Università Mercatorum

Università telematica delle Camere di Commercio Italiane

PROGRAMMA DEL CORSO DI PROGRAMMAZIONE PER LA STATISTICA E I BIG DATA

SETTORE SCIENTIFICO

ING-INF/05

CFU

9

OBIETTIVI

Python è un linguaggio intuitivo e un software open source, con una molteplicità di librerie disponibili per i diversi campi applicativi.

L’obiettivo del corso è fornire allo studente gli strumenti basilari per poter programmare con tale linguaggio, fornendo in particolare una conoscenza approfondita delle librerie adatte nell’ambito della Data Science.

MODALITÀ DI ESAME ED EVENTUALI VERIFICHE DI PROFITTO IN ITINERE

L’esame può essere sostenuto sia in forma scritta che in forma orale.

L’esame orale consiste in un colloquio nel corso del quale il docente formula di solito tre domande. L’esame scritto consiste nello svolgimento di un test con 30 domande. Per ogni domanda lo studente deve scegliere una di 4 possibili risposte. Solo una risposta è corretta.

Sia le domande orali che le domande scritte sono formulate per valutare sia il grado di comprensione delle nozioni teoriche sia la capacità di ragionare utilizzando tali nozioni. Le domande sulle nozioni teoriche consentiranno di valutare il livello di comprensione. Le domande che richiedono l’elaborazione di un ragionamento consentiranno di valutare il livello di competenza e l’autonomia di giudizio maturati dallo studente.

Le abilità di comunicazione e la capacità di apprendimento saranno valutate attraverso le interazioni dirette tra docente e studente che avranno luogo durante la fruizione del corso (videoconferenze, e-tivity report, studio di casi elaborati) proposti dal docente o dal tutor.

MODALITÀ DI ISCRIZIONE E DI GESTIONE DEI RAPPORTI CON GLI STUDENTI

L’iscrizione ed i rapporti con gli studenti sono gestiti mediante la piattaforma informatica che permette l’iscrizione ai corsi, la fruizione delle lezioni, la partecipazione a forum e tutoraggi, il download del materiale didattico e la comunicazione con il docente.

Un tutor assisterà gli studenti nello svolgimento di queste attività.

ATTIVITÀ DI DIDATTICA EROGATIVA (DE)

- 54 Videolezioni + 54 test di autovalutazione

Impegno totale stimato: 54 ore

ATTIVITÀ DI DIDATTICA INTERATTIVA (DI) ED E-TIVITY CON RELATIVO FEED-BACK AL SINGOLO STUDENTE DA PARTE DEL DOCENTE O DEL TUTOR

- Applicazione di Python per risolvere un problema pratico
- Partecipazione a una web conference
- Svolgimento delle prove in itinere con feedback
- Forum sui diversi software statistici

Totale 12 ore

ATTIVITÀ DI AUTOAPPRENDIMENTO

- 240 ore per lo studio individuale

LIBRO DI RIFERIMENTO

- Dispense del docente.
- Testi di riferimento principali:

1. Gaddis, T. (2016). Introduzione a Python. Ediz. mylab. Con espansione online. Pearson.
2. Zinoviev, D. (2017). Data science con Python. Apogeo.

- Testi di riferimento secondari:

1. Maggi, G. (2020). Data science con Python. Edizioni LSWR.
2. Marin, I., Shukla, A. Sarang, VK. (2019). L’analisi dei Big Data con Python. Tecniche nuove.
3. Moore, A. D. (2018). Python GUI Programming with Tkinter. Packt.
4. Botto, M., Raganelli, V. (2021). Programmazione e analisi statistica con R. in riga edizioni.
5. Di Franco, G. (2009). L'analisi dei dati con SPSS. FrancoAngeli.

RISULTATI DI APPRENDIMENTO ATTESI

- Conoscenza e capacità di comprensione

Il corso intende fornire le conoscenze utili per implementare programmi con Python, utili per l'analisi statistica dei Big data. Esse includono: conoscenza e comprensione di fondamenti del linguaggio, degli oggetti e funzioni usate, delle principali librerie statistiche e delle capacità grafiche e di visualizzazione di Python. L'attenzione è rivolta essenzialmente ai metodi applicati in caso di big data.

- Capacità di applicare conoscenza e comprensione

Il corso permette allo studente di effettuare elaborazioni statistiche su dati standard e Big data, dando la capacità allo stesso di attingere alle opportune librerie/moduli offerte dal programma, per risolvere in modo più appropriato il problema e l'oggetto sotto studio.

- Autonomia di giudizio

Autonome capacità di giudizio sulla qualità dei risultati applicativi ottenuti, nonché autonome capacità di giudizio, di valutazione comparativa e scelta di soluzioni, moduli e software, avendo una visione coplessiva anche su altri software statistici (R, SAS, SPSS).

