

## *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

### **Presupposti teorici del Data Mining e mappa dell'innovazione in corso**

*Lorenza Saitta - Università del Piemonte Orientale, Dipartimento di Informatica*

Secondo una prospettiva che sta influenzando le politiche di ricerca in ambito Europeo, l'Era della nuova informazione darà forma a una nuova società in cui l'informazione sarà accessibile ad ogni singolo individuo. Per raggiungere questo obiettivo, l'ubiquità dei dati e delle informazioni deve essere accoppiata all'ubiquità degli strumenti utilizzati per far emergere la conoscenza. Questa visione è una fonte di nuovi bisogni e di molte aspettative, alcune delle quali riguardano gli *individui*, altre le *imprese* e, infine, altre sono di interesse per la società tutta.

Grazie alla grande quantità di dati disponibili, l'informazione estratta può diventare il nucleo di sistemi di conoscenza adattabili, robusti e scalabili, facili da integrare con i database esistenti. Naturalmente i database non sono l'unica risorsa di dati di cui ha bisogno la *conoscenza nascosta*: documenti di testo, immagini, dati audio, il Web. Il mantenimento dei software esistenti diventa sempre più importante, visto che il software ha rimpiazzato i processi meccanici.

Le nuove sfide per la *scoperta della conoscenza* sono la natura diffusa delle data warehouses, delle risorse di calcolo e delle applicazioni: un Data Mining diffuso otterrà vantaggi dalle reti di comunicazioni wireless e dalle reti di computer.

## ***Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente***

### **Data Mining: dall'automazione delle procedure all'automazione di processo**

*Flavio Bonifacio - Metis Ricerche S.r.l*

L'obiettivo della relazione è quello di attirare l'attenzione sulle peculiarità del Data Mining come strumento importante dell'Analisi Dati. Quest'ultima viene vista come il campo specifico di applicazione di professionalità nuove, che uniscono competenze diverse: la statistica, l'informatica, la matematica, la filosofia. La specificità di questa nuova professione viene individuata proprio nella sua doppia trasversalità, tra i domini e tra le professionalità.

Nella relazione si illustrano brevemente alcuni antefatti dell'attuale situazione e si focalizza l'attenzione su tre aspetti intorno a cui si costruisce l'evoluzione e il campo di applicazione del data Mining: l'evoluzione degli strumenti di analisi statistica, l'evoluzione delle interfacce, l'evoluzione dell'organizzazione del lavoro e delle competenze professionali.

Si definisce in primo luogo che cosa si intende per automazione delle procedure: come si passa dalla programmazione della statistica al suo utilizzo con linguaggi evoluti.

In secondo luogo si mostrano esempi di come l'evoluzione delle interfacce grafiche (Graphical User Interface) abbia influito sullo sviluppo del software medesimo e come le competenze richieste si siano modificate.

In terzo luogo si costruiscono esempi di analisi statistica e di Data Mining al fine di rendere evidenti le peculiarità dei due percorsi che vengono infine enucleate. Sono altresì elencate le tecniche del Data Mining.

Infine viene riformulata la definizione del Data Mining come insieme complesso di procedure statistiche, utilizzabili in modo semiautomatico con interfacce grafiche su grandi masse di dati. Sono poste in rilievo le caratteristiche che fanno diventare il Data Mining uno strumento forte di dislocazioni professionali alternative su percorsi organizzativi non consueti.

## *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

### **L'analisi dei dati nel marketing assicurativo**

*Oreste Porreca - Reale Mutua Assicurazioni*

Il primo problema dell'Analisi dei Dati in generale e del Data Mining in particolare è il reperimento del bro oggetto.

Vi è innanzitutto la necessità di una comprensione dei contenuti delle basi dati: molto spesso, anzi quasi sempre, le necessità di tipo organizzativo e gestionale per cui i dati sono stati raccolti non coincidono con gli scopi in cui altre funzioni aziendali li vorrebbero utilizzare. Ciò è vero sia nel marketing assicurativo come in altri campi: nella pubblica amministrazione o in ambito sanitario ad esempio.

