Course guides
270615 - ADM - Algorithmics for Data Mining

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: 2021  ECTS Credits: 6.0  Languages: English

LECTURER

Coordinating lecturer: JOSÉ LUIS BALCÁZAR NAVARRO

Others:
Segon quadrimestre:
MARTA ARIAS VICENTE - 10
JOSÉ LUIS BALCÁZAR NAVARRO - 10
JOSEP CARMONA VARGAS - 10
JORGE CASTRO RABAL - 10

PRIOR SKILLS

Thorough understanding of computing in general; good command of several programming languages; basic ability to formalize mathematically issues in informatics engineering.

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.

General:
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. APPROPRIATE 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
Theory sessions, problem solving sessions with or without a programming component, practical sessions with commercial data mining software, development of a case study.

LEARNING OBJECTIVES OF THE SUBJECT
1. To 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

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes</td>
<td>18,0</td>
<td>12.00</td>
</tr>
<tr>
<td>Laboratory classes</td>
<td>36,0</td>
<td>24.00</td>
</tr>
<tr>
<td>Self study</td>
<td>96,0</td>
<td>64.00</td>
</tr>
</tbody>
</table>

Total learning time: 150 h

CONTENTS

Main models and algorithms for Data Mining
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

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

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 4 written documents on topics related to the course. Coauthorship is allowed but the sets of coauthors in each pair of documents must be disjoint. Final grade will be the average of the grades of the documents.

The topic of each document is to be individually agreed with the lecturer by each student. Many suggestions will be provided along the teaching but individual initiative and open-minded approaches are particularly encouraged. The topics of the documents may be different or, alternatively, several documents may deepen successively on the same or closely related topics.

The only condition is that at least one of the documents focuses on the practical usage of data mining algorithms, preferably as implemented in a standard platform, on a dataset of the student’s choice.

In normal times, the last document may be replaced by an oral presentation. In times of remote teaching due to confinement or otherwise, this option will not be available.

BIBLIOGRAPHY

Basic: