Logo Image not found or type unknown

PROGRAMMA DEL CORSO DI DATA MINING E BIG DATA

SETTORE SCIENTIFICO

SECS-S/01 (STAT-01/A)

CFU

9

SETTORE DISCIPLINARE

/**/
SECS-S/01

ANNO DI CORSO

/**/
I Anno

TIPOLOGIA DI ATTIVITÀ FORMATIVA

/**/

Base q

Caratterizzante X

Affine q

Altre attività q

DOCENTE

/**/
Raffele Miele

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.

OBIETTIVI FORMATIVI SPECIFICI

/**/

L’obiettivo del corso è di far acquisire le necessarie conoscenze tecniche di analisi e trattamento dei dati, soprattutto con applicazioni pratiche su Big data, attraverso l’utilizzo di opportuni software per il Data Mining.

RISULTATI DI APPRENDIMENTO ATTESI

/**/

Conoscenza e capacità di comprensione

Il corso mira a far conoscere gli aspetti metodologici e applicativi del Data Mining, partendo dalla classica esplorazione dei dati, per arrivare a metodi più complessi, quali il machine learning e le reti neurali. La conoscenza teorica e applicativa dei differenti metodi viene presentata parallelamente, con particolare attenzione a casi studio con l’uso di Big data.

Capacità di applicare conoscenza e comprensione

Lo studente sarà in grado di applicare le conoscenze apprese durante il corso a casi pratici, valutando quali tecnica applicare e interpretando in maniera appropriata i risultati ottenuti.

Autonomia di giudizio

Il corso intende far acquisire allo studente la capacità di formulare in maniera autonoma la tecnica di analisi più appropriata alla tipologia di informazione disponibile e agli obiettivi di analisi.

Le conoscenze apprese nel corso permetteranno di prendere iniziative e decisioni in termini di estrazione di conoscenza dai dati, in un’ottica di risoluzione di problemi pratici a partire dai dati e dalle informazioni estratte da banche dati classiche oppure da fonti tipo Big data.

Abilità comunicative

Gli strumenti messi a disposizione del corso permettono di comunicare i risultati e le analisi nella maniera più efficace e appropriata. Gli argomenti teorici e le esercitazioni pratiche svolte nel corso supportano tale capacità.

Capacità di apprendimento

Il corso mira a far acquisire allo studente la capacità di apprendimento e aggiornamento continuo, necessario per affrontare problemi di Data Mining per l'analisi dei dati nei diversi contesti applicativi.

PROGRAMMA DIDATTICO

/**/

DATA MINING, APPRENDIMENTO AUTOMATICO. DEFINIZIONI

2 - COS'E' UN MODELLO DI DATA MINING E COME SI USA

3 - L'IMPORTANZA DEL DATA MINING IN AZIENDA. CASI D'USO

4 - DATABASE & DATA WAREHOUSE

5 - BIG DATA

6 - DATA MINING. BUONE PRATICHE IN AZIENDA. IL MODELLO CRISP-DM

7 - ANALISI ESPLORATIVA UNIVARIATA

8 - ANALISI ESPLORATIVA BIVARIATA

9 - ANALISI ESPLORATIVA. TRASFORMARE I DATI

10 - ANALISI ESPLORATIVA. ASPETTI PRATICI.

11 - INTRODUZIONE ALL'SQL. ESTRAZIONE DEI DATI

12 - INTRODUZIONE ALL'SQL. ESTRAZIONE DEI DATI DA PIU TABELLE

13 - INTRODUZIONE AL LINGUAGGIO R

14 - LE STRUTTURE DATI E GLI OGGETTI R PIU' IMPORTANTI

15 - GLI STRUMENTI R PER LA MANIPOLAZIONE DEI DATI

16 - GLI STRUMENTI R PER LA VISUALIZZAZIONE DEI DATI

17 - ANALISI ESPLORATIVA CON R. ASPETTI PRATICI

18 - COSTRUIRE IL DATASET DI TRAINING

19 - LA COSTRUZIONE DELLE FEATURE

20 - TASSONOMIA DEI MODELLI DI DATA MINING

21 - MODELLI DI CLASSIFICAZIONE

22 - ALBERI DI CLASSIFICAZIONE

23 - REGRESSIONE

24 - MISURAZIONE DELLE PERFORMANCE DEI MODELLI SUPERVISIONATI

25 - CLASSIFICAZIONE: PROBLEMA DEL TARGET IMBALANCE

26 - OVERFITTING E IPERPARAMETRI

27 - ENSEMBLE LEARNING

28 - ESERCITAZIONE: MARKETING CHURN. ANALISI ESPLORATIVA

29 - ESERCITAZIONE: MARKETING CHURN. MODELING

30 - ESERCITAZIONE: MARKETING CHURN. EVALUATION E DEPLOYMENT

31 - ESERCITAZIONE: REGRESSIONE. DEMAND FORECASTING.

