

Analisi critica di metodi e strumenti a supporto della marketing intelligence

Survey

Nel presente documento si riassumono i contenuti del deliverable *D-D.1_2 Analisi critica di metodi e strumenti a supporto della Marketing Intelligence* relativo al pacco di lavoro D1. Nel documento sono riportati i risultati delle attività svolte in InViMall riguardanti l'analisi e lo studio di metodologie, strumenti e tecniche di Marketing Intelligence che saranno poi sfruttate dalla piattaforma InViMall per innovare tanto la raccolta, quanto l'analisi delle informazioni, al fine di migliorare le interazioni con gli utenti ed ottimizzare le azioni di marketing.

I mutamenti che hanno interessato il contesto socioeconomico degli ultimi anni hanno spinto le imprese ad assumere comportamenti innovativi per adattare i propri modelli di business ai cambiamenti. Tale situazione si è andata affermando anche grazie allo sviluppo dell'ICT che ha consentito un migliore sfruttamento della rete Internet per la vendita online e per la realizzazione di nuove politiche di marketing ed innovative attività di profilazione ed analisi della clientela. In particolare, per supportare i processi decisionali, si è avuto un notevole sviluppo delle tecniche e degli strumenti di **Business Intelligence (BI)**; tali strumenti permettono, non solo di supportare i processi decisionali, ma anche di aumentare l'efficacia delle campagne di marketing, fidelizzare i clienti ed aumentare la produttività. Inoltre, la diffusione capillare delle reti sociali virtuali ha permesso di considerare al centro dell'intero sistema economico non più l'impresa, ma il consumatore, concepito come un soggetto attivo, ed ha portato allo sviluppo di nuove strategie e processi di marketing orientati alla costruzione di relazioni con i clienti; nasce così il concetto di **marketing relazionale**, che svolge un ruolo rilevante nel progettare e gestire un'interazione di lungo periodo con i clienti acquisiti e potenziali.

Il marketing relazionale permette di migliorare e rafforzare la relazione diretta con i clienti, attraverso una continua identificazione dei loro specifici bisogni ed esigenze. Dal punto di vista operativo, il marketing di relazione si identifica attraverso particolari procedure quali la Customer Experience Management (CEM) e la Customer Relationship Management (CRM) che permettono all'azienda di incrementare il numero di clienti, aumentarne la redditività, massimizzare il grado di fedeltà e fidelizzazione ed evitarne il più possibile l'abbandono.

Tali strategie, secondo cui ogni cliente va considerato in modo "unico e speciale", si concretizzano nella personalizzazione dell'offerta verso il cliente in modo da instaurare un rapporto uno-a-uno con il venditore realizzando in questo modo azioni di **marketing one-to-one**. Il web ha cambiato il modo di fare marketing, sono cambiate le metodologie, le tipologie di campagne e promozioni realizzabili sul web, e gli indicatori di marketing necessari per valutare l'efficacia e l'efficienza delle azioni intraprese. Risulta possibile, ad esempio, monitorare in tempo reale l'andamento di una campagna pubblicitaria, contando il numero di click effettuati su un banner oppure valutare la cronologia di navigazione di un utente, analizzando la sequenza di pagine visitate; possedere queste informazioni precise e dettagliate permette così di modificare la strategia di marketing in qualsiasi momento,

riducendo i costi. Si riportano delle metodologie che, conformemente alla nuova filosofia di mercato, consentono al marketing di perseguire alcuni principali obiettivi come: raccogliere informazioni per capire su che segmento di mercato si vuole operare e di cosa hanno bisogno i vari attori che si vuole far interfacciare; creare visibilità sul sito in modo da attrarre i clienti; profilare i consumatori per comprendere analiticamente il comportamento di acquisto e di consumo, al fine di capirne i bisogni, accrescere il raggio d'azione dell'impresa stessa e creare uno stretto legame di fiducia con il cliente; alimentare alleanze in modo tale da raggiungere posizioni di leadership nel mercato nel più breve tempo possibile.

Per il raggiungimento di tali obiettivi risultano essenziali le informazioni racchiuse all'interno delle basi dati aziendali, quasi mai adeguatamente utilizzate; l'adozione di tool e soluzioni di Business Intelligence che permettano di trasformare dati ed informazioni in "asset di conoscenza" utili per prendere decisioni e migliorare la performance aziendale diventa quindi un obiettivo primario.

La Business Intelligence comprende le aree di financial intelligence, marketing intelligence, accounting intelligence e management intelligence, come mostrato in Figura 1. Per quel che concerne l'area "**Marketing Intelligence**" possono essere individuate due sottoaree principali, marketing research e customer relationship marketing/database marketing.

