

Napoli Federico II si avvale del sistema di Assicurazione Qualità (AQ)²⁰, sviluppato in conformità al documento “Autovalutazione, Valutazione e Accreditamento del Sistema Universitario Italiano” dell’ANVUR, utilizzando:

- indagini sul grado di inserimento dei laureati nel mondo del lavoro e sulle esigenze post-lauream;
- dati estratti dalla somministrazione del questionario per la valutazione della soddisfazione degli studenti per ciascun insegnamento presente nel piano di studi, con domande relative alle modalità di svolgimento del corso, al materiale didattico, ai supporti didattici, all’organizzazione, alle strutture.

I requisiti derivanti dall’analisi dei dati sulla soddisfazione degli studenti, discussi e analizzati dalla Commissione di Coordinamento Didattico e dalla Commissione Paritetica Docenti Studenti (CPDS), sono inseriti fra i dati di ingresso nel processo di progettazione del servizio e/o fra gli obiettivi della qualità.

3. L’organizzazione dell’AQ sviluppata dall’Ateneo realizza un processo di miglioramento continuo degli obiettivi e degli strumenti adeguati per raggiungerli, facendo in modo che in tutte le strutture siano attivati processi di pianificazione, monitoraggio e autovalutazione che consentano la pronta rilevazione dei problemi, il loro adeguato approfondimento e l’impostazione di possibili soluzioni.

Art. 21 **Norme finali**

1. Il Consiglio di Dipartimento, su proposta della Commissione di Coordinamento Didattico, sottopone all’esame del Senato Accademico eventuali proposte di modifica e/o integrazione del presente Regolamento.

Art. 22 **Pubblicità ed entrata in vigore**

1. Il presente Regolamento entra in vigore il giorno successivo alla pubblicazione all’Albo ufficiale dell’Università; è inoltre pubblicato sul sito d’Ateneo. Le stesse forme e modalità di pubblicità sono utilizzate per le successive modifiche e integrazioni.
2. Sono parte integrante del presente Regolamento l’Allegato 1 (Struttura CdS) e l’Allegato 2 (Schedina insegnamento/attività).

Allegato 1 - Struttura del CdS

Il primo anno di corsi è comune per tutti gli studenti. Il secondo anno si articola invece secondo quattro percorsi curriculari finalizzati all’apprendimento di strumenti specifici per il dominio applicativo prescelto. I 4 curricula previsti sono:

- Fundamental Sciences
- Intelligent Systems
- Information Technologies
- Data Science for Public Administration, Economy and Management

²⁰ Il sistema di Assicurazione Qualità, basato su un approccio per processi e adeguatamente documentato, è progettato in maniera tale da identificare le esigenze degli studenti e di tutte le parti interessate, per poi tradurle in requisiti che l’offerta formativa deve rispettare.

I Curriculum sono finalizzati a fornire al laureato in *Data Science* la capacità di comprendere la struttura dati e le problematiche di specifici domini applicativi. Per ogni curriculum/percorso lo studente dovrà scegliere un esame curriculare da 12 cfu ed un esame curriculare da 6 CFU coerenti con il curriculum/percorso prescelto (vedere Tabelle 5.1, 5.2, 5.3, 5.4), inoltre dovrà scegliere un esame a scelta suggerito dalla tabella 6 tra quelli relativi al percorso scelto, ed uno a scelta completamente libera (purché coerente con il percorso formativo) tra tutti (inclusi quelli elencati nelle tabelle 5 e 6) gli insegnamenti offerti dalla Scuola Politecnica e delle Scienze di Base. Piani individuali potranno ricorrere ad altri esami dell'offerta didattica dell'ateneo

L'esame finale consisterà nella discussione di un elaborato originale risultante da un'attività di progettazione o di ricerca o di analisi di caso, che dimostri la padronanza degli argomenti e degli strumenti utilizzati, nonché la capacità di operare in modo autonomo nel campo delle organizzazioni che usano strumenti della data science. Tale elaborato potrà essere il risultato di una lavoro svolto in ambito accademico o presso un ente/azienda ospitante il tirocinio.

Nella sua struttura generale il corso può essere schematizzato come nelle seguenti tabelle. I corsi da 12 CFU sono costituiti da due moduli da 6 CFU ciascuno.

| I Anno | Insegnamento | SSD | CFU | CFU aula | CFU lab | TAF |
|---------|---|---------------------|-------------------------------------|----------|---------|-----------------|
| 1.1.1 | Mathematical methods for Data science | MAT-08 | 6 | 6 | 0 | Caratterizzante |
| 1.1.5 | Fundamentals of programming & Data management | ING-INF/05 | 12 | 8 | 4 | Caratterizzante |
| 1.1.3 | Statistical Learning and Data Analysis | SECS-S/01 | 12 | 8 | 4 | Caratterizzante |
| 1.1.4 | Data Mining & Machine Learning | SECS-S/01 INF/01 | 12 | 8 | 4 | Caratterizzante |
| 1.2.2 | Hardware and software for Big data | ING-INF/05 | 12 | 8 | 4 | Caratterizzante |
| 1.2.7 | Theory and Ethics of Big Data and AI | M-FIL/03 | 6 | 6 | 0 | Caratterizzante |
| | Totale corsi primo anno: 6 | | Totale CFU primo anno 60 | | | |
| II Anno | Insegnamento | SSD | CFU | CFU aula | CFU lab | TAF |
| 2.1.8 | Insegnamento curriculare I | II | 12 | | | |
| 2.1.9 | Insegnamento curriculare II | II | 6 | | | |
| 2.2.10 | Attività a scelta | II | 6 | | | scelta autonoma |

| | | | | | | |
|--------|-------------------------------------|----|----|--------------------------------|--|-----------------|
| 2.2.11 | Attività a scelta | II | 6 | | | scelta autonoma |
| 2.2.12 | Tirocinio-stage o progetto | | 8 | | | |
| 2.2.13 | Altre conoscenze * | | 6 | | | |
| 2.2.14 | Esame di laurea | | 16 | | | |
| | Totale corsi secondo anno: 5 | | | Totale CFU secondo anno | | |
| | | | | 60 | | |
| | Totale esami: 11 | | | Totale CFU complessivi | | |
| | | | | 120 | | |

* Per i soli studenti stranieri questi CFU saranno concessi dopo il conseguimento di un certificato di conoscenza della lingua italiana (corsi erogati dal Centro Linguistico di Ateneo – CLA). Per gli italiani, i crediti potranno essere conseguiti acquisendo competenze tecniche esterne al corso di studi (e.g. partecipazione a seminari, congressi, attività di tutoraggio, ecc.)

