



Carlo Erminero & Co.
RICERCA E CUSTOMER INTELLIGENCE



Social media listening, mobile e online.

Un mondo digitale a disposizione del marketing strategico

Luca Massaron

Direttore di Ricerca e Resp. Centro Studi - Carlo Erminero & Co.

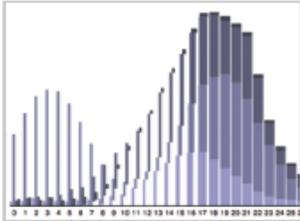
**Insight, Web & Big Data,
il ricercatore e la sua expertise come fonte di
apprendimento
per agenti artificiali intelligenti**



Big Data?

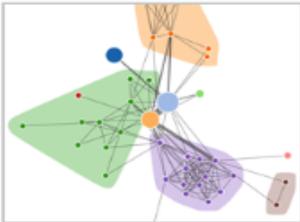
Visualization of all editing activity by user "Pearle" on Wikipedia (Pearle is a robot). To find out more about this project, go to: www.research.ibm.com/visual/projects/chromogram.html

Big Data, 3V



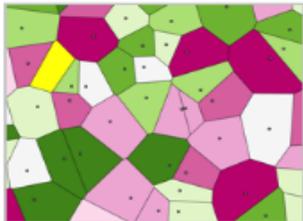
- **VOLUME**

Quotidianamente vengono prodotti nel mondo 2,5 quintilioni di nuovi byte di dati (10^{18} byte)



- **VELOCITY**

Il 90% dei dati esistenti in questo momento nel mondo è stata registrata negli ultimi 24 mesi



- **VARIETY**

l'80% dei *Big Data* non è strutturato in alcun modo



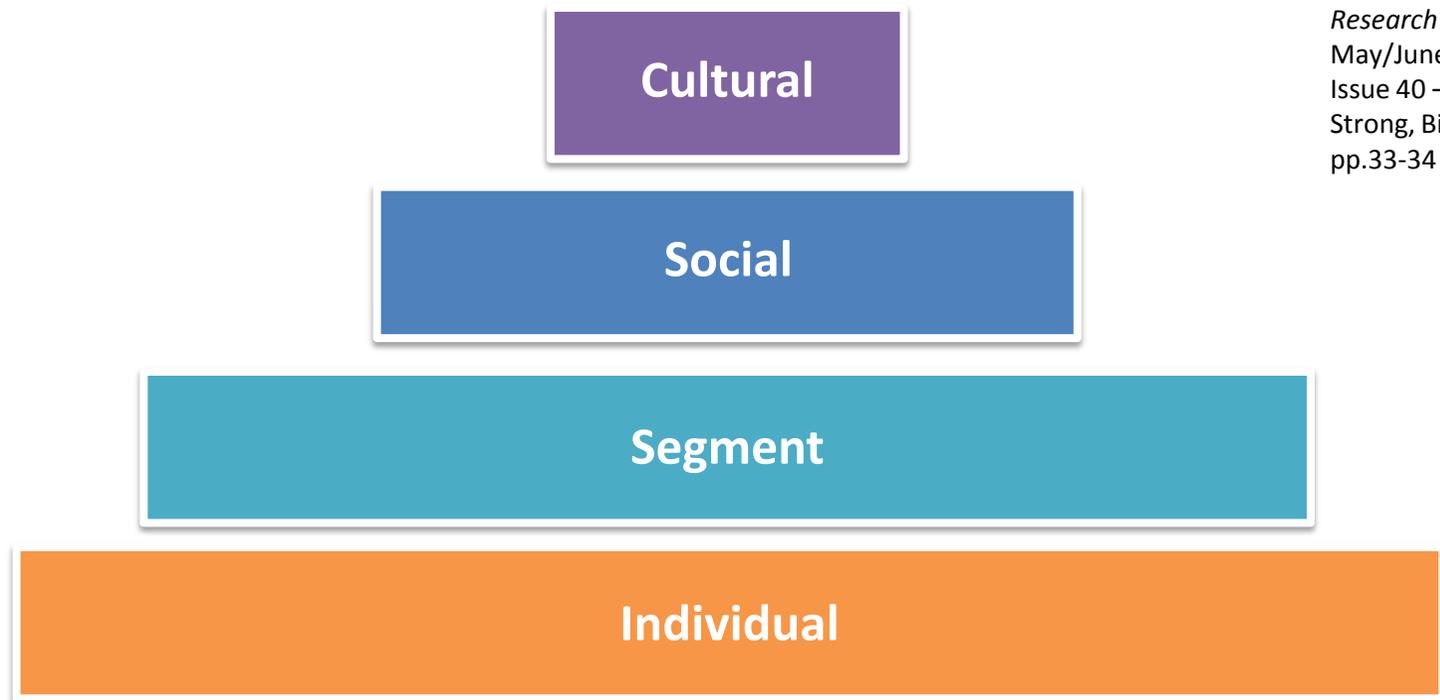
*Sistema informativo di marketing
inondato da ogni tipo e qualità di
informazioni*



Big Data, data sources

Research World,
 May/June 2013 -
 Issue 40 – Colin
 Strong, Big Data,
 pp.33-34

Societal trends
 ↑
 Scale of insight
 ↓
 Individual response



*Proprietary
 Brand Data*

*Aggregated
 Brand Data*

*Third Party
 Brand Data*

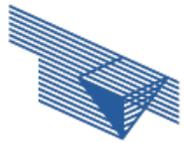
*Digital Social
 Media*

*Cultural
 Artifacts*

*Government
 Data*

Proprietary

Open



Paradigmi a confronto

RICERCHE DI MERCATO

- Dati su campione
- Costosi
- In numero limitato, al massimo stanno nella RAM del PC
- Selezione accurata dei casi meglio su base probabilistica
- Pulizia e certificazione dei dati
- Dati strutturati o facilmente strutturabili
- Ricerca di senso (insight) condotta in ogni fase dell'elaborazione



BIG DATA

- Enormi moli di dati
- Poco costosi, spesso gratuiti
- Non sono gestibili in RAM, servono memorie di massa capienti e più computer in parallelo
- Nessun controllo o selezione dei dati
- Dati destrutturati, rumore, garbage in
- Le elaborazioni sono scatole nere focalizzate sul risultato (azione), non sul senso

Ipotesi e tesi

Quale sarà il futuro delle ricerche di mercato?

Ipotesi:

Dovranno inserirsi in un sistema informativo di marketing altamente tecnologico, sempre più computazionale, decisamente data-driven.

Tesi:

Possiamo gettare un ponte fra le nuove tecniche analitiche pensate per i BIG DATA e la tradizione delle ricerche di mercato di dare senso all'universo sociale e dei consumi e di comunicare utilmente ed efficacemente l'informazione nelle organizzazioni complesse e ramificate, come le aziende.

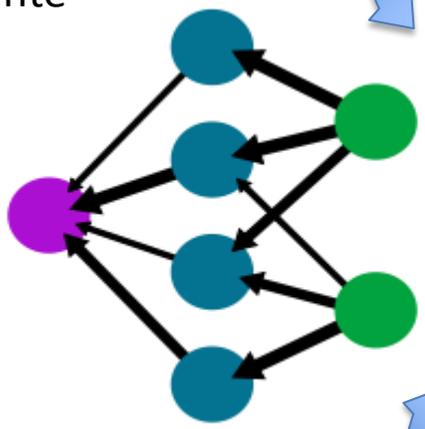


Idea agente intelligente

- WEB
- CRM
- BIG DATA repository

Azione
(Classificazione
Reperimento
Integrazione
Modifica
...
)

Software costituito da un modello matematico ed interfaccia con l'ambiente



Feedback / nuovi dati

Big Data repository



Esemplificazione di teorie ed idee del ricercatore



Dataset



<http://times.cs.uiuc.edu/~wang296/Data/>

Hotel Name ← Four Seasons Hotel George V Paris

Overall Rating ←

Reviewer ID ← By trollydollySydney

Review Text {
Lovely location, however, for 820 euros this was really bad value.
The room was nice, but you could have been anywhere in the world- it felt like a chain hotel in the worst sense. The room was tiny! Normally Four Seasons have mind blowing service and although they were nice it was not amazing. We had just been to Claridge's in London which was fantastic and half the price. It wasn't bad , but it wasn't great and not worth the money. A coke was 10 euros! There was no free wireless- all in all very average.

TripAdvisor dataset

Overall rating | Review text

Scala risposta: 1-5, ordinale
206MB, 244.007 recensioni

Wang, H., Lu, Y., & Zhai, C. (2010, July). Latent aspect rating analysis on review text data: a rating regression approach. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 783-792). ACM.



MSE = 1,24



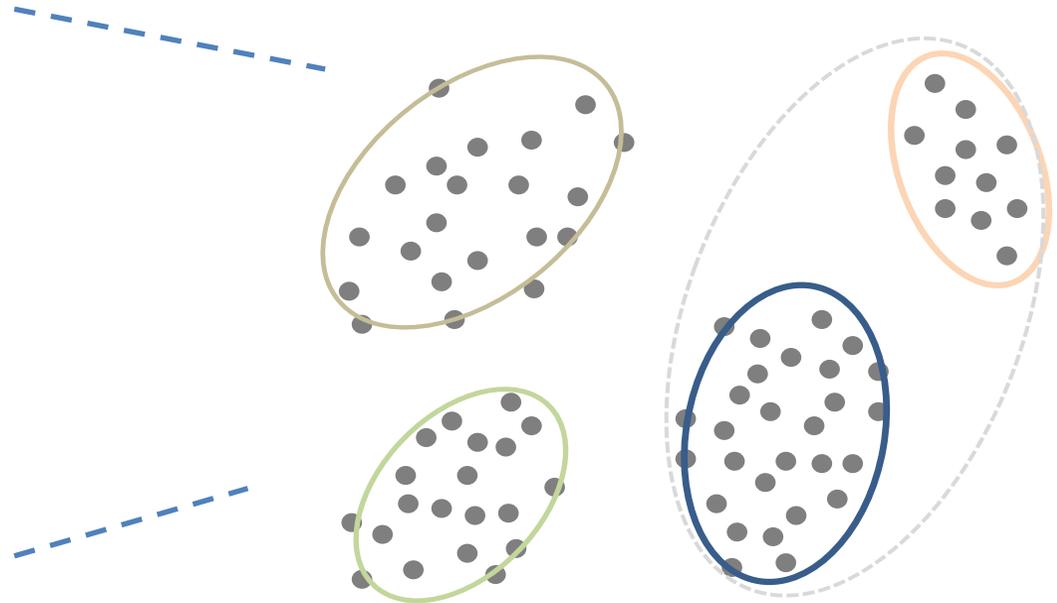
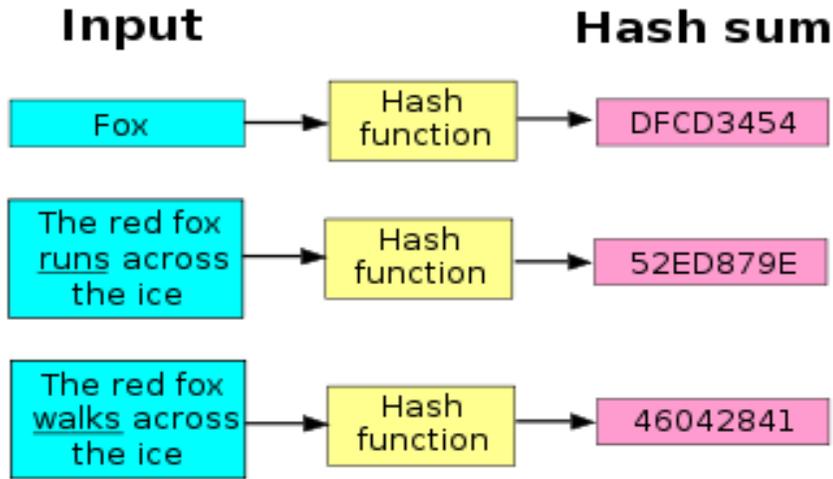
Modello supervisionato

$$\text{Rating} = \beta_1 p_1 + \beta_2 p_2 + \dots + \beta_n p_n + \beta_{1\&2} p_1 * p_2 + \beta_{1\&n} p_1 * p_n + \beta_{2\&n} p_2 * p_n + \dots + \beta_{n-1\&n} p_{n-1} * p_n$$

Caratteristiche del modello:

