

# Analisi Dati per la Business Intelligence e Data Science

V edizione 2016/2017



UNIVERSITÀ DEGLI STUDI  
DI TORINO

Dipartimento di Culture, Politica e Società  
Dipartimento di Informatica  
Dipartimento di Matematica "G. Peano"  
Dipartimento di Economia e Statistica  
"Cognetti de Martiis"

Gestito da:



in collaborazione con



---

aggiornato al 8 luglio 2016

# **PROGRAMMA DIDATTICO**

## **1. Software per la Business Intelligence**

*36 ore di lezione frontale, 89 ore di studio individuale*

*Crediti Formativi: 5*

I temi trattati nel presente insegnamento saranno:

Business Intelligence: dal DB Management e Data Quality agli analytics e alla comunicazione dei risultati: la toolbox
Backend:Strumenti per archiviazione e aggiornamento dei dati, flate tables, DW e DataMart
Introduzione agli strumenti:
Excel
QlikView
SAS
Seminario KNIME
Verifica di fine modulo

## **2. Data Management e trattamento dei dati non strutturati**

*20 ore di lezione frontale, 55 ore di studio individuale*

*Crediti Formativi: 3*

I temi trattati nel presente insegnamento saranno:

Esplorare le fonti informative interne ed esterne all'azienda
"Lessons learned" nella costruzione di basi dati decisionali
Come costruire basi dati decisionali (modellare i dati decisionali, la costruzione della base dati decisionale)
Costruire le basi informative a partire dai bisogni di conoscenza del cliente
Costruire le basi informative a partire dai bisogni di conoscenza del cliente
Verifica di fine modulo

### **3. Raccolta dei dati e data quality**

*32 ore di lezione frontale, 68 ore di studio individuale*

*Crediti Formativi: 4*

I temi trattati nel presente insegnamento saranno:

Strumenti per la raccolta dei dati e data quality: Access data
Strumenti per la raccolta dei dati e data quality: Manage data 1
Strumenti per la raccolta dei dati e data quality: Manage data 2
Strumenti per la raccolta dei dati e data quality: Analyze data
Introduzione a SAS Data Integration Studio
Semantic Web: strumenti e formati; intro al semantic web, con cenni a strumenti e formati
Esempio di estrazione in laboratorio di informazioni da testi (come articoli di giornale), loro codifica in RDF e interrogazione del triple store risultante tramite SPARQL
Esercitazione: integrazione dei dati
Strumenti per la raccolta dei dati e data quality: Present data
La gestione del data anagrafico per il riconoscimento e la contattabilità del Cliente - non PA
L'assessment di una qualità di una base dati - non PA
Prestare attenzione alla qualità delle informazioni (le dimensioni di qualità dei dati) - PA
Verifica di fine modulo

### **4. Statistica per la BI**

*24 ore di lezione frontale, 51 ore di studio individuale*

*Crediti Formativi: 3*

I temi trattati nel presente insegnamento saranno:

Eventi e probabilità
Distribuzioni di probabilità
Campionamento e intervalli di confidenza
Test di ipotesi
Linguaggio SQL
SAS Proc SQL
Verifica di fine modulo

## 5. Mining I – Statistica Descrittiva

*24 ore di lezione frontale, 51 ore di studio individuale*

*Crediti Formativi: 3*

I temi trattati nel presente insegnamento saranno:

I dati, le funzioni, i grafici, altri oggetti: statistica descrittiva
Test statistici parametrici e non parametrici
La correlazione e la regressione
ANOVA a una via
Tabelle di contingenza e indipendenza statistica
Verifica di fine modulo

## 6. Mining II – Relazioni Statistiche

*40 ore di lezione frontale, 85 ore di studio individuale*

*Crediti Formativi: 5*

I temi trattati nel presente insegnamento saranno:

Introduzione, preprocessing, classificazione
Esercitazioni con Knime
Clustering e pattern
Esercitazioni con Knime
Sperimentazione su metodi predittivi con software open source
Esercitazioni con Knime
Natural Language Processing; introduzione (sintassi, semantica); l'interpretazione semantica e il problema della Word Sense Disambiguation.
Risorse per l'NLP: WordNet; algoritmi per la disambiguazione.
Estrazione di informazioni da testi (come articoli di giornale), estrazione di keywords e riassunto automatico.
Rappresentazione della conoscenza e semantic web
Semantic Web: strumenti e formati; intro al semantic web, con cenni a strumenti e formati.
Verifica di fine modulo

## **7. Mining III – Sintesi dell'informazione statistica**

*24 ore di lezione frontale, 51 ore di studio individuale*

*Crediti Formativi: 3*

I temi trattati nel presente insegnamento saranno:

Individuazione di differenti tipologie di clienti: analisi cluster
Conoscere i propri clienti costruendone il profilo attraverso la Segmentazione Comportamentale: identificazione dell'obiettivo e impostazione dell'analisi
Sintesi dei fenomeni multidimensionali: analisi fattoriale e analisi delle corrispondenze
Interpretazione dei risultati, passaggio in produzione e supporto alle strategie di marketing dell'azienda
Esempi di sintesi dell'informazione statistica in ambito Pubblica Amministrazione (indicatori e cruscotti di performance, data e text mining, visualizzazione di informazioni)
Verifica di fine modulo

## **8. Modelli e modelli di equazioni strutturali**

*24 ore di lezione frontale, 51 ore di studio individuale*

*Crediti Formativi: 3*

I temi trattati nel presente insegnamento saranno:

Esercitazione: integrazione dei dati
Esercitazione: SAS Enterprise Miner
Modelli di regressione lineare e regressione logistica
Scoring models: esempi
Modelli PATH 1
Modelli PATH 2
Verifica di fine modulo

## 9. Tecniche di simulazione

24 ore di lezione frontale, 51 ore di studio individuale  
Crediti Formativi: 3

I temi trattati nel presente insegnamento saranno:

Scelta dei modelli (lineari, non lineari)
Stima dei modelli
Tuning dei modelli
Previsioni - What if analysis 1
Previsioni - What if analysis 2
Verifica di fine modulo

## 10. Interpretazione e comunicazione delle relazioni statistiche

52 ore di lezione frontale, 123 ore di studio individuale  
Crediti Formativi: 7

I temi trattati nel presente insegnamento saranno:

Come si scrive un report: regole pratiche e finalità
Come si presenta un report: modi, strumenti e finalità
Media, Pubblici e stakeholder della ricerca scientifica: la loro rilevanza nell'impostazione teorica e nella comunicazione dei risultati e nella comunicazione del metodo (probabilismo, falsificazionismo)
Responsabilità legale della ricerca, proprietà dei dati del finanziatore, copy right e open access nella sua diffusione
Informazioni per la Business Intelligence in impresa
I numeri nella comunicazione
Numeracy: valutare i numeri e ciò che ci viene detto con i numeri e con le rappresentazioni grafiche dei numeri
Banco di prova (benchmarking): costruire riferimenti empirici per la valutazione dei numeri
Intensionality: il contesto in cui i numeri vengono prodotti e i modelli "teorici" da cui derivano
Verifica di fine modulo