Curriculum DS for Fundamental Sciences

Sono previsti tre percorsi. *DS Methodologies* (Metodologie di Base) strettamente collegato con il DMRC e che mutua esami delle lauree magistrali in Matematica ed in Ingegneria Matematica); *DS for Physics* (Data Science per la Fisica), strettamente collegato con il DFEP e che mutua esami delle lauree magistrali in Fisica ed in Scienze e Tecnologie quantistiche; *DS for life sciences* (Data Science per le Scienze della Vita).

Tabella 5.1 - Insegnamenti affini e integrativi curriculari - Curriculum n 1 - Fundamental Sciences.

| Nome del corso | SSD | CFU | STATUS | Note | Percorso |
|--|------------------|------|-------------------|----------------|----------------------|
| Computational Methods for Physics | FIS-01 | 6 ** | active DFEP | Curriculare I | DS for physics |
| Physically Inspired Machine Learning + Astroinformatics | FIS-05 FIS-05 | 12 | Nuova Istituzione | Curriculare II | DS for physics |
| | | | | | |
| Algorithms and Parallel computing + Computational complexity | INF-01 | 12 | active DMRC | Curriculare I | DS Methodologies |
| Operational Research | MAT-09 | 6 | active DMRC | Curriculare II | DS Methodologies |
| | | | | | |
| Molecular Biology and biological data | BIO-11 | 12 | nuova istituzione | Curriculare I | DS for life sciences |

| Nome del corso | SSD | CFU | STATUS | Note | Percorso |
|--------------------------------------|--------|-----|--------------|----------------|----------------------|
| Biologia computazionale e Statistica | Inf-01 | 6 | active DIETI | curriculare II | DS for life sciences |

Curriculum DS for Information Technologies

Questo curriculum è strettamente collegato con il DIETI. Esso prevede cinque percorsi: DS for Text and Speech Processing (*Elaborazione dei Testi e del Parlato*), Signal and Video Processing (*Elaborazione dei Segnali e dei Video*), Data Security (*Sicurezza dei Dati*), Industrial Applications (*Applicazioni Industriali*), Statistics and Robotics for Health (*Statistica e Robotica per la Salute*).

Tabella 5.3 - Insegnamenti affini e integrativi curriculari - Curriculum n 2 - Information Technologies

| Nome del corso | SSD | CFU | STATUS | Note | Percorso |
|---|----------------------|-----|-------------------------|----------------|-----------------------------|
| Multimedia Information Retrieval & Text Mining | ING-INF/05 | 12 | active DIETI | Curriculare I | Text and speech processing |
| Natural Language Processing | INF/01 | 6 | active DIETI | Curriculare II | Text and speech processing |
| | | | | | |
| Information Theory & Signal Theory | ING-INF/03 | 12 | active DIETI | Curriculare I | Signal and video processing |
| Image and Video processing for autonomous driving | ING-INF/03 | 6 | active DIETI | Curriculare II | Signal and video processing |
| | | | | | |
| Data Security & Computer Forensics | ING-INF/05 INF/01 | 12 | active DIETI | Curriculare I | Data Security |
| Algorithm Design | INF/01 | 6 | active DIETI | Curriculare II | Data Security |
| | | | | | |
| Advanced Statistical learning and modeling | SEC-S/01 | 12 | Nuova attivazione DIETI | Curriculare I | Industrial Application |
| Statistical Methods for Industrial Process Monitoring | SEC-S/02 | 6 | Active DII | curriculare II | Industrial applications |
| | | | | | |

| Nome del corso | SSD | CFU | STATUS | Note | Percorso |
|--|------------|-----|-------------------------|----------------|------------------------------------|
| Multimedia Information Retrieval & Text Mining | ING-INF/05 | 12 | active DIETI | Curriculare I | Text and speech processing |
| Advanced Statistical learning and modelling | SEC-S/01 | 12 | Nuova attivazione DIETI | Curriculare I | Statistics and robotics for health |
| Robotics for bio-engineering | ING-INF/04 | 6 | Active DIETI | curriculare II | Statistics and robotics for health |

Curriculum DS for Public Administration, Economy and Management

Questo curriculum non prevede percorsi.

Tabella 5.3 - Insegnamenti affini e integrativi curriculari - Curriculum n 3 Public Administration, Economy and Management

| Nome del corso | SSD | CFU | STATUS | Note |
|---|-----------|-----|--------------|------------------|
| Computational statistics and Generalized linear models* | SECS-S/01 | 12 | active DISES | Curriculare I ** |
| Introduction to Economy and econometrics | SECS-S/03 | 12 | active DISES | Curriculare I ** |
| Statistical methods for evaluation | SECS-S/05 | 6 | active DISES | Curriculare II |

** I due corsi curriculari da 12 CFU sono offerti in alternativa ma è anche possibile sostenerli come 2 esami indipendenti da 6 CFU l'uno (in questo caso si configurerebbero come un esame curriculare un esame a scelta oppure come due esami a scelta)

Curriculum Intelligent systems

Questo curriculum non prevede percorsi.

Tabella 5.4 - Insegnamenti affini e integrativi curriculari - Curriculum n 4 Intelligent Systems

| Nome del corso | SSD | CFU | STATUS | Affini e integrativi |
|---|--------|-----|-------------|----------------------|
| Computational Intelligence and Machine learning for Physics | INF-01 | 12 | active DFEP | Curriculare I |
| Computational Neurosciences | BIO-09 | 6 | active DFEP | Curriculare II |

Tabella 6 - Elenco corsi a scelta suggeriti strutturati per curriculum/percorso.

| Nome del corso | ssd | CFU | STATUS | Note |
|----------------|-----|-----|--------|------|
|----------------|-----|-----|--------|------|

