



LUMSA  
UNIVERSITÀ

Dip. di Scienze Umane – Comunicazione, Formazione, Psicologia  
Corso di laurea magistrale in Marketing & digital communication

---

Project work

# Data Mining e IT

## Introduzione al corso

Nicola Barbieri

# Contatti e informazioni

---

**Nicola Barbieri**

n.barbieri@lumsa.it

Ricevimento studenti: **su appuntamento** (richiesto via e-mail)

[Pagina web](#) su lumsa.it

- Istruzioni per l'esame
- Materiale del corso
  - ✓ Teoria (slide)
  - ✓ Esercizi e soluzioni
  - ✓ Link utili

## Percorso sul sito

www.lumsa.it

Corsi

Corsi di laurea

Scienze umane – Comunicazione...

Attività didattiche integrative

Project work (Data mining e IT)



**Nessun libro consigliato:** il materiale fornito è sufficiente e in caso di bisogno la rete abbonda di materiale gratuito da consultare.

# Il corso

---

## **10 incontri da 3 ore – ogni mercoledì dalle 9 alle 12**

In aula: quando permesso dalle disposizioni per l'emergenza sanitaria

Da remoto: <https://meet.google.com/twz-zpuy-wfb>

Il corso è un **laboratorio**: la frequenza è molto **importante**.  
Assenze solo per **validi motivi** (lavoro, malattia...).

- Teoria (quanto basta)
- Pratica (più possibile)

*Ciò che senti, dimentichi.*  
*Ciò che vedi, impari.*  
*Ciò che fai, comprendi.*

*Life is now: feedback, please!*

# L'esame

---

## Progetto individuale pratico

- da **richiedere** via e-mail alla fine del corso (le istruzioni dettagliate saranno pubblicate sulla pagina web del laboratorio)
- da **consegnare** alcuni giorni prima dell'appello in cui si intende sostenere l'esame
- Il giorno dell'esame: **discutere** il progetto svolto col docente + **rispondere** a domande sul programma
- **valutazione**: idoneità + giudizio
- **Appelli**: uno per sessione, come previsto dall'Ateneo per i Project Work.

# Data Mining e IT: obiettivi del corso

---

- Eliminare paure, dubbi e incertezze davanti a un PC
- Conoscere problematiche e tecniche legate alla **scelta** e all'**ottimizzazione** dei **dati**
- Acquisire una conoscenza di base sugli **strumenti tecnologici** per la **conservazione** e l'**elaborazione** dei dati
- Sperimentare le **potenzialità** del **data mining** tramite lo studio e l'applicazione di tecniche e algoritmi

# Data Mining

## Date una definizione...

- attività automatica o semi-automatica

che prevede:

- estrazione di informazioni da grandi quantità di dati

allo scopo di:

- ricercare relazioni ricorrenti (pattern) non note a priori

**I dati sono il nuovo petrolio**



# Valore dei dati

---

**Nel 2016 la General Electric ha speso circa un miliardo di dollari** per analizzare i dati provenienti da sensori posti su turbine a gas, motori a reazione, oleodotti.

Alcuni possibili benefici:

- Scoprire dettagli per far lavorare le macchine con più **efficienza**
- Combinando i dati con quelli dei clienti, **predire la domanda** e conoscere le **modalità d'uso** più comuni per ciascun prodotto
- **Ottimizzare** gli aspetti tecnici e logistici della **produzione**

# Potenzialità del Data Mining: il caso Target

La catena di supermercati **Target** è all'avanguardia nello studio dei dati a fini di marketing.

Un uomo di Minneapolis ha trovato nella posta della figlia adolescente dei **buoni sconto** di Target per acquistare culle e abiti per **neonati**.

Furioso, ha protestato col direttore del negozio Target più vicino: «Mia figlia è ancora al liceo, non vorrete spingerla a restare incinta!» Il direttore, sorpreso e incredulo, si è scusato, e dopo alcuni giorni ha telefonato a quel cliente per rinnovare le sue scuse.

L'uomo, questa volta, era più calmo e imbarazzato: aveva parlato con **la figlia**, che aveva ammesso di essere rimasta **incinta**.





# Potenzialità del Data Mining: il caso Target

---

Ciascun **cliente** di Target è descritto da:

- Un numero identificativo
- Nome o indirizzo e-mail
- Sesso (M/F)
- Indirizzo di residenza
- Dati della carta di credito
- Lista di tutti i beni che ha acquistato nel tempo

In USA esiste un **registro pubblico** delle nascite.

Analizzando gli acquisti delle **clienti che hanno avuto bambini**, si è scoperto che le donne in gravidanza tendono a comprare cose come:

- Lozioni non profumate
- Integratori di calcio, magnesio e zinco
- Saponi e gel igienizzanti
- Batuffoli di cotone

# Provenienza dei dati

---

**Da quali fonti vengono raccolti i dati che alimentano il data mining?**

- Social network
- Cellulari, dispositivi mobili e app
- Software
- Siti web (motori di ricerca, e-commerce, portali, siti privati)
- Automobili, mezzi e macchine industriali, scatole nere
- Telecamere stradali, caselli, sensori di traffico
- Smart TV, decoder connessi, console per videogiochi
- Domotica/IoT (elettrodomestici, stereo, caldaie, lampadine...)
- Sistemi informatici di banche, poste, finanziarie, assicurazioni
- Sistemi inf. di operatori telefonici, fornitori di gas, acqua, luce...
- Sistemi informatici di negozi e supermercati (carte fedeltà)
- Sistemi informatici di alberghi, ristoranti, locali
- È finita la pagina...

# Evoluzione dell'analisi dei dati

<b>Passi evolutivi</b>	<b>Domande tipiche</b>	<b>Tecnologie disponibili</b>	<b>Caratteristiche</b>
Raccolta dati (dagli anni '60)	"Quanto abbiamo fatturato in totale negli ultimi 3 anni?"	Computer, nastri, dischi	Analisi del passato, output statico
Accesso ai dati (dagli anni '80)	"Quanti prodotti abbiamo venduto in Italia a marzo?"	Database relazionali (RDBMS), linguaggio di interrogazione (SQL), driver di connessione (ODBC)	Analisi del passato, output dinamico
Supporto alle decisioni e Data Warehousing (dagli anni '90)	"Quanti prodotti abbiamo venduto in Italia a marzo? E quanti nel Lazio?" (drill down)	On-line analytic processing (OLAP), database multidimensionali, data warehouse	Analisi del passato, output dinamico a più livelli
Data Mining	"Cosa prevediamo che accada alle vendite nel Lazio il prossimo mese?" Perché?	Algoritmi avanzati, computer più potenti, database massivi	Analisi delle prospettive future, output utile per azioni proattive