Università degli Studi di Roma "Tor Vergata"

Scheda Insegnamento

Mod. Scheda Insegnamento v. 1.1

Docente responsabile dell'insegnamento/attività formativa

Nome	Andrea
Cognome	Clementi

Denominazione insegnamento/attività formativa

Italiano	ALGORITMI E STRUTTURE DATI PER BIG DATA
Inglese	ALGORITHMS AND DATA STRUCTURES FOR BIG DATA

Informazioni insegnamento/attività formativa

A.A.	2025/26

(•) L ( ) LM ( ) LM CU

CdS	Metodi e Modelli per Data Science
Codice	8067757
Canale	
CFU	9
Lingua	Italiano

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"

Scheda Insegnamento

Obiettivi formativi e risultati di apprendimento attesi

Italiano	OBIETTIVI FORMATIVI: Il corso ha come obiettivi l'apprendimento dei principali metodi e strumenti algoritmici per l'analisi e la gestione efficiente di enormi quantità di dati. In particolare, si studiano ed implementano modelli dinamici (streaming, avversariali e gerarchici) per la rappresentazione dei dati e la realizzazione di strutture dati che permettono queries di natura statistica/combinatorica in modo efficiente sia come costi di memoria che come tempo di esecuzione. Istanze importanti di queste strutture che si studieranno sono le Locality Hasing Function, il Min-Hashing, la Shingling Technique, i modelli di dati Streaming, il sampling su streaming, gli algoritmi ranodmizzati di counting, e l'algoritmo di page-rank per il WEB. Verranno anche studiati modelli e tecniche parallele su Map-Reduce. CONOSCENZA E CAPACITÀ DI COMPRENSIONE: Comprendere questioni e processi complessi attinenti gli aspetti algoritmici e di performance evalutation in Data Mining CAPACITÀ DI APPLICARE CONOSCENZA E COMPRENSIONE: Capacità di applicare modelli e tecniche in contesti applicativi che riguardano diversi sistemi di data mining e information retrieval.
Inglese	LEARNING OUTCOMES: The main learning goal of the course is the study of models and algorithmic tools for the management and analysis (mining) of large sets of data (Big Data). In particular, the course focuses on dynamic digital models (streaming, adversarial, hierarchical) for the raw input data and describe advanced data structures that allow space/time efficient combinatorial and statistical queries. Typical examples of such structures that will be described are: Locality Sensitive Hashing, Shingling Technique, Data Streaming, Sampling over Data Streaming, probabilistic counting, Page-Rank algorithm for WEB Information Retrieval. It will also give an introduction to fast parallel techniques on the Map-Reduce Architecture. KNOWLEDGE AND UNDERSTANDING: Cope with complex questions and processes in Data Mining Algorithms and Performance Evalutations APPLYING KNOWLEDGE AND UNDERSTANDING: Ability to apply models and

Prerequisiti

Italiano	Algoritmi e Strutture di Dati Probabilità I
Inglese	Algorithms and Data Structures Probability I

Programma

Italiano	PART A: Probability and Computing: Randomized Algorithms for Big Data Basic Notions of Probability (Chapt 1 of [2]) Verifying Polynomial Identities (Chapt 1.1 of [2], Verifying Matrix Multiplications (Chapt. 1.3 of [2], Contention Resolutions (Chapt 13 of [3])
Inglese	PART A: Probability and Computing: Randomized Algorithms for Big Data Basic Notions of Probability (Chapt 1 of [2]) Verifying Polynomial Identities (Chapt 1.1 of [2], Verifying Matrix Multiplications (Chapt. 1.3 of [2], Contention Resolutions (Chapt 13 of [3])

Modalità di valutazione

- ☐ Prova scritta
- ☒ Prova orale
- ☐ Valutazione in itinere
- ☒ Valutazione di progetto
- ☐ Valutazione di tirocinio
- ☐ Prova pratica
- ☐ Prova di laboratorio

Descrizione delle modalità e dei criteri di verifica dell'apprendimento

Italiano	Verifica orale in fase di esame Domande ed esercizi svolti a lezione Svolgimento di progetti sia teorici che pratici (sviluppo e sperimentazione SW) - Facoltativo.
Inglese	Oral Tests (Final Exams) Questions, Exercises Theoretical and Practical Projects (not mandatory)

Testi adottati

Italiano	1] Leskovec, Rajaraman, and Ullamann: Mining of Massive Data Sets (+ slides in preparazione) [2] Mitzenmacher and Upfal: Probability and Computing (Cambridge) [3] Kleinberg and Tardos: Algorithm Design
Inglese	1] Leskovec, Rajaraman, and Ullamann: Mining of Massive Data Sets (+ slides in preparazione) [2] Mitzenmacher and Upfal: Probability and Computing (Cambridge) [3] Kleinberg and Tardos: Algorithm Design

Bibliografia di riferimento

Italiano	1] Leskovec, Rajaraman, and Ullamann: Mining of Massive Data Sets (+ slides in preparazione) [2] Mitzenmacher and Upfal: Probability and Computing (Cambridge) [3] Kleinberg and Tardos: Algorithm Design
Inglese	1] Leskovec, Rajaraman, and Ullamann: Mining of Massive Data Sets (+ slides in preparazione) [2] Mitzenmacher and Upfal: Probability and Computing (Cambridge) [3] Kleinberg and Tardos: Algorithm Design

Modalità di svolgimento

- [x] Modalità in presenza
- [ ] Modalità a distanza

Descrizione della modalità di svolgimento e metodi didattici adottati

Italiano	Svolgimento di Lezioni ed Esercitazioni in aula con videoproiettore e lavagna Verifica ed esposizione dei Progetti presentati da parte degli studenti (facoltativo)
Inglese	Lectures in class with slides and balckboard Presentation of Student's Projects (not mandatory)

Modalità di frequenza

- (•) Frequenza facoltativa
- ( ) Frequenza obbligatoria

Descrizione della modalità di frequenza

Italiano	frequenza fortemente consigliata ma non obbligatoria
Inglese	not mandatory