

# PROGRAMMA DEL CORSO DI STATISTICAL LEARNING E ANALISI DEI BIG DATA

## SETTORE SCIENTIFICO

SECS-S/01

## CFU

9

## DESCRIZIONE

/\*\*/

1 - INTRODUZIONE AI BIG DATA

2 - PROCESSI DI ESPLORAZIONE DEI DATI 3 - ELABORAZIONE DEI DATI 4 - INTRODUZIONE A R 5 - GLI INDICATORI STATISTICI 6 - COSTRUZIONE DI UN INDICATORE COMPOSITO 7 - IL CONCETTO DI QUALITÀ 8 - MODELLI PER LA CUSTOMER SATISFACTION 9 - L'INDAGINE STATISTICA DI CUSTOMER SATISFACTION 10 - LA RILEVAZIONE DEI DATI 11 - IL MODELLO DI REGRESSIONE LINEARE SEMPLICE 12 - I MODELLI STATISTICI DI PREVISIONE 13 - L'ANALISI MULTIDIMENSIONALE DEI DATI (AMD) 14 - CONSIDERAZIONI PRELIMINARI SU SCALARI E VETTORI 15 - ALGEBRA DELLE MATRICI 16 - L'ANALISI DELLE COMPONENTI PRINCIPALI (ACP) 17 - LA MATRICE DEI DATI E LA MATRICE DI BASE NELLA ACP 18 - IL SOFTWARE SPSS 19 - "I BIG DATA NEL TURISMO" INTERVISTA AL PROF. FABRIZIO ANTOLINI 20 - ANALISI IN COMPONENTI PRINCIPALI CON SPSS 21 - METODI STATISTICI PER IL CREDIT SCORING 22 - L'ANALISI FATTORIALE DISCRIMINANTE 23 - "I BIG DATA PER LA GESTIONE DELLE RISORSE UMANE" INTERVISTA 24 - L'ANALISI DELLE CORRISPONDENZE BINARIE (AC) 25 - I RISULTATI DELL'ANALISI DELLE CORRISPONDENZE BINARIE (AC) 26 - "I BIG DATA IN SANITÀ" INTERVISTA AL DOTT. STEFANO DOMENICO CICALA 27 - L'ANALISI DELLE CORRISPONDENZE MULTIPLE (ACM) 28 - CASO STUDIO: UN ESEMPIO DI ACM 29 - TEXT MINING 30 - I METODI DI CLASSIFICAZIONE AUTOMATICA 31 - LA CLUSTER ANALYSIS 32 - LA QUALITÀ NEL CONTROLLO STATISTICO DI PROCESSO 33 - I SETTE STRUMENTI MANAGERIALI DELLA QUALITÀ 34 - LE CARTE DI CONTROLLO 35 - I BIG DATA NELLA PUBBLICA AMMINISTRAZIONE - INTERVISTA 36 - INTRODUZIONE ALLA LINGUISTICA COMPUTAZIONALE. INTERVISTA CON MICHELE GALLO

## OBIETTIVI FORMATIVI PER IL RAGGIUNGIMENTO DEI RISULTATI DI APPRENDIMENTO PREVISTI

/\*\*/

Il corso ha lo scopo di offrire agli studenti conoscenze avanzate riguardo ai metodi e modelli per estrarre informazioni rilevanti da grandi moli di dati, con particolare attenzione all'apprendimento

statistico (statistical learning) in contesti di classificazione supervisionata (supervised classification).

## RISULTATI DI APPRENDIMENTO ATTESI

/\*\*/

## Conoscenza e capacità di comprensione

Il corso consentirà allo studente di conoscere le principali metodologie e modelli di apprendimento automatico per l'analisi dei dati con i loro punti di forza e di debolezza; sarà in grado di identificare un modello di analisi dei dati adeguato a un dato problema, di valutare le prestazioni empiriche e teoriche di diversi modelli; conoscere le principali piattaforme, linguaggi di programmazione e soluzioni per sviluppare questa tipologia di modelli che siano efficienti alla risoluzione dei problemi presentati.

