

FACOLTÀ: Lettere

CORSO DI LAUREA: Lingue e letterature moderne e traduzione interculturale-LM37 – curr. Traduzione e processi interlinguistici

INSEGNAMENTO: Data Mining

CFU: 9

EVENTUALE ARTICOLAZIONE IN MODULI: n/a

ANNO DI CORSO: SECONDO

NOME DOCENTE: Michela Antonelli

Indirizzo e-mail: michela.antonelli@uniecampus.it

I docenti possono essere contattati attraverso la sezione *Ricevimento docenti*, presente nell'area riservata del sito di Ateneo, che comprende *Ufficio virtuale*, *Sistema di messaggistica* e *Ricevimento Telefonico*.

Per le comunicazioni scritte bisogna utilizzare il *Sistema di Messaggistica*.

Orario ricevimento on line: mercoledì dalle 14.00 alle 15.00

É molto apprezzato dal docente essere contattato in anticipo tramite il sistema di messaggistica per avvertirlo della necessità di ricevimento e delle tematiche da approfondire.

RISULTATI DI APPRENDIMENTO DELL'INSEGNAMENTO

L'insegnamento ha l'obiettivo di far conseguire allo studente i seguenti risultati di apprendimento

1. Con riferimento alla conoscenza e capacità di comprensione, il corso si propone di fornire ai partecipanti:
 - a. Gli strumenti basilari (modelli ed algoritmi) per l'analisi e l'estrazione della conoscenza dai dati;
 - b. Un insieme di competenze specialistiche per l'utilizzo di strumenti software per il data mining;
 - c. Una panoramica di esempi pratici e reali di applicazione delle tecniche di data mining.
2. Con riferimento alla conoscenza e capacità di comprensione applicate, al termine del corso lo studente sarà in grado di:
 - a. Risolvere semplici problemi di data mining;
 - b. Formalizzare problemi di estrazione di conoscenza dai dati ed identificare gli opportuni modelli teorici (descritti e/o predittivi) e gli algoritmi per la loro soluzione;
 - c. Progettare ed implementare applicazione software in grado di risolvere problemi reali di data mining.

3. Con riferimento all'autonomia di giudizio
 - a. Saper impostare correttamente un progetto di estrazione della conoscenza, integrandolo all'interno di una reale applicazione informatica e scegliendo in maniera opportuna le strategie di progetto e di implementazione.
4. Con riferimento alle abilità comunicativa
 - a. Utilizzo del lessico e della terminologia appropriata per presentare, in forma scritta o verbale, le argomentazioni e i risultati del proprio studio.
5. Con riferimento all'abilità ad apprendere
 - a. Autonomia nell'utilizzo del materiale e negli strumenti didattici in piattaforma.

PROGRAMMA DETTAGLIATO

Le lezioni saranno organizzate secondo i seguenti nuclei tematici:

1. Introduzione al corso e al data mining
2. Trattamento preliminare dei dati
3. Misure di distanza e similarità
4. Analisi descrittiva tramite clustering
5. Analisi descrittiva tramite regole associative
6. Analisi predittiva: classificazione
7. Il tool WEKA e uso delle sue API
8. Realizzazione del progetto

EVENTUALI PROPEDEUTICITÀ CONSIGLIATE

Si consiglia di aver almeno seguito il corso di Linguaggi di Programmazione e di essere a conoscenza dei fondamenti di Ingegneria del Software e Basi di Dati.

MODALITÀ DI SVOLGIMENTO ESAME

L'esame si svolge a scelta dello studente in modalità scritta, attraverso una prova costituita da 23 domande a risposta chiusa e 2 domande a risposta aperta con eventuale orale integrativo, o in modalità orale, in base a quanto previsto dal *Regolamento per lo svolgimento degli esami di profitto* consultabile sul sito dell'Ateneo, al seguente link.

https://www.uniecampus.it/fileadmin/user_upload/regolamenti/Regolamento_per_lo_Svolgimento_degli_esami_di_profitto.pdf

Lo studente per essere ammesso alla prova **scritta (o alla prova orale ove previsto) deve realizzare preventivamente un progetto didattico** o una esercitazione alternativa. Sia il progetto sia l'esercitazione devono essere discussi (e approvati), almeno una settimana prima della prova scritta, con il docente nel suo ufficio virtuale. Vedere di seguito (didattica Interattiva) le **modalità di valutazione** del progetto didattico o dell'esercitazione alternative.