- Abilità comunicative

Capacità di comunicare in maniera appropriata le elaborazioni svolte tramite Python, sia attraverso gli strumenti di visualizzazione e grafici, che in forma più approfondita, come output delle funzioni e applicazioni eseguite, con particolare attenzione ai casi di utilizzo di Big data. Lo sviluppo di abilità comunicative, sia orali che scritte, sarà anche stimolata attraverso la didattica interattiva (con la redazione di elaborati da parte dello studente) e i momenti di videoconferenza attivati, ivi compreso la prova finale di esame.

- Capacità di apprendimento

Capacità di apprendere, sperimentare ed usare, in modo autonomo, linguaggi di programmazione anche di nuova definizione ed implementazione, rivolti all'uso di big data. Capacità di seguire l'evoluzione della programmazione Python in ogni suo aspetto attraverso la letteratura e la documentazione tecnica, rimanendo aggiornato sulle nuove librerie rese disponibili in forma open-source. Capacità di aggiornamento ed autoaggiornamento.

PROGRAMMA DIDATTICO

1. MODULO: Introduzione al linguaggio Python (n. 4 lezioni)

2. MODULO: Script in linguaggio Python (n. 5 lezioni)

3. MODULO: Struttura dati e oggetti (n. 3 lezioni)

4. MODULO: Utilizzo di database (n. 4 lezioni)

5. MODULO: Programmazione strutturata (n. 6 lezioni)

6. MODULO: I moduli e le librerie di Python (n. 5 lezioni)

7. MODULO: Librerie per i Big data e Data science (n. 6 lezioni)

8. MODULO: Utilizzo dei dati delle reti (n. 5 lezioni)

9. MODULO: L’interfaccia di Python (n. 4 lezioni)

10. MODULO: Software alternativi: R, SAS e SPSS (n. 6 lezioni)

11. MODULO: Esempi applicativi con Big Data (n. 6 lezioni)

Il docente si riserva il diritto di modificare il titolo delle lezioni.

VIDEOLEZIONI

1 - IL LINGUAGGIO PYTHON

2 - FONDAMENTI DEL LINGUAGGIO PYTHON

3 - OPERAZIONI SUI DATI

4 - STRUTTURE DECISIONALI E ITERATIVE

5 - FUNZIONI

6 - CASO DI STUDIO: PENSARE CON LE FUNZIONI

7 - FILE

8 - ECCEZIONI

9 - CASO DI STUDIO: PROGRAMMARE UTILIZZANDO FILE E GESTENDO LE ECCEZIONI

10 - LISTE

11 - DIZIONARI E SET

12 - CLASSI E OGGETTI

13 - DATABASE RELAZIONALI

14 - SQL

15 - SQL AVANZATO

16 - CONFIGURAZIONE E USO DI UN DATABASE MYSQL CON PYTHON

17 - OPERARE SULLE STRINGHE

18 - ESERCITAZIONI DI PROGRAMMAZIONE STRUTTURATA

19 - ESPRESSIONI REGOLARI

20 - APPROFONDIMENTO SULLE ESPRESSIONI REGOLARI

21 - APPROFONDIMENTO SUI FILE

22 - OPERAZIONI SULLE SEQUENZE

23 - ANACONDA E LA LIBRERIA NUMPY

24 - ARRAY IN NUMPY

25 - OPERAZIONI SUI DATI

26 - RAPPRESENTAZIONE GRAFICA

27 - PROBABILITA' E STATISTICA

28 - PANDAS

29 - OPERAZIONI CON PANDAS

30 - OPERAZIONI AVANZATE CON PANDAS

31 - TECNOLOGIE PER I BIG DATA

32 - PYSPARK

33 - BIG DATA CON PYSPARK

34 - I GRAFI E LE RETI

35 - UTILITY PER IL RECUPERO E L'ANALISI DEI DATI DELLE RETI

36 - "NETWORKX"

37 - VISUALIZZAZIONE E ANALISI DI GRAFI

38 - ESEMPIO DI ANALISI DEI DATI DELLE RETI

39 - INTRODUZIONE ALLA GUI DI PYTHON CON TKINTER

40 - PROGETTAZIONE DEI FORM CON TKINTER

41 - WIDGET TKINTER E CLASSE LABELINPUT

42 - PROGRAMMAZIONE DEL FORM

43 - INTRODUZIONE A R

44 - FATTORI E MATRICI IN R

45 - DATAFRAME E LISTE IN R

46 - OPERAZIONI E RAPPRESENTAZIONE GRAFICA IN R

47 - SPSS

48 - SAS

49 - PROBLEMI DI DATA SCIENCE: DEFINIZIONE, VARIABILE TARGET E METRICHE

50 - GESTIONE DEI DATI MANCANTI E RIPRODUCIBILITÀ COMPUTAZIONALE

51 - IL FENOMENO BIG DATA

52 - BIG DATA, SOCIAL NETWORK E SISTEMA DELL'INFORMAZIONE

53 - AMAZON WEB SERVICES (AWS)

54 - DOMINI APPLICATIVI DEI BIG DATA E INTEGRAZIONE CON L'AI