Università degli Studi di Roma "Tor Vergata"

Facsimile Scheda Insegnamento

Mod. Scheda Insegnamento v. 1.1

Docente responsabile dell'insegnamento/attività formativa

Nome	Mauro
Cognome	De Sanctis

Denominazione insegnamento/attività formativa

Italiano	
Inglese	Information Theory and Data Science

Informazioni insegnamento/attività formativa

A.A.	2025/26	○ L	◉ LM	○ LM CU
CdS	ICT and Internet Engineering			
Codice	80300052			
Canale				
CFU	6			
Lingua	Inglese			

Docente del modulo didattico (compilare solo per attività formative articolate in moduli)

Nome	
Cognome	

Denominazione modulo didattico (compilare solo per attività formative articolate in moduli)

Italiano	
Inglese	

Università degli Studi di Roma “Tor Vergata”

Facsimile Scheda Insegnamento

Obiettivi formativi e risultati di apprendimento attesi

Italiano	OBIETTIVI FORMATIVI: Questo corso ha l’obiettivo di fornire una visione unificata di informazione, conoscenza e analisi dei dati, estendendo le applicazioni della teoria dell'informazione oltre la teoria classica delle comunicazioni. In particolare, il corso si concentra su metodi statistici per il data science. Verranno presentate applicazioni di teoria dell'informazione e data science in diverse aree, tra cui comunicazioni digitali, economia, marketing, biologia, medicina, meteorologia. CONOSCENZA E CAPACITÀ DI COMPRENSIONE: Al termine del corso lo studente sarà in grado di: guadagnare una conoscenza approfondita della teoria dell’informazione e delle sue applicazioni nelle tecniche di apprendimento statistico; comprendere le sfide che si presentano nei problemi di analisi dei dati.
Inglese	LEARNING OUTCOMES: This course will give a unified view of Information, Knowledge and Data Analysis, extending the applications of Information Theory beyond the classical theory of communications. In particular, the course focuses on statistical methods for data analysis. Applications of information theory and data science in several areas will be presented including: digital communications, economics, marketing, biology, medicine, meteorology. KNOWLEDGE AND UNDERSTANDING: After the course, the student will be able to: gain an in depth knowledge of the information theory and its applications to statistical learning techniques; understand the challenges that arise in data science tasks. APPLYING KNOWLEDGE AND UNDERSTANDING: After the course. the student will be

Università degli Studi di Roma “Tor Vergata”

Facsimile Scheda Insegnamento

Prerequisiti

Italiano	Teoria della probabilità (o equivalente).
Inglese	Probability Theory (or equivalent).

Programma

Italiano	Elementi di teoria della probabilità: variabili e processi aleatori di tipo continuo e discreto, densità di probabilità, massa di probabilità, valore atteso. Teoria dell'informazione: concetto di informazione, autoinformazione, entropia di Shannon, misure alternative di entropia, entropia relativa, divergenza di Kullback-Leibler, divergenza di Jensen-Shannon, entropia condizionale, entropia congiunta, informazione reciproca, correlazione totale, entropia differenziale, misure di informazione normalizzate. Applicazioni al data science: concetto base di data science, definizione di dataset e attributi/feature, train set e test set, tipi di dati, analisi multivariata, descrizione statistica dei dataset, case study, metriche teoriche delle informazioni in attività di data science, preparazione dei dati, pulizia dei dati, discretizzazione degli attributi, riduzione della
Inglese	Elements of probability theory: discrete and continuous random variables and processes, probability distribution function, probability mass function, expected value. Information Theory: concept of information, self-information, Shannon entropy, alternative entropy measures, relative entropies, Kullback-Leibler divergence, Jensen-Shannon divergence, conditional entropy, joint entropy, mutual information, total correlation, differential entropy, normalized measures of information. Data science: basic concept of data science, definition of dataset and attribute/feature, train set and test set, data types, multivariate analysis, basic statistical description of data, case studies, information theoretic metrics in data science tasks, data preparation, data cleaning, discretization of attributes, dimensionality reduction (Singular Value Decomposition)

