UNIVERSITÀ DI PISA

DATA MINING AND MACHINE LEARNING

FRANCESCO MARCELLONI

Academic year	2019/20
Course	ARTIFICIAL INTELLIGENCE AND DATA ENGINEERING
Code	878II
Credits	12

Modules	Area	Type	Hours	Teacher(s)
ADVANCED DATA MINING AND MACHINE LEARNING	ING-INF/05	LEZIONI	60	PIETRO DUCANGE FRANCESCO MARCELLONI
FUNDAMENTALS OF DATA MINING AND MACHINE LEARNING	ING-INF/05	LEZIONI	60	PIETRO DUCANGE FRANCESCO MARCELLONI

Obiettivi di apprendimento

Conoscenze
Gli studenti che completeranno con successo l'insegnamento avranno una solida conoscenza delle principali tecniche per pre-processare i dati, frequent pattern mining, frequent sequential pattern mining, graph mining, classificazione, predizione, clustering, e outlier detection. Questa conoscenza permetterà loro di affrontare ogni tipo di problema inerente il data mining e di identificare la tecnica più adatta per risolverlo.

Modalità di verifica delle conoscenze
Durante la verifica delle conoscenze, gli studenti devono dimostrare di aver appreso le diverse tecniche insegnate durante lo svolgimento del corso e devono essere capaci di identificare la soluzione più adatta per problemi di data mining specifici.
I metodi sono:

- esame orale
- report e presentazione del progetto

Ulteriori informazioni: allo studente è richiesto di sviluppare un progetto in cui vengono utilizzate tecniche di data mining. I risultati del progetto vengono discussi durante una presentazione.

Capacità
Al termine del corso,

- lo studente saprà affrontare i più comuni problemi di data mining, trovando le soluzioni più idonee per risolverli
- lo studente saprà valutare e confrontare più soluzioni e scegliere la più adatta

Modalità di verifica delle capacità
Lo studente dovrà preparare e presentare una relazione scritta che riporti i risultati dell'attività di progetto

Comportamenti
Lo studente potrà acquisire un metodo per affrontare problemi di data mining e per selezionare le migliori soluzioni da adottare

Modalità di verifica dei comportamenti
Durante le sessioni di laboratorio saranno valutati il grado di accuratezza e precisione delle attività svolte dallo studente
Durante lo sviluppo del progetto saranno verificate le modalità di gestione e organizzazione delle fasi progettuali

Prerequisiti (conoscenze iniziali)
Conoscenze di base di matematica
Conoscenze di linguaggi di programmazione (preferibilmente Java)

Indicazioni metodologiche

Le lezioni verranno svolte frontalmente con l'ausilio di lucidi
Le esercitazioni verranno svolte in laboratorio con l'ausilio di lucidi
Durante il corso, verrà sviluppato dallo studente un progetto che costituirà parte della valutazione finale

L'intero corso è tenuto in Inglese

Programma (contenuti dell'insegnamento)

Data Preprocessing: data cleaning, integration, reduction, transformation and discretization.
Frequent pattern mining: basic concepts, A-priori algorithm, Pattern-Growth approach, vertical data format, pattern evaluation methods, constraint-based frequent pattern mining, colossal pattern.
Classification: basic concepts, decision tree induction, Bayes classification methods, rule-based classification, lazy learners, techniques for improving accuracy, model evaluation and selection.
Clustering: basic concepts, partitioning methods, hierarchical methods, density-based methods, grid-based methods, model evaluation and selection, clustering with constraints.
Outlier detection: statistical, proximity-based, clustering-based and classification-based approaches.
Sequential Pattern Mining: basic concepts, AprioriAll, AprioriSome, AprioriDynamicSome
Graph Mining: basic concept, geodesic distance, SimRank, Density-based approaches to graph clustering.
Distributed frameworks: basic concepts, Hadoop, MapReduce paradigm, Spark, some examples of data mining algorithms implemented by using MapReduce

Bibliografia e materiale didattico

Slides
Libro: J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd ed., 2011 Papers on the different algorithms described during the course Slides of the lectures
Articoli forniti dal docente

Modalità d'esame

L'esame è composto dalla discussione del progetto e una prova orale.
La discussione del progetto viene tipicamente tenuta qualche giorno prima dell'esame orale. Il candidato deve presentare come il progetto è stato sviluppato, motivare le sue scelte progettuali e discutere i risultati ottenuti. Il progetto viene valutato positivamente se il candidato mostra di aver seguito un approccio corretto e di aver valutato in modo critico le possibili soluzioni, scegliendo la più appropriata
La prova orale consiste in un colloquio tra il candidato e il docente su alcune domande che sono assegnate in forma scritta al candidato.
La prova orale è superata se il candidato mostra padronanza degli argomenti trattati, si esprime in modo chiaro e con terminologia corretta, mostra capacità di analisi e sintesi.

Ultimo aggiornamento 16/11/2019 10:56