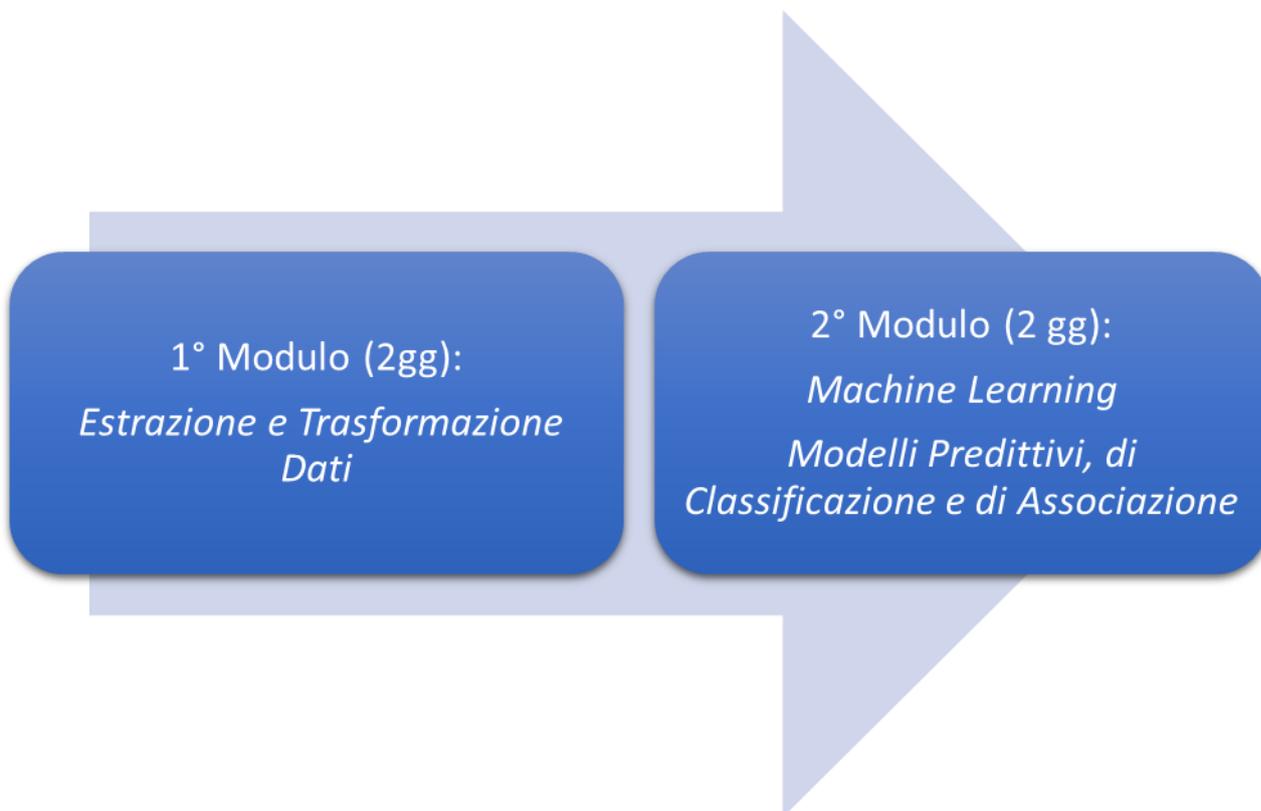


# Percorso Data Analytics



## Il perché della Data Analytics

Per essere competitivi al giorno d'oggi non è più sufficiente erogare prodotti e servizi di qualità, ma è necessario avere **capacità previsionali** per capire e anticipare le nuove tendenze del mercato che vanno ben oltre la semplice osservazione dei report aziendali.

La *Data Analytics (BA)*, applicata al *Business*, è un insieme di competenze e tecnologie atte a ottenere informazioni ad alto valore aggiunto a partire dai **grandi volumi di dati** disponibili in azienda e in rete, in modo tale che si possa intervenire efficacemente su ogni aspetto legato alle performance aziendali.

A differenza della *Business Intelligence*, che si limita a valutare "cosa è successo", la *Data Analytics* spiega il "perché", indicando anche "cosa si dovrebbe fare" per ottenere sempre migliori risultati.

Si riconoscono tre tipi di Data Analytics: **Descrittiva**, che mira a comprendere lo stato attuale del business; **Predittiva**, che utilizza tecniche di apprendimento

automatico per determinare risultati futuri; **Prescrittiva**, che indica le decisioni da prendere in base ai risultati ottenuti.

La *Data Analytics* può essere utilizzata con successo in vari settori aziendali, ad esempio:

**CRM** per indentificare ciò che crea insoddisfazione da parte del cliente con conseguente abbandono, e attuare strategie per il mantenimento dei clienti più profittevoli.

**Marketing** per suddividere i clienti in micro-segmenti ai quali indirizzare offerte mirate.

**Produzione** per individuare i fattori che influenzano la bottom-line, per contenere i costi e aumentare l'efficienza dei processi produttivi e di tutta la supply-chain.

**Risorse Umane** per analizzare le competenze disponibili e anticipare le richieste future ottimizzando proattivamente la composizione del personale.

Per questi motivi sono sempre più numerose le aziende che avvertono la necessità di **adottare tecnologie emergenti** e **sviluppare le competenze** per poterle utilizzare in modo efficace.

Il presente percorso formativo si rivolge a soddisfare tale necessità.

