

CORSO DI STUDIO *Strategie di Impresa e Management (SIM)*
ANNO ACCADEMICO 2023-2024

DENOMINAZIONE DELL'INSEGNAMENTO *Metodi Statistici Multivariati (Multivariate Data Analysis)*

Principali informazioni sull'insegnamento	
Anno di corso	2° anno
Periodo di erogazione	11/09 – 22/12 2023
Crediti formativi universitari (CFU/ETCS):	8
SSD	Statistica (SECS-S/01)
Lingua di erogazione	Italiano (English on demand)
Modalità di frequenza	Facoltativa

Docente	
Nome e cognome	Massimo Bilancia
Indirizzo mail	massimo.bilancia@uniba.it
Telefono	N/A
Sede	Dipartimento Jonico in "Sistemi Giuridici ed Economici del Mediterraneo: società, ambiente, culture"
Sede virtuale	Microsoft teams (codice: a3ixcpa)
Ricevimento	Qualsiasi giorno, previo appuntamento

Organizzazione della didattica			
Ore			
Totali	Didattica frontale	Pratica (laboratorio, campo, esercitazione, altro)	Studio individuale
200	64		136
CFU/ETCS			
8	8		

Obiettivi formativi	Fornire agli studenti gli elementi di base, sia teorici che pratici, per le tecniche di Data Science e le loro applicazioni in campo economico e aziendale
Prerequisiti	Conoscenze di base di Statistica e Matematica

Metodi didattici	Lezioni frontali sugli argomenti teorici + lezioni di laboratorio in R
-------------------------	--

Risultati di apprendimento previsti	
DD1 Conoscenza e capacità di comprensione	Il corso si prefigge di fornire gli elementi di base delle tecniche di Data Science
DD2 Conoscenza e capacità di comprensione applicate	Lo studente apprenderà a stimare ed utilizzare in pratica i modelli appresi durante la parte teorica del corso, mediante l'utilizzo del più diffuso software di analisi dei dati.

<p>DD3-5 Competenze trasversali</p>	<ul style="list-style-type: none"> • <i>Autonomia di giudizio</i> <ul style="list-style-type: none"> ○ lo studente sarà in grado di decidere il modello più appropriato da utilizzare per estrarre nuova conoscenza negli svariati ambiti di applicazione previsti (applicazioni campo aziendale, business intelligence e marketing). • <i>Abilità comunicative</i> <ul style="list-style-type: none"> ○ lo studente, al termine del corso, avrà acquisito la preparazione necessaria a generare reportistica, negli ambiti di applicazione previsti, sulla nuova conoscenza estratta. • <i>Capacità di apprendere in modo autonomo</i> <ul style="list-style-type: none"> ○ il corso si prefigge di fornire gli elementi di base delle tecniche di Data Science, conoscenze sulle quali fondare la possibilità di seguire corsi di natura più avanzata mirati alla preparazione della figura di Data Scientist.
<p>Contenuti di insegnamento (Programma)</p>	<p>Modulo I. Elementi di probabilità ed inferenza statistica</p> <ol style="list-style-type: none"> 1. Introduzione di base 2. Probabilità elementare 3. Variabili aleatorie discrete e continue 4. Variabili aleatorie doppie e multiple 5. Elementi di inferenza statistica 6. Matrici dati <p>Modulo II. Modelli di Data Mining e Knowledge Discovery per l'azienda</p> <ol style="list-style-type: none"> 7. Apprendimento supervisionato e non supervisionato 8. Regole associative e Market Basket Analysis 9. Classificazione Naïve Bayes 10. Alberi di decisione 11. Clustering I: algoritmi gerarchici 12. Clustering II: k-means <p>Modulo III. Laboratorio</p> <p>Il laboratorio è parte integrante del corso. Sarà utilizzato il software di analisi dei dati R, liberamente disponibile sul network del CRAN all'indirizzo http://cran.r-project.org. Durante il corso di laboratorio sarà erogata un'introduzione approfondita a R, e verranno mostrati casi pratici di studio relativi a tutti gli argomenti teorici riguardanti la seconda parte.</p>
<p>Testi di riferimento</p>	<p>M. Bilancia (2020) <i>Dispense per il Corso di Metodi Statistici Multivariati – Versione 1.2 Settembre 2020</i>. Dipartimento Ionico, liberamente distribuite sotto Licenza Creative Commons 4.0 CC BY-NC-ND</p>
<p>Note ai testi di riferimento</p>	<p>N/A</p>
<p>Materiali didattici</p>	<p>Tutto il materiale didattico è reperibile sulla classe Teams del corso</p>
<p>Valutazione</p>	
<p>Modalità di verifica dell'apprendimento</p>	<ol style="list-style-type: none"> 1. PROVA PRATICA AL CALCOLATORE (durata 1 ore), consistente nell'analisi di un dataset reale mediante i metodi esposti durante le lezioni di laboratorio: il candidato dovrà produrre un documento Microsoft™ Word™ (oppure utilizzando il word processor di Libre Office) contenente tutti i comandi di R utilizzati, tutto l'output testuale e grafico prodotto e i relativi commenti esplicativi. Per lo svolgimento della prova pratica è consentita la consultazione delle dispense di laboratorio 4-5-6-7, stampate su cartaceo: avrete a disposizione, a tal fine, le dispense di queste lezioni riportate su un formato più pratico di 4 slides per foglio. Queste slides devono essere stampate così come le avete ricevute, ossia non devono essere da voi commentate o chiosate con i vostri appunti (inoltre, non è possibile utilizzare il pdf delle dispense: la violazione di tutte queste prescrizioni comporterà l'impossibilità dell'utilizzo delle medesime durante la prova