Il percorso di reinterpretazione dei dati, che spesso implica una loro ricostruzione storica, dovrebbe dar luogo ad una *definizione condivisa* nell'ambito del nuovo dominio di utilizzo e, possibilmente, anche nel vecchio. Anche se questo problema sarebbe, a rigor di logica, di pertinenza dell'area *warehousing* non è inopportuno sottolinearlo qui. Molto spesso infatti il warehouse si scontra con il problema definitorio di cui sopra, data l'intrinseca necessità di operare astrazioni generalizzanti per arrivare a repository d'informazione general purpose, polifunzionali. Cosa che ne ritarda l'operatività e lo fa essere quasi sempre "non ancora pronto". Per questa ragione il problema del Data Mining non è in principio un problema del Data Mining, ma un problema del data warehouse.

Un altro problema che si pone all'altro capo del percorso è la descrizione degli obiettivi. Il Data Mining ha compiti facilitati quando sa per che cosa scavare. Ma quest'attività non gli deve essere estranea o sopraggiungere dall'esterno: è parte integrante delle proprie funzioni. Pertanto, prima di essere un bagaglio di tecniche statistiche, il Mining deve consistere in attività concettuali, di affinamento e perfezionamento "di idee", che adotta metodiche interattive di lavoro. Il Miner *aiuta* a reperire i dati ed a strutturare un percorso logico finalizzato: egli sa però che entrambi esistono già in qualche forma nell'organizzazione che lo ospita. In una forma che bisognerà scoprire.

Quest'attività *maieutica* si esprime proprio in gruppi di lavoro votati al contenuto, per così dire, prima che alle tecniche. Oltre che Data Miner, l'esperto di tecniche dovrà diventare un *concepts mixer*, un frullatore di concetti.

In terzo luogo il Data Mining deve scoprire il proprio valore aggiunto: un'attività di Mining che inizi nel 2004 e termini nel 2008, per quanto interessante e foriera di buoni risultati, alla fine risulterà inutile. Beninteso possono essere concepite ancor oggi attività di lunga durata. Il tempo dipende infine dalla dimensione dei problemi. Non è concepibile però che non siano previsti punti di emersione e di verifica intermedi. La misura dei risultati è infatti un terzo, importantissimo compito dell'attività di Data Mining. Le previsioni di vendita elaborate dall'ufficio del marketing devono poter trovare un riscontro nelle campagne di vendita sia a fini di validazione che di affinamento dei modelli. Tra l'altro è proprio attraverso l'analisi dei risultati delle diverse attività di marketing e di vendita che, in un contesto commerciale, il Data Mining misura sé stesso.

# *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

## **L'analisi dei dati in ambito epidemiologico**

*Roberto Zanetti – Centro Prevenzione Oncologica*

Io mi sono figurato il *Data mining* come un punto del labirinto delle conoscenze contemporanee ed ho pensato che il mio contributo avrebbe potuto consistere nel cercare insieme a voi risposte a queste due domande:

1. Come siamo arrivati a questo punto del labirinto?  
Affronto questa domanda dividendola in tre temi:
  - a. Come è cambiato lo “status” di ciò che chiamiamo dati
  - b. Come è evoluta la Statistica
  - c. Come è evoluta l'Informatica (almeno in relazione alla Statistica)

Che affronterò succintamente con enunciati corredati di riferimenti (almeno sulle epoche e gli autori)

2. Come “uscire” dal Data mining

## *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

### **Data Mining e insegnamento dell'analisi dati**

*Luca Ricolfi - Università degli studi di Torino, Facoltà di Psicologia*

Parlare di Data Mining nei corsi universitari che si occupano di analisi dei dati è reso difficile da tre circostanze:

- a) L'insegnamento dell'analisi dei dati non è mai stato completamente istituzionalizzato nell'università italiana, e viene tuttora confuso sistematicamente con la metodologia, la statistica e l'informatica.
- b) Sono rarissimi i casi di corsi di laurea che prevedono una successione di insegnamenti abbastanza articolata e lunga da permettere di arrivare a un vero e proprio corso di Data Analysis (dopo quelli di statistica e di metodologia).
- c) È inverosimile che nel nuovo ordinamento questa situazione possa migliorare.

Queste circostanze sfavorevoli contrastano con la crescente importanza che il Data Mining riveste, o potrebbe rivestire, negli sviluppi attuali dell'analisi dei dati.

# *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

## **La Ricerca in ambito Text Mining**

*Maurizio Lana - Università del Piemonte Orientale*

Lo studio di testi per mezzo del computer è un'attività dalle molteplici facce: aziendale, commerciale, universitaria, di ricerca, di controllo. Oggi in generale non esiste un contesto universitario specifico e definito dove tutti questi vari aspetti siano presenti. Così chi voglia studiare testi per mezzo del computer si trova ad occupare una posizione *borderline* sia rispetto agli studiosi di testi sia rispetto agli informatici, dagli uni e dagli altri guardato con un misto di curiosità, scetticismo, incredulità.

Il punto di vista che verrà presentato non è quello di chi lavora nel campo della ricerca sociale, ma molti dei temi e delle problematiche sono interdisciplinari e quindi, ritengo, ugualmente interessanti.

Primo risultato è che occorre fare molto da sé per imparare, non esistendo curricula predefiniti. Secondo risultato è che il software che si usa è di solito software commerciale e non software sviluppato in un contesto di ricerca misto (studiosi di testi e informatici).

Nella mia esperienza le ricerche che ho svolto hanno sempre impiegato software commerciale, mai software sviluppato ad hoc. E ho sempre lavorato (e lavoro tutt'ora) con un esperto del mezzo, cioè un esperto degli strumenti informatici per l'analisi dei testi. Questo perché i software sono spesso ostici anche solo dal punto di vista dell'interfaccia utente e delle modalità operative.

Un percorso come quello appena vagamente delineato qui sopra ha però delle ragioni di interesse per le quali val la pena di affrontare e – si spera – superare le difficoltà: primo, utilizzare metodi noti di analisi dei testi, accelerandone enormemente i tempi di esecuzione; secondo, utilizzare metodi di analisi dei testi specificamente resi possibili dall'uso del computer. I testi da analizzare possono essere dei tipi più vari: un'indagine sulla lettura realizzata presentando a 2000 studenti di scuola media superiore; le relazioni di commento al bilancio prodotte da un'azienda negli ultimi 15 anni per gli azionisti; tutti i documenti prodotti da un'azienda negli ultimi 5 anni; gli articoli pubblicati dai quotidiani italiani nel luglio 2001 sul G8 di Genova; e altro ancora.

Due progetti di ricerca verranno presentati con un certo livello di dettaglio (ma essendo in pieno sviluppo, non ne verranno esposte conclusioni): una ricerca sul linguaggio usato nei principali quotidiani italiani per parlare del G8 di Genova; una gara di attribuzione di testi ("Ad-Hoc Authorship Attribution Competition"), promossa da P. Juola, Dipartimento di Matematica e Informatica della Duquesne University di Pittsburgh.

Di una terza ricerca più semplice, nata e sviluppatasi in contesto eminentemente didattico, si potrà parlare compiutamente in quanto conclusa: uno studio sul personaggio della *Fata Turchina*, o *Fata dai capelli turchini*, ne "Le Avventure di Pinocchio" da cui emerge con chiarezza quanto piccola sia la conoscenza che si crede di avere dei testi, quanto grande sia la parte di ignoranza, e quanto aiuto possano dare gli strumenti informatici, anche elementari, di analisi dei testi per approfondire le proprie conoscenze portandole fino a scoperte talvolta sorprendenti.

## *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

### **Il Data Mining nella Pubblica Amministrazione: un esempio**

*Giuliana Bonello - CSI-Piemonte*

I dati descrivono oggettivamente eventi o situazioni, mentre le informazioni danno un significato ai dati. Il processo che porta il dato a diventare informazione e quindi conoscenza è un modello tipico di Business Intelligence.

Business Intelligence significa quindi trasformare il patrimonio di informazioni, organizzato in opportuni modelli, in conoscenza, che permette di interpretare andamenti, formulare previsioni, eseguire simulazioni per ottenere un quadro operativo e strategico.

La Pubblica Amministrazione piemontese ha avviato da anni la realizzazione di sistemi informativi decisionali trasversali agli Enti per integrare banche dati e per offrire strumenti di Business Intelligence di accesso ed analisi semplici e flessibili.

Dal 1997 il CSI-Piemonte ha avviato per la Pubblica Amministrazione piemontese la costruzione dello strato informativo - decisionale (Data Warehouse) e degli strumenti per accedervi in modalità web (Business Intelligence) su una infrastruttura tecnologica e applicativa basata sulla Server Farm centralizzata presso il CSI.