| Curriculum 1 - Fundamental sciences | | | | |
|--|-------------|----|-------------------|--------------------------------------|
| X-informatics | FIS-05 | 6 | nuova istituzione | Percorso Fisica/metodologie |
| Computational physics | FIS-01 | 6 | active DFEP | Percorso Fisica |
| Computational Thermodynamics | FIS-03 | 6 | active DFEP | Percorso Fisica |
| Machine learning for Physics | INF-01 | 6 | active DFEP | Percorso Fisica |
| Quantum Computing Systems | FIS-03 | 6 | attivo DFEP | Percorso Fisica |
| Artificial Intelligence and Quantum Computing | INF-01 | 6 | attivo DFEP | Percorso Fisica/metodologie |
| Mathematical Physics Models | MAT-07 | 6 | active DMRC | Percorso Fisica |
| Real and Functional Analysis | MAT-05 | 6 | active DMRC | Percorso Metodologie |
| Signals Theory | ING-INF /03 | 6 | Active DIETI | Percorso Metodologie |
| Computational complexity | INF-01 | 6 | Active DIETI | Percorso Metodologie |
| Information Theory | ING-INF /03 | 6 | Active DIETI | Percorso Metodologie |
| Algorithms and Parallel computing | INF-01 | 12 | Active DMRC | Percorso Metodologie |
| Probability Theory | MAT-06 | 6 | Nuova istituzione | Percorso Metodologie |
| Curriculum 2 - Information Technologies | | | | |
| Techniques of Text analysis and Computational Linguistic | L/LIN-01 | 6 | active DIETI | Percorso text processing |
| Speech Processing | L/LIN-01 | 6 | nuova istituzione | Percorso text processing |
| Text Mining | ING-INF /05 | 6 | active DIETI | Percorso text processing |
| Natural Language Processing | INF-01 | 6 | active DIETI | Percorso text processing |
| Algebraic Methods for Cryptography | MAT/02 | 6 | active DIETI | Percorso Data Security |
| Biometric Systems | INF/01 | 6 | active DIETI | Percorso signal and video processing |

| | | | | |
|--|----------------|---|-------------------------|--------------------------------------|
| Computer Graphics | ING-INF /05 | 6 | active DIETI | Percorso Signal and Video Processing |
| Signals Theory | ING-INF /03 | 6 | Active DIETI | Percorso Signal and Video Processing |
| Environmental risk monitoring and evaluation | SEC-S/0 2 | 6 | Nuova istituzione DII | Percorso Industrial Applications |
| Methods for Artificial Intelligence | INF/01 | 6 | Active DIETI | Statistics and robotics for health |
| Human robot interaction | INF/01 | 6 | Active DIETI | Statistics and robotics for health |
| Advanced non linear methods for industrial signal processing | ING-IND /13 | 6 | Nuova Istituzione DII | Percorso Industrial Applications |
| Curriculum n. 3 - Public Administration, Economy and Management | | | | |
| Mathematics for economics and finance | SEC-S/0 6 | 6 | active DISES | n.a |
| Statistics for monetary and financial markets | SEC-S/0 1 | 6 | active DISES | n.a |
| SW and methods for statistical analysis of economic data | SEC-S/0 1 | 6 | active DISES | n.a |
| Preference learning | SEC-S/0 1 | 6 | Nuova istituzione DISES | n.a |
| Curriculum n.4 - Intelligent Systems | | | | |
| Quantum Computing Systems | FIS/03 | 6 | Active DFEP | n.a |
| Artificial Intelligence and Quantum Computing | INF/01 | 6 | Active DFEP | n.a |
| Machine learning for physics | INF/01 | 6 | Active DFEP | n.a |

Allegato 2 – Schede insegnamento

Si allegano le schede degli insegnamenti fondamentali del primo anno e degli insegnamenti di nuova istituzione (le schede degli insegnamenti mutuati sono già presenti nel database di ateneo).

Per tutti i corsi, l'attribuzione del voto avverrà secondo i criteri riportati nella seguente Tabella:

| | |
|----------------------|--|
| <18 Insufficiente | Conoscenze frammentarie e superficiali dei contenuti, errori nell'applicare i concetti, prova scritta insufficiente ed esposizione carente |
| 18-20 | Conoscenze dei contenuti sufficienti ma generali, esposizione semplice, incertezze nell'applicazione di concetti teorici |
| 21-23 | Conoscenze dei contenuti appropriate ma non approfondite, capacità di applicare i concetti teorici, capacità di presentare i contenuti in modo semplice |
| 24-25 | Conoscenze dei contenuti appropriate ed ampie, discreta capacità di applicazione delle conoscenze, capacità di presentare i contenuti in modo articolato. |
| 26-27 | Conoscenze dei contenuti precise e complete, buona capacità di applicare le conoscenze, capacità di analisi, esposizione chiara e corretta |
| 28-29 | Conoscenze dei contenuti ampie, complete ed approfondite, buona applicazione dei contenuti, buona capacità di analisi e di sintesi, esposizione sicura e corretta |
| 30 30 e lode | Conoscenze dei contenuti molto ampie, complete ed approfondite, capacità ben consolidata di applicare i contenuti, ottima capacità di analisi, di sintesi e di collegamenti interdisciplinari, padronanza di esposizione |

SCHEDA DELL'INSEGNAMENTO

STATISTICAL LEARNING AND DATA ANALYSIS

SSD SEC-S/01

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: I

SEMESTRE: I

MODULO: A

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 6

INSEGNAMENTI PROPEDEUTICI

Nessuno

EVENTUALI PREREQUISITI

Nessuno

OBIETTIVI FORMATIVI

Il corso fornisce agli studenti la metodologia statistica per l'apprendimento dai dati, come trasformare i problemi reali in sfide statistiche, come esplorare i dati ed estrarre importanti tipologie, come costruire modelli per il processo decisionale e predittivo, come validare i risultati, come interpretare e comunicare i risultati dell'analisi statistica dei dati. Si fa riferimento al paradigma classico della Statistica e a quello moderno dello Statistical Learning: Data Mining, Inferenza e Predizione. I domini applicativi sono scelti nell'Ingegneria e Scienze di Base, nell'Economia e nelle Scienze della vita.

Nello specifico, gli studenti vengono esposti e formati sui metodi fondamentali per l'analisi esplorativa dei dati e la modellazione statistica per l'inferenza e la previsione nei problemi di classificazione e regressione. Il ritmo di apprendimento è mantenuto da esercitazioni pratiche svolte con linguaggi di programmazione open-source, in modo che l'acquisizione di tali metodi e competenze sia rafforzata attraverso lo sviluppo di casi di studio basati su dati del mondo reale. La partnership con importanti aziende assicura la possibilità di partecipare ad un challenge per lo sviluppo di un progetto in team, così da completare la formazione con "soft skills" (capacità di lavorare in gruppo, abilità nella comunicazione e presentazione, autonomia e condivisione delle scelte, etc.) per un vero e proprio apprendimento esperienziale della statistica. Il primo Modulo copre gli argomenti di statistical learning.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente dovrà dimostrare di aver acquisito familiarità con i fondamenti della statistica, e di essere capace di orientarsi tra i vari metodi e modelli imparati. Dovrà inoltre mostrare di aver compreso in linea generale potenzialità, limiti e ambiti di applicazione degli strumenti di Statistical Learning in diversi contesti applicativi.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di saper tradurre concetti di statistica in termini matematici e informatici, e di saper scegliere i metodi più adatti (e costruire algoritmi) per esplorare ipotesi e fornire risposte a specifiche domande in ambito applicativo.