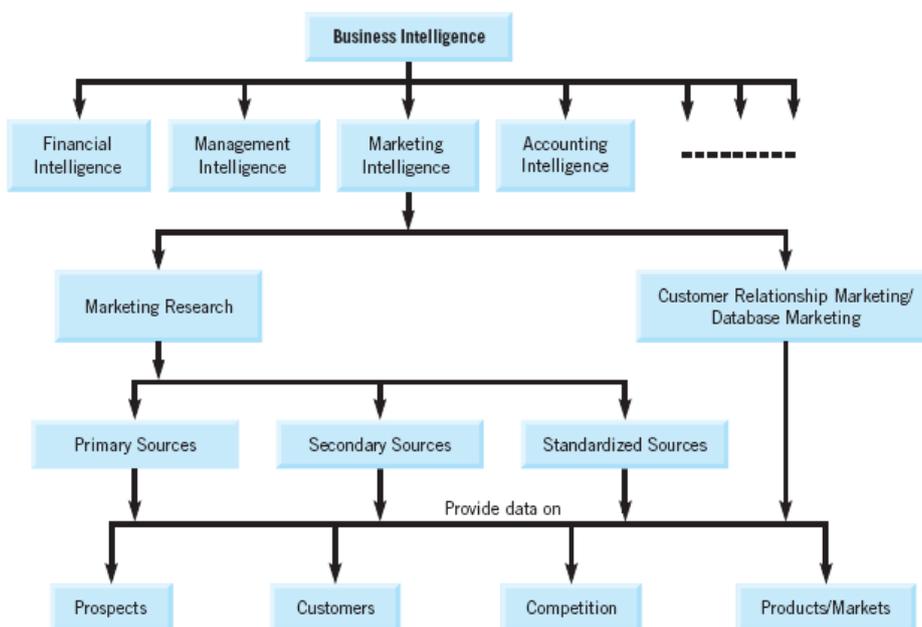


Figura 1: Aree della Business Intelligence

La **Marketing Intelligence (MI)** è l'insieme delle strategie, degli strumenti e delle applicazioni che supportano le aziende al fine di migliorare le relazioni con la propria clientela, definire il posizionamento ottimale per l'acquisizione dei prospect, concepire offerte e prodotti/servizi congeniali. MI permette di contribuire allo sviluppo del business attraverso ricerca strategica, analisi, rating, pubblicazione e reporting; si occupa, quindi, della comprensione, dell'analisi e valutazione dell'ambiente interno ed esterno relativo ai clienti dell'azienda, i concorrenti, i mercati. La disponibilità di dati provenienti dalla gestione operativa e dai canali di contatto consente alle aziende di riorganizzare i processi in un'ottica "client-centric" e di pianificare le iniziative commerciali basandosi sul supporto decisionale

generato da strumenti di intelligence quali data warehousing, data mining, geomarketing e campaign management.

Il **Data Warehousing** rappresenta la metodologia che supporta in modo completo le attività di analisi di business e di decision making relative a diverse aree aziendali. Il data warehouse gestisce e usa nel modo più efficace la memoria storica dell'azienda consentendo di interpretare le informazioni cruciali affinché l'azienda risponda tempestivamente ed efficacemente ai cambiamenti delle condizioni di business.

Il **Data Mining (DM)** è una delle attività cruciali per la comprensione, la navigazione e lo sfruttamento dei dati. Si tratta del processo automatico di scoperta ed individuazione di strutture all'interno dei dati, dove per struttura si intendono pattern, modelli e relazioni. Si riporta, in Figura 2, lo schema del processo di estrazione della conoscenza.

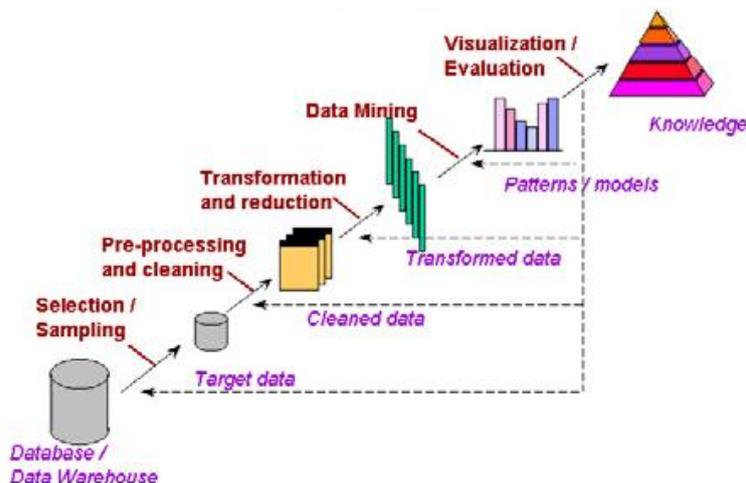


Figura 2: Processo di estrazione della conoscenza

Il **Geomarketing** è un sistema di supporto alle decisioni, in grado di orientare le scelte di marketing operativo, in funzione del contesto territoriale nel quale operano le singole realtà produttive e distributive, fornendo informazioni sulla domanda attuale e potenziale del mercato, sul posizionamento rispetto alla concorrenza, sul grado di copertura della rete commerciale.

Marketing Automation è l'insieme di tutte le tecnologie volte ad automatizzare le principali componenti del processo di marketing permettendo all'azienda di avviare automaticamente campagne di marketing e di comunicazione personalizzate allo scopo di promuovere notizie, eventi, offerte commerciali verso specifici target di utenti.

Per quanto concerne le strategie di Marketing Intelligence, le principali sono la segmentazione, la soddisfazione e la fedeltà, la personalizzazione; in dettaglio, per quanto riguarda la *segmentazione*, essa rappresenta una fase del marketing strategico basata sull'individuazione e definizione di molteplici segmenti di mercato basati sull'analisi dei comportamenti della clientela.

I criteri fondamentali di segmentazione sono: il criterio descrittivo che si basa sull'utilizzo di variabili di tipo geografico, demografico e psicografico ed il criterio della strumentalità che, invece, si basa su variabili comportamentali legate al rapporto tra cliente e prodotto. Individuati i criteri di segmentazione, l'azienda elabora strategie per la costruzione dei segmenti utilizzando due approcci principali, quello aggregativo, basato sui modelli a priori, e quello disaggregativo basato su modelli di clustering. Relativamente al contesto del

marketing in rete, la segmentazione viene realizzata attraverso l'uso delle tecniche di Data Mining, in particolare con gli algoritmi di clustering. La segmentazione in rete è il processo che mira ad individuare la struttura logica che sottende l'attività di navigazione/acquisto del consumatore e di definire le fondamentali tipologie di comportamento che scaturiscono da questa attività. *La soddisfazione e la fedeltà (loyalty)*, invece, permettono alle aziende di comprendere i bisogni e le aspettative dei clienti, per poi realizzare offerte competitive, oltre a definire un piano strategico che aumenti la produttività dell'azienda. Per realizzare l'analisi della soddisfazione dei clienti si può ricorrere a diverse tecniche di raccolta delle informazioni, tra cui lo strumento più utilizzato è il questionario. L'indagine di customer satisfaction deve consentire di rilevare le opinioni dei clienti con lo scopo di individuare delle azioni da realizzare per migliorare le performance aziendali. Analizzare i dati di un'indagine di customer satisfaction significa trasformare i punteggi e le altre risposte del questionario in informazioni chiare e concrete al fine di individuare gli interventi di miglioramento attraverso l'utilizzo di scale di valutazione. Un problema più complesso è quello riguardante la misurazione del grado di fedeltà dei clienti, in quanto essa non è legata ad una singola esperienza di contatto del cliente con l'azienda e quindi ad una soddisfazione "istantanea" relativa ad un unico evento ma è necessario affiancare alla soddisfazione istantanea una sequenza di esperienze positive che permettono di determinare una soddisfazione "cumulata, consecutiva ed ininterrotta". È necessario focalizzare l'attenzione sul concetto di fidelizzazione di un cliente: questo è il risultato di un processo di costruzione che ha come linee guida esperienze d'acquisto pienamente soddisfacenti, mai interrotte da elementi negativi che potrebbero ridurre il livello di fedeltà di acquisto. Si deduce quindi l'importanza di ricavare il valore atteso ed il valore percepito da parte del cliente, il *gap* fra i due e il valore stesso del cliente, al fine di valutare la sua soddisfazione per poter, successivamente, intraprendere delle azioni di CRM specifiche in base ai risultati ottenuti oppure attuare una strategia di *loyalty*.

La definizione di fedeltà, attraverso l'analisi RFM, richiede di considerare altri tre aspetti:

- **Recency:** è l'intervallo temporale trascorso dall'ultimo acquisto (azione) del cliente.
- **Frequency:** è la frequenza con cui il cliente effettua acquisti (il numero di volte che il cliente ha completato un'azione).
- **Monetary:** è il valore monetario speso in acquisti dal cliente (valore creato dal completamento di azioni).

A tal proposito, per eseguire l'analisi RFM, occorre rintracciare i dati relativi alla frequenza d'acquisto (per calcolare il grado di fedeltà della clientela) e quelli relativi all'ammontare della spesa (per calcolare il *budget ratio*); tale analisi permette all'azienda di avere una metrica per la misurazione del valore potenziale del cliente. Alla tecnica RFM nel tempo si è aggiunta la tecnica RFMT (Recency, Frequency, Monetary, Type) che, oltre alle variabili considerate dalla RFM, comprende la quarta variabile *T* (type: categoria o tipo di bene/servizio acquistato). *La personalizzazione*, infine, fornisce ad ogni individuo informazioni, prodotti o servizi ritagliati su misura, ed occupa un ruolo significativo nelle applicazioni intelligenti di e-marketing. La personalizzazione consiste nel processo di raccolta e memorizzazione delle informazioni relative al traffico generato dal sito web, nell'analisi volta a comprendere i modelli di comportamento degli utenti, le abitudini e le preferenze, al fine di fornire ad ogni visitatore, suggerimenti personalizzati. La personalizzazione può essere utilizzata per gestire la relazione con i clienti e lo sviluppo di campagne pubblicitarie rivolte ad un determinato target e per promuovere prodotti o servizi. Le fasi del processo di personalizzazione comprendono la raccolta dei dati, l'elaborazione e l'output personalizzato. Il Web Mining è l'applicazione di algoritmi di Data Mining utilizzata per scoprire ed estrarre informazioni da

documenti e servizi web; tale tecnica si applica nell'ambito della fase di raccolta ed elaborazione del processo di personalizzazione. Il Web Mining può essere suddiviso in tre categorie:

- *Web Content Mining*: è il processo che si occupa della scoperta di informazioni e del recupero della conoscenza da pagine Internet (testi, link o profili utente che possono essere inferiti o inseriti esplicitamente dagli utenti).
- *Web Structure Mining*: è l'applicazione di tecniche di DM per ricostruire la struttura di uno o più siti web; deduce conoscenza da strutture web e link.
- *Web Usage Mining*: è l'applicazione di tecniche di DM per scoprire modelli di comportamento degli utenti a partire dai dati di accesso al sito web.

Nel settore del commercio elettronico la personalizzazione ha permesso di realizzare sistemi di raccomandazione (*Recommendation System* o *RS*), in grado di aiutare l'utente, suggerendogli come navigare nell'enorme quantità di prodotti e informazioni disponibili. Essi sono sistemi capaci di predire l'interesse di un utente per un certo item, ovvero per un qualsiasi oggetto, fisico o digitale. I sistemi che raccomandano elementi come ad es. pagine web, film e canzoni, basandosi sulle opinioni delle persone con gusti simili sono detti di *filtraggio collaborativo (ACF - Automated Collaborative Filtering)*; tali sistemi eliminano la necessità per i computer di analizzare e comprendere il contenuto dell'elemento. Il termine *collaborativo* indica che il processo di raccomandazione si basa sui rating degli altri utenti e sulle loro preferenze. Tali sistemi, infatti, memorizzano i rating che gli utenti danno ad un certo item per poi successivamente consigliare ad un utente gli item che hanno avuto buone valutazioni da utenti simili. Un'altra tecnica di raccomandazione è il **Content Based Filtering** il cui obiettivo è quello di individuare una classificazione specifica per ogni utente, sulla base dei suoi rating e degli attributi di ogni elemento. Un sistema basato su tale tecnica di filtraggio, suggerisce ad un utente elementi (come contenuti, servizi e prodotti) simili a quelli che ha preferito in passato. Allo scopo di aumentare l'efficienza e l'efficacia del processo decisionale ed individuare una soluzione per un problema non completamente strutturato è possibile utilizzare un *Decision Support System (DSS)*; questo è un sistema software che mette a disposizione dell'utente, il "decisore", una serie di funzionalità di analisi dei dati e permette l'utilizzo di modelli in maniera interattiva ed estremamente semplice.

Le soluzioni di BI prevedono diverse tecnologie per la raccolta e analisi dei dati, quali DW, OLAP, EIS, ETL e DM. In Figura 3 è riportato lo schema architetturale di una soluzione completa di Business Intelligence.

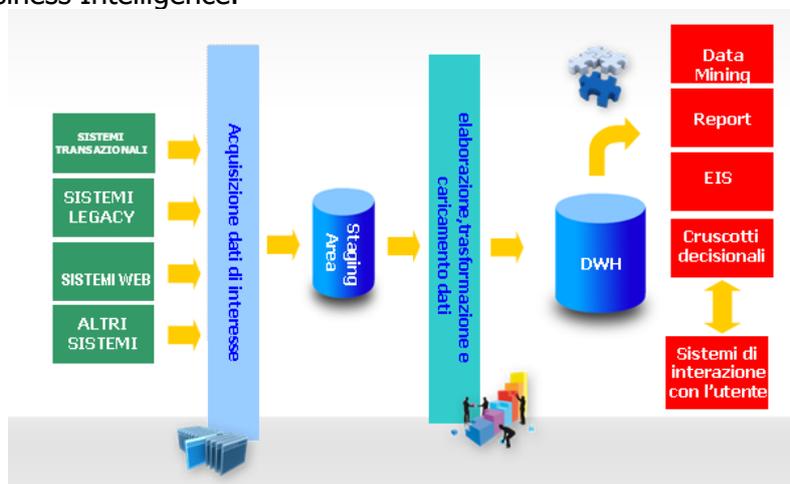


Figura 3: Schema architetturale di una soluzione completa di Business Intelligence.

Data Warehouse: sono grandi database usati per la memorizzazione di dati "storici", provenienti da diverse fonti, siano esse interne o esterne all'azienda. Deve mantenere le informazioni nel tempo e deve permettere di ricostruire la storia delle interazioni tra il cliente e l'azienda. È organizzato al fine di rendere possibile l'analisi dei dati e l'elaborazione di informazioni finalizzate alla pianificazione ed al processo decisionale.

OLAP: è l'insieme di tecniche software che riguardano l'analisi interattiva e veloce di grandi moli di dati storici, che è possibile esaminare attraverso operazioni di aggregazione, espansione e selezione. Questa è la componente tecnologica base del data warehouse e, ad esempio, serve alle aziende per analizzare i risultati delle vendite, l'andamento dei costi di acquisto merci, al marketing per misurare il successo di una campagna pubblicitaria, e i risultati possono fornire un utile supporto alle decisioni che dovranno essere prese. Le tecnologie OLAP possono essere realizzate con strutture multidimensionali reali e permanenti o con strutture relazionali, organizzate in modo da offrire una vista multidimensionale.

EIS: sono sistemi che forniscono informazioni in tempo reale attraverso un'interfaccia intuitiva combinando le informazioni disponibili all'interno dell'organizzazione e quelle relative all'ambiente esterno e fornendo un facile accesso alle informazioni utili al raggiungimento degli obiettivi strategici dell'azienda. L'utente di un sistema EIS può vedere rapidamente lo stato e l'andamento delle sue attività; il compito di un EIS è quello di gestire risorse di marketing, analizzare il rischio di un progetto e il suo impatto sull'azienda nel breve e lungo termine, fornire un approccio per le previsioni di vendita, fornire un metodo per stabilire il prezzo dei prodotti.

ETL: è il processo di estrazione, trasformazione e caricamento dei dati in un sistema di sintesi. I dati vengono estratti da sistemi sorgenti quali database transazionali (OLTP), comuni file di testo o altri sistemi informatici (sistemi ERP o CRM); questi subiscono un processo di trasformazione che ha lo scopo di consolidarli, renderli omogenei e memorizzarli nel Data Warehouse. Lo scopo del processo ETL è rendere disponibili i dati raccolti in azienda, provenienti da fonti eterogenee, ai soggetti incaricati di assumere le decisioni, nella forma e secondo le tempistiche più idonee a supportare il processo decisionale.

Strumenti di Data Mining: il DM è un processo strutturato di creazione di modelli a supporto delle decisioni di business, che permette di scoprire relazioni, similitudini, sequenze e tendenze all'interno di basi dati di grandi dimensioni contenenti informazioni eterogenee. Esistono quindi varie applicazioni del DM in ambito aziendale, dove ogni contesto più efficacemente trae vantaggio da una tecnica (o un insieme di tecniche) piuttosto che da un'altra. I metodi di DM supportano sia la decisione a livello di formulazione strategica dei quesiti, sia l'ottimizzazione di campagne di marketing diverse nel corso del ciclo di vita del cliente. Il Data Mining consente, grazie alle tecniche interpretative e predittive basate sull'analisi statistica (modelli di clustering, regressivi, ecc) e sul calcolo automatico (reti neurali, alberi decisionali, ecc), di identificare pattern direttamente dai dati osservati nel datawarehouse e di costruire indicatori che consentano di applicare uno scoring ai clienti. Tali indicatori sono detti Predictive Analytics; i principali sono: il Life Time Value, il Churn Risk, il Risk Scoring ed il Redemption Likelihood. Il Data Mining, quindi, è utile in tutti gli stadi della relazione con il cliente e costituisce valore aggiunto, insieme con l'esperienza pregressa, in tutti gli strumenti del relationship management e nelle tradizionali fasi del ciclo di vita di un cliente. Le predictive analytics hanno un ruolo determinante anche nel supportare l'azienda nel processo di rinnovo della fiducia e della soddisfazione dei clienti, nonché nel prevenire la propensione all'abbandono. I progetti dove le predictive analytic hanno un ruolo centrale sono assimilabili in generale ad un progetto di MI. In letteratura

esistono numerose metodologie per la schematizzazione del processo di DM; uno di questi è il modello noto con il nome di Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM è un progetto realizzato da IBM, NCR, Daimler-Chrysler, ISL, Ohra e altre importanti realtà industriali europee ed è finanziato dalla Commissione Europea con l'obiettivo di definire un approccio metodologico standardizzato ai progetti di DM. Scopo del progetto è definire e convalidare uno schema d'approccio indipendente dalla tipologia di business, dall'area di applicazione e dalle tecnologie utilizzate.

Tra le diverse tecniche di Data Mining presenti in letteratura, sono state analizzate le tecniche per la classificazione, la regressione e le regole associative, in riferimento al contesto del marketing online e col fine di individuare nuove tecniche di Marketing Intelligence.

I **classificatori** consentono di derivare un modello per la classificazione dei dati secondo un insieme di classi predefinite (ad esempio permettono di stimare se il destinatario di una campagna di mailing acquisterà un prodotto). Sono stati analizzati gli alberi decisionali, i classificatori Bayesiani e le reti neurali artificiali; per quanto riguarda gli alberi decisionali, questi sono particolari classificatori che permettono di identificare, in ordine di importanza, le cause che portano al verificarsi di un evento; un esempio implementativo di algoritmo è il C4.5. I classificatori bayesiani invece rappresentano un approccio probabilistico alla risoluzione di problemi di classificazione, applicabile nei casi in cui la relazione tra i valori degli attributi e quello della classe è non deterministica. Diversi approcci ai classificatori Bayesiani sono rappresentati da Naive Bayes, Flexible Bayes e Bayesian Belief Network. Le reti neurali artificiali (RNA) rappresentano, invece, uno strumento matematico-statistico ispirato al processo di apprendimento del cervello umano. Mediante modelli matematici viene rappresentata l'interconnessione tra elementi definiti *neuroni artificiali*, ossia costrutti matematici che in qualche misura imitano le proprietà dei neuroni viventi; un esempio applicativo delle reti neurali è rappresentato dal Multi Layer Perceptron (MLP).

Le **tecniche di Regressione** hanno l'obiettivo di costruire un modello predittivo basato su dati storici che viene poi utilizzato per stimare l'uscita, in termini di valori numerici, relativa a nuovi casi. La regressione lineare consente di analizzare la relazione causale tra una variabile dipendente (quantitativa) ed una o più variabili indipendenti. Tale relazione può essere descritta dall'equazione della retta in cui la variabile da stimare è legata a quella indipendente da parametri che devono essere stimati. Il metodo più utilizzato per la stima dei pesi è il criterio dei minimi quadrati *Ordinary Least Square*. La Support Vector Regression, invece, classifica i dati mediante un separatore lineare eseguendo una mappatura basata su funzioni kernel (ad es. gaussiane, polinomiali, lineari). Si basa sulla ricerca di una funzione ignota $y=f(x)$ che sia caratterizzata da un errore basso rispetto ai dati di training (ϵ -insensitive loss function). Tale tecnica viene risolta dall'algoritmo *Sequential Minimal Optimization* (SMO).

Le **Regole associative** permettono di determinare le regole di implicazione logica presenti nella base di dati, quindi di trovare i gruppi di affinità tra oggetti (ad esempio permettono di utilizzare un'analisi di mercato sugli acquisti per suggerire ad un cliente ulteriori prodotti da acquistare). Sono stati trattati tre algoritmi principali quali Apriori, Predictive Apriori ed FPGrowth. L'algoritmo Apriori è fondato sul principio che, dato un itemset frequente, allora anche tutti i suoi sotto-insiemi devono esserlo; è un algoritmo iterativo e basato sulla generazione dei candidati, Predictive Apriori, invece, usa l'algoritmo Apriori e combina le metriche di confidenza e supporto fornendo una singola misura di accuratezza. FPGrowth invece si basa sulla strategia divide-et-impera, usa strutture dati ottimizzate, chiamate FP-Tree.

La cluster analysis consente di identificare gruppi di cluster, basandosi su un determinato set di attributi, affinché le unità appartenenti allo stesso gruppo presentino valori degli attributi simili. I metodi di clustering possono essere suddivisi in gerarchici (agglomerativi e divisivi) e non gerarchici. In base ad un'ulteriore classificazione, si parla di hard clustering (clustering esclusivo) quando un'unità può appartenere solo ad un unico cluster e viene calcolata una singola probabilità per l'appartenenza di ogni unità a tale cluster: i cluster sono quindi disgiunti e non si sovrappongono. Nel soft clustering (clustering non-esclusivo o fuzzy clustering) un'unità appartiene a più cluster e viene calcolata una probabilità per ogni combinazione unità-cluster. Il metodo non gerarchico più diffuso è il *K-means*, applicato in presenza di variabili continue e prevede di individuare dei punti di aggregazione, i centroidi, aggiornati mentre l'algoritmo procede, attorno a cui quindi si costruiscono i gruppi, attribuendo ad essi le osservazioni più vicine ai centri individuati. Per quel che concerne i metodi gerarchici, questi consentono di costruire una struttura ad albero binario, il *dendrogramma*, dove i nodi corrispondono ai sottoinsiemi dei punti mentre le foglie alle unità (Figura 4). Questi metodi si basano su un processo suddiviso in più fasi, che consente di rendere la procedura computazionalmente più efficiente, ma non offre la possibilità di correggere eventuali errori di classificazione commessi nelle fasi precedenti. Tali metodi possono essere agglomerativi e divisivi, basati su due opposti approcci, rispettivamente bottom-up e top-down. Sono stati analizzati due diversi algoritmi: *TwoStep Cluster Analysis* e *Self Organizing Maps* (SOM).

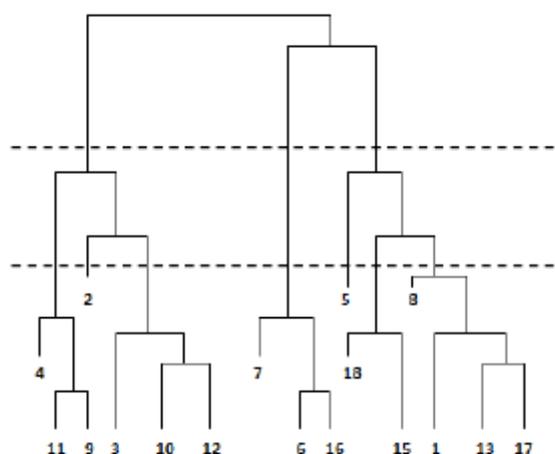


Figura 4: Esempio di dendrogramma

Tali metodi di clustering, tuttavia, si basano quasi esclusivamente sui dati senza una formale costruzione di modelli, cioè non vengono fatte assunzioni o creati modelli per la distribuzione dei dati all'interno dei gruppi. Inoltre, esistono altri metodi di clustering, tra cui quelli basati su modelli probabilistici per determinare l'appartenenza di un'unità al cluster. L'algoritmo EM (Expectation-Maximization), infatti, si basa su tale logica e determina la probabilità che un'unità appartenga ad un determinato cluster. Se durante il processo vengono generati cluster vuoti o se l'appartenenza ad uno o più cluster risulta inferiore ad una data soglia, l'algoritmo riparte da una nuova iterazione. L'algoritmo EM si basa sull'utilizzo dei "mixture models", impiegati quando non si hanno sufficienti informazioni sulla specifica forma parametrica da adottare per il modello di densità per una data variabile. Un modello "mixture" può essere considerato come una variabile casuale la cui funzione di probabilità o funzione di densità di probabilità è data da una somma pesata di funzioni di probabilità o di densità di altre variabili casuali. Altri algoritmi analizzati sono il Clustering In QUEst (CLIQUE)

basato sulla riduzione dimensionale, il *Minimal Spanning Tree* (MST) e *Association Rules Hypergraph Partitioning* (ARHP) basati sulla teoria dei grafi ed infine algoritmi basati su alberi probabilistici (COBWEB) dove ogni foglia è costituita da una classe e per ogni possibile valore di ogni "feature", vengono definite le percentuali di probabilità all'interno della classe.

L'applicazione delle tecniche di DM alle azioni di marketing, permette di sviluppare tecniche innovative di Marketing Intelligence con l'obiettivo di comprendere e valutare a pieno il mercato, i clienti ed i prodotti al fine di ottimizzare i processi decisionali. In particolare sono state analizzate la tecnica di Scoring Analysis volta ad individuare i prospect interessati all'acquisto di prodotti target, la Customer Lifetime Value e la Churn Analysis per stimare il valore dei clienti e prevenirne l'abbandono, ed infine la Market Basket Analysis per dedurre le correlazioni tra i prodotti.

La **Scoring Analysis** è una tecnica che si applica a diversi contesti per la risoluzione di problemi più disparati, ad esempio è possibile calcolare il rischio creditizio di un cliente di una banca, in tal caso si parla di credit-scoring, piuttosto che scoprire e prevenire comportamenti fraudolenti di utenti di telefonia, oppure lo scoring può essere usato per definire campagne di marketing ed individuare i migliori prospect.

Per quanto riguarda l'ambito *marketing*, lo scoring può essere usato per la determinazione efficace e precisa di strategie per la realizzazione di campagne di marketing, tenendo in considerazione gli utenti che hanno la probabilità maggiore di acquisto. In letteratura viene presentato un sistema di Business Intelligence chiamato EMCA orientato all'automazione delle strategie per le campagne di marketing. Il sistema generando mailing list verso uno specifico gruppo di clienti, riduce i costi per il marketing, evitando di coinvolgere utenti non interessati; per la realizzazione di EMCA vengono utilizzate tecniche di Data Mining quali le regole associative, gli alberi decisionali, Naive Bayes e le serie temporali. La Scoring Analysis permette, non solo di individuare i migliori prospect tra tutti gli utenti del sistema, ma soprattutto di ottimizzare le campagne di marketing in termini di costi, evitando di colpire gli utenti non interessati al prodotto offerto, che potrebbero essere disturbati e maturare un avversione verso il brand; essa si può realizzare con l'uso di tecniche di Data Mining che vanno sotto il ramo della Predictive Analytic. In particolar modo si prevede la possibilità di misurare il risultato delle campagne promozionali, collezionando i dati storici, al fine di creare i response model ed attraverso questi modelli, prevedere la probabilità di acquisto dei prospect contattati. Gli utenti oggetto della campagna di marketing saranno selezionati sulla base dell'indice di utilità che questi evidenziano in riferimento al prodotto oggetto della campagna, ancora prima della fase di stima della probabilità di acquisto.

Le tecniche di DM utilizzate per l'implementazione della "parte core" della scoring analysis sono le tecniche di classificazione; in particolare l'attenzione verrà focalizzata sui classificatori di tipo Bayesiano trattati in precedenza, in modo da considerare la classificazione in termini probabilistici.

Il **Customer Lifetime Value** è un indicatore finanziario dinamico e dipende dai comportamenti del cliente; tale indicatore rappresenta una metrica attrattiva per la definizione delle strategie di marketing in quanto fornisce una previsione futura sui costi e ricavi generati dal cliente. Il Customer Lifetime Value è un indice che può essere utilizzato per stimare il valore dei clienti, esso è spesso impiegato per predire i ricavi futuri dei clienti ed organizzare i clienti in diverse categorie a cui rivolgere differenti offerte promozionali. È molto importante identificare quali sono i clienti più redditizi e quali meno in quanto da tali informazioni è possibile riuscire a ricavare informazioni utili all'azienda al fine di intraprendere azioni di marketing opportune che possono permettere di "conservare" a lungo i propri

clienti. Nell'ambito dell'e-commerce il CLV può essere definito come la somma del *valore retrospettivo* dato dalla differenza tra ricavi generati e costi sostenuti relativi ad un cliente fino al momento considerato e dal *valore prospettico* dato dalla differenza tra ricavi previsti e costi previsti per il futuro. Tale tecnica permette ad un "merchant" di individuare clienti a *più alto valore*, con l'obiettivo di offrire eventuali sconti o coupon, oppure clienti a *più basso valore* a cui indirizzare campagne di marketing di up-selling o cross-selling al fine di aumentarne il valore. Per quanto riguarda il calcolo delle componenti del CLV, è possibile ricavare il valore retrospettivo a partire dai dati storici presenti nel database, mentre per stimare i costi e ricavi futuri relativi ad ogni cliente (valore prospettico) si utilizzano tecniche di DM quali i modelli predittivi. In particolare le tecniche utilizzate sono le Regressioni e le Reti Neurali Artificiali. Operativamente il CLV è dato da:

$$\text{Valore Retrospettivo (VR) + Valore Prospettico (VP)}$$

dove:

$VR = \text{Ricavi generati} - \text{Costi sostenuti relativi ad un cliente fino al momento considerato}$

$VP = \text{Ricavi previsti} - \text{Costi previsti per il futuro}$

L'obiettivo del CLV è selezionare i termini di VR e predire i termini di VP attraverso l'uso delle tecniche della predictive analytic.

La **Churn Analysis** è una tecnica che mira a comprendere il comportamento degli utenti al fine di prevedere il momento in cui avverrà l'evento di "churn" e quindi l'utente cambierà le sue abitudini; il tipico evento di abbandono è legato al fatto che, un cliente che usufruisce di un determinato servizio, interrompe la fruizione, scegliendo un altro servizio, tipicamente concorrente. La Churn Analysis può quindi essere applicata a diversi contesti, come ad es. quello dei mercati finanziari, dei fornitori di servizi e-mail, dei commercianti al dettaglio di prodotti industriali, nel settore delle telecomunicazioni. Fattore comune ai vari contesti è il miglioramento della customer relationship management (CRM) che l'analisi, basata sul churn, apporta.