Questo patrimonio informativo è una fonte molto ricca per provare ad utilizzare tecniche di analisi specializzate quali quelle del Data Mining.

Nel corso del 2003 è stato possibile avviare alcune sperimentazioni in questo ambito grazie all'iniziativa promossa dal Settore Sistemi Informativi ed Informatica della Regione Piemonte, denominata Laboratorio di Business Intelligence.

Sono state sperimentate tecniche di Data Mining su alcuni contesti di interesse regionale, attingendo da alcuni bacini informativi particolarmente ricchi. In particolare le esperienze riguardano:

- analisi del rispetto delle linee guide nel trattamento dell'ipertensione
- segmentazione delle scuole materne del Piemonte
- modello di previsione del successo della formazione professionale

Inoltre è stata avviata un'analisi di Web Mining di un sito regionale nell'ambito del progetto Catalogo dei servizi decisionali.

Sulla base di queste esperienze è possibile fare alcune considerazioni.

Nell'ambito privato ci sono già molte esperienze consolidate e specializzate di uso di tecniche statistiche avanzate, mentre nell'ambito della pubblica amministrazione le esperienze sono meno diffuse: questo contrasta col fatto che invece il patrimonio informativo è molto più grande nell'ambito pubblico rispetto a quello disponibile in ambito privato.

C'è quindi molta strada da fare per sfruttare al meglio le opportunità offerte dal Data Mining; se da un lato occorre continuare ad organizzare le basi dati esistenti in ottica di uso decisionale, dall'altro è opportuno far crescere la conoscenza delle potenzialità di queste tipologie di analisi come strumenti di conoscenza e di supporto alle decisioni nell'ambito della pubblica amministrazione.

## *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

### **Data Mining e CRM: l'evoluzione del contesto analitico nelle aziende**

*Guido Cuzzocrea - Nunatac*

Quali prospettive di inserimento in azienda aveva un neo-laureato o diplomato in statistica, non più di quindici anni fa? Oppure, volendo affrontare il tema dal lato della domanda, quali aziende e perché andavano alla ricerca di competenze analitico-quantitative da inserire nel proprio contesto organizzativo?

Se escludiamo i centri di ricerca e gli uffici studi, per un motivo o per l'altro non direttamente integrati nei processi di business delle rispettive aziende, le sole realtà in cui la figura dello statistico assumeva un ruolo centrale erano le società per le ricerche di mercato. Poche erano le eccezioni a questa regola generale.

L'attività svolta dal team degli statistici di una società di marketing research si basava su alcune solide fondamenta: definizione degli obiettivi di misurazione e conseguente redazione di un apposito questionario, sottoposto ad un campione di individui (disegnato per poter rappresentare il fenomeno); raccolta dei nella forma e al livello di dettaglio corretti per essere analizzati; controlli di coerenza (che beneficiavano delle dimensioni limitate della base dati e della conoscenza a priori della struttura logica del questionario); fase inferenziale che si poggiava su un percorso pre-definito in cui risultava determinante la verifica di talune ipotesi modellistiche; valutazione di bontà dei risultati basata sulla validità del metodo, la veridicità delle ipotesi e la rappresentatività campionaria.

Cosa è cambiato da allora?

La statistica e l'analisi del dato per le ricerche sociali e di mercato hanno seguito una propria evoluzione, mantenendo le loro caratteristiche di metodo e rigore.

Nel frattempo, il mutato contesto di mercato e lo sviluppo della tecnologia hanno spinto le aziende verso un utilizzo massivo dei dati prodotti dalle rispettive gestioni operative (milioni di transazioni se pensiamo ad esempio a banche o telefoniche), rivolto alla conoscenza di caratteristiche, comportamenti e aspettative dei propri clienti.

In questo affonda le proprie radici il Customer Relationship Management (CRM) o, quantomeno, l'Analytical CRM.

Soltanto per elencare alcuni aspetti salienti che l'intervento cercherà di chiarire, diciamo che:

- la mole di dati a disposizione è enorme, ma la loro natura è strettamente connessa all'attività gestionale dell'azienda, il livello di dettaglio e la qualità sono variabili;
- alcuni elementi di metodo della statistica classica lasciano il posto alle esigenze di interpretazione dei risultati e di costruzione di processi decisionali integrati nell'operatività dell'azienda;
- la bontà delle regole e degli indicatori predittivi stimati viene misurata empiricamente;
- la funzione "Analytics" assume sempre più un ruolo strategico e trasversale rispetto alle funzioni di business;
- in questo scenario, il ruolo del software è quello di strumento facilitatore.

## *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

### **Un esempio di analisi dei testi: le guide ai ristoranti italiani**

*Roberto Fontana - Agenzia Regionale per la Promozione Turistica del Piemonte Osservatorio Turistico*

Creare una mappa della ristorazione piemontese, confrontarla con quella italiana ed acquisire informazioni sulla sua clientela. Questo, in sintesi, l'obiettivo del progetto avviato nel 2002 dall'Osservatorio Turistico della Regione Piemonte. Questa relazione presenta alcuni sviluppi recenti dello studio ed integra quanto illustrato nel corso dell'ultimo convegno "SAS-I Days", svoltosi a Milano il 22 e 23 ottobre 2003.

Il turismo enogastronomico sta conquistando, con il passare del tempo, una rilevanza sempre maggiore in generale e, per il Piemonte, in particolare. I ristoranti tipici e di qualità, che fanno della cucina una vera e propria arte, ne costituiscono uno degli elementi fondamentali.

Una parte rilevante del lavoro è stata dedicata allo studio delle "guide ai ristoranti", che sono disponibili nelle librerie e che vengono consultate da chi è alla ricerca di un posto in cui apprezzare la buona tavola. Si tratta infatti di una fonte molto rilevante di informazioni sia per l'elevato grado di copertura del territorio sia per la cura con cui sono redatte. Sarebbe risultato tuttavia molto difficile e avrebbe richiesto tempi proibitivi esplorare in modo sistematico questa notevole quantità di dati se non si fosse avuto a disposizione il formato elettronico di alcune delle guide più importanti e se non si fosse utilizzato un software idoneo all'effettuazione di analisi statistiche approfondite (Enterprise Miner e SAS Enterprise Guide).

In particolare un ruolo chiave è stato svolto dal modulo per il Text Mining, che ha permesso di elaborare la parte testuale delle guide, fornendo in questo modo indicazioni in grado di tener conto dell'intera descrizione del locale e cioè non soltanto della parte quantitativa (prezzi, numero di coperti) ma anche di quella qualitativa, che è in forma discorsiva.

Ad integrazione dello studio delle guide è stata condotta un'indagine presso i titolari dei ristoranti di eccellenza per acquisire informazioni sulla clientela e sulle professionalità di questo tipo di locali.

L'esperienza dimostra come la Business Intelligence fornisca un valido aiuto anche per addentrarsi in un settore ricco e variegato come quello del turismo in generale e della ristorazione in particolare.

Si ringraziano gli editori Slowfood, Touring Club e Veronelli per aver fornito il formato elettronico di alcune sezioni della propria guida ed i titolari dei ristoranti di eccellenza che hanno risposto al questionario.

Una parte del lavoro è stata condotta in collaborazione con Epat-Fipe Piemonte e SAS.

## *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

### **Un esempio di analisi dei testi**

*Gianluca Bo - Metis Ricerche S.r.l*

Negli ultimi anni si è assistito ad un crescente interesse da parte della Business Intelligence alle tecniche di analisi dei testi. La crescente quantità di informazioni di tipo testuale facilmente disponibili per le organizzazioni (si pensi ad esempio allo sviluppo di Internet) ha di fatto reso prezioso l'affiancamento di tecniche di Text Mining ai più tradizionali strumenti di analisi dati.

Così come le tecniche di Data Mining consentono di estrarre informazioni da database strutturati (*Knowledge Discovery in Database*), le tecniche di Text Mining rendono possibile l'estensione di tale approccio ad informazioni non strutturate (documenti di testo) in possesso di un'organizzazione (*Knowledge Discovery in Text*).

Nel processo di produzione di conoscenza il Text Mining rappresenta dunque un risorsa in più a disposizione del *Knowledge Information Worker*, uno strumento il cui ruolo può essere non solo alternativo a quello del più tradizionale Data Mining, ma in taluni casi complementare.

Un esempio classico di elaborazione di dati testuali con tecniche multivariate è quello proposto sin dagli anni '80 da Lebart e Salem nell'ambito della scuola di Benzecri (*Analyse statistique des données textuelles: question ouvertes et lexicométrie* di L. Lebart, A. Salem Paris Dunod, 1998).

