UAB
Universitat Autònoma de Barcelona

2020/2021

Data Science

Code: 104540
ECTS Credits: 6

Degree	Type	Year	Semester
2503743 Management of Smart and Sustainable Cities	OB	2	2

The proposed teaching and assessment methodology that appear in the guide may be subject to changes as a result of the restrictions to face-to-face class attendance imposed by the health authorities.

Contact

Name: Dimosthenis Karatzas

Email: Dimosthenis.Karatzas@uab.cat

Use of Languages

Principal working language: spanish (spa)

Some groups entirely in English: No

Some groups entirely in Catalan: No

Some groups entirely in Spanish: No

Other comments on languages

Written material of the subject will be prepared in English.

Prerequisites

To have completed the first-year subjects of Computer Science (Informàtica), Mathematics, Internet applications programming, and the second year subject of Databases.

Objectives and Contextualisation

This subject must allow the student to discover the existing technologies and the different ways for managing and analysing the data generated in the city on a daily basis.

Students will learn techniques for visualization, analysis and modelling of data that will allow them to generate new knowledge and intuitions from the city data.

Competences

- Demonstrate creativity, initiative and sensitivity in the different social and environmental topic areas.
- Solve urban management problems using knowledge, methodology and procedures for the design and implementation of computer applications for different types of environment (web, mobile, cloud) and different paradigms.
- Students must be capable of applying their knowledge to their work or vocation in a professional way and they should have building arguments and problem resolution skills within their area of study.
- Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
- Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
- Work cooperatively in complex and uncertain environments and with limited resources in a multidisciplinary context, assuming and respecting the role of the different members of the group.

Learning Outcomes

1. Apply automated decision-making techniques.
2. Demonstrate creativity, initiative and sensitivity in the different social and environmental topic areas.
3. Students must be capable of applying their knowledge to their work or vocation in a professional way and they should have building arguments and problem resolution skills within their area of study.
4. Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
5. Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
6. Work cooperatively in complex and uncertain environments and with limited resources in a multidisciplinary context, assuming and respecting the role of the different members of the group.

Content

- Data preparation
    - Data visualization
    - Normalization
    - Unknown values
    - Reduction of dimensionality
    - Feature selection
- Classification and regression (supervised techniques)
    - Linear and polynomial regression
    - Logistic regression
    - Probabilities, Naive Bayes Classifier
    - Decision trees and "random forests"
    - Hierarchical classification
- Generation of knowledge (unsupervised techniques)
    - Rules of association
    - Recommendation systems

Methodology

Data science is defined by the types of problems that it aims to solve; therefore, it will be that typology of problems that will direct the organization of all the contents.

There will be three types of sessions:

Theory classes: The objective of these sessions is for the teacher to explain the theoretical background of the subject. For each one of the topics studied, the theory and mathematical formulation is explained, as well as the corresponding algorithmic solutions.

Exercise sessions: They will be sessions that facilitate interaction. In these sessions, the aim is to reinforce the comprehension of the topics seen in the theory classes by proposing practical cases that require the design of a solution in which the methods seen in the theory classes are used.

Practical laboratory sessions: They will be sessions in which different types of activities related to the realization of two projects by teams of students will be carried out. During the practical sessions the projects to be solved will be presented and a series of activities will be carried out in teams of students in collaborative work mode. The identification of the problem, the discussion of the design, the distribution and organization of the work to be carried out, the development of the solution and the presentation of the results to the teacher and the rest of the students will be addressed.

All the information of the subject and the related documents that the students need will be found in the virtual campus.

The teacher will give individualized comments for each of the activities delivered by the students. A system of tutoring and consultations outside class hours will be established and students will be encouraged to make use of it.

The transversal competence T01 will be put into practice through teamwork and the collaborative exchange that involves the development of the two projects, which is accompanied by supervised activities in the laboratory of practices. The evaluation of theprojects includes an oral presentation of each team to the rest of the class, during which the students will have to present their work and also explain the organization of the team during the development of the project.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Exercise sessions	12	0.48	1, 5
Theory classes	26	1.04	5, 3, 4
Type: Supervised			
Practical laboratory sessions	12	0.48	1, 5, 3, 4, 6
Tutoring	5	0.2	1, 5, 3, 4
Type: Autonomous			
Dedication to resolve exercises	12	0.48	1, 3, 4
Further reading and study of the material	40	1.6	1
Work on practicals (projects)	37	1.48	1, 3, 4, 6

Assessment

To assess the level of student learning, a formula is established that combines knowledge acquisition, the ability to solve problems and the ability to work as a team, as well as the presentation of the results obtained.

Final note

The final grade is calculated weighted in the following way and according to the different activities that are carried out:

Final grade = 0.4 * Theory Grade + 0.1 * Exercises Grade + 0.5 * Laboratory Grade

This formula will be applied as long as the theory and the laboratory grades, are higher than 5. There is no restriction on the exercises grade. If doing the calculation of the formula yields >= 5 but does not reach the minimum required in any of the evaluation activities, then a final grade of 4.5 will be given.

Theory Grade

The theory grade aims to assess the individual abilities of the student in terms of the theoretical content of the subject, this is done continuously during the course through two partial exams:

Theory Grade = 0.5 * Grade Exam 1 + 0.5 * Grade Exam 2

Exam 1 is done in the middle of the semester and serves to eliminate part of the subject if it is passed.

Exam 2 is done at the end of the semester and serves to eliminate part of the subject if it is passed.

These exams aim to assess the abilities of each student in an individualized manner, both in terms of solving exercises using the techniques explained in class, as well as evaluating the level of conceptualization that the student has made of the techniques seen. In order to obtain a final pass theory grade, it will be required for the partial exam grades 1 and 2 to be both higher than 4.

Recovery exam. In case the theory grade does not reach the adequate level to pass, the students can take a recovery exam, destined to recover the failed part (1, 2 or both) of the continuous evaluation process.

Exercises Grade

The aim of the exercises is for the student to train with the contents of the subject continuously and become familiar with the application of the theoretical concepts. As evidence of this work, the presentation of a portfolio is requested in which the exercises worked out will be collated:

Exercises Grade = Portfolio evaluation

Practical Laboratory Sessions Grade

The part of laboratory based practical sessions carries an essential weight in the overall mark of the subject. Laboratory sessions aim for the student to design a solution to a problem that is set out in a contextualized way. Such problems will require the design of an integral solution, from the exploration of available techniques to data modelling. In addition, the students must demonstrate their teamwork skills and present the results to the class convincingly.

Laboratory sessions are structured around two projects. Each of the two projects is evaluated through its deliverable, an oral presentation that students will make in class, and an self-evaluation process. The grade is calculated as follows:

Project Grade = 0.5 * Grade Deliverables + 0.3 * Grade Presentation + 0.2 * Grade Self-evaluation

Laboratory Grade = 0.5 * Grade Project 1 + 0.5 * Grade Project 2

In case of not passing any of the projects, the recovery of the part of the deliverables of the unsuccessful projects will be allowed, restricted to a maximum grade of 7/10. The oral presentation cannot be recovered.

Important notes

Notwithstanding other disciplinary measures deemed appropriate, and in accordance with the academic regulations in force, evaluation activities will be suspended with zero (0) whenever a student commits any academic irregularities that may alter such evaluation (for example, plagiarizing, copying, letting copy, ...). The evaluation activities qualified in this way and by this procedure will not be recoverable. If you need to pass any of these assessment activities to pass the subject, this subject will be failed directly, without opportunity to recover it in the same course.

In case there the studentdoes not deliver any exercise solutions, does not attend any project presentation session during the laboratory sessions and does not take any exam, the corresponding grade will be a "non-evaluable". In another case, the "no shows" count as a 0 for the calculation of the weighted average.

In order to pass the course with honours, the final grade obtained must be equal or higher than 9 points. Because the number of students with this distinction cannot exceed 5% of the total number of students enrolled in the course, it is given to whoever has the highest final marks. In case of a tie, the results of the partial exams will be taken into account.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Exams	40	5	0.2	5, 3
Exercises deliverables	10	0	0	1, 4
Project deliverables	25	0	0	1, 2, 3, 4, 6
Project presentations	15	1	0.04	5, 3, 6
Self-evaluation	10	0	0	6

Bibliography

- Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media, 2015, 1st Ed.
- Python Data Science Handbook, Jake VanderPlas, O'Reilly Media, 2016, 1st Ed.
- Pattern Recognition and Machine Learning, Christopher Bishop, Springer, 2011
- Model-Based Machine Learning, J. Winn, C. Bishop, early access: http://mbmlbook.com/
- Computational and Inferential Thinking: The Foundations of Data Science, Ani Adhikari and John DeNero, online: https://ds8.gitbooks.io/textbook/content/