Nel contesto dell'e-commerce dire churn analysis significa evitare che i clienti possano abbandonare il "merchant" attuale e rivolgersi ad un concorrente. L'idea alla base di tale tecnica consiste nell'evidenziare i clienti che sono più a rischio "churn" ed effettuare azioni di marketing volte ad aumentare la loyalty, cioè azioni di fidelizzazione del cliente, riducendo la probabilità dell'evento di "churn". Nell'ambito delle diverse tecniche di Data Mining che è possibile applicare, si distinguono differenti approcci, quali ad esempio considerare il churn come un evento dicotomico impostando due livelli differenti, fedele o churn; in questo caso si applicano le tecniche di classificazione quali gli alberi decisionali o le reti neurali artificiali. Un ulteriore approccio consiste nella stima del tempo per cui un cliente resterà fedele, in tal caso si potranno usare le tecniche di regressione, oppure considerare il churn come una sequenza di eventi ed in tal caso trovare le combinazioni di stato o eventi che condurranno il cliente al churn; quest'ultimo caso può essere affrontato attraverso le tecniche di associazione, generando regole che descrivono la dinamica delle relazioni che hanno condotto il cliente al churn. La Churn Analysis ha l'obiettivo di comprendere i comportamenti dei clienti al fine di evitare che questi possano abbandonare il "merchant" attuale e quindi rivolgersi a "merchant" concorrenti. L'idea alla base di tale tecnica consiste nell'evidenziare i clienti che sono più a rischio "churn" ed effettuare azioni di marketing volte ad aumentare la loyalty, cioè azioni di fidelizzazione del cliente, riducendo la probabilità dell'evento di "churn". La **Market Basket Analysis** ha lo scopo di analizzare ed individuare i comportamenti e le abitudini di acquisto dei consumatori, al fine di progettare opportune azioni di marketing.

L'obiettivo principale della Market Basket Analysis è migliorare l'offerta dei prodotti, ad es. la disposizione dei prodotti all'interno della vetrina, nonché incrementare le vendite di prodotti tramite offerte su quelli ad essi associati ad es. proponendo coupons ed inoltre è possibile rendere più efficaci le azioni di marketing e merchandising ad es. attraverso azioni di cross-selling, up-selling, pianificazione delle campagne promozionali, schedulazione degli approvvigionamenti di magazzino. Tra i possibili approcci, la Market Basket Analysis può considerare anche i dati sociali e culturali degli utenti al fine di dedurre ulteriori informazioni utili. In questo ambito le classiche tecniche di Data Mining, quali le regole associative, vengono integrate con tecniche come la Principal Component Analysis e la Mixture Density Analysis inoltre, al fine di ridurre il numero di regole ottenute in output e non interessanti, si può usare la cluster analysis; le tecniche di DM utilizzate per realizzare analisi di tipo Market Basket sono le regole associative.

L'utilizzo di **piattaforme di E-Commerce o Business Intelligence** che presentano servizi di Marketing Intelligence consente di accrescere il vantaggio economico e competitivo dell'azienda, in quanto è possibile ricavare una serie di informazioni utili a partire dalla gran mole di dati disponibili ed intraprendere correttamente decisioni strategiche per l'azienda.

Si sottolinea che le piattaforme di BI, rispetto a quelle di e-commerce, offrono servizi più accurati per eseguire strategie di MI poiché i dati raccolti vengono opportunamente elaborati e vengono utilizzati per comprendere l'andamento delle performance dell'azienda, generare stime previsionali, ipotizzare scenari futuri e nuove campagne di marketing. A tal proposito, a partire dalle principali tecniche di Marketing Intelligence individuate, è stata eseguita una fase di scouting per verificare quali piattaforme di Business Intelligence e di E-commerce presentino funzionalità di MI.

Le principali piattaforme di BI open source analizzate sono: *Spago BI, Pentaho e Jasper Software*. Dall'analisi effettuata è risultato che solo le prime due piattaforme prevedono tool di Data Mining offrendo però solo servizi base per azioni/operazioni di Marketing Intelligence. L'attenzione si è concentrata poi sulle piattaforme commerciali di BI; sono state individuate MicroStrategy, IBM Cognos BI v10.1, Microsoft Business Intelligence, SAP Business Objects Business Intelligence 4.0, SAS Business Analytics e Oracle Business Intelligence Enterprise Edition 11g. Dall'analisi condotta è emerso che tutte le piattaforme prevedono servizi standard di BI ma solo la piattaforma SAS presenta servizi di BI avanzati che potrebbero permettere l'utilizzo di alcune delle tecniche di Marketing Intelligence individuate (come ad es. la Scoring Analysis e la Churn Analysis).

Il passo successivo è stato quello di analizzare le piattaforme di e-commerce. Per quel che concerne le piattaforme open source quelle analizzate sono state: Apache OFBiz, CubeCart, Ez Publish, LiteCommerce, Magento, osCommerce, PrestaShop, VirtueMart, ZenCart. Dallo studio fatto si è rilevato che Magento risulta essere la principale piattaforma open source che offre dei servizi minimali di Marketing Intelligence (come ad es. la segmentazione della clientela volta ad eseguire azioni mirate di marketing).

Infine per quanto riguarda le piattaforme e-commerce commerciali quelle analizzate sono state: ATG Commerce Server 10, AZIMOOOTH custom Commerce, ESCALATE Blue Martini, IBM Websphere Commerce Enterprise, iCongo, Microsoft Commerce Server 2009, NEW CART, PosteCommerce, THE NetSuit Ecommerce, FRY, ElasticPath. Si è osservato che ATG Commerce Server, IBM Websphere Commerce Enterprise e Microsoft Commerce Server 2009 risultano essere le uniche piattaforme commerciali che offrono alcuni dei servizi di Marketing Intelligence (come ad es. la possibilità di attuare campagne di marketing mirate basate sull'analisi degli acquisti).



InViMall - Intelligent Virtual Mall
Progetto MSE MI01 – 123



L'intera analisi eseguita ha dimostrato che allo stato dell'arte non esistono piattaforme che forniscono servizi integrati di Marketing Intelligence così come pensati in InViMall e che opportunamente implementati permetteranno di sfruttare a pieno le innumerevoli informazioni collezionate e disponibili ai merchant consentendo di migliorare le interazioni con i customer ed ottimizzare in termini di costi ed efficacia le campagne di marketing.