti, le caratteristiche comuni dei quali possono definirne il comportamento e quindi determinare strategie di vendita e marketing appropriate.

Una volta che è stata effettuata una segmentazione della clientela, gli stessi segmenti possono essere analizzati, trattandosi di dimensioni, ricorrendo a strumenti di analisi multidimensionale; questo dà agli utenti meno specializzati l'opportunità di esaminare i modelli che sono stati creati mediante analisi avanzate senza il bisogno di dover effettuare queste analisi in prima persona.

In questo manuale modellazione e segmentazione dei dati vengono considerate nello stesso paragrafo perché, pur comportando tecnologie sottostanti differenti, rientrano entrambe nell'ambito delle analisi di marketing classiche e richiedono livelli di capacità simili. Nella maggior parte dei casi, la modellazione e la segmentazione utilizzano del software specializzato, per questa ragione l'analista che esegue queste procedure, oltre a essere in grado di comprendere i dati a un più elevato livello di dettaglio, molto probabilmente è stato addestrato per saper utilizzare strumenti sviluppati appositamente.

Knowledge Discovery: il potere dell'ignoto

Le aziende hanno speso milioni di dollari per cercare di individuare i clienti più opportuni cui rivolgere determinate campagne di marketing, alcune volte ricorrendo alla tattica meno complicata, sebbene anche meno remunerativa, di coinvolgere qualsiasi possibile cliente noto. La Knowledge Discovery¹ (scoperta della conoscenza) è rappresentata da

¹ Con il termine Knowledge Discovery (KD) si indica quell'aspetto del processo di acquisizione della conoscenza (Knowledge Process) che, nell'ambito della gestione della conoscenza (Knowledge Management), consente al personale di reperire le informazioni necessarie per svolgere con profitto il proprio lavoro e raggiungere gli obiettivi aziendali dei singoli reparti dell'impresa. Il Knowledge Process

infiniti algoritmi molto potenti che ricercano particolari elementi nell'ambito di database di grandi dimensioni. A differenza della modellazione, questi elementi non vengono specificati anticipatamente ed è per questo che viene sfruttata in tutti quei casi nei quali l'utente sta cercando risposte a domande che non è in grado di formulare. Ne risulta che è il Data Warehouse a indicare all'impresa dove si trovino gli elementi, le relazioni e i "punti caldi" più interessanti.

Comunque se sono disponibili dati dettagliati, è possibile localizzare esattamente i comportamenti significativi, i modelli e le sequenze senza che sia necessario ricorrere a un analista dati che li esamini. Dato che gli strumenti di Knowledge Discovery individuano determinati modelli e relazioni che non sono state specificate dall'utente, gli esempi in proposito spostano l'attenzione su ciò che è stato trovato piuttosto che su ciò che si stava cercando.

Mentre le vere e proprie attività di Knowledge Discovery non prevedono che sia stata formulata alcuna ipotesi, molte restano comunque limitate a particolari obiettivi aziendali o ad aree dati legate a un determinato soggetto; per questa ragione un dato progetto di Knowledge Discovery rimarrà comunque focalizzato su una determinata area funzionale dell'impresa (per esempio la clientela e il comportamento che assumerà in futuro) indipendentemente dall'informazioni che potrebbe trovarvi.

La promessa del Data Warehousing, come ben sapeva il nostro amico dell'esempio relativo all'istituto di credito, risiede nella capacità di trovare "risposte sconosciute" e

coinvolge, oltre alla KD, anche l'organizzazione dell'informazione (Knowledge Organization, la gestione di tutti i tipi di informazione) e lo sviluppo della conoscenza cooperativa (Collaborative Knowledge Development), ovvero la collaborazione dei diversi reparti aziendali nell'ambito dello sviluppo, in tutti i suoi stadi, di progetti e processi informativi. [N.d.T]

nell'utilizzarle per accrescere le entrate in modo nuovo e innovativo.

Alcuni esempi tratti dalla vita reale

Dopo aver introdotto i quattro livelli di analisi vedremo di esaminare in questo paragrafo alcuni esempi che ne caratterizzano l'uso.

La piramide riportata nella Figura 2 rappresenta una "evoluzione" lineare dell'analisi dati per il supporto alle decisioni; si tratta di una sorta di tassonomia del supporto alle decisioni.