L'Analisi delle Corrispondenze Lessicali (ACL) consente infatti di visualizzare l'associazione tra elementi testuali (parole o segmenti) contenuti in una particolare tabella di contingenza (tabella di Burt). L'analisi delle corrispondenze consente inoltre di studiare l'associazione tra elementi testuali e categorie di risposta rilevate per mezzo di domande chiuse fornendo un valido aiuto nell'interpretazione delle categorie di queste ultime.

Verranno presentati alcuni esempi di ricerca. Il primo illustra appunto l'analisi congiunta di domande chiuse e di domande aperte raccolte nel corso di una indagine telefonica (CATI) sull'atteggiamento dei residenti a Torino nei confronti dei prossimi Giochi Olimpici Invernali. L'elaborazione congiunta di questi due tipi di dati (utilizzo complementare di tecniche di Data Mining e di Text Mining) consentirà di comprendere le diverse sfumature che caratterizzano l'atteggiamento dei torinesi; informazioni preziose per chi (*decision maker*) ha la responsabilità di curare la campagna di informazione dei Giochi Olimpici Invernali 2006.

Quindi, rimanendo in tema, si cercherà di fornire una panoramica dell'immagine che i frequentatori la Fiera del Libro hanno dato delle Alpi attraverso tre domande aperte.

Infine verrà presentato un caso nel quale l'elaborazione di domande aperte è stata utilizzata nell'analisi della *customer satisfaction* dei clienti di una azienda.

## *Data Mining e metodologia della ricerca sociale: la creazione di valore aggiunto per l'utente*

### **SAS e il Data Mining. Principali ambiti di applicazione e prospettive**

*Elena Fabbris - SAS*

Il termine Data Mining identifica un processo di esplorazione e di modellazione del Customer Data Base (CDB) finalizzato all'estrazione dell'informazione in esso contenuta ed all'identificazione di comportamenti non facilmente individuabili ma utili ai fini del business.

La capacità degli attuali strumenti informatici di raccogliere e memorizzare quantità sempre maggiori di dati ha fatto aumentare notevolmente la disponibilità di informazioni da sottoporre ad analisi statistiche.

Di fronte ad una tale esplosione di dati, l'uso singolo e separato delle diverse tecniche di analisi statistica risulta spesso incapace di gestire il patrimonio informativo in essi contenuto.

La discriminante tra Data Mining ed Analisi statistica risiede proprio nella possibilità di rendere processo l'analisi statistica. Gli applicativi di Data Mining infatti prevedono una trascrizione delle tecniche statistiche nell'ottica di renderle concatenabili tra loro, attraverso un flusso di input-output dal contenuto il più possibile omogeneo. In questa accezione la singola tecnica perde di importanza mentre il processo, ossia l'intero percorso gestionale/informativo/decisionale, diventa il "focus".

Si introduce quindi il concetto di "**strategia di analisi**" che concretizza l'idea di una possibile integrazione tra differenti tecniche di analisi allo scopo di estrarre l'informazione contenuta nei dati.

Nel portafoglio delle soluzioni SAS esistono da tempo componenti dedicate all'analisi dei dati largamente impiegate da aziende di tutti i settori merceologici per effettuare analisi del CDB.

Lo strumento SAS per il Data Mining si pone come obiettivo quello di supportare al meglio lo svolgimento della attività di analisi dei dati presenti nei vari ambiti aziendali.

Volendo classificare in quattro principali aree le applicazioni del Data Mining potremo ipotizzare una suddivisione di questo tipo:

- il **Predictive Modeling**, si pone come obiettivo quello di prevedere il comportamento del cliente rispetto ad un target di interesse, fissato in base all'obiettivo di business da raggiungere.
- la **Customer Segmentation** ha come obiettivo quello di identificare raggruppamenti di clienti con comportamenti simili rispetto alle variabili di analisi.
- Il **Web Mining** consiste nell'applicare tecniche di Data Mining ai dati di navigazione sul Web.
- Il **Text Mining** si colloca nel panorama della Knowledge Discovery in Text (KDT), che si differenzia dalla Knowledge Discovery in Database (KDD), della quale abbiamo parlato nei punti precedenti, per il fatto che le informazioni analizzate sono in formato testuale e quindi non strutturate.