

## PROGRAMMA DEL CORSO DI DATA MINING E BIG DATA

### SETTORE SCIENTIFICO

SECS-S/01

### CFU

9

### OBIETTIVI

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.

## **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**

- Elaborato con applicazione su un argomento del corso
- Partecipazione a una web conference

- Forum sulle diverse tipologie dei Big Data da utilizzare nel data mining
- Svolgimento della simulazione del test finale

Totale 12 ore

### **ATTIVITÀ DI AUTOAPPRENDIMENTO**

- 216 ore per lo studio individuale

### **LIBRO DI RIFERIMENTO**

- Dispense del docente.
- Testo di riferimento non necessario

### **PROGRAMMA DIDATTICO**

- Data Mining e Big data: concetti fondamentali ( 6 - Lezioni)
- Analisi esplorativa. (4 - Lezioni)
- Gli strumenti tecnici necessari: I linguaggi R e SQL (7- Lezioni)
- Data Mining: approfondimenti (5- Lezioni)
- Apprendimento supervisionato su dati strutturati (11- Lezioni)
- Apprendimento non supervisionato su dati strutturati (6- Lezioni)
- Apprendimento da dati non strutturati (2 - Lezioni)
- Applicazioni a casi reali: dati strutturati (11- Lezioni)
- Applicazione a casi reali: dati non strutturati (2- Lezioni)

Il Docente si riserva di modificare il programma

Lezioni

1 - 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 PIÙ 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 DIMENSIONALITÀ
- 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