PROGRAMMA – SYLLABUS

- Introduzione alla Statistica, Technè-Logia, Analysis of Data
- Analisi Esplorativa dei Dati (Data Pre-processing, Descrittive, Analisi Bivariata, Visualizzazione)
- Modelli di probabilità (Elementi di probabilità, Modelli discreti, Modelli continui)
- Analisi Confermativa dei Dati (Inferenza classica, Stima e Test statistici parametrici e non parametrici)
- Modelli Lineari (ANOVA, Regressione lineare semplice)
- Modelli Lineari per l'analisi delle serie temporali (Approccio classico, approccio Box Jenkins)
- Analisi Multidimensionale dei Dati

Applicazioni in Python o R a:

1. Dati complessi (immagini e serie temporali) di origine varia

MATERIALE DIDATTICO

E' possibile utilizzare qualunque testo attinente al corso. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni.

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;
- Pseudocodici, templates e scripts di riferimento.
- Corso MOOC Statistical Data Analysis (Module A) su Federica WEB Learning.

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- A) L'esame finale mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati nella sezione contenuti del programma
- B) verrà interrogato sugli argomenti del corso e verranno valutati il grado di completezza della sua risposta, il livello di integrazione tra i vari contenuti del corso, il raggiungimento da parte dello studente di una visione organica dei temi affrontati, la padronanza espressive e la proprietà nel linguaggio scientifico.
- C) La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti metodologici che nella discussione di un elaborato progettuale.

SCHEDA DELL'INSEGNAMENTO

STATISTICAL LEARNING AND DATA ANALYSIS

SSD SEC-S/01

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: I

SEMESTRE: I

MODULO: B

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 6

INSEGNAMENTI PROPEDEUTICI

Statistical Learning and Data Analysis – Module A

EVENTUALI PREREQUISITI

Programmazione in R o in Python

OBIETTIVI FORMATIVI

Il corso fornisce agli studenti la metodologia statistica per l'apprendimento dai dati, come trasformare i problemi reali in sfide statistiche, come esplorare i dati ed estrarre importanti tipologie, come costruire modelli per il processo decisionale e predittivo, come validare i risultati, come interpretare e comunicare i risultati dell'analisi statistica dei dati. Si fa riferimento al paradigma classico della Statistica e a quello moderno dello Statistical Learning: Data Mining, Inferenza e Predizione. I domini applicativi sono scelti nell'Ingegneria e Scienze di Base, nell'Economia e nelle Scienze della vita.

Nello specifico, gli studenti vengono esposti e formati sui metodi fondamentali per l'analisi esplorativa dei dati e la modellazione statistica per l'inferenza e la previsione nei problemi di classificazione e regressione. Il ritmo di apprendimento è mantenuto da esercitazioni pratiche svolte con linguaggi di programmazione open-source, in modo che l'acquisizione di tali metodi e competenze sia rafforzata attraverso lo sviluppo di casi di studio basati su dati del mondo reale. La partnership con importanti aziende assicura la possibilità di partecipare ad un challenge per lo sviluppo di un progetto in team, così da completare la formazione con "soft skills" (capacità di lavorare in gruppo, abilità nella comunicazione e presentazione, autonomia e condivisione delle scelte, etc.) per un vero e proprio apprendimento esperienziale della statistica. Questo modulo approfondisce e estende i contenuti del Modulo A.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente dovrà dimostrare di aver acquisito familiarità con i fondamenti della statistica, e di essere capace di orientarsi tra i vari metodi e modelli imparati. Dovrà inoltre mostrare di aver compreso in linea generale potenzialità, limiti e ambiti di applicazione degli strumenti di Statistical Learning in diversi contesti applicativi.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di saper tradurre concetti di statistica in termini matematici e informatici, e di saper scegliere i metodi più adatti (e costruire algoritmi) per esplorare ipotesi e fornire risposte a specifiche domande in ambito applicativo.

PROGRAMMA – SYLLABUS

- Modelli lineari per la regressione multipla
- Diagnostica sui residui
- Modelli lineari per la classificazione supervisionata
- Inferenza Computazionale
- Metodi di regolarizzazione
- Selezione del modello
- Modelli non lineari e stima non parametrica
- Classificazione e Regressione ad Albero
- Metodi ensembles
- Support Vector Machines
- Regressione procustiana
- Analisi di Sopravvivenza
- Test Multipli

Applicazioni a:

Dati complessi (immagini e serie temporali) di origine varia

MATERIALE DIDATTICO

E' possibile utilizzare qualunque testo attinente al corso. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni.

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;)
- Pseudocodici, templates e scripts di riferimento.

- Corso MOOC Statistical Data Analysis (Module B) su Federica WEB Learning.

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- D) L'esame finale mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati nella sezione contenuti del programma
- E) Lo studente verrà interrogato sugli argomenti del corso e verranno valutati il grado di completezza della sua risposta, il livello di integrazione tra i vari contenuti del corso, il raggiungimento da parte dello studente di una visione organica dei temi affrontati, la padronanza espressive e la proprietà nel linguaggio scientifico.
- F) La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti metodologici che nella discussione di un elaborato progettuale.

SCHEDA DELL'INSEGNAMENTO
MATHEMATICAL METHODS FOR DATA SCIENCE
SSD MAT-08

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: I

SEMESTRE: I

MODULO: --

LINGUA: INGLESE

MODALITA': lezioni frontali

CFA: 6

INSEGNAMENTI PROPEDEUTICI

nessuno

EVENTUALI PREREQUISITI

Conoscenze basilari di algebra e analisi

OBIETTIVI FORMATIVI

Il corso fornisce agli studenti i fondamenti matematici necessari ad affrontare lo studio della data science.

Nello specifico, gli studenti vengono esposti e formati sui metodi fondamentali per il calcolo vettoriale e matriciale, per la differenziazione di matrici e la risoluzione di sistemi di equazioni differenziali e per la risoluzione di basilari problemi di ottimizzazione.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente dovrà dimostrare di aver acquisito familiarità con i fondamenti matematici della Data Science e di essere capace di applicarli.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di saper utilizzare i concetti acquisiti nella risoluzione di problemi legati al calcolo matriciale e all'ottimizzazione.

