Course guide
270615 - ADM - Algorithmics for Data Mining

Last modified: 02/02/2024

Unit in charge: Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science.

Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Optional subject).
MASTER'S DEGREE IN DATA SCIENCE (Syllabus 2021). (Optional subject).

Academic year: 2023 **ECTS Credits:** 6.0 **Languages:** English

LECTURER

Coordinating lecturer: JOSÉ LUIS BALCÁZAR NAVARRO - LUIS ANTONIO BELANCHE MUÑOZ

Others: Segon quadrimestre:
LUIS ANTONIO BELANCHE MUÑOZ - 10

PRIOR SKILLS

Adequate understanding of computing in general, especially algorithms; good level of various programming languages (such as R, python, Julia) or willingness to achieve it; basic to average ability to mathematically formalize concepts in computing, statistics, etc.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CEE3.1. Capability to identify computational barriers and to analyze the complexity of computational problems in different areas of science and technology as well as to represent high complexity problems in mathematical structures which can be treated effectively with algorithmic schemes.
CEE3.2. Capability to use a wide and varied spectrum of algorithmic resources to solve high difficulty algorithmic problems.
CEE3.3. Capability to understand the computational requirements of problems from non-informatics disciplines and to make significant contributions in multidisciplinary teams that use computing.

Generical:
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CG5. Capability to apply innovative solutions and make progress in the knowledge to exploit the new paradigms of computing, particularly in distributed environments.

Transversal:
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Basic:
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

TEACHING METHODOLOGY

Theoretical classes, exercises and problems with or without a programming component and development of case studies.

LEARNING OBJECTIVES OF THE SUBJECT

1.Te be aware of the theoretical and practical set of problems that constitute Data Mining, and to understand the main models and algorithms to tackle it: both at the conceptual level and at the level of their application through commercial tools, preferably open-source.
2.To acquire and demonstrate an ability to put to work the knowledge obtained in the autonomous, team-wise deployment of a practical data mining case, including a public presentation of the work developed.

STUDY LOAD

Type	Hours	Percentage
Self study	96,0	64.00
Hours small group	36,0	24.00
Hours large group	18,0	12.00

Total learning time: 150 h

CONTENTS

Selected techniques and algorithms for Data Mining

Description:
Algorithms and techniques are representative of the good and the best a data practitioner needs to know, among which:

backpropagation
expectation-maximization
association rules
pagerank
GLMs

Each topic of study is focused in 3 aspects:

theoretical
algorithmic
practical

ACTIVITIES

Theoretical and conceptual study of the main data mining algorithms.

Description:
Theoretical and conceptual study of the main data mining algorithms.

Specific objectives:
1

Related competencies :
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.
CG5. Capability to apply innovative solutions and make progress in the knowledge to exploit the new paradigms of computing, particularly in distributed environments.
CEE3.1. Capability to identify computational barriers and to analyze the complexity of computational problems in different areas of science and technology as well as to represent high complexity problems in mathematical structures which can be treated effectively with algorithmic schemes.
CEE3.3. Capability to understand the computational requirements of problems from non-informatics disciplines and to make significant contributions in multidisciplinary teams that use computing.
CEE3.2. Capability to use a wide and varied spectrum of algorithmic resources to solve high difficulty algorithmic problems.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

Full-or-part-time: 30h
Theory classes: 18h
Practical classes: 6h
Self study: 6h

Deploy of a practical case study

Description:
Deploy of a practical case study

Specific objectives:
1, 2

Related competencies :
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.
CG5. Capability to apply innovative solutions and make progress in the knowledge to exploit the new paradigms of computing, particularly in distributed environments.
CEE3.1. Capability to identify computational barriers and to analyze the complexity of computational problems in different areas of science and technology as well as to represent high complexity problems in mathematical structures which can be treated effectively with algorithmic schemes.
CEE3.3. Capability to understand the computational requirements of problems from non-informatics disciplines and to make significant contributions in multidisciplinary teams that use computing.
CEE3.2. Capability to use a wide and varied spectrum of algorithmic resources to solve high difficulty algorithmic problems.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.

Full-or-part-time: 54h
Laboratory classes: 36h
Self study: 18h

GRADING SYSTEM

Evaluation is fully offline and there will be no exams. Each person must contribute with three (3) case studies, solved exercises or studied problems (written report plus eventually code) on topics related to the course; each of these must be worked out by three (3) people, as follows:

- 2 people do the exercise per se
- 1 person evaluates the work done

The evaluator role will be taken by each member exactly once. The order will be left to the group members to decide. The lecturer(s) evaluate both the work done and the evaluation itself. Rubrics will be available showing the precise way in which all evaluations are carried out, all of them publicly available at all times. Additional information as delivery dates, document format, etc will be given at due time.

The final grade will be computed as follows. Let

Ri = evaluation of work 'i' by the lecturer
SEi = evaluation of work 'i' by the student
LEi = evaluation of evaluation 'i' by the lecturer

FGi = final grade of work 'i' = 1/2*(Ri + 10 - |SEi - LEi|)

FS = final grade = [3*FG1 + 3*FG2 + 3*FG3 + SS]/10

where SS is the soft skills grade. (vegi's la guia docent de l'assignatura per més informació).

The topic of each work is to be agreed with the lecturer(s) by each group of students. Many suggestions will be provided along the lectures. That said, individual initiative and open-minded approaches are particularly encouraged. The topics of the works may be different or, alternatively, chained work can deepen successively on the same or closely related topics.

BIBLIOGRAPHY

Basic:

- Wu, X.; Kumar, V. (eds.). The top ten algorithms in data mining. Boca Raton: CRC Press, 2009. ISBN 9781420089646.