Se si segue questa piramide dal livello inferiore a quello superiore si potrà sviluppare di pari passo la capacità dell'impresa di effettuare analisi, nonché accrescerne gradualmente l'infrastruttura Data Warehouse. In altri termini, alla crescita e maturazione delle metodologie e tecnologie di analisi corrisponde un'evoluzione dell'hardware, del software e dei dati del Data Warehouse. Proprio come un bambino impara qualcosa di nuovo in ogni successiva classe scolastica, così un'impresa incrementa sempre più i vantaggi offerti dal proprio Data Warehouse aggiungendo progressivamente tecnologie supplementari e facendo crescere la comunità di utenti che le sfruttano nel corso del tempo.

Con questo non vogliamo dire che dovete assolutamente iniziare dalle query standard, tuttavia una buona valutazione della qualità e del grado di fruibilità del vostro Data Warehouse è data proprio dalla possibilità di effettuare query standard: se non siete in grado di effettuare questo tipo di analisi, sicuramente non siete pronti per passare a quelle di tipo multidimensionale. Analogamente, se non riuscite a eseguire analisi multidimensionali non siete sufficientemente dotati per poter passare alla modellazione.

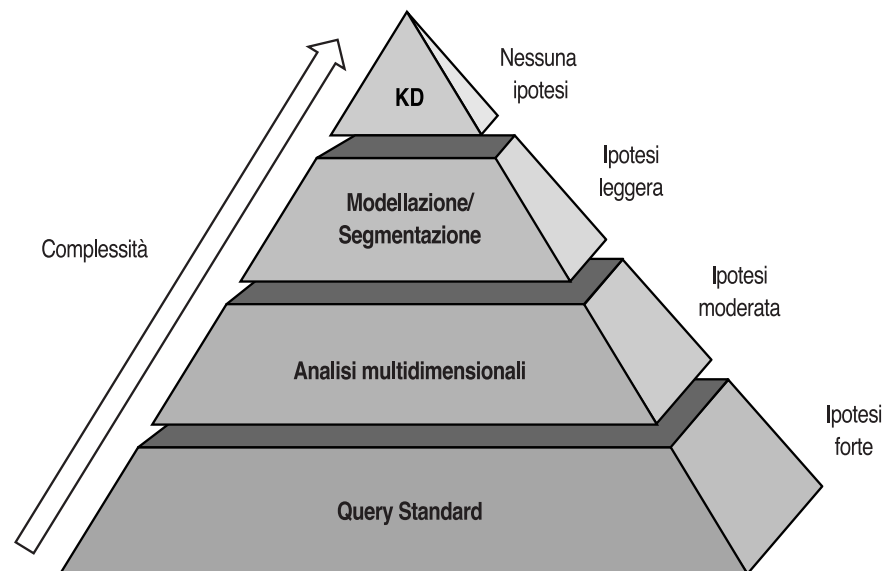


Figura 2. La piramide Baseline (Fonte: Baseline Consulting Group).

All'evoluzione e alla maturazione dei dati contenuti nel Data Warehouse corrisponde un affinamento delle abilità di analisi dei dati da parte degli utenti, oltre a una loro maggiore consapevolezza delle tecnologie utilizzate. Più spesso un utente finale accede al Data Warehouse e raccoglie e-Data significativi e più il Data Warehouse diventa una materia prima. Inoltre, l'utente finale che grazie agli e-Data ha potuto realizzare dei miglioramenti aziendali (si tratti di risparmi sui tempi o sui costi, oppure di un consolidamento del processo) sarà la persona che appoggerà il prossimo progetto o che gli darà persino vita.

Se i nostri amici dell'istituto di credito avessero seguito l'approccio evolutivo suggerito dalla piramide, implementando lentamente le aree di soggetto verticali e acquisendo diversi strumenti, avrebbero potuto essere sulla giusta via per realizzare la modellazione di cui necessitavano. Allo stato attuale delle cose, devono spendere tempo e denaro per creare la propria infrastruttura dati prima di poter

effettivamente sperare di capitalizzare qualcosa in base alle promesse dei modelli di previsione, senza considerare che nel frattempo il maggior concorrente dell'istituto di credito sta di fatto acquisendo del vantaggio concorrenziale nei confronti dei suoi clienti.

Le sezioni che seguono illustrano gli usi che utenti reali hanno fatto dei singoli livelli della piramide, potrete così paragonarli a quelli che ne fate in azienda o agli obiettivi che vi prefiggete di raggiungere.

Query standard

La Figura 3 illustra il modo in cui una piccola società telefonica classifica i tipi di query effettuate dai suoi utenti.

Questa società telefonica offre report sulle vendite, informazioni sui costi dei prodotti, report sull'utilizzo della rete e sulle entrate, oltre a report di redditività a fine mese, il tutto utilizzando query standard. Grazie a queste query il personale, ovvero addetti alle vendite, Product Manager e analisti finanziari compresi, può esaminare gli e-Data e prendere decisioni immediate.