PROGRAMMA – SYLLABUS

- Richiami di analisi matematica: derivate, derivate parziali, funzioni complesse.
- Fondamenti di teoria delle probabilità
- Algebra lineare: sistemi di equazioni differenziali, matrici, matrici inverse, spazi vettoriali, indipendenza lineare, sistemi di basi, rank, mapping lineare, trasformazioni affini, cambi di variabili,
- Geometria analitica: Norme, prodotti interni, lunghezze e distanze, angoli e ortogonalità, prodotto interno di funzioni, proiezioni ortogonali, rotazioni
- Decomposizione di matrici: tracce e determinante, autovettori e autovalori, decomposizione di cholewsky, diagonalizzazione, single value decomposition, filogenia delle matrici
- Calcolo vettoriale: differenziazione di funzioni univariate, derivate parziali e gradiente, gradienti di funzioni vettoriali, gradienti di matrici, backpropagation, derivate di ordine superiore
- Ottimizzazione: gradient descent, moltiplicatori di Lagrange.

MATERIALE DIDATTICO

E' possibile utilizzare qualunque testo attinente al corso. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni.

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;)

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica sarà svolta in maniera interattiva e gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- G) L'esame finale mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati nella sezione contenuti del programma
- H) Lo studente verrà interrogato sugli argomenti del corso e verranno valutati il grado di completezza della sua risposta, il livello di integrazione tra i vari contenuti del corso, il raggiungimento da parte dello studente di una visione organica dei temi affrontati, la padronanza espressive e la proprietà nel linguaggio scientifico.
- I) La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti del corso.

SCHEDA DELL'INSEGNAMENTO

DATA MINING AND MACHINE LEARNING – MODULO A

SSD SEC-S/01

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: I

SEMESTRE: II

MODULO:A

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 6

INSEGNAMENTI PROPEDEUTICI

nessuno

EVENTUALI PREREQUISITI

Conoscenze matematiche di base

OBIETTIVI FORMATIVI

Il corso introdurrà gli studenti al paradigma classico della Statistica e a quello moderno dello Statistical Learning: Data Mining, Inferenza e Predizione. I domini applicativi sono scelti nell'Ingegneria e Scienze di Base, nell'Economia e nelle Scienze della vita. Lo studente acquisirà conoscenza operativa di metodi di Data Mining sia supervisionati che non supervisionati. Tale conoscenza sarà poi approfondita con specifiche applicazioni a diverse tipologie di dati.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente dovrà dimostrare di aver acquisito familiarità con i fondamenti del Data Mining, e di essere capace di orientarsi tra i vari metodi e modelli imparati. Dovrà inoltre mostrare di aver compreso in linea generale potenzialità, limiti e ambiti di applicazione degli strumenti e dei metodi appresi in diversi contesti applicativi.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di saper applicare i metodi appresi per estrarre informazioni utili dai dati, utilizzando i metodi più adatti (e costruire i relativi algoritmi) per esplorare ipotesi e fornire risposte a specifiche domande in ambito applicativo.

PROGRAMMA – SYLLABUS

- Introduction to data mining, vision, basic data types, application scenario
- Data Editing, similarities and dissimilarities, fundaments of data visualization
- Principal Component Analysis, correspondence analysis and multiple correspondence analysis, multidimensional scaling and latent semantic analysis
- Hierarchical clustering and non-hierarchical clustering. Clustering validation
- Soft clustering, density based algorithms and graph based algorithms
- Support vector clustering
- Frequent pattern mining models, Association rules generator, frequent set mining algorithms
- Vapnik statistical learning theory.
- Decision trees, multi-class learning
- Regression modeling, semi-supervised learning
- Ensemble methods
- Mining data streams, Text data, time series, graph data, web data
- Outlier detection

Applicazioni a:

Dati complessi (immagini e serie temporali) di origine varia

MATERIALE DIDATTICO

E' possibile utilizzare qualunque testo attinente al corso. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni.

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;)
- Pseudocodici, templates e scripts di riferimento.
- Corso MOOC Statistical Data Anlysis (Module B) su Federica WEB Learning.

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- J) L'esame finale mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati nella sezione contenuti del programma
- K) Lo studente verrà interrogato sugli argomenti del corso e verranno valutati il grado di completezza della sua risposta, il livello di integrazione tra i vari contenuti del corso, il raggiungimento da parte dello studente di una visione organica dei temi affrontati, la padronanza espressive e la proprietà nel linguaggio scientifico.
- L) La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti metodologici che nella discussione di un elaborato progettuale.

SCHEDA DELL'INSEGNAMENTO

DATA MINING AND MACHINE LEARNING – MODULO B

SSD: INF-01

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: I

SEMESTRE: II

MODULO: B – Machine Learning

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 6

INSEGNAMENTI PROPEDEUTICI

Statistical Learning and Data Analysis – Mod A, Mathematical Methods for Data Science, Data Mining – Module A.

EVENTUALI PREREQUISITI

Conoscenze di statistica, algebra lineare e analisi matematica

OBIETTIVI FORMATIVI

Il corso introdurrà gli studenti ai fondamenti dell'apprendimento automatico (machine learning) esplorandone sia gli aspetti teorici che quelli applicativi, facendo loro comprendere sia le potenzialità che i limiti dei vari metodi. Il corso, inoltre, introdurrà lo studente all'utilizzo critico di librerie e software di ML open source.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente acquisirà conoscenza teorica e operativa dei fondamenti del machine learning e delle principali famiglie di modelli. Inoltre acquisirà una conoscenza teorica ed operativa di qualimodelli utilizzare per la risoluzione di specifici problemi.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di saper applicare i metodi appresi per estrarre informazioni utili dai dati, utilizzando i metodi più adatti (e costruire i relativi algoritmi) per esplorare ipotesi e fornire risposte a specifiche domande in ambito applicativo.

PROGRAMMA – SYLLABUS

- Storia del machine learning
- Utilizzo supervisionato, non supervisionato e semi supervisionato
- I primi modelli: perceptrone, support Vector machines
- Il neurone sigmoide e il Perceptrone Multi Strato
- Funzioni di attivazione e funzioni costo
- Utilizzo delle MLP e analisi del loro addestramento
- Le reti neurali profonde
- Reti CNN
- Reti RESNET
- Reti Ricorrenti
- Autoencoder e energy based models
- GAN

Applicazioni a:

Dati complessi (immagini e serie temporali) di origine varia

MATERIALE DIDATTICO

Il docente utilizzerà i seguenti testi di riferimento: Agarvall, Data Mining; I. Goodfellow et al., Deep Learning; F. Chollet, Deep Learning with Python. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni. Inoltre sarà possibile avvalersi di:

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;
- Pseudocodici, templates e scripts di riferimento su libreria github dedicata
- Corso MOOC Data Mining (Module B) su Federica WEB Learning.

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

Il corso si articolerà nella realizzazione di progetti di gruppo (max 3-4 persone) su problemi specifici concordati con il docente. Il lavoro prevederà una prima fase iniziale in cui gli studenti dovranno comprendere il problema facendo uso di articoli scientifici. Dovranno quindi definire le metodologie più adatte a risolvere il problema loro assegnato e implementare un codice (o più codici) atto a ricavare i risultati desiderati.