32 - CALIBRAZIONE, SELEZIONE DELLE VARIABILI E REGOLARIZZAZIONE

33 - ESERCITAZIONE: RISCHIO DI CREDITO

34 - BIAS E FAIRNESS DEI MODELLI. MODEL EXPLAINABILITY

35 - ESERCITAZIONE: BIAS E FAIRNESS

36 - RIDUZIONE DELLA DIMENSIONALITA

37 - ESERCITAZIONE: DATA PREPARATION. ONG

38 - ESERCITAZIONE: SCORING. ONG.

39 - CLUSTER ANALYSIS

40 - ESERCITAZIONE: MARKETING CLUSTERING

41 - ANOMALY DETECTION

42 - ESERCITAZIONE: CYBER SECURITY ANOMALY DETECTION

43 - ASSOCIATION RULES: ALGORITMO APRIORI

44 - ESERCITAZIONE: REGOLE ASSOCIATIVE

45 - ESERCITAZIONE: FRAUD DETECTION. CLASSIFICAZIONE

46 - SOCIAL NETWORK ANALYSIS: INTRODUZIONE

47 - SOCIAL NETWORK ANALYSIS: CONCETTI AVANZATI

48 - ESERCITAZIONE: GRAFI. MEMBER GET MEMBER

49 - ESERCITAZIONE: GRAFI. INVESTIGAZIONE AML

50 - ESERCITAZIONE: GRAFI. RECOMMENDATION ENGINE

51 - NATURAL LANGUAGE PROCESSING. INTRODUZIONE

52 - COMPUTER VISION. RETI NEURALI. CENNI

53 - ESERCITAZIONE: NLP. MARKETING

54 - ESERCITAZIONE: COMPUTER VISION

TIPOLOGIE DI ATTIVITÀ DIDATTICHE PREVISTE E RELATIVE MODALITÀ DI SVOLGIMENTO

/**/

Ogni Macro-argomento è articolato in 15-17 videolezioni da 30 min. corredate da dispense, slide e test di apprendimento.

Per ogni insegnamento sono previste sino a 6 videolezioni (n.1 CFU) di didattica innovativa secondo modalità definite dal docente di riferimento.

Le videolezioni sono progettate in modo da fornire allo studente una solida base di competenze culturali, logiche e metodologiche atte a far acquisire capacità critiche necessarie ad esercitare il ragionamento matematico, anche in una prospettiva interdisciplinare, a vantaggio di una visione del diritto non meramente statica e razionale, bensì quale espressione della società e della sua incessante evoluzione.

Il modello didattico adottato prevede sia didattica erogativa (DE) sia didattica interattiva (DI):

La didattica erogativa (DE) prevede l'erogazione in modalità asincrona delle videolezioni, delle dispense, dei test di autovalutazioni predisposti dai docenti titolari dell'insegnamento; la metodologia di insegnamento avviene in teledidattica. La didattica interattiva (DI) comprende il complesso degli interventi didattici interattivi, predisposti dal docente o dal tutor in piattaforma, utili a sviluppare l'apprendimento online con modalità attive e partecipative ed è basata sull'interazione dei discenti con i docenti, attraverso la partecipazione ad attività didattiche online.
Sono previsti interventi brevi effettuati dai corsisti (ad esempio in ambienti di discussione o di collaborazione, in forum, blog, wiki), e-tivity strutturate (individuali o collaborative), sotto forma tipicamente di produzioni di elaborati o esercitazioni online e la partecipazione a web conference interattive.