Per esempio alcuni utenti hanno accesso a report preconfezionati e possono quindi effettuare regolarmente questa query:

Visualizza un elenco di tutti quei clienti le cui entrate relative all'utilizzo del servizio nelle fasce orarie di punta sono diminuite per lo meno del 20 percento².

Il risultato di questa query può dare vita a una serie di differenti decisioni aziendali: dall'offerta di sconti sui prodotti

² Nonostante questa query sia preconfezionata l'utente può modificare dinamicamente la percentuale, trasformando il 20 percento in un 35 percento, a seconda delle circostanze. Questo tipo di query è detto anche query "parametrizzata".

e sulle fasce orarie, all'interruzione di un servizio le cui entrate sono in diminuzione.

La maggior parte delle query standard si basa su ipotesi alquanto sicure, così, per esempio, la seguente query effettuata da una società di credito fondiario presume che il cliente X abbia pagato il proprio debito in ritardo, magari ripetutamente:

Mostra tutti i pagamenti ritardati del cliente X con le relative date di pagamento, per i casi in cui i termini di pagamento erano scaduti da oltre due settimane.



Figura 3. Un esempio di classificazione di possibili query.

Potremmo persino sapere quando sono stati saldati questi debiti. L'utente che avvia questa query potrebbe semplicemente effettuare un controllo o verificare un sospetto nutrito da molti, al fine di inviare al cliente X una lettera per il recupero dei crediti. Dato che questa query conferma un sospetto è un'ipotesi forte.

La query semplice può costituire un vantaggio di ampio respiro per l'azienda, considerate il seguente esempio:

Mostra tutti gli utenti di cellulari le cui chiamate interne sono state interrotte per più del 20 percento delle volte nel corso di una settimana.

La risposta a questa richiesta dati potrebbe individuare quegli utenti che sfruttano molto il servizio e potrebbero aver

bisogno di apparecchiature aggiuntive, possibili concorrenti sul mercato, candidati per prodotti potenziati, sospetti relativi a frodi, clienti potenzialmente insoddisfatti e molto di più. In breve può consentire di mantenere e acquisire nuovi clienti.

Analisi multidimensionali

Supponiamo ora che la risposta alla query succitata porti a formulare un'altra domanda, potremo per esempio voler sapere:

Di quei clienti (le cui chiamate interne sono state interrotte per più del 20 percento delle volte nel corso di una settimana) quali effettuano chiamate al di fuori della rete?

Questa richiesta di dati raggiunge un maggiore livello di dettaglio, fornendo informazioni rilevanti in merito a parametri relativi all'origine e alla destinazione delle chiamate che potrebbero spronare a proporre servizi aggiuntivi ai clienti più importanti o persino portare allo sviluppo di nuovi prodotti o servizi. Un esempio di classificazione di possibili analisi multidimensionali è rappresentato in Figura 4.

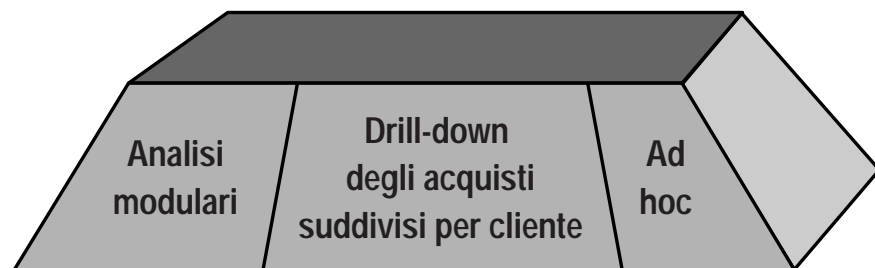


Figura 4. Un esempio di classificazione delle analisi multidimensionali.

Prendiamo ora in considerazione un esempio di drill-down effettuato con la tecnologia OLAP: sulla seguente query standard per il supporto alle decisioni si basano successive analisi multidimensionali.

Mostra le entrate registrate trimestralmente suddivise per grandi clienti aziendali nelle zone settentrionali, nord-occidentali e sud-occidentali per gli anni 1997 e 1998.

La risposta alla presente richiesta dati avrebbe una rappresentazione tabulare simile a quella riportata nella Tabella 1.

Osservando questa tabella è possibile notare che le entrate nelle regioni nord-occidentali sono inferiori rispetto alle altre e quindi un Product Manager potrebbe decidere di scendere a un ulteriore livello di dettaglio:

Tabella 1. Esempio di report base per il supporto alle decisioni.

	1997				1998	
	Q1	Q2	Q3	Q4	Q1	Q2
Nord	100	70	88	190	90	66
Nord-Ovest	76	86	86	100	76	45
Sud-Ovest	68	92	92	144	60	28

Entrate registrate (in milioni di lire)

Mostra gli stessi dati suddivisi per aree delle regioni nord-occidentali.

Supponiamo che la risposta a questa richiesta sia simile a quella riportata in Tabella 2:

Tabella 2. Emissione risultati OLAP.

		1997				1998	
		Q1	Q2	Q3	Q4	Q1	Q2
Nord		100	70	88	190	90	66
Nord-Ovest	Regione A	26	30	28	36	22	28
	Regione B	20	24	24	26	32	32
	Regione C	30	32	34	38	38	30
Sud-Ovest		68	92	92	144	60	28

Entrate registrate (in milioni di lire)

Dato che il distretto che presenta le entrate minori è quello B l'utente potrebbe voler esaminare i risultati dell'analisi relativi a questa singola area geografica per comprendere le ragioni del problema.

La Tabella 3 mostra chiaramente che le entrate di Città 3 sono inferiori a quelle registrate nel resto del distretto. Incrementando il livello di dettaglio l'utente ha quindi potuto collegare un problema nelle entrate a una specifica località non commettendo l'errore di considerarlo una questione di livello aziendale o territoriale.

Potrebbe esservi capitato di sentir parlare del termine slicing and dicing (affettare e ridurre a cubetti) che consiste nella capacità di plasmare le risposte a seconda delle preferenze dell'utente. Gli strumenti OLAP sono straordinari proprio per la facilità con cui consentono agli utenti di richiedere i dati in modi diversi. Dopo tutto una lista di prodotti con le relative date di vendita non è di aiuto al manager che ne sta analizzando il ciclo di vita e necessita che queste informazioni siano invece suddivise per zone geografiche, per codici area o per segmenti demografici.

Tabella 3. Emissione risultati OLAP.

		1997				1998	
		Q1	Q2	Q3	Q4	Q1	Q2
Nord		100	70	88	190	90	66
Nord-Ovest	Regione B	20	24	24	26	16	32
	Città 1	4	6	8	8	4	8
	Città 2	12	14	12	12	8	18
	Città 3	4	4	4	6	4	6
Sud-Ovest		68	92	92	144	60	28

Entrate registrate (in milioni di lire)

Modellazione e segmentazione

I modelli offrono alle aziende la possibilità di determinare in modo semplice e costante nel tempo il futuro comportamento della clientela, oltre a stabilire la redditività economico-finanziaria a lungo termine; la segmentazione invece consente di classificare e riclassificare la clientela in base alle caratteristiche demografiche, alle costanti di acquisto, alla propensione all'acquisto e ad altre categorie simili. Riportiamo di seguito alcuni esempi di possibili segmenti:

- clienti che rispondono a nuove promozioni;
- clienti che rispondo agli sconti;
- clienti che rispondo a offerte di nuovi prodotti;
- clienti che non rispondono alle promozioni.

La segmentazione offre suggerimenti su come un'azienda dovrebbe trattare una determinata fascia di clienti e consente di rispondere a domande come queste:

- Inizialmente a quale gruppo di clienti dovrebbe essere mirato un determinato nuovo servizio?

- Chi è più probabile che chiami per richiedere un servizio dopo che è stato messo a disposizione un nuovo servizio?
- Quali clienti potrebbero assumere più facilmente comportamenti fraudolenti?
- Quali clienti potrebbero rispondere con più probabilità ai saldi?