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- L'esame finale mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati nella sezione contenuti del programma
- L'esame finale, quindi consisterà in una prova orale in cui lo studente dovrà dimostrare di avere compreso nel dettaglio le metodologie utilizzate e le specificità del problema, presentare i risultati ottenuti e discuterli.
- La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti metodologici che nella presentazione e discussione dei risultati di un progetto concordato con il docente.

SCHEDA DELL'INSEGNAMENTO

HARDWARE AND SOFTWARE FOR BIG DATA

ING-INF/05

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: I

SEMESTRE: I

MODULI: A & B

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 12

INSEGNAMENTI PROPEDEUTICI

nessuno

EVENTUALI PREREQUISITI

Conoscenze di matematica

OBIETTIVI FORMATIVI

Il corso si prefigge di introdurre lo studente ai principali aspetti legati all'ingestione e gestione dei Big data dove per Big Data si intendono dati che, per dimensioni, complessità o velocità di acquisizione pongono problemi non facilmente risolvibili con tecniche tradizionali. Lo studente quindi, acquisirà familiarità con strumenti quali Hadoop, Spark ecc. e con il calcolo in ambiente distribuito o nel cloud. Lo studente disporrà, al termine del corso delle nozioni e conoscenze fondamentali sulle diverse tipologie di Big data disponibili empiricamente nelle analisi aziendali. Queste, integrate con i dati ufficiali, rappresentano la base per fare scelte e programmazioni manageriali opportune. In tale ottica la Business Intelligence e Analytics rappresentano due chiavi di sviluppo fondamentali, che saranno presentati attraverso casi empirici.

Il corso, inoltre si prefigge di dare agli studenti una comprensione dei fondamentali aspetti della sicurezza dei dati.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente acquisirà conoscenza teorica e operativa della ingestione, gestione e utilizzo dei big data e dei principali strumenti software a tal fine utilizzati. Inoltre acquisirà una conoscenza teorica ed operativa di quali strumenti utilizzare per la risoluzione di specifici problemi.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di saper applicare i metodi appresi per gestire ed estrarre informazioni utili da big data in DB statici e dinamici, stand Alone e distribuiti.

PROGRAMMA – SYLLABUS

- Metodi di raccolta, classificazione e qualificazione dei dati
- Popolazione statistica e popolazione con i Big data
- I database
- SQL, non SQL e Relazionali
- Database distribuiti
- Fondamenti di sicurezza dei dati.
- Hadoop & Spark
- Mongo DB e data visualization con Tableau
- La strategia europea dei dati e gli open data
- Business Analytics and business intelligence
- Big data e processi decisionali: l'analisi SWOT
- Esempi di utilizzo di archivi e banche dati per l'impresa (lezioni previ

Applicazioni a:

Dati complessi (immagini e serie temporali) di origine varia

MATERIALE DIDATTICO

Il docente fornirà testi di riferimento. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni. Inoltre sarà possibile avvalersi di:

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;
- Pseudocodici, templates e scripts di riferimento su libreria github dedicata

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

Il corso si articolerà nella realizzazione di progetti di gruppo (max 3-4 persone) su problemi specifici concordati con il docente. Il lavoro prevederà una prima fase iniziale in cui gli studenti dovranno comprendere il problema facendo uso

di articoli scientifici. Dovranno quindi definire le metodologie più adatte a risolvere il problema loro assegnato e implementare un codice (o più codici) atto a ricavare i risultati desiderati.

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- L'esame finale mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati nella sezione contenuti del programma
- L'esame finale, quindi consisterà in una prova orale in cui lo studente dovrà dimostrare di avere compreso nel dettaglio le metodologie utilizzate e le specificità del problema, presentare i risultati ottenuti e discuterli.
- La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti metodologici che nella presentazione e discussione dei risultati di un progetto concordato con il docente.

SCHEDA DELL'INSEGNAMENTO

FUNDAMENTALS OF PROGRAMMING AND DATA MANAGEMENT

ING-INF/05

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: I

SEMESTRE: I

MODULO: A Fundamentals of Programming

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 6

INSEGNAMENTI PROPEDEUTICI

nessuno

EVENTUALI PREREQUISITI

Conoscenze di matematica

OBIETTIVI FORMATIVI

Fornire le nozioni di base per le discipline informatiche, introducendo lo studente allo studio dei fondamenti teorici dell'informatica, dell'architettura dei calcolatori e dei linguaggi di programmazione ad alto livello. Fornire le conoscenze necessarie per lo sviluppo di programmi in Python per la risoluzione di problemi di limitata complessità.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente acquisirà conoscenza teorica e operativa dei fondamenti dell'informatica e della programmazione in Python. Inoltre acquisirà una conoscenza teorica ed operativa di quali strumenti utilizzare per la risoluzione di specifici problemi.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di sapere utilizzare Python per risolvere problemi connessi con l'ingestione, uso ed analisi di dati.

PROGRAMMA – SYLLABUS

- General introduction to programming, algorithms and debugging.
- Simple Python data: variables, expressions and statements; values and data types; variable names and keywords; operators and operands; the function input; reassignment and updating variables.
- Debugging: types of errors; how to reduce debugging.
- Turtle graphics: the turtle module; the for loop; the range function.
- Modules: math and random.
- Functions: unit testing; scope of variables and parameters in functions; the Accumulator pattern; nested functions; the main function.
- Selection: Boolean values and expressions; logical operators; conditional execution; nested and chained conditionals; Boolean functions.
- Iteration: for and while loops; other uses of while: sentinel values and validating input; printing simple tables.
- Strings: collection data types; operations of strings; the index operator; string methods; the slice operator; string comparison; mutable and immutable data types; traversal by item and by index; optional parameters in functions.
- Lists: operations on lists; list methods; differences between mutable and immutable data types; objects and references; aliasing and cloning; using lists as parameters; pure functions; list comprehension; nested lists; tuples and mutability; tuples as return values.
- Files: reading and writing files.
- Dictionaries: operations and methods; aliasing and copying; dictionary based implementation of sparse matrices.
- Numpy: the numpy array object; creating arrays; basic data types; basic visualization; indexing and slicing; copies and views; fancy indexing; numerical operations on arrays: elementwise operations; basic reductions; broadcasting; array shape manipulation; sorting data;
- Matplotlib: introduction; simple plot; figures, subplots, axes and ticks.
- Scipy: introduction and some examples

Applicazioni a:

Dati tabulari e serie temporali di origine varia

MATERIALE DIDATTICO

Il docente fornirà testi di riferimento. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni. Inoltre sarà possibile avvalersi di:

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;
- Pseudocodici, templates e scripts di riferimento su libreria github dedicata

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- L'esame finale mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati nella sezione contenuti del programma
- L'esame finale, quindi consisterà in una prova orale in cui lo studente dovrà dimostrare di avere compreso nel dettaglio le metodologie utilizzate e le specificità del problema, presentare i risultati ottenuti e discuterli.
- La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti metodologici che nella risoluzioni di specifici problemi assegnati dal docente.

