UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Data Science

2526-2-E1803M119

Obiettivi formativi

Il corso si propone di familiarizzare gli studenti con i fondamenti della programmazione e fornire loro una panoramica delle tecniche di data science e intelligenza artificiale utilizzate in un contesto aziendale. Il programma comprende un'introduzione alla programmazione (Python), ai database (SQL) e alle tecniche di machine learning. **Le lezioni e il materiale didattico saranno in lingua inglese.**

Il corso è alternativo al corso 'Informatica Generale e Laboratorio Informatico' che si concentra maggiormente sui sistemi informativi aziendali e delle loro implicazioni organizzative e gestionali.

Obiettivi formativi (secondo i Descrittori di Dublino)

1. Conoscenza e capacità di comprensione (Knowledge and understanding)
Al termine del corso, lo studente avrà acquisito:

- la conoscenza dei principi fondamentali della programmazione imperativa;
- la comprensione dei fondamenti dei database relazionali e del linguaggio SQL;
- una panoramica delle principali tecniche di data science, machine learning e intelligenza artificiale applicabili al contesto aziendale;
- la comprensione del ruolo dei dati e dell'innovazione digitale nei processi aziendali, in particolare nel settore finanziario.

2. Capacità di applicare conoscenza e comprensione (Applying knowledge and understanding)
Al termine del corso, lo studente sarà in grado di:

- progettare e implementare semplici programmi in linguaggio Python;
- interrogare ed elaborare dati mediante SQL;
- proporre soluzioni operative per la gestione e l'analisi dei dati all'interno di team interdisciplinari.

3. Autonomia di giudizio (Making judgements)

Al termine del corso, lo studente sarà in grado di:

- valutare criticamente l’adeguatezza e i limiti delle soluzioni informatiche adottate in contesti aziendali;
- individuare le tecnologie emergenti più rilevanti e valutarne l’impatto strategico sulle imprese;
- integrare conoscenze di natura tecnica, organizzativa e gestionale per supportare processi decisionali aziendali data-driven.

4. Abilità comunicative (Communication skills)
Al termine del corso, lo studente sarà in grado di:

- utilizzare un linguaggio tecnico appropriato per comunicare efficacemente con specialisti IT, data scientist e stakeholder aziendali;
- presentare risultati di analisi e proposte progettuali a interlocutori sia tecnici che non specialistici.

5. Capacità di apprendimento (Learning skills)
Al termine del corso, lo studente avrà sviluppato:

- capacità autonome di aggiornamento continuo rispetto alle tecnologie emergenti nel campo dell’informatica applicata all’azienda;
- abilità di apprendere nuovi strumenti e tecniche di programmazione e analisi dei dati in modo autonomo e proattivo.

Contenuti sintetici

Il corso di "Data Science" si propone di introdurre gli studenti ai concetti fondamentali della programmazione e offrire loro una visione generale delle tecniche di data science e intelligenza artificiale applicate nell'ambito aziendale. Il programma del corso comprende una introduzione sulle tecniche di intelligenza artificiale, una panoramica introduttiva della programmazione utilizzando il linguaggio Python, l'apprendimento dei concetti base dei database con l'utilizzo del linguaggio SQL, e l'acquisizione di competenze pratiche nelle tecniche di machine learning. **Le lezioni e il materiale didattico saranno in lingua inglese.**

Programma esteso

Introduzione al ruolo delle tecnologie nel mondo del business come componente fondamentale per le organizzazioni di successo.

Analisi delle nuove tendenze digitali: processi di digital transformation in azienda ed evoluzione tecnologica nella società dell’informazione.

L'intelligenza artificiale: inquadramento storico e prospettive evolutive dell’IA con approfondimento sulle più recenti evoluzioni, inclusi i grandi modelli di linguaggio (large language models).

Cenni di programmazione: concetti di base della programmazione imperativa e sviluppo di semplici programmi in Python.

Metodi di gestione e analisi dati: database relazionali, SQL, data warehouse, e knowledge graph.

Machine learning e data mining: soluzioni supervisionate e non supervisionate tipicamente utilizzate nell’ambito d’impresa con particolare attenzione alle reti neurali.