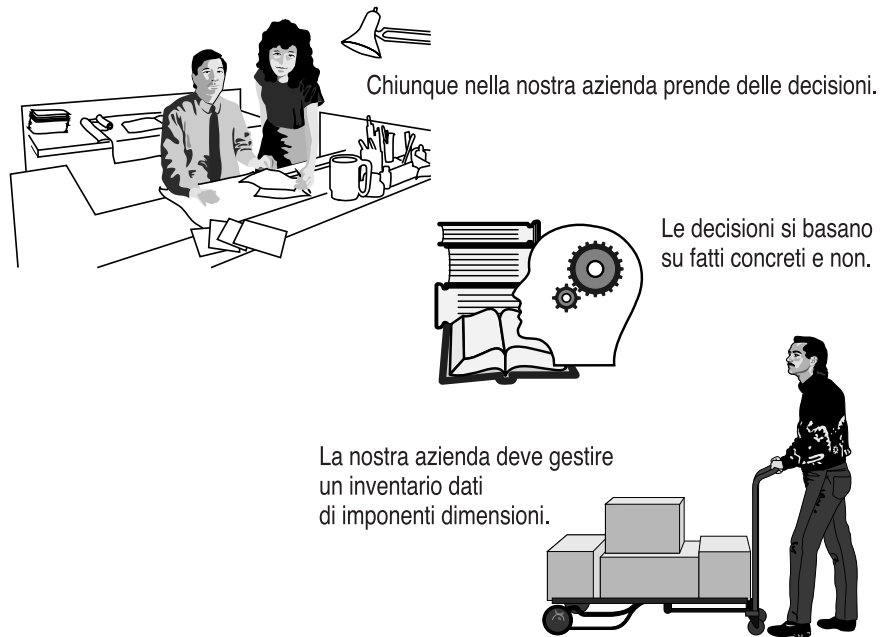


Figura 5. Un esempio di classificazione delle analisi effettuate mediante modellazione/segmentazione.

Un istituto di credito utilizza la modellazione per la valutazione del valore medio sia della perdita di clienti sia della durata della relazione di clientela (Figura 5), in entrambi i casi partendo dai dati storici contenuti nel Data Warehouse. Alcuni esempi di modellazione potrebbero essere i seguenti:

- **Durata media della relazione di clientela.** Quali sono gli attributi che influenzano il valore a lungo termine di un cliente per l'azienda.
- **Perdita della clientela.** Quante probabilità ci sono che un cliente lasci l'azienda? In quanto tempo lascerà l'impresa? Vi sono altre caratteristiche (configurazioni degli acquisti, rallentamenti nella prestazione dei servizi, o altro) che dovrebbero essere prese in considerazione quando si esamina la perdita della clientela?
- **Modellazione predittiva.** In che modo le vendite di un determinato prodotto saranno influenzate dal cattivo tempo di questa primavera? Che effetto avrà l'economia sulle vendite?

La categoria delle analisi mediante modellazione e segmentazione si colloca in una posizione intermedia tra le analisi multidimensionali e la Knowledge Discovery per una ragione molto valida: dopo essere stati creati, i segmenti e i modelli possono essere esaminati più approfonditamente ricorrendo a strumenti di tipo multidimensionale oppure possono essere ridefiniti in modi nuovi e imprevedibili utilizzando la Knowledge Discovery.

Knowledge Discovery

A differenza degli altri tipi di analisi, la Knowledge Discovery è una analisi che parte da "zero ipotesi" e dato che non sappiamo quali domande porre non possiamo prevedere quali saranno le risposte che otterremo. La Knowledge Discovery, nota anche come scoperta non guidata della conoscenza (undirected knowledge discovery), scopre i modelli nascosti nei dati che riflettono il comportamento della clientela, le disdette, gli acquisti futuri o altri eventi aziendali. Questi modelli sono troppo particolari e apparentemente arbitrari per essere specificati e quindi l'analista sarebbe costretto a un continuo "gioco degli indovinelli" per cercare di immaginare i possibili modelli

esistenti in un database. Fortunatamente il procedimento è esattamente l'opposto: speciali strumenti di Knowledge Discovery scoprono questi schemi e li comunicano all'analista specificando di cosa si tratta e dove reperirli.

Naturalmente, poche aziende possono giustificare una simile ricerca randomizzata tra i propri e-Data, specialmente se si considerano i costi degli strumenti e la complessa preparazione necessaria per analizzare i dati. Per queste ragioni prima di attuare una simile ricerca dati è comunque opportuno stabilire un, seppur vago, obiettivo aziendale.

Una casa farmaceutica utilizza la Knowledge Discovery per esaminare cosa accade quando certe medicine vengono prese contemporaneamente, rendendo possibile l'analisi dei modelli di affinità dei prodotti. Inoltre l'azienda analizza i successivi acquisti di prodotti in modo da determinare se certi medicinali portino a necessitarne degli altri (Figura 6).

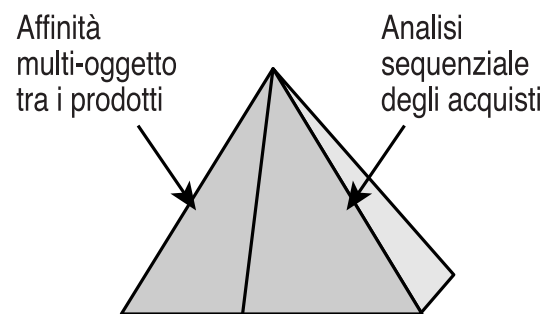


Figura 6. Un esempio di classificazione di analisi effettuate mediante la Knowledge Discovery.