### Gli obiettivi del corso

Il corso è finalizzato a fornire ai partecipanti conoscenze relative alla

*Data Analytics* per:

- **Estrarre e trasformare dati** interni (aziendali) ed esterni (internet)
- **Comprendere e valutare le informazioni** ottenute con strumenti di *Data Analysis* e *Data Visualization*
- **Applicare modelli previsionali** basati su algoritmi di Machine Learning
- **Interpretare, presentare e mettere in opera** i risultati ottenuti

## I plus del corso

- Docenti che sono anche consulenti, in grado di portare in aula **l'esperienza diretta di numerosi progetti di DA** realizzati con successo
- Opportunità di confrontarsi con persone di altre realtà aziendali e **condividere esperienze** e metodologie
- Possibilità di conoscere numerosi **casi pratici**, per comprendere l'applicabilità dei contenuti teorici
- Acquisizione delle competenze di base per l'utilizzo della piattaforma open-source **KNIME® Analytics Platform**

## Il corso

### 1° Modulo (2 giornate)

*Estrazione e Trasformazione Dati*

### 2° Modulo (2 giornate)

*Machine Learning: Modelli Predittivi, di Classificazione e di Associazione*

- Il 1° Modulo (2 giornate) è dedicato ai concetti base di *Data Analytics*, pulizia e trattamento dei dati per la costruzione di dataset tematici e loro analisi con tecniche di statistica descrittiva. Verrà introdotto l'utilizzo della piattaforma **KNIME®** attraverso esercitazioni guidate.
- Il 2° Modulo (2 giornate) tratta il Machine Learning attraverso modelli Predittivi, di Classificazione e di Associazione

I moduli vengono presentati attraverso l'ausilio di casi pratici ed esercitazioni svolte in aula.

## Avvertenze e Prerequisiti

---

- A ciascun partecipante è richiesto di portare un laptop
- Il corso prevede l'utilizzo intensivo del software open-source *KNIME Analytics Platform* di cui non è richiesta conoscenza pregressa.

Per partecipare al corso è necessario avere installato ed installato la piattaforma open-source *KNIME*®, dal seguente link:

<https://www.knime.com/>

- Si richiede di conoscere la statistica descrittiva (indicatori sintetici: media, mediana, varianza, deviazione standard; significato di distribuzione statistica) e possibilmente, seppure non strettamente richiesto, le basi della statistica inferenziale (intervallo di confidenza, test statistico)

## Quota di iscrizione

---

La quota di iscrizione prevista è di Euro 2.500,00 + IVA e comprende la partecipazione a tutti e due i moduli e il materiale didattico e documentale. L'iscrizione al corso non è nominativa: una volta versata la quota d'iscrizione, l'Azienda può decidere di iscrivere persone diverse a ciascun modulo. Sono previste quote scontate per le aziende che iscrivono più di un partecipante.

## Le metodologie didattiche: la formula del workshop

---

Sono previste metodologie didattiche attive che alternano momenti di comunicazione a momenti di esercitazione/workshop applicativi.

I partecipanti apprenderanno concetti e nuove metodologie attraverso casi, lavori di gruppo, esercitazioni, test di apprendimento, simulazioni.

## 1° Modulo

## Estrazione, Trasformazione ed esplorazione dei dati

Durata: 2 giornate

---

### Il perché della Data Analytics

- Cos'è la Data Analytics
- Quattro tipi di Data Analytics
- Chi ne è coinvolto e quale tecnologia
- Applicazioni e Strumenti
- La catena del valore del dato
  - CRISP\_DM
  - Data Quality

### Il software utilizzato: *KNIME® Analytics Platform*

- Installazione
- Ambiente operativo
- Principi di funzionamento

### Il processo di ETL (Extract, Transformation, Loading)

#### *Extract*

- File di testo (txt, csv) ed Excel
- Tabelle di Database

#### *Transformation e Data Quality*

- Trasformazione dei dati
- Dati mancanti (Missing data)

#### *Loading*

- Costruzione del dataset tematico (Customer Table)
- Esercitazione di ETL

### Exploratory Data Analysis (EDA)

- Cosa si intende per esplorazione dei dati

- Gli strumenti principali: Line-Plot, Grafici di dispersione, Istogrammi, Box-plot, Ricerca dei valori estremi (Outlier)

## 2° Modulo      **Machine Learning: Modelli Predittivi, di Classificazione e di Associazione**

Durata: 2 giornate

---

### **Cosa si intende per Machine Learning**

- Cosa si intende per Machine Learning
- Costruzione, Validazione e Deployment dei modelli
  - Due tipi di algoritmi per il Machine Learning
  - *Supervised Learning*
  - *Unsupervised Learning*

### **Supervised Learning**

- Regressione Lineare multipla con variabili numeriche e categoriali
- Regressioni Logistica per funzione di risposta dicotomica
- Classificazione: anticipare le preferenze per prendere decisioni mirate
  - Decision Tree
  - Random Forest
- Presentazione e discussione di esempi

### **Unsupervised Learning**

- Cluster Analysis: riconoscere pattern all'interno dei dati e raggrupparli in gruppi omogenei
  - Clustering gerarchico
  - K-Means
- Presentazione e discussione di esempi

### **Regole di associazione (*affinità di prodotti/servizi*)**

- Market Basket Analysis. Regole associative del tipo: chi compra A, anche B  
compra
- Misure di qualità  
e



affidabilità delle regole

- Presentazione e discussione di esempi

Via Vittor Pisani, 8  
20124 Milano  
Per informazioni  
contattare il numero:  
331.477.04.67  
oppure scrivere a:  
[nives.boncristiano@galganogroup.com](mailto:nives.boncristiano@galganogroup.com)