- Elevato numero di coefficienti (min. 140.000, max. $643 * 10^6$) – corrisponde ad un language model
- Parole non trattate (solo lowercase, interpunzioni non rimosse)
- Bag of words di parole singole e bigrammi
- Regolarizzazione L2 – smorzamento coefficienti irrilevanti, mediamente ne utilizza effettivamente dai 6.000 ai 10.000
- Non è una scatola nera, si può ispezionare e comprendere
- Pochi passaggi sui dati (stochastic gradient descent)
- Elevata velocità di configurazione del modello (<180sec)
- Modello aggiornabile con successivi batch di esempio o in modalità interattiva, con continui feedback agente/ricercatore

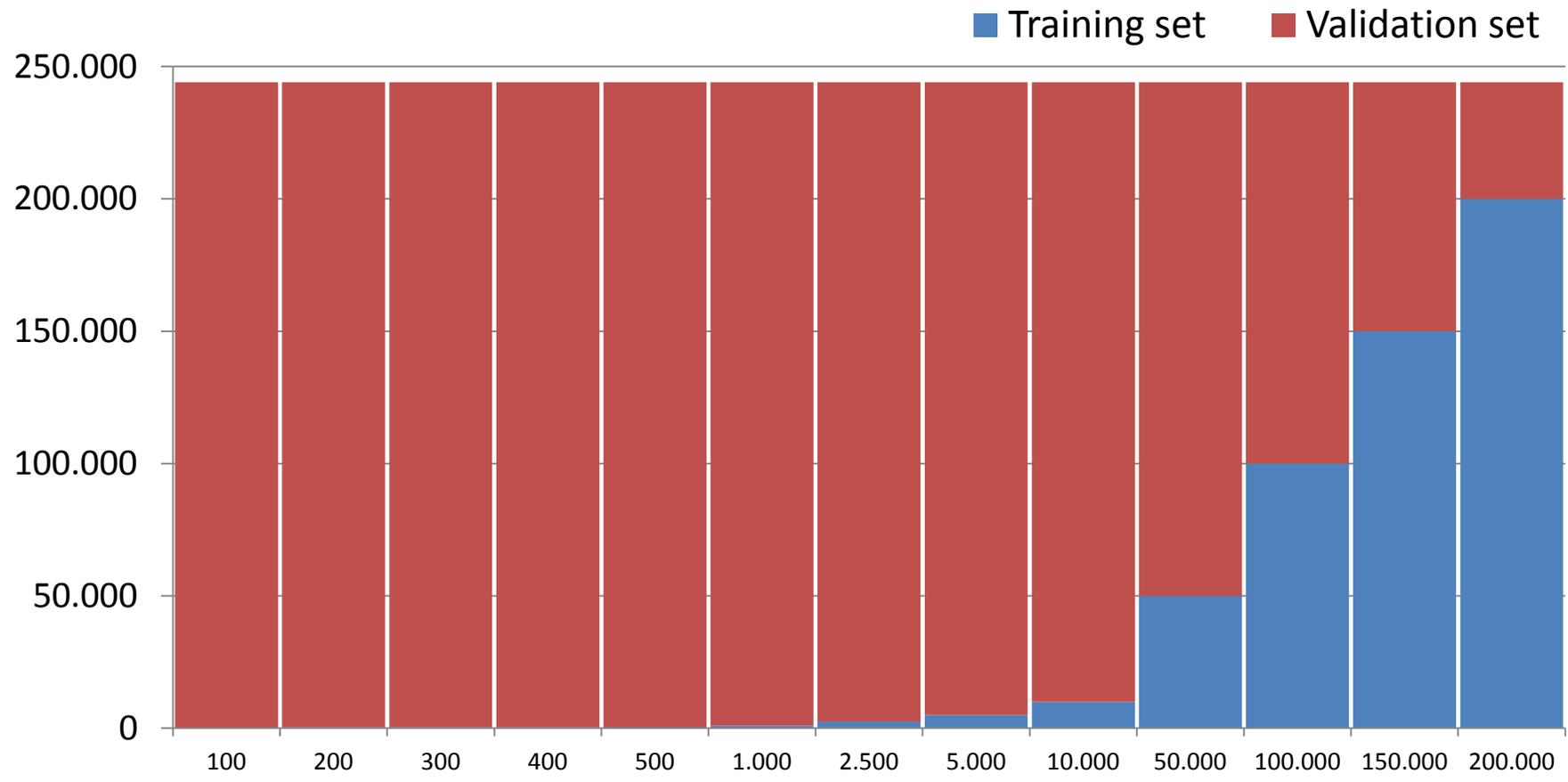
Modello semi-supervisionato



parola, frase \rightarrow hashing \rightarrow k-cluster $\in \mathcal{D}$

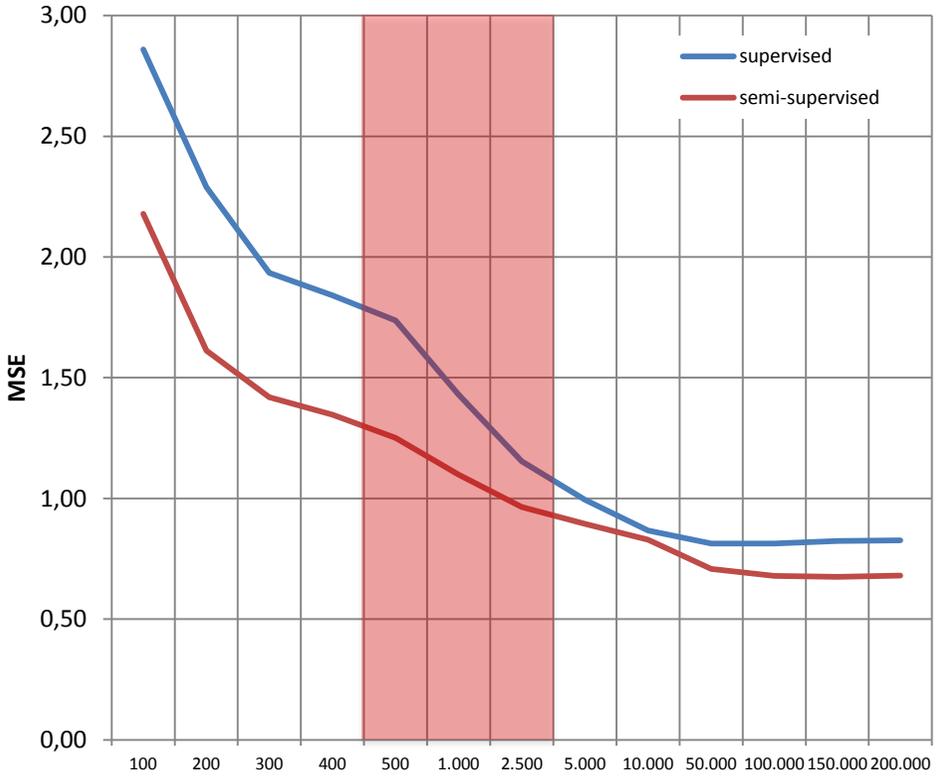
- I testi disponibili sono mappati in uno spazio di dimensione pre-ordinata, secondo una funzione univoca di assegnazione (hashing semantico)
- Dalla cluster analysis dello spazio semantico, è ricavata l'appartenenza ai cluster che costituiscono un nuovo dizionario, \mathcal{D} , in grado di facilitare l'apprendimento delle caratteristiche dei testi con un minor numero di esempi