L'analisi delle affinità è una delle molte analisi di Knowledge Discovery possibili. Un potente algoritmo delle affinità legge tutti gli elementi informativi contenuti nel Data Warehouse e collega ogni attributo con ogni altro oggetto informativo prescindendo dal loro significato. Sebbene questa possa sembrare a prima vista una analisi "forzata", il suo valore effettivo è dato dai risultati cui porta. Il risultato è

un insieme di regole che descrive determinate relazioni tra i dati e queste regole possono portare sorprese interessanti.

Considerate il valore dei risultati della seguente analisi di affinità:

Dei clienti che acquistano patatine, il 63 per cento acquista anche chupa-chups.

Questa è una scoperta interessante soprattutto se si considera che un'azienda non aveva modo altrimenti di provare questa verità senza ricorrere alla Knowledge Discovery. Un analista non avrebbe mai potuto giungere con assoluta certezza a una simile conclusione considerando le innumerevoli relazioni esistenti tra i dati e la capacità di elaborazione necessaria per collegare esponenzialmente ogni attributo con ogni altro.

Tuttavia, le potenzialità dell'analisi delle affinità diventa palese quando consente di rilevare affinità "multi-oggetto":

Dei clienti che acquistano patatine e chupa-chups, il 74 per cento acquista anche vino rosso.

Grazie a questo tipo di informazione, che non avrebbe potuto essere raccolta servendosi di query e analisi multidimensionali ma che risiede nel Data Warehouse come una pepita d'oro, il proprietario di un alimentari potrebbe attuare una serie di azioni, dalla spedizione di coupons ai clienti che acquistano più spesso uno di questi tre prodotti, all'eliminazione di sconti sul vino rosso, ad una più efficace disposizione dei prodotti nel negozio.

Riportiamo di seguito un sorprendente elenco di dati che possono essere acquisiti mediante la Knowledge Discovery.

- Si possono individuare i prodotti che innescano altri acquisti. Per esempio, la probabilità che una agenzia di intermediazione riesca con successo a commercializzare una carta di debito ai clienti che hanno già un conto corrente bancario è relativamente alta.

- Si può prevedere quale sarà il prossimo acquisto. Il tipo di analisi sequenziale degli eventi combina il comportamento della clientela con lo storico delle vendite dei prodotti per stabilire delle costanti negli acquisti in modo da prevedere quale sarà l'acquisto successivo di un determinato cliente o di un segmento di clienti prima che questo venga effettuato.
- Si possono rilevare costanti nelle flessioni delle vendite o nelle disdette di un servizio. In altri termini, la Knowledge Discovery consente di determinare le ragioni, i fattori, che hanno portato a queste flessioni o alla disdetta di un servizio. Per fare un esempio, questo tipo di analisi può consentire a una Pay TV di prevedere eventuali disdette del servizio in modo più efficace che ricorrendo ai modelli di perdita standard.
- Si possono determinare le caratteristiche del cliente che influenzano la durata della vita di un prodotto. In un caso è stato possibile determinare che le famiglie nelle quali i figli si erano appena diplomati erano più propense a permutare una vecchia auto con una nuova.

Nell'ambito della Knowledge Discovery i dati non mentono mai, l'unico aspetto discutibile resta l'attendibilità dell'interpretazione di questi dati. Proprio come accade nel caso della modellazione e della segmentazione, anche i risultati delle analisi di Knowledge Discovery possono essere presentati in modi diversi e devono essere analizzati da specialisti che abbiano una certa familiarità sia con gli strumenti principali sia con i dati che risiedono nel Data Warehouse.

Perché si ricorre al Data Mining?

Notate che fino a questo momento non abbiamo mai utilizzato il termine Data Mining per una ragione molto valida: il Data Mining non è una vera e propria tipologia di

analisi quanto piuttosto una classificazione delle diverse tipologie di analisi dei dati.

Molte case produttrici di software, decise a incrementare le vendite, hanno confuso le idee della clientela presentando i propri prodotti come veri e propri strumenti di Data Mining. Alcuni fornitori piuttosto ambigui, ma soprattutto che hanno mostrato la tendenza a sostituire il termine "supporto alle decisioni" con "Data Mining", si sono sbilanciati fino al punto di affermare che questo termine comprende tutti i livelli della piramide.