CRITERI DI VALUTAZIONE DELL'APPRENDIMENTO

1. Con riferimento alla conoscenza e capacità di comprensione
 - Domande a risposta chiusa e aperta nella prova scritta.
2. Con riferimento alla conoscenza e capacità di comprensione applicate
 - Esercitazioni applicative durante il corso con consegna in ePortfolio;
 - Esercizi in sede di esame scritto;
 - Realizzazione del progetto didattico;
 - In sede di esame orale, allo studente sarà data la possibilità di discutere problematiche teorico/pratiche relative agli argomenti trattati nel corso.
3. Con riferimento all'autonomia di giudizio:
 - Lo studente dovrà essere in grado di impostare in maniera ottimale il progetto didattico. Ciò è valutato sulla base della documentazione fornita e della discussione del progetto in aula virtuale.
4. Con riferimento all'abilità comunicativa:
 - L'abilità viene valutata sulla base della proprietà di linguaggio tecnico utilizzato nelle risposte alle domande a risposta aperta e durante l'eventuale prova orale.
5. Con riferimento all'abilità ad apprendere:
 - L'abilità è valutata essenzialmente sulla base delle esercitazioni svolte in itinere, in modo da verificare che lo studente sia effettivamente in grado di utilizzare gli strumenti offerti in piattaforma di e-learning.

CRITERI DI ATTRIBUZIONE DEL VOTO FINALE

Sulla base dei criteri di valutazione sopra indicati, l'attribuzione del voto finale avviene attraverso i seguenti criteri:

- 1) 1. Criteri di attribuzione del voto alla prova scritta:
 - a) le risposte alle domande chiuse sono valutate su scala 0-1 punti, secondo i seguenti criteri:
 - 0 = risposta errata;
 - 1 = risposta corretta;
 - b) le risposte alle domande aperte sono valutate su scala 0-2 punti, secondo i seguenti criteri:
 - 0 = risposta mancante, errata o priva di elaborazione personale;
 - 1 = contestualizzazione della risposta corretta, ma con presenza di elementi non corretti o esposta in modo non efficace o incompleto;
 - 2 = risposta corretta, ben esposta.

Alla prova scritta può essere attribuito un punteggio massimo di 27/30.

È possibile sostenere una prova orale integrativa per il raggiungimento di un punteggio superiore come dettagliato nel Regolamento per lo svolgimento degli esami di profitto.

2. Criteri di attribuzione del voto alla prova orale:

0/30 – 17/30: prevalenza di argomentazioni non corrette e/o incomplete e scarsa capacità espositiva;
18/30 – 21/30: prevalenza di argomentazioni corrette adeguatamente esposte;
22/30 – 26/30: argomentazioni corrette e ben esposte;
27/30 – 30/30 e lode: conoscenza approfondita della materia ed elevata capacità espositiva, di approfondimento e di rielaborazione.

3. Criteri di attribuzione del voto al progetto o esercitazione finale:

Durante il corso, viene proposta la realizzazione di un progettinio didattico. Le specifiche del progetto devono essere concordate con il docente con adeguato anticipo. Il docente non valuterà progetti di cui non saranno inviate ed approvate le specifiche.

É **obbligatorio** caricare nell'ePortfolio dello studente un certo numero di artefatti relativi al progetto didattico. Tali artefatti sono relativi a tutte le fasi di realizzazione del progetto e **saranno valutati ai fini dell'esame finale**.

É obbligatorio discutere il progetto o l'esercitazione alternativa, fissando un appuntamento in aula virtuale del docente, entro la settimana che precede l'appello d'esame. In caso di mancata discussione, lo **studente non sarà ammesso alla prova di esame**.

