MATHEMATICS (LM39)

(Lecce - Università degli Studi)

Teaching DATA MINING

GenCod A004898

Owner professor Massimo CAFARO

Teaching in italian DATA MINING

Teaching DATA MINING

SSD code ING-INF/05

Reference course MATHEMATICS

Course type Laurea Magistrale

Credits 6.0

Teaching hours Ore-Attivita-frontale: 42.0

For enrolled in 2017/2018

Taught in 2017/2018

Course year 1

Language INGLESE

Curriculum PERCORSO COMUNE

Location Lecce

Semester Secondo-Semestre

Exam type Orale

Assessment Voto-Finale

Course timetable https://easyroom.unisalento.it/Orario

BRIEF COURSE DESCRIPTION

Introduzione al corso. Map-Reduce. (2 ore) Mining data streams. Frequent Items. (6 ore) Frequent Itemsets ed association rules.(4 ore) Mining similar items e Locality-Sensitive Hashing. (2 ore) Analisi di grafi. Link analysis e PageRank. (2 ore) Clustering. (4 ore) Recommendation systems. (4 ore) Mining Social-Network Graphs. (4 ore) Dimensionality reduction. (2 ore) Classification. (6 ore) Esercitazioni. (6 ore).

REQUIREMENTS

Analisi Matematica. Probabilità e statistica. Algebra lineare. Programmazione ed analisi di algoritmi.

COURSE AIMS

Il corso fornisce una moderna introduzione al data mining, un insieme di tecniche, algoritmi e metodologie per scoprire la struttura, patterns e relazioni in insiemi di dati (tipicamente, quelli più grandi) e fare previsioni. Le applicazioni del data mining stanno già accadendo intorno a noi, e se ben fatte, possono a volte anche passare inosservate. Come funziona la ricerca sul web di Google? Come fa Shazam a riconoscere una canzone? Come fa Netflix a raccomandare film a ciascuno dei suoi utenti? I principi del data mining forniscono le risposte di base a queste e ad altre domande simili. Il data mining abbraccia i campi dell'informatica, dello statistical machine learning e dei database. Obiettivo del corso è mettere in grado gli studenti di esplorare, analizzare e sfruttare i dati disponibili al fine di trasformarli in informazioni quantitative e qualitative di valore ed interesse per una azienda, ad esempio ai fini di un processo di decision-making.

Risultati di apprendimento

Dopo aver seguito il corso, lo studente dovrebbe essere in grado di:

- descrivere ed utilizzare le principali tecniche di data mining;
- comprendere le differenze tra algoritmi diversi che risolvono uno stesso problema e riconoscere quale algoritmo è il migliore rispetto a condizioni diverse;
- affrontare nuovi problemi di data mining scegliendo i metodi più appropriati e giustificando le proprie scelte;
- affrontare nuovi problemi di data mining progettando appositi algoritmi e valutando i risultati;
- spiegare i risultati ottenuti sperimentalmente anche a persone con un background teorico diverso da statistica e/o informatica.

ASSESSMENT TYPE	L'esame è orale. Durante l'esame, allo studente viene chiesto di illustrare argomenti teorici per verificare la sua conoscenza e comprensione degli argomenti scelti.
REFERENCE TEXT BOOKS	Mining of Massive Datasets J. Leskovec, A. Rajaraman and J. Ullman Disponibile gratuitamente online: http://www.mmds.org Data Mining and Analysis M. J. Zaki and W. Meira Disponibile gratuitamente online: http://dataminingbook.info