Noi preferiamo considerare il Data Mining una categoria di analisi che comprende sia la modellazione/segmentazione sia la Knowledge Discovery come illustrato in Figura 7.

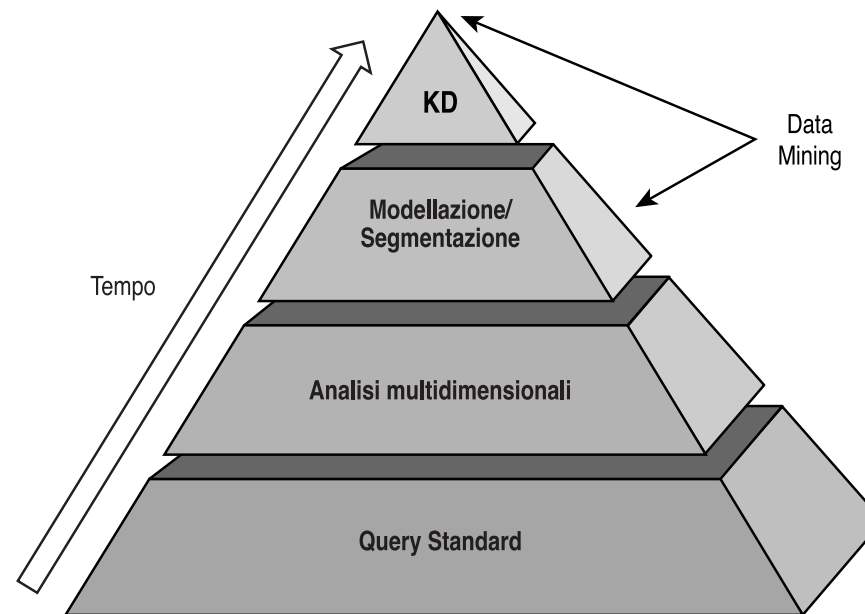


Figura 7. Dove è corretto collocare il Data Mining.

Il Data Mining da sempre è stato sinonimo di analisi statistica realizzata a livello accademico per sviluppare

un'accurata comprensione dei dati dettagliati. Sempre più spesso, anche gli utenti non esperti di statistica, grazie agli sforzi dei fornitori, possono utilizzare con minori difficoltà gli strumenti di Data Mining riuscendo più facilmente a rilevare costanti interessanti nell'ambito dei dati, ma anche a sfruttarne efficacemente i risultati.

Nel corso del manuale vedremo come il Data Mining sia stato implementato nelle diverse realtà aziendali mettendo in evidenza i benefici che le singole imprese ne hanno tratto. Quando più avanti menzioneremo nuovamente il termine "Data Mining" lo faremo per indicare l'effettiva attuazione di avanzate analisi dati, ma avremo cura di specificare di volta in volta se ci si riferisce in particolare alla modellazione, alla segmentazione o alla Knowledge Discovery.

Il Data Warehousing nell'effettiva realtà aziendale

La maggior parte dei Data Warehouse oggi esistenti viene utilizzata per lo più per effettuare comuni funzioni operazionali; questa affermazione non vuole assolutamente essere critica quanto piuttosto evidenziare un aspetto particolarmente proficuo dell'utilizzo di Data Warehouse operazionali. Sebbene le funzioni operazionali, quali per esempio la realizzazione di resoconti sulle entrate o di analisi delle vendite, possano non essere complesse o affascinanti quanto, per esempio, le analisi sul valore della durata della relazione di clientela, va sottolineato che si tratta comunque di procedure aziendali estremamente frequenti. In effetti, la maggior parte delle aziende di cui si tratterà in questo manuale, si sono buttate nel Data Warehousing iniziando dalla realizzazione di semplici report che riportavano le entrate mensili o gli indirizzi dei clienti; per quanto questi report potessero essere semplici consentivano comunque all'utente di avere una prima impressione sui dati trattati. Questi inizi furono forieri di un mutamento ben più rilevante.

Una relazione del 1999 elaborata dal Palo Alto Management Group ha confermato questa tendenza generalizzata di sfruttare il Data Warehousing essenzialmente per report basilari; questa ricerca ha infatti determinato che le ricerche dati effettuate più di frequente sono quelle legate alle analisi finanziarie o alla redazione di report generici. In Figura 8 è rappresentato lo sfruttamento delle diverse azioni di supporto alle decisioni in base alla popolazione che usufruisce delle tecnologie del Data Warehousing.