	<p>d'esame). Non è consentita la consultazione di alcun altro testo o documento (scritto o in formato elettronico), pena l'esclusione immediata dalla prova d'esame.</p> <p>2. PROVA SCRITTA (durata 1 ore), vertente sulla produzione di un elaborato su una domanda aperta (a libero svolgimento) selezionate tra tutti gli argomenti di natura teorica svolti in aula. Per lo svolgimento dell'elaborato non è consentita la consultazione di alcun testo, documento o supporto informatico/telefonico di qualsiasi natura, pena l'esclusione immediata dalla prova d'esame.</p>
Criteri di valutazione	<ul style="list-style-type: none"> • <i>Conoscenza e capacità di comprensione:</i> <ul style="list-style-type: none"> ○ Capacità di maturare una conoscenza nel campo della teoria della probabilità e dell'analisi multivariata dei dati, con un'ottica in vista delle applicazioni in campo economico e business. • <i>Conoscenza e capacità di comprensione applicate:</i> <ul style="list-style-type: none"> ○ Capacità di applicare nella pratica mediante un software di analisi dei dati dedicato le competenze teoriche apprese. • <i>Autonomia di giudizio:</i> <ul style="list-style-type: none"> ○ Capacità di saper selezionare la tecnica di analisi dei dati adeguata alla natura del problema proposto. • <i>Abilità comunicative:</i> <ul style="list-style-type: none"> ○ Capacità di produrre una relazione efficace sul problema proposto. • <i>Capacità di apprendere:</i> <ul style="list-style-type: none"> ○ Dimostrare di saper efficacemente esporre le competenze apprese.
Criteri di misurazione dell'apprendimento e di attribuzione del voto finale	<p>La prova di laboratorio viene valutata mediante una idoneità (SI/NO). La prova scritta viene valutata in 30esimi. Il voto attribuito al candidato che ha superato entrambe le prove è quello della prova scritta.</p> <p>Il candidato che non ha superato una delle due prove può ripeterla in una degli appelli successivi.</p> <p>Ogni candidato può rifiutare selettivamente il voto della sola prova scritta, e conservare la prova di laboratorio (se superata) per gli appelli successivi.</p>
Altro	N/A
	.