Due delle domande a risposta aperta **saranno relative al progetto** e consentiranno di valutare conoscenze e le capacità **di comprensione applicate** dello studente e **il livello di autonomia di giudizio** raggiunto dallo studente.

Gli studenti che decideranno di non svolgere il progetto, non potranno dimostrare completamente di aver acquisito capacità **di comprensione applicate**. Inoltre, il livello di **autonomia di giudizio** non sarà completamente valutabile. Per questo motivo, se avranno svolto correttamente la loro esercitazione alternativa (da concordare con il docente almeno 10 giorni prima dell'esame), **potranno raggiungere al massimo la valutazione di 24/30** (rispondendo correttamente alle 21 domande a crocette e alla domanda aperta).

Gli studenti che avranno svolto il progetto, potranno ambire al massimo dei voti.

In tutti i test si ritroveranno sempre le seguenti domande:

- 1) Quale è stata la soluzione proposta per soddisfare i requisiti del progetto?
- 2) Quali sono stati i risultati raggiunti con la soluzione proposta per soddisfare i requisiti del progetto?

Sarà premura del **docente valutare** le risposte fornite, facendo riferimento a quanto svolto nel progetto.

Ovviamente, ad un **progetto mediocre** seguirà **una valutazione mediocre** delle risposte fornite (anche se verranno scritte delle cose apparentemente bellissime). Ad un **progetto di ottimo** livello verrà assegnata la **massima valutazione**.

Maggiori informazioni relative al progetto didattico e alla esercitazione alternativa si trovano nella prima lezione, sulle FAQ predisposte dal docente e nei documenti pubblicati sulla pagina personale del docente.

MATERIALE DIDATTICO

Susi Dulli, Sara Furini e Edmondo Peron "Data mining. Metodi e strategie", Springer, 2009 (**Testo Obbligatorio** da cui approfondire gli argomenti, molte delle lezioni fanno riferimento a questo libro)

J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2nd ed., Morgan Kaufmann Publishers, 2006 (Testo Consigliato, In Inglese, uno dei libri più completi sul data mining).

Ian H. Witten, Eibe Frank, Mark A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition", Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2011 (Testo Consigliato, In Inglese, fa riferimento all'uso del programma WEKA).

ATTIVITÀ DIDATTICHE

Attività di Didattica Erogativa (ore di impegno stimato per lo studente):

- 9.5 ore di Videolezioni (4.75 ore effettive, da vedere due volte);
- 6 ore di Webinar;
 - Introduzione al corso (Livello Base)
 - Dati e preprocessing (Livello Base)
 - Algoritmi di classificazione (Livello Intermedio)
 - Algoritmi di clustering (Livello Intermedio)
 - Algoritmi di generazione di regole di associazione (Livello Intermedio)
 - Simulazione prova di esame (Livello Base)

Attività di Didattica Interattiva (ore di impegno stimato per lo studente):

- 30.5 ore di approfondimento con **esercizi** e alle sessioni di **esempi di domande di esame**;
- 8 ore di aula virtuale.

Attività di autoapprendimento (ore di impegno stimato per lo studente):

- 171 ore (slide del corso, dispense, libri, articoli, sitografia, testi d'esame).

L'insegnamento prevede la sostituzione dell'**apprendimento in situazione**, dunque del tradizionale rapporto in presenza tra il docente e i discenti, attraverso le attività di Webinar (aula virtuale) e esercizi.

I Webinar costituiscono delle lezioni svolte in diretta dal docente tramite la piattaforma telematica dell'Ateneo, della durata di circa 1 ora.

I Webinar sono attività organizzate in periodici cicli. Lo studente può visionare il calendario dei Webinar del CdS nella sezione Webinar. I Webinar offerti, sia erogativi che interattivi, sono flessibilmente aperti alla differenziazione e personalizzazione didattica.

CONSIGLI DEL DOCENTE

Si invitano gli studenti a leggere attentamente la prima lezione del corso e le FAQ predisposte dal docente sulla sua pagina.