Laboratorio informatico: la parte di laboratorio è finalizzata a conferire agli studenti una conoscenza di livello intermedio nell'utilizzo dell'applicativo Microsoft Excel.

Prerequisiti

Conoscenze di tipo matematico e logico acquisite nella scuola superiore. Conoscenza della lingua inglese.

Metodi didattici

Data Science - 42 ore

- 5 lezioni da 2 ore svolte in modalità erogativa in presenza;
- 16 lezioni da 2 ore svolte in modalità erogativa nella parte iniziale che è volta a coinvolgere gli studenti in modo interattivo nella parte successiva (prevalentemente Python e SQL) .

Laboratorio Informatico (Excel) - 12 ore

- 6 attività di laboratorio da 2 ore svolte in modalità interattiva in presenza.

Modalità di verifica dell'apprendimento

La verifica dell'apprendimento comprende una prova scritta e, opzionalmente, una prova orale integrativa. La prova scritta, che prevede domande a scelta multipla e domande aperte, si svolgerà nei laboratori didattici per poter valutare le abilità dello studente nell'utilizzo del linguaggio Python e dell'applicativo Microsoft Excel.

La prova finale si articola in due parti.

Parte di Data Science:

- 16 domande a selezione multipla (1 risposta corretta su 4), che assegnano 1 punto ciascuna (0 punti in caso di risposta errata o mancante). Queste saranno estratte in maniera casuale da una lista di domande pre-definite. In particolare ci saranno tre tipi di domande: 1) Teoria di Data Science e Informatica (coperta dalle slide del corso), 2) Teoria di Python, 3) Interpretazione di codice Python (in cui si mostra un piccolo frammenti di codice e si chiede l'output).
- 1 esercizio di Python che assegna un massimo di 8 punti (0 punti in caso di risposta del tutto o in larga parte mancante). L'esercizio richiede di scrivere un frammento di codice che esegua un determinato compito (per esempio: iterare su una lista e eliminare tutti i numeri pari, trovare il minimo in una lista numerica).
- 1 query SQL che assegna un massimo di 8 punti (0 punti in caso di risposta del tutto o in larga parte mancante; si ricorda che la query deve essere eseguibile). L'esercizio (simile a quello riportato in sezione 'Esercitazioni SQL- Esempio di domanda SQL nella prova finale') mostrerà un database composto da alcune tabelle e richiederà di formulare una certa query (di solito combinando più tabelle).

Questa parte attribuisce la votazione massima conseguibile nella prova.

Parte di laboratorio Informatico (Excel):

- 3 domande aggiuntive a selezione multipla (1 risposta corretta su 4) sui contenuti del Laboratorio Informatico (la risposta a tali domande richiede l'utilizzo di Excel). Le domande saranno sia di teoria che di pratica.

Errori o omissioni nella parte di Laboratorio Informatico determinano le seguenti penalizzazioni da applicare alla votazione conseguita nella parte di Informatica Generale:

- 3 risposte corrette: nessuna penalizzazione;
- 1 risposta sbagliata o mancante: - 1 punto;
- 2 risposte sbagliate o mancanti: - 3 punti;
- 3 risposte sbagliate o mancanti: - 5 punti.

La votazione finale è quella della parte di Informatica Generale meno le eventuali penalizzazioni di Laboratorio Informatico (esempio: 27 punti nella parte di informatica generale, -3 punti in quella di Laboratorio Informatico - ovvero, 2 risposte errate o mancanti - totale della prova 24/30).

Testi di riferimento

- Slide delle lezioni e altro materiale condiviso sulla piattaforma di e-Learning.
- W3C Python tutorial (https://www.w3schools.com/python/)
- Downey, A. B. (2015). Think Python: How to Think Like a Computer Scientist, Version 2.4. Green Tea Press. (Disponibile gratuitamente online con licenza Creative Commons, https://greenteapress.com/thinkpython2/thinkpython2.pdf)

Si consiglia inoltre il seguente testo facoltativo:

- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition. Aurélien Géron. O’Reilly, 2019.

Periodo di erogazione dell'insegnamento

Primo semestre.

Lingua di insegnamento

Inglese.

Sustainable Development Goals

IMPRESE, INNOVAZIONE E INFRASTRUTTURE