

Contesto

Nel mondo aziendale e dell'industria risulta sempre crescente la necessità di competenze volte all'analisi delle grandi moli di dati raccolti e a disposizione per il monitoraggio, la pianificazione e l'ottimizzazione delle attività. In questo quadro, il management delle aziende deve saper formulare le proprie strategie e implementare le proprie attività avendo consapevolezza del valore associato alla raccolta e alla gestione dei dati.

A tal fine, il corso intende sviluppare competenze di base ed avanzate relative all'analisi dei dati per il supporto delle decisioni aziendali e per lo sviluppo del business. In particolare, i temi trattati riguarderanno tre ambiti principali: l'esplorazione dei dati, l'uso dei dati a fine di classificazione lo sviluppo di modelli predittivi.

Il corso oltre prevede lezioni frontali e momenti di approfondimento mediante casi studio.

Obiettivi

- Comprendere gli strumenti di base per la raccolta e l'elaborazione dei dati in chiave descrittiva
- Comprendere le tecniche in grado di supportare la profilazione efficace dei prodotti e degli utenti
- Comprendere gli strumenti da utilizzare per la quantificazione dell'incertezza e il suo utilizzo all'interno di previsioni di breve/medio/lungo periodo

Destinatari

- Pianificazione strategica
- Area Marketing & Comunicazione
- Area Ricerca e Sviluppo

Durata e luogo

Il corso si compone di 4 giornate formative da svolgersi on line mediante la piattaforma di Microsoft Teams e sarà così strutturato:

martedì 23 marzo 2021 (Unsupervised Learning 1)

9.30 – 17.00

Mercoledì 24 marzo 2021 (Unsupervised Learning 2)

9.30 – 17.00

Lunedì 19 aprile 2021 (Supervised Learning 1)

9.30 – 17.00

Martedì 20 aprile 2021 (Supervised Learning 2)

9.30 – 17.00



Programma dettagliato

Martedì 23 marzo 9.30 – 17.00

9.30 - 13.00

- Data exploration: cruscotti di monitoraggio delle attività
 - *Acquisizione e utilizzo di dati multivariati*
 - *Riflessione su “Data Deluge” e Big Data*
 - *Il ruolo del data scientist nel processo decisionale*

- *Razionale per l'introduzione di modelli statistici in grado di gestire la complessità del dato*

Pausa caffè

- *Data visualization: rationale per la creazione di grafici e la presentazione efficace dei risultati*

13.00 *Pausa pranzo*

14.00 – 17.00

- *Descrizione di dati univariati e multivariati: indici di posizione e dispersione, correlazione e covarianza, loro rappresentazione*

- *Indici robusti*

Pausa caffè

WORKING SESSION

- *Introduzione al software R: linguaggio semi compilato, importazione/esportazione dataset, determinazione variabili numeriche e categoriche*

WORKING SESSION

- *Manipolazione dei database, comandi di base per l'indagine statistica esplorativa*
- *Applicazione dei contenuti trattati a dataset reali*

Mercoledì 24 marzo 9.30 – 17.00

09.30 -13.00

- Clustering e riduzione dimensionale: profilazione di clienti e prodotti
 - *Introduzione: riduzione dimensionale e classificazione non supervisionata*

- *K-means algorithm e algoritmi gerarchici*

Pausa caffè

- *Principal Component Analysis (PCA)*

13.00 *Pausa pranzo*

14.00 - 17.00

WORKING SESSION

- *Esempi di clustering e riduzione dimensionale*

WORKING SESSION

- *Applicazione dei contenuti trattati a dataset reali*

Pausa caffè

- **Testimonianza aziendale (Lorenzo Montelatici , Head of DataLab, Edison)**



Lunedì 19 aprile 9.30 – 17.00

9.30 – 13.00

- Presentazione in aula dei risultati del contest

- *Confronto in aula relativamente alle soluzioni del contest!*

- *Confronto aperto su temi presentati nella prima parte del corso*

Pausa caffè

- **Regressione e classificazione supervisionata (approccio model-based)**

- *Modelli statistici per la descrizione e la previsione in ambito aziendale*

- *Il concetto di stimatore, test di ipotesi e intervallo di confidenza*

- *Errori di I e II tipo*

13.00 *Pausa pranzo*

14.00 - 17.00

- *Il modello di regressione lineare : stima dei coefficienti e loro interpretazione*

- *Inferenza sul modello*

- *Verifica delle ipotesi e analisi dei punti influenti*

- *Selezione delle variabili*

Pausa caffè

- *Modelli di regressione logistica e multinomiale: stima dei coefficienti e loro interpretazione*

- *Inferenza sul modello*
- *Verifica delle ipotesi*

Martedì 20 aprile 9.30 – 17.00

9.30 – 13.00

- **Regressione e classificazione supervisionata (approccio model-based-continued)**

- *Modello di regressione logistica e multinomiale: stima dei coefficienti e loro interpretazione*

- *Inferenza sul modello*

- *Verifica delle ipotesi*

Pausa caffè

- **Regressione e classificazione supervisionata parametrica (Approccio Machine Learning)**

- *KNN Algorithms*

- *Alberi di classificazione e regressione (CART)*

- *Random Forests*

13.00 *Pausa pranzo*

14.00 – 17.00

WORKING SESSION

- *Applicazione dei contenuti trattati a dataset reali*

- *Lezione riassuntiva dei contenuti visti durante gli incontri*

- *Confronto aperto sui temi presentati nella seconda parte del corso*

Pausa caffè

- **Testimonianza aziendale
(Paolo Cerioli – Resp. Digital Scouting & Open Innovation ENI)**



Simone Vantini

Professore Associato di Statistica al Politecnico di Milano. Ha pubblicato diversi articoli e testi sull'Analisi dei dati funzionali e orientati agli oggetti. I suoi attuali interessi di ricerca includono: test di permutazione, previsioni non parametriche, controllo di processo, dati non euclidei e in generale metodi e applicazioni statistici motivati da problemi aziendali o industriali.