SCHEDA DELL'INSEGNAMENTO

FUNDAMENTALS OF PROGRAMMING AND DATA MANAGEMENT

ING-INF/05

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: I

SEMESTRE: II

MODULO: B - Data Management

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 6

INSEGNAMENTI PROPEDEUTICI

Fundamentals of Programming and Data Management – Module A, Hardware and Software for Big Data – Module A, Mathematical Methods for Data Science

EVENTUALI PREREQUISITI

nessuno

OBIETTIVI FORMATIVI

Il corso si prefigge di fornire agli studenti una conoscenza approfondita delle tecniche usate per acquisire, immagazzinare, mantenere e accedere a dati complessi. Lo studente, inoltre, comprenderà quante e quali sono le componenti fondamentali per fare vero e proprio data management nelle organizzazioni.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente acquisirà conoscenza teorica e nell'immagazzinamento ed uso di dati di interesse aziendale acquisendo le competenze necessarie al mantenimento della qualità e dell'integrità del dato durante tutto il suo ciclo di vita.

Capacità di applicare conoscenza e comprensione

Lo studente saprà utilizzare i più comuni strumenti software per risolvere problemi connessi con l'ingestione, uso ed analisi di dati.

PROGRAMMA – SYLLABUS

- Data Management: cosa significa il termine
- Dati strutturati e non strutturati
- Data Governance and data Excellence
- Data Architecture
- Data Modeling and design
- Data Fusion
- Data Integration and interoperability
- Document & Content Management

Applicazioni a:

Dati tabulari, immagini e serie temporali di origine varia

MATERIALE DIDATTICO

Il docente fornirà testi di riferimento. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni. Inoltre sarà possibile avvalersi di:

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;
- Pseudocodici, templates e scripts di riferimento su libreria github dedicata

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- L'esame finale mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati nella sezione contenuti del programma
- L'esame finale, quindi consisterà in una prova orale in cui lo studente dovrà dimostrare di avere compreso nel dettaglio le metodologie utilizzate e le specificità del problema, presentare i risultati ottenuti e discuterli.
- La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti metodologici che nella risoluzioni di specifici problemi assegnati dal docente.

SCHEDA DELL'INSEGNAMENTO
THEORY AND ETHICS OF BIG DATA AND AI
SSD M-FIL/03

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: I

SEMESTRE: II

MODULO: --

LINGUA: INGLESE

MODALITA': lezioni frontali

CFA: 6

INSEGNAMENTI PROPEDEUTICI

nessuno

EVENTUALI PREREQUISITI

nessuno

OBIETTIVI FORMATIVI

Il corso fornisce agli studenti i fondamenti etici necessari per comprendere le responsabilità personali e sociali insite nell'utilizzo dei big data e dei dati da essi derivati. In particolare il corso porrà lo studente dinanzi a problemi etici reali inducendolo ad affrontarli da diversi punti di vista.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente acquisirà consapevolezza delle proprie responsabilità etiche nell'acquisizione e sfruttamento dei dati.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà saper utilizzare i concetti acquisiti nella risoluzione di problemi etici legati allo sfruttamento dei big data.

PROGRAMMA – SYLLABUS

- Radici antiche del problema.
- Implicazioni sociali e etiche dei big data
- Il cattivo uso dei dati come "arma di distruzione matematica"
- Il dato: suo significato filosofico
- Bias espliciti e impliciti nei dati e loro implicazioni
- I dark data come strumento per nascondere la verità
- La privacy dei dati: motivazioni e limiti. Legislazione corrente e la normativa europea.

MATERIALE DIDATTICO

E' possibile utilizzare qualunque testo attinente al corso. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni.

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;)

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica sarà svolta in maniera interattiva e gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- L'esame finale mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati nella sezione contenuti del programma
- Lo studente verrà interrogato sugli argomenti del corso e verranno valutati il grado di completezza della sua risposta, il livello di integrazione tra i vari contenuti del corso, il raggiungimento da parte dello studente di una visione organica dei temi affrontati, la padronanza espressive e la proprietà nel linguaggio scientifico.
- La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti del corso.

SCHEDA DELL'INSEGNAMENTO

PREFERENCE LEARNING

SSD SEC-S/01

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: II

SEMESTRE: I

MODULO: none

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 6

INSEGNAMENTI PROPEDEUTICI

nessuno

EVENTUALI PREREQUISITI

nessuno

OBIETTIVI FORMATIVI

Il preference learning riguarda l'acquisizione di modelli di preferenza dai dati: implica l'apprendimento da osservazioni che rivelano informazioni sulle preferenze di un individuo, di una classe di individui o di sistemi automatizzati da cui si possono ricavare preferenze e la costruzione di modelli che si generalizzano oltre tali dati di addestramento.

Il corso si propone di fornire conoscenze sull'utilizzo di tecnologie avanzate di statistical e machine learning per la risoluzione di problemi di preference learning, tra cui rank aggregation, distance-based models, clustering e classificazione.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente dovrà dimostrare di aver acquisito familiarità con concetti fondamentali del preference learning , e di essere capace di orientarsi tra le risorse di dominio. Dovrà inoltre mostrare di aver compreso in linea generale potenzialità, limiti e ambiti di applicazione degli strumenti appresi.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di saper tradurre un problema in termini matematici e informatici, e di saper assemblare pipelines (e costruire algoritmi) in grado di esplorare ipotesi e fornire risposte a domande relative ai contenuti dell'esame.

PROGRAMMA – SYLLABUS

- Introduzione a preference learning e al framework dei preference rankings
- Distance-based models
- Rank aggregation
- Model-based clustering per preference data
- Metodi di classificazione per preference data
- Laboratori applicativi

MATERIALE DIDATTICO

Il docente utilizzerà testi di riferimento indicati all'inizio del corso. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni. Inoltre sarà possibile avvalersi di:

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;
- Pseudocodici, templates e scripts di riferimento su libreria github dedicata

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

Il corso si articolerà nella realizzazione di progetti di gruppo (max 3-4 persone) su problemi specifici concordati con il docente. Il lavoro prevederà una prima fase iniziale in cui gli studenti dovranno comprendere il problema facendo uso di articoli scientifici. Dovranno quindi definire le metodologie più adatte a risolvere il problema loro assegnato e implementare un codice (o più codici) atto a ricavare i risultati desiderati.