Università degli Studi di Roma “Tor Vergata”

Facsimile Scheda Insegnamento

Modalità di valutazione

- [x] Prova scritta
- [ ] Prova orale
- [ ] Valutazione in itinere
- [x] Valutazione di progetto
- [ ] Valutazione di tirocinio
- [ ] Prova pratica
- [ ] Prova di laboratorio

Descrizione delle modalità e dei criteri di verifica dell’apprendimento

Italiano	L'esame consiste in un project work. Il project work consiste nell'implementazione di una tecnica di data science (codice sorgente e un documento descrittivo che contiene eventualmente risultati sotto forma di tabelle o figure). Il project work deve essere adeguatamente documentato dallo studente. In fase di esame orale, ogni studente presenta e discute il suo progetto con il docente. Inoltre, le domande orali sono finalizzate a verificare la preparazione dello studente e la sua capacità di stabilire collegamenti critici tra i diversi temi affrontati durante il corso. Il voto finale dell’esame, espresso in trentesimi, sarà assegnato secondo i seguenti criteri: - Non idoneo: gravi lacune nella comprensione degli argomenti, scarsa capacità di analisi e giudizio, esposizione incoerente e linguaggio inadeguato
Inglese	The exam consists in a project work. The project work consists in the implementation of a data science technique (source code and a descriptive document eventually containing results in the form of Tables or Figures). The project work needs to be properly documented by the student. During the oral exam, each student presents and discusses their project with the instructor. In addition, the oral questions are aimed at assessing the student's preparation and his/her ability to make critical connections between the various topics covered throughout the course. The final exam grade, expressed out of thirty, will be assigned according to the following criteria:

Università degli Studi di Roma “Tor Vergata”

Facsimile Scheda Insegnamento

Testi adottati

Italiano	Slides fornite dal docente
Inglese	Slides provided by the lecturer

Bibliografia di riferimento

Italiano	E. Cianca, M. De Sanctis, M. Ruggieri, ``Information and Coding: theory overview, Design, Applications and Exercises'', ARACNE Editrice, 2007. M. J. Zaki, W. Meira, ``Data Mining and Analysis - Fundamental Concepts and Algorithms'', Cambridge University Press, 2014.
Inglese	E. Cianca, M. De Sanctis, M. Ruggieri, ``Information and Coding: theory overview, Design, Applications and Exercises'', ARACNE Editrice, 2007. M. J. Zaki, W. Meira, ``Data Mining and Analysis - Fundamental Concepts and Algorithms'', Cambridge University Press, 2014.

Università degli Studi di Roma “Tor Vergata”

Facsimile Scheda Insegnamento

Modalità di svolgimento

☒ Modalità in presenza
☐ Modalità a distanza

Descrizione della modalità di svolgimento e metodi didattici adottati

Italiano	Il corso si svolge con lezioni frontali tradizionali per l’80% e per il 20% con lezioni pratiche (“hands-on”) che richiedono l’utilizzo di un PC da parte dello studente. Nelle lezioni “hands-on” lo studente è guidato dal docente nello svolgimento di attività legate alla teoria dell’informazione e alle tecniche di data science.
Inglese	The course is based on 80% of traditional frontal lectures and on 20% of “hands-on” practical lectures. The “hands-on” lectures require the use of a PC by the student. During the “hands-on” lectures the student is led by the teacher through a set of activities concerning information theory and data science techniques.

Modalità di frequenza

◉ Frequenza facoltativa
○ Frequenza obbligatoria

Descrizione della modalità di frequenza

Italiano	La frequenza è facoltativa ma fortemente consigliata per poter raggiungere pienamente gli obiettivi formativi.
Inglese	The attendance is not mandatory but it is strongly suggested to achieve all the planned learning outcomes.