

Metodi e linguaggi per il trattamento dei dati (MTD)

PROGRAMMA DEL CORSO (potrà subire piccole variazioni)

Informazioni generali:

- 48 ore di lezione frontale
- Parte delle lezioni dedicate alla *teoria*, parte ad *esercitazioni* pratiche che richiedono la disponibilità di un portatile personale. Nel corso della prima lezione procederemo all'installazione del software necessario per l'avvio delle lezioni.
- Email docente: matteo.re@unimi.it
- ricevimento studenti:
concordare via email , Dipartimento di Informatica, via Celoria 18, ufficio 3010
- sito del corso: <https://homes.di.unimi.it/re/mtd1920.html>

Mailing list del corso:

<https://docs.google.com/spreadsheets/d/1Yjs8RQGi3UCw-O7MxDyPhVoqafkh-jnMtErrxg8IFes/edit?usp=sharing>

Obiettivi didattici: Fornire agli studenti le competenze necessarie per l'analisi, l'elaborazione e la manipolazione di dati di natura eterogenea. Acquisizione di competenze di programmazione procedurale.

Competenze acquisite:

Linguaggi di programmazione orientati all'analisi dei dati. Linguaggio R (statistica), linguaggio PERL (elaborazione testi ed accesso a banche dati remote), linguaggio SQL (interrogazione a banche dati), linguaggio Python, analisi dei dati ed apprendimento automatico.

Struttura corso: Il corso è diviso in *tre moduli*, ognuno dedicato ai tre linguaggi di programmazione principali trattati nel corso: R, PERL, PYTHON. Ogni modulo contiene una parte di introduzione al linguaggio seguita da esempi di applicazione per un totale di 16 ore.

Orario lezioni:

Martedì: 13.30-15.30	aula 310
Giovedì 13.30-15.30	aula 306

(settore didattico via Celoria)

Modalita' esame: Esame composto da tre parti (ognuna da 10 punti) dedicate ai tre linguaggi di programmazione trattati. Per le parti di R e PERL e' previsto un esame in aula di calcolo. La parte dell'esame dedicata a PYTHON consiste in un piccolo progetto da realizzare a casa e da consegnare prima di sostenere l'esame finale. I progetti saranno avviati in classe nel corso delle ultime lezioni e potrete avere l'assistenza del docente durante la realizzazione del progetto.

PROGRAMMA DEL CORSO (potra' essere soggetto a piccole variazioni)

PARTE 0: Introduzione

- Algoritmi e linguaggi di programmazione. Linguaggi di programmazione a basso ed alto livello. Linguaggi interpretati e compilati.

PARTE I: DESCRIZIONE STATISTICA DEI DATI (linguaggio R)

- L'interfaccia grafica per l'utente di R
- Identificatori e variabili; tipi di dati base; operatori, espressioni e istruzioni
- Strutture dati fondamentali in R: vettori, fattori, matrici, array, liste, data frame ed environment
- Strutture di controllo del flusso di esecuzione: blocchi, istruzioni condizionali, iterazioni
- Funzioni e script
- Operazioni di I/O
- L'ambiente grafico di R

ESERCITAZIONI R

- Grafici (istogrammi, grafici a torta)
- Regressione lineare
- Distribuzioni. Distribuzione gaussiana. Plot grafico di distribuzione. Integrali.
- Calcolo di P-value
- Stima intervallo di confidenza

PARTE II: Manipolazione di dati locali e remoti (linguaggio PERL)

- Basi di linguaggio PERL, tipi di dati, strutture di controllo
- Manipolazione di dati in forma testuale
- Espressioni regolari I
- Espressioni regolari II

ESERCITAZIONI PERL

- Accesso a banche dati remote: il linguaggio SQL
- Interrogazione diretta banche dati remote via SQL
- Utilizzo di Application Programming Interfaces (APIs) per l'accesso diretto a banche dati remote

PARTE III: Classificazione ed inferenza (linguaggio PYTHON)

- Basi di PYTHON
- Introduzione all'apprendimento automatico
- Esempi di classificazione ed inferenza