Nelle suddette attività convergono molteplici strumenti didattici, che agiscono in modo sinergico sul percorso di formazione ed apprendimento dello studente. La partecipazione attiva alle suddette attività ha come obiettivo quello di stimolare gli studenti lungo tutto il percorso didattico e garantisce loro la possibilità di ottenere una valutazione aggiuntiva che si sommerà alla valutazione dell'esame finale.

Per le attività di autoapprendimento sono previste 162 ore di studio individuale.

L'Ateneo prevede 7 h per ogni CFU articolate in 6 h di didattica erogativa (DE) e 1 h di didattica interattiva (DI).

Nel computo delle ore della DI sono escluse le interazioni a carattere orientativo sui programmi, sul cds, sull'uso della piattaforma e simili, che rientrano un semplice tutoraggio di orientamento. Sono altresì escluse le ore di tutorato didattico disciplinare, cioè la mera ripetizione di contenuti già proposti nella forma erogativa attraverso colloqui di recupero o approfondimento one-to-one.

MODALITÀ E CRITERI DI VALUTAZIONE DELL'APPRENDIMENTO

/**/

La partecipazione alla didattica interattiva (DI) ha la finalità, tra le altre, di valutare lo studente durante l'apprendimento in itinere.

L'esame finale può essere sostenuto in forma scritta o in forma orale; lo studente può individuare, in autonomia, la modalità di svolgimento della prova, sempre rispettando la calendarizzazione predisposta dall'Ateneo.

L'esame orale consiste in un colloquio nel corso del quale il docente formula almeno tre domande.

L'esame scritto consiste nello svolgimento di un test a risposta multipla con 31 domande. Per ogni domanda lo studente deve scegliere una delle 4 possibili risposte. Solo una risposta è corretta.

Sia i quesiti in forma orale che i quesiti in forma scritta sono formulati per valutare il grado di comprensione delle nozioni teoriche e la capacità di sviluppare il ragionamento utilizzando le nozioni acquisite. I quesiti 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.

CRITERI DI MISURAZIONE DELL'APPRENDIMENTO E ATTRIBUZIONE DEL VOTO FINALE

Sia lo svolgimento dell'elaborato, sia la presenza attiva durante le web conference prevedono un giudizio, da parte del docente, fino a un massimo di 2 punti. Lo studente può prendere parte ad entrambe le attività ma la votazione massima raggiungibile è sempre di 2 punti.

La valutazione proveniente dallo sviluppo dell'elaborato può essere pari a 0, 1 o 2 punti.

La valutazione derivante dalle web conference è strutturata tramite lo svolgimento, al termine della stessa, di un test finale a risposta multipla che può garantire da 0 a 1 punto.

È data facoltà allo studente di partecipare o meno alla didattica interattiva.

La valutazione finale ha lo scopo di misurare il raggiungimento degli obiettivi di apprendimento definiti alla base dell'insegnamento. Il giudizio riguarda l'intero percorso formativo del singolo insegnamento ed è di tipo sommativo.

Il voto finale dell'esame di profitto tiene conto del punteggio che lo studente può aver ottenuto partecipando correttamente alla didattica interattiva e deriva, quindi, dalla somma delle due valutazioni. Il voto derivante dalla didattica interattiva verrà sommato al voto dell'esame se quest'ultimo sarà pari o superiore a diciotto trentesimi.

Il voto finale è espresso in trentesimi. Il voto minimo utile al superamento della prova è di diciotto trentesimi.

Ciascun test dovrà essere composto da 31 domande, così da garantire la possibilità di conseguire la lode, in ottemperanza alle norme Europee sul Diploma Supplement. L'attribuzione della lode è concessa esclusivamente allo studente che ha risposto positivamente alle prime 30 domande.

ATTIVITÀ DI DIDATTICA EROGATIVA (DE)

/**/

- è 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

/**/

- è Elaborato con applicazione su un argomento del corso
- è Partecipazione a una web conference
- è Forum sulle diverse tipologie dei Big Data da utilizzare nel data mining
- è Svolgimento della simulazione del test finale

Totale 9 ore

MATERIALE DIDATTICO UTILIZZATO

/**/

- è Videolezioni
- è Dispense predisposte dal docente e/o slide del docente
- è Testo di riferimento suggerito dal docente (facoltativo)

Il materiale didattico è sempre disponibile in piattaforma e consultabile dallo studente nei tempi e nelle modalità ad egli più affini.