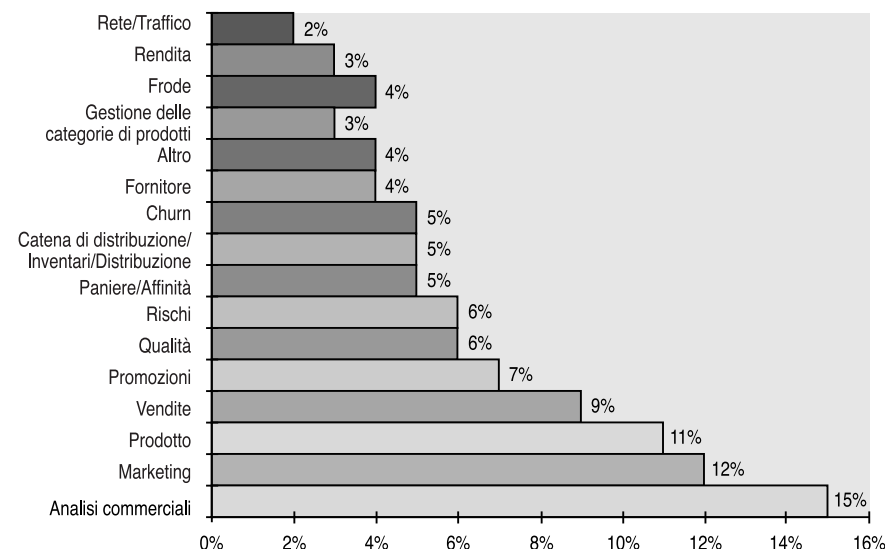


Figura 8. Quote di applicazione corrente del Data Warehousing (Fonte: Palo Alto Management Group).

Sempre in base a questo studio i report di Business Intelligence starebbero abbandonando gradualmente l'area di influenza degli utenti esperti per essere accolti dall'intero

organico aziendale³. In altri termini la punta della piramide (query effettuate da utenti aziendali di medio livello) rappresenta la tendenza principale nell'ambito dell'utilizzo di Data Warehouse e Data Mart.

Cosa è necessario fare per raggiungere la cima

Torniamo ai nostri amici dell'istituto di credito che, come abbiamo visto, senza nemmeno essere in grado di effettuare query standard, pretendevano di passare direttamente alla modellazione e sicuramente alla Knowledge Discovery, senza approntare l'opportuna infrastruttura tecnologica.

Cerchiamo di capire ciò di cui avrebbero avuto bisogno se fossero stati pronti per passare alla modellazione predittiva. La maggior parte delle aziende che hanno scalato la piramide fino alla cima presentano tratti in comune:

- I dati che hanno individuato per effettuare le analisi esistono sul minor numero possibile di piattaforme, idealmente su un Data Warehouse o un Data Mart centralizzati. Questo accorgimento consente di risparmiare tempo nel consolidamento e nella combinazione dei dati.
- I loro utenti aziendali sono abituati ad accedere al Data Warehouse in prima persona per ottenere le informazioni di cui necessitano.
- Hanno nel loro organico utenti di diversa esperienza, dai neofiti non specializzati a utenti estremamente esperti.
- Hanno un insieme di strumenti applicativi diversi mediante i quali è possibile analizzare i dati. Uno strumento non può avere la stessa importanza per tutti gli utenti, pertanto una serie di strumenti consente alle

differenti categorie di utenti di utilizzare quello a loro più congeniale, oltre a mostrare al personale la volontà da parte dell'azienda di utilizzare il Data Warehouse in modo eterogeneo.

- Queste aziende credono fermamente nel valore aziendale di un Data Warehouse e mettono a disposizione fondi per progetti in fase di sviluppo.
- Uno o più membri dell'esecutivo di formazione economica sostiene le attività di supporto alle decisioni future o già in uso.
- Sono aziende determinate a modificare le proprie procedure e consuetudini aziendali. Il Data Warehousing può comportare per esempio miglioramenti nella produttività, trasformazioni culturali, un incremento della redditività: tutti cambiamenti che l'azienda deve essere in grado di supportare adeguando la propria organizzazione.

Questi sono alcuni degli aspetti che è necessario approntare se si vuole che la propria azienda faccia il grande passo verso analisi avanzate. Tuttavia non disperatevi se state ancora eseguendo query standard, potreste comunque essere ugualmente un passo più avanti rispetto ai vostri concorrenti. Come dice un vecchio proverbio cinese: "Un viaggio di migliaia di chilometri inizia con un solo passo".

³ Palo Alto Management Group, *Business Intelligence and Data Warehousing: Crossing the Millennium*, 1992, 12. Per un sunto della ricerca visitate il sito Internet www.pamg.com