Modalità di esame

L'esame finale, quindi consisterà in una prova orale in cui lo studente dovrà dimostrare di avere compreso nel dettaglio le metodologie utilizzate e le specificità del problema, presentare i risultati ottenuti e discuterli

SCHEDA DELL'INSEGNAMENTO
ADVANCED STATISTICAL LEARNING AND MODELING
SSD SEC-S/01

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: II

SEMESTRE: I

MODULO: A & B

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 12

INSEGNAMENTI PROPEDEUTICI

Statistical Learning and Data Analysis, Data Mining

EVENTUALI PREREQUISITI

Conoscenza operativa di R e Python

OBIETTIVI FORMATIVI

Il corso ha l'obiettivo di acquisire metodologie avanzate per le decisioni e le previsioni. Il modulo A propone la classe di modelli lineari generalizzati, a effetti fissi e casuali, i modelli multi-livello.. Il modulo B propone le metodologie di apprendimento statistico moderno quali i modelli grafici, i modelli mistura, le reti bayesiane, i modelli di sopravvivenza multi-stato.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente dovrà dimostrare di aver acquisito familiarità con concetti fondamentali del preference learning , e di essere capace di orientarsi tra le risorse di dominio. Dovrà inoltre mostrare di aver compreso in linea generale potenzialità, limiti e ambiti di applicazione degli strumenti appresi.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di saper tradurre un problema in termini matematici e informatici, e di saper assemblare pipelines (e costruire algoritmi) in grado di esplorare ipotesi e fornire risposte a domande relative ai contenuti dell'esame.

PROGRAMMA – SYLLABUS

Modulo A: Advanced Statistical Models

1. Introduzione ai modelli statistici per le decisioni
2. Modelli Lineari Generalizzati
3. Modelli a risposta dicotomica
4. Modelli a risposta politomica
5. Modelli per dati di conteggio
6. Modelli Lineari Generalizzati a effetti fissi e casuali
7. Modelli Multi-Livello

Modulo B: Advanced Statistical Learning

1. Introduzione alle metodologie moderne di apprendimento statistico per le decisioni
2. Modelli grafici
3. Modelli mistura
4. Metodi di approssimazione variazionale
5. Inferenza causale
6. Reti bayesiane per l'analisi di dati spaziali e temporali
7. Modelli di sopravvivenza multi-stato

MATERIALE DIDATTICO

Il docente utilizzerà testi di riferimento indicati all'inizio del corso. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni. Inoltre sarà possibile avvalersi di:

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;
- Pseudocodici, templates e scripts di riferimento su libreria github dedicata

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

Il corso si articolerà nella realizzazione di progetti di gruppo (max 3-4 persone) su problemi specifici concordati con il docente. Il lavoro prevederà una prima fase iniziale in cui gli studenti dovranno comprendere il problema facendo uso di articoli scientifici. Dovranno quindi definire le metodologie più adatte a risolvere il problema loro assegnato e implementare un codice (o più codici) atto a ricavare i risultati desiderati.

Modalità di esame

L'esame finale, quindi consisterà in una prova orale in cui lo studente dovrà dimostrare di avere compreso nel dettaglio le metodologie utilizzate e le specificità del problema, presentare i risultati ottenuti e discuterli.

SCHEDA DELL'INSEGNAMENTO

X-INFORMATICS

SSD: FIS-01/FIS-05

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN DATA SCIENCE

ANNO ACCADEMICO 2023-2024

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: TBD

TELEFONO:

EMAIL: *@unina.it

INFORMAZIONI GENERALI – ATTIVITA'

ANNO DI CORSO: II

SEMESTRE: I

MODULO: none

LINGUA: INGLESE

MODALITA': lezioni frontali e di laboratorio

CFA: 6

INSEGNAMENTI PROPEDEUTICI

Statistical Learning and Data Analysis, Data Mining

EVENTUALI PREREQUISITI

Conoscenza operativa di Python

OBIETTIVI FORMATIVI

Il corso si prefigge di fornire agli studenti una conoscenza approfondita delle tecniche di machine learning and deep learning per l'analisi di dati scientifici complessi prodotti dai grandi strumenti e reti di sensori oggi disponibili in Fisica delle particelle, astrofisica, geofisica e altre discipline.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente dovrà dimostrare di aver acquisito familiarità con tecniche avanzate di machine learning e deep learning , e di essere capace di utilizzarle per estrarre conoscenza utile dalle grandi basi di dati distribuiti che caratterizzano il mondo scientifico. Dovrà inoltre mostrare di aver compreso potenzialità, limiti e ambiti di applicazione degli strumenti appresi.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di saper tradurre un problema reale in termini matematici e informatici, e di saper assemblare pipelines (e costruire algoritmi) in grado di esplorare ipotesi e fornire risposte a domande relative ai contenuti dell'esame.

PROGRAMMA – SYLLABUS

2. I dati scientifici. Generalità e identificazione del perché essi differiscono dai big data tradizionali, formati specifici (FITS, json, ecc.). Importanza dei metadati.
3. Data preprocessing: metodi specifici
4. Gestione delle incertezze e degli errori nel ML
5. Metodi di Deep Learning e loro principali applicazioni
7. Analisi di immagini, data cubes, hypercubes e serie temporali
8. Feature analysis and validation of results
9. Casi d'uso

MATERIALE DIDATTICO

Il docente utilizzerà testi di riferimento indicati all'inizio del corso. Il docente è disponibile a fornire chiarimenti sugli argomenti del programma, nell'orario di ricevimento. Saranno inoltre fornite le diapositive delle lezioni. Inoltre sarà possibile avvalersi di:

- Appunti del corso;
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso;
- Collezione di link di riferimento;
- Pseudocodici, templates e scripts di riferimento su libreria github dedicata

MODALITA' DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

Il corso si articolerà nella realizzazione di progetti di gruppo (max 3 persone) su problemi specifici concordati con il docente. Il lavoro prevederà una prima fase iniziale in cui gli studenti dovranno comprendere il problema facendo uso di articoli scientifici. Dovranno quindi definire le metodologie più adatte a risolvere il problema loro assegnato e implementare un codice (o più codici) atto a ricavare i risultati desiderati.

Modalità di esame

L'esame consisterà nella discussione sia dei contenuti metodologici che nella presentazione e discussione dei risultati di un progetto concordato con il docente.