Training & validation



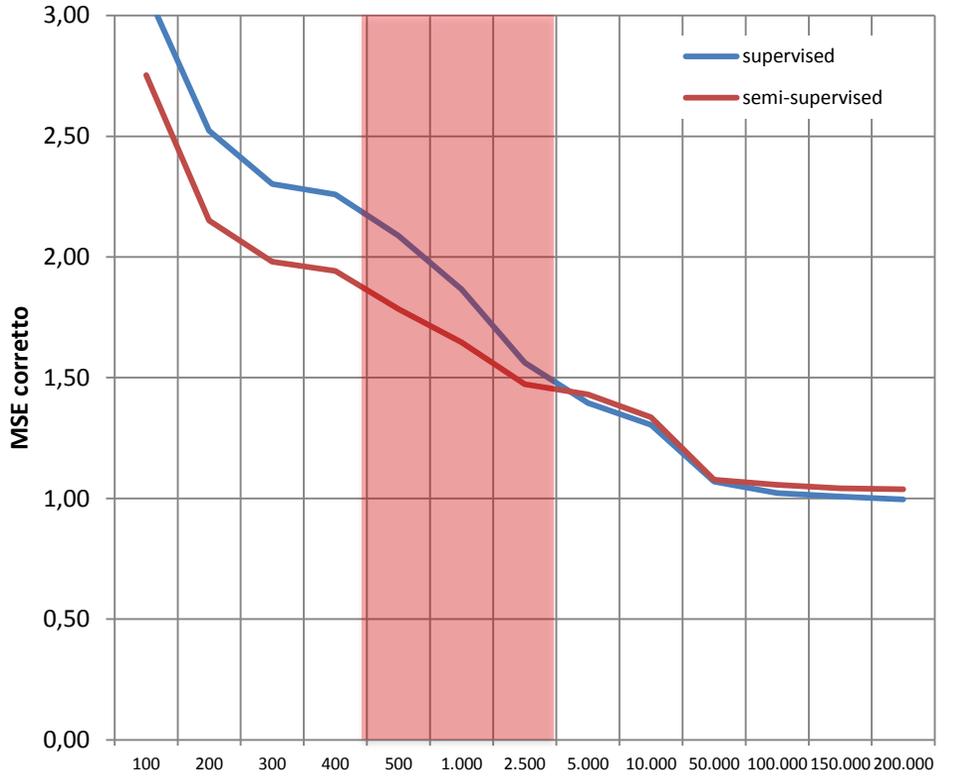
Risultati test sperimentale

MSE → nessuna variabilità quote classi



Dimensioni campione di addestramento

MSE_{corretto} → quote classi variabili



Dimensioni campione di addestramento

Applicazioni

- Classificazione e ricerca di testi su Internet (blog, tweets, social data, estesi repository di informazioni e dati) da utilizzare in funzione dei propri studi e ricerche.
- Analisi delle verbalizzazioni di soddisfazione ed insoddisfazione presenti nei CRM aziendali, integrazione della parte quantitativa e qualitativa delle ricerche di customer satisfaction.



Applicazione su dati digital pubblici

Ricerca di documenti Web per realizzare uno studio qualitativo su un argomento molto specifico e preciso

Sviluppo di un agente intelligente per la ricerca Web e la text classification

Attivazione di una ricerca estesa che integra ricerca per keyword, Web crawling e classificazione testuale

Analisi qualitativa del ricercatore sui testi selezionati dal software per trasformare i Big Data in **Smart Data**

1



2



3



4

Un esempio ricorrente: motori di ricerca e db non forniscono testi attinenti, occorre visionare migliaia e migliaia di pagine

Il lavoro del ricercatore si limita a fornire al software quanto reperito ed a costruire dei testi campione attraverso cui il software può capire quello che il ricercatore vuole

Il software attiva una ricerca in parallelo (attraverso un sistema map-reduce), recupera e classifica centinaia di migliaia di testi digitali

Il ricercatore riceve solo i testi selezionati dal software e può verificare la loro attinenza, utilizzandoli per il proprio studio

Applicazione su dati digital proprietari

Verbalizzazioni
soddisfazione clientela
in numero elevato
(oltre 10^5 per anno)
grazie ai nuovi canali
digitali

Estrazione
campionaria di circa
3.000-5.000
verbalizzazioni, analisi
approfondita,
classificazione dei testi

Sviluppo di un agente
intelligente per la text
classification e
l'anomaly detection

Deployment
dell'agente nel sistema
informativo
(marketing data mart)
per trasformare i Big
Data in **Smart Data**

1



2



3



4

*Un esempio ricorrente:
il numero elevato di
dati suggerisce di
consultare specialisti
non appartenenti al
mondo delle ricerche*

*Il lavoro del
ricercatore si limita a
campioni più o meno
grandi, dove si
esplicita tutta la sua
competenza
metodologica e
professionale*

*Il lavoro di ricerca non
si ferma ad indicazioni
di massima, ma viene
opportunamente
tradotto in un
software capace di
gestire l'interezza dei
dati*

*Dopo opportuna
calibratura, il software
viene facilmente
integrato nel sim, dove
agisce ed informa
puntualmente il
marketing*



Carlo Erminero & Co.
RICERCA E CUSTOMER INTELLIGENCE

**DIGITAL
WORLD
2013**



Social media listening, mobile e online.

Un mondo digitale a disposizione del marketing strategico

Grazie per l'attenzione!

luca.massaron@ce-co.it