



UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO	Scienze Economiche, Aziendali e Statistiche		
ANNO ACCADEMICO OFFERTA	2015/2016		
ANNO ACCADEMICO EROGAZIONE	2016/2017		
CORSO DILAUREA MAGISTRALE	SCIENZE STATISTICHE		
INSEGNAMENTO	METODI ESPLORATIVI PER "BIG DATA"		
TIPO DI ATTIVITA'	C		
AMBITO	21031-Attività formative affini o integrative		
CODICE INSEGNAMENTO	18165		
SETTORI SCIENTIFICO-DISCIPLINARI	SECS-S/01		
DOCENTE RESPONSABILE	PLAIA ANTONELLA	Professore Ordinario	Univ. di PALERMO
ALTRI DOCENTI			
CFU	8		
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	140		
NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA	60		
PROPEDEUTICITA'			
MUTUAZIONI			
ANNO DI CORSO	2		
PERIODO DELLE LEZIONI	1° semestre		
MODALITA' DI FREQUENZA	Facoltativa		
TIPO DI VALUTAZIONE	Voto in trentesimi		
ORARIO DI RICEVIMENTO DEGLI STUDENTI	PLAIA ANTONELLA Mercoledì 15:00 17:00 La modalita, in studio o su Teams, va concordata col docente		

DOCENTE: Prof.ssa ANTONELLA PLAIA

PREREQUISITI	
RISULTATI DI APPRENDIMENTO ATTESI	Conoscenza e capacità di comprensione Conoscenza dei metodi e delle procedure statistiche per analisi esplorative di data-set di grandi dimensioni. Capacità di applicare conoscenza e comprensione Lo studente dovrà essere in grado di: 1.formulare correttamente un problema 2.scegliere soluzioni di analisi statistico-informatiche appropriate, 3.tradurre i risultati in decisioni operative. Autonomia di giudizio Lo studente dovrà essere in grado di: 1.tradurre in termini statistici una esigenza conoscitiva sorta in differenti campi applicativi 2. intervenire con attività di pulizia, riorganizzazione, analisi descrittiva e interpretazione, dei dati, 3.elaborare e comunicare coerentemente i risultati delle proprie analisi. Abilità comunicative Lo studente dovrà essere in grado di: 1. comunicare con chiarezza, sia agli specialisti che ai non specialisti della materia, i concetti, e le tecniche di analisi dei dati studiati, 2.giustificare le scelte degli strumenti utilizzati per l'analisi, 3.comunicare i risultati con linguaggio appropriato. Capacità d'apprendimento Lo studente avrà sviluppato le abilità di apprendimento necessarie per approfondire autonomamente lo studio delle tecniche statistiche più comunemente utilizzate nell'analisi di grandi moli di dati.
VALUTAZIONE DELL'APPRENDIMENTO	Presentazione di una Tesina e Prova Orale
OBIETTIVI FORMATIVI	Il corso illustra metodi statistici di apprendimento da dati empirici complessi. L'obiettivo principale è l'analisi di grandi database al fine di trovare pattern, associazioni, cambiamenti, anomalie e strutture di particolare interesse. Alla fine del corso il discente sarà in grado di individuare gli strumenti adeguati per l'analisi che deve svolgere e applicarli, sintetizzando i risultati in modo efficace.
ORGANIZZAZIONE DELLA DIDATTICA	Lezioni frontali, Esercitazioni
TESTI CONSIGLIATI	Dispense rese disponibili dal docente sul portale di Ateneo. Risorse on-line. Breiman, L. Friedman, J. H. Olshen, R. A. Stone, C. J. (1984) Classification and regression trees, Chapman & Hall. Brian S. Everitt, Sabine Landau, Morven Leese, Daniel Stahl (2011). Cluster Analysis, 5th Edition. Wiley G. James, D. Witten, T. Hastie, R. Tibshirani . (2013) An Introduction to Statistical Learning, with applications in R. Springer.

PROGRAMMA

ORE	Lezioni
2	Analisi esplorativa dei dati
2	Data Visualization
2	Trasformazioni di variabili
2	Origine e Caratteristiche dei Big Data
6	Trattamento dei dati mancanti
12	Metodi di classificazione: Alberi, bagging, boosting, random forest
10	Cluster analysis

ORE	Esercitazioni
2	Data visualization in R
2	Trasformazioni di variabili
4	Trattamento dei dati mancanti
8	Metodi di classificazione: Alberi, bagging, boosting, random forest
8	Cluster analysis