## Capacità di applicare conoscenza e comprensione

Il corso vuole, oltre alla comprensione degli aspetti teorici, presentare delle problematiche che possano essere immediatamente affrontate dagli studenti tramite i metodi e le tecniche apprese nella durata dell'insegnamento e che possano consequenzialmente applicare nelle attività giornaliere.

## Autonomia di giudizio

Al completamento con successo di questo corso, gli studenti svilupperanno un atteggiamento critico nei confronti dei modelli e delle valutazioni empiriche che gli vengono presentati e saprà valutare quale delle tecniche di analisi statistica sia l'ideale per il raggiungimento dei propri scopi.

## Abilità comunicative

Lo studente saprà comunicare in modo chiaro e lineare idee originali, risultati sperimentali e principi alla base di tecniche avanzate di analisi dei dati in forma scritta e orale. Saranno inoltre in grado di offrire critiche costruttive alle presentazioni di lavori non propri.

## Capacità di apprendimento

Lo studente svilupperà le competenze necessarie per una comprensione efficace e per lo sviluppo di nuove metodologie di modellazione statistica insieme alla capacità di una loro effettiva implementazione.

## Programma didattico

Programma didattico (per macro aree + numero lezioni previste)

Introduzione ai principali modelli di statistical learning (apprendimento statistico) (lezioni previste n. 12)

Problemi di previsione e classificazione: richiami su regressione lineare e sui principali metodi di classificazione non supervisionata (lezioni previste n. 18)

1. Supervised classification: K-Nearest-Neighbours;
2. Errore di misclassificazione;
3. Metodi di ricampionamento: cross validation e bootstrap;
4. Metodi basati su alberi decisionali: regression trees, classification trees, bagging, random forests, boosting.

Introduzione ai metodi di classificazione Semi-supervisionata (lezioni previste n. 12)

Uso dell'ambiente statistico R (lezioni previste n. 12)

Modalità di raccordo con altri insegnamenti (indicare le modalità e gli insegnamenti con i quali sarà necessario raccordarsi)

Nessun raccordo

### **MODALITÀ DI ESAME ED EVENTUALI VERIFICHE DI PROFITTO IN ITINERE**

*/\*\*/*

L'esame può essere sostenuto sia in forma scritta che in forma orale.

L'esame orale consiste in un colloquio nel corso del quale il docente formula di solito tre domande. L'esame scritto consiste nello svolgimento di un test con 31 domande. Per ogni domanda lo studente deve scegliere una di 4 possibili risposte. Solo una risposta è corretta.

Sia le domande orali che le domande scritte sono formulate per valutare sia il grado di comprensione delle nozioni teoriche sia la capacità di ragionare utilizzando tali nozioni. Le domande sulle nozioni teoriche consentiranno di valutare il livello di comprensione. Le domande che richiedono l'elaborazione di un ragionamento consentiranno di valutare il livello di competenza e l'autonomia di giudizio maturati dallo studente.

Le abilità di comunicazione e la capacità di apprendimento saranno valutate attraverso le interazioni dirette tra docente e studente che avranno luogo durante la fruizione del corso (videoconferenze, e-tivity report, studio di casi elaborati) proposti dal docente o dal tutor.

### **MODALITÀ DI ISCRIZIONE E DI GESTIONE DEI RAPPORTI CON GLI STUDENTI**

*/\*\*/*

L'iscrizione ed i rapporti con gli studenti sono gestiti mediante la piattaforma informatica che permette l'iscrizione ai corsi, la fruizione delle lezioni, la partecipazione a forum e tutoraggi, il download del materiale didattico e la comunicazione con il docente.

Un tutor assisterà gli studenti nello svolgimento di queste attività.

Attività di didattica erogativa (DE) 54 Videolezioni + 54 test di autovalutazione

Impegno totale stimato: 54 ore.

Didattica Interattiva

Attività di didattica interattiva (DI) ed e-tivity con relativo feed-back al singolo studente da parte del docente o del tutor  
Redazione di un elaborato

Partecipazione a una web conference

Svolgimento delle prove in itinere con feedback

Svolgimento della simulazione del test finale, Totale 9 ore.

Attività di autoapprendimento 162 ore per lo studio individuale.

Libro di riferimento Dispense del docente.