Course guides
270957 - ML - Machine Learning

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

Degree: ERASMUS MUNDUS MASTER’S DEGREE IN BIG DATA MANAGEMENT AND ANALYTICS (BDMA) (Syllabus 2021). (Compulsory subject).
MASTER’S DEGREE IN DATA SCIENCE (Syllabus 2021). (Compulsory subject).

Academic year: 2021 ECTS Credits: 6.0 Languages: English

LECTURER

Coordinating lecturer: MARTA ARIAS VICENTE

Others: Segon quadrimestre:
MARTA ARIAS VICENTE - 11
MARIO MARTÍN MUÑOZ - 11
RAQUEL LEANDRA PÉREZ ARNAL - 11

PRIOR SKILLS

Elementary notions of probability and statistics.
Elementary linear algebra and real analysis
Good programming skills in a high-level language

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact

General:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats

Transversal:
CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.
Basic:
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
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.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
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

The course introduces the most important concepts in machine learning and its most relevant techniques with a solid foundation in math. All the theory and concepts are illustrated and accompanied by real-world examples and code using open source libraries.

The theory is introduced in lectures where the teacher exposes the concepts, and during the lab sessions students will see many examples on how to apply the methods and theory learned, as well as code their own solutions to exercises proposed by the teacher.

Students have to work on a course project using a real-world dataset.

LEARNING OBJECTIVES OF THE SUBJECT

1. Formulate the problem of (machine) learning from data, and know the different machine learning tasks, goals and tools.
2. Ability to decide, defend and criticize a solution to a machine learning problem, arguing the strengths and weaknesses of the approach. Additionally, ability to compare, judge and interpret a set of results after making a hypothesis about a machine learning problem
3. To be able to solve concrete machine learning problems with available open-source software

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laboratory classes</td>
<td>24,0</td>
<td>16.00</td>
</tr>
<tr>
<td>Guided activities</td>
<td>6,0</td>
<td>4.00</td>
</tr>
<tr>
<td>Theory classes</td>
<td>24,0</td>
<td>16.00</td>
</tr>
<tr>
<td>Self study</td>
<td>96,0</td>
<td>64.00</td>
</tr>
</tbody>
</table>

Total learning time: 150 h

CONTENTS

<table>
<thead>
<tr>
<th>Introduction to Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
</tr>
<tr>
<td>General information and basic concepts. Overview to the problems tackled by machine learning techniques. Supervised learning (classification and regression), unsupervised learning (clustering and density estimation) and semi-supervised learning (reinforcement and transductive). Examples.</td>
</tr>
<tr>
<td>Topic</td>
</tr>
<tr>
<td>--------------------------------------------</td>
</tr>
<tr>
<td><strong>Kernel functions and support vector machines</strong></td>
</tr>
<tr>
<td><strong>Ensemble methods</strong></td>
</tr>
</tbody>
</table>
# ACTIVITIES

<table>
<thead>
<tr>
<th>Theory lectures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>Theory lectures</td>
</tr>
<tr>
<td><strong>Specific objectives:</strong></td>
</tr>
<tr>
<td>1, 2</td>
</tr>
<tr>
<td><strong>Related competencies:</strong></td>
</tr>
<tr>
<td>CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats</td>
</tr>
<tr>
<td>CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact</td>
</tr>
<tr>
<td>CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats</td>
</tr>
<tr>
<td>CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation</td>
</tr>
<tr>
<td>CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats</td>
</tr>
<tr>
<td>CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem</td>
</tr>
<tr>
<td>CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes</td>
</tr>
<tr>
<td>CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management</td>
</tr>
<tr>
<td>CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market</td>
</tr>
<tr>
<td>CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context</td>
</tr>
<tr>
<td>CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous</td>
</tr>
<tr>
<td>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</td>
</tr>
<tr>
<td>CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments</td>
</tr>
<tr>
<td>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</td>
</tr>
</tbody>
</table>

**Full-or-part-time:** 47h
Theory classes: 27h
Self study: 20h
Lab lectures

Description:
Lab lectures

Specific objectives:
2, 3

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
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.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
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: 54h
Laboratory classes: 27h
Self study: 27h
Mid-term exam (test)

**Description:**
Mid-term exam (test)

**Specific objectives:**
1, 2

**Related competencies:**
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
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.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
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:** 9h 30m
Guided activities: 1h 30m
Self study: 8h
### Final exam

**Description:**
Final exam

**Specific objectives:**
1, 2

**Related competencies:**
- CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
- CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
- CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
- CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
- CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
- CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
- CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
- CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
- CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.

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.

CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

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:** 18h
Guided activities: 2h
Self study: 16h
Course project

Description:
Course project

Specific objectives:
1, 2, 3

Related competencies:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CE7. Identify the limitations imposed by data quality in a data science problem and apply techniques to smooth their impact
CE13. Identify the main threats related to ethics and data privacy in a data science project (both in terms of data management and analysis) and develop and implement appropriate measures to mitigate these threats
CE12. Apply data science in multidisciplinary projects to solve problems in new or poorly explored domains from a data science perspective that are economically viable, socially acceptable, and in accordance with current legislation
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE10. Identify machine learning and statistical modeling methods to use and apply them rigorously in order to solve a specific data science problem
CT1. ENTREPRENEURSHIP AND INNOVATION: Know and understand the organization of a company and the sciences that govern its activity; Have the ability to understand labor standards and the relationships between planning, industrial and commercial strategies, quality and profit. Being aware of and understanding the mechanisms on which scientific research is based, as well as the mechanisms and instruments for transferring results among socio-economic agents involved in research, development and innovation processes.
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
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.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
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: 25h
Self study: 25h

GRADING SYSTEM

The course is graded as follows:

P = Grade of mid-term test-type exam
F = Score of the final exam
L = Score for the practical work

final grade = 20% P + 40% F + 40% L
BIBLIOGRAPHY

Basic:

Complementary:

RESOURCES

Hyperlink: