

anno di corso	insegnamento	modulo	cfu	settore scientifico disciplinare	tipo	semestre	tipo attività didattica	cfu	ore	docente	
1	F9101Q005 Machine learning and decision models	Machine Learning (6 cfu)	12	INF/01	obbligatorio	Primo semestre	lezione	4	32	Stella Fabio	
							esercitazione	2	20	Stella Fabio	
		Decision models (6 cfu)		MAT/09		obbligatorio	Secondo semestre	lezione	4	32	Messina Enza
								esercitazione	2	20	Ciavotta Michele

CV docenti: <http://www.unimib.it/go/176181440>)

Contenuti

L'insegnamento è diviso in due moduli: il primo modulo riguarda le tecniche di data mining per l'analisi di dati e la seconda presenta i modelli decisionali evidenziando il valore dei dati all'interno del processo decisionale. In particolare verranno affrontati i seguenti argomenti: Il ciclo del data mining. Esplorazione preliminare e tecniche di pre-processamento dei dati. Classificazione supervisionata e non supervisionata. Misure di validità e validazione. Modelli decisionali e tecniche di ottimizzazione. Il valore dell'informazione: analisi di scenari e di sensibilità.

Obiettivi formativi

Lo studente sarà in grado di progettare e sviluppare cicli di data mining per risolvere problemi di classificazione supervisionata e non supervisionata, così come problemi di analisi delle associazioni. Lo studente sarà in grado di capire il valore dei dati nel processo decisionale per trasformare le informazioni in conoscenza di supporto alle decisioni, anche attraverso l'utilizzo di esempi e casi di studio. Lo studente apprenderà l'uso di KNIME, una piattaforma software open source, che integra la potenza e l'espressività di applicativi di data mining come Weka e RapidMiner con linguaggi per l'analisi dei dati come R, Java e Python.

Prerequisiti

Conoscenza base di informatica, probabilità e statistica.

Testi di riferimento

- Pang-Ning Tan, Michael Steinbach and Vipin Kumar (2006). Introduction to Data Mining, Pearson International.
- Bertsimas, Dimitris, and Robert Freund. *Data, Models, and Decisions: The Fundamentals of Management Science*. Dynamic Ideas, 2004. ISBN: 9780975914601.

Metodi didattici

Il corso si basa su una diffusa applicazione pratica dei contenuti. In particolare, diversi dataset sono utilizzati per mostrare work-flow di analisi con lo scopo di formulare e risolvere diverse tipologie di problemi.

Modalità di verifica dell'apprendimento

Progetto ed esercizi durante l'erogazione del Corso. Inoltre, esame scritto ed orale.

Programma esteso

Modulo Machine Learning

- 1) Introduzione al Data Mining
 - Esplorazione dei dati
 - Preprocessing
- 2) Classificazione supervisionata e non supervisionata
 - Classificatore e misure di prestazione (accuratezza, errore, precisione and richiamo): schemi di stima, k-folds cross validation, overfitting and underfitting, curse of dimensionality
 - Matrice di costo, receiver operating characteristic curve, grafici lift e del guadagno cumulato
 - Attributi irrilevanti e ridondanti, selezione delle feature
 - Problema di classificazione non supervisionata: Metodi di partizione, gerarchici, basati sulla densità e sul concetto di grafo
 - Misure di validità e validazione
- 3) Analisi delle associazioni
 - Regola di associazione, itemset ed itemset frequente, supporto e confidenza
 - Algoritmo Apriori
 - Itemset massimali e itemset frequenti chiusi
 - Criteri per valutare la qualità delle regole associative

Modulo Decision Models

- 1) Modelli decisionali
 - Tipi di decisioni
 - Decisioni ed incertezza
 - Modelli data driven
- 2) Dai dati alle decisioni
 - Modelli di ottimizzazione
 - Modelli di ottimizzazione in condizioni di incertezza
 - Principi di simulazione
- 3) Il valore dell'informazione
 - Analisi degli scenari
 - Analisi di sensitività

Master of science in Data Science a.a. 2017-18

year	code and course name	module	ECTS	type	semester	educational activity type	ECTS	hours	faculty
1	F9101Q005 Machine learning and decision models	Machine Learning (6 ECTS)	12	mandatory	First semester	lecture	4	32	Stella Fabio
						practice exercise	2	20	Stella Fabio
		Decision models (6 ECTS)		mandatory	Second semester	lecture	4	32	Messina Enza
						practice exercise	2	20	Ciavotta Michele

CV: <http://www.unimib.it/go/176181440>

Content of the course

The course consists of two modules: the first one, Machine Learning, is about machine learning and data mining techniques, while the second is about decision models and emphasizes the relevance of data in decision making.

The following topics are covered: the data mining cycle. Data exploration and pre-processing. Supervised classification, un-supervised classification and association rules extraction. Validity measures and model validation. Decision analysis and optimization models. Information Value: scenario analysis and sensitivity analysis.

Educational goals

Students will understand the relevance of transforming data into information to enhance decision-making effectiveness. Students will be able to choose and use the most effective data mining and optimization models. Students will be able to design and develop data mining workflows to solve supervised and un-supervised classification problems as well as association rule learning problems. Students will learn by using practical examples. Finally, students will learn to use KNIME, an open source platform that integrates power and expressiveness of Weka, RapidMiner and programming languages such as R, Java and Python.

Prerequisites

Basic knowledge of informatics, probability and statistics

Teaching material

- Pang-Ning Tan, Michael Steinbach and Vipin Kumar (2006). *Introduction to Data Mining*, Pearson International.
- Bertsimas, Dimitris, and Robert Freund (2000). *Data, Models, and Decisions: The Fundamentals of Management Science*. Dynamic Ideas, 2004. ISBN: 9780975914601.

Teaching methods

The course is hands-on. In particular, we use different data sets to show workflows to formulate and solve different types of problems.

Assessment methods

Project and running course exercises. Written and oral exam.

Extended program

Machine Learning Module

1. Introduction to Data Mining

- Data Exploration
- Pre-processing

2 Supervised and Unsupervised Classification

- Classifier and performance measures
- Performance estimation. Hold-out, k-folds cross validation,
- Overfitting and underfitting, curse of dimensionality
- Cost matrix, receiver operating characteristic curve, lift and cumulative gain charts
- Non relevant and redundant attributes, feature selection
- Unsupervised classification. Partitioning, Hierarchical, density-based and graph-based
- Validity measures and validation

3 Association Analysis

- Association rule, itemset, frequent itemset, support and confidence
- Apriori algorithm
- Maximal Itemset and closed itemset
- Evaluation criteria

Decision Models Module

1. Decision Analysis

- Types of decision
- Decision and uncertainty
- Data driven models

2. From data to decision

- Optimization models
- Optimization models under uncertainty
- Simulation concepts

3. Information value

- Scenario analysis
- Sensitivity analysis