IDENTIFYING DATA

Data analysis

Subject	Data analysis
Code	V05M145V01322
Study programme	Máster Universitario en Ingeniería de Telecomunicación

Descriptors	ECTS Credits	Choose	Year	Quadmester
	5	Optional	2nd	1st

Teaching language	#EnglishFriendly Spanish
Department	
Coordinator	González Castaño, Francisco Javier
Lecturers	García Méndez, Silvia González Castaño, Francisco Javier
E-mail	javier@det.uvigo.es
Web	http://moovi.uvigo.gal/
General description	Data analysis with a practical approach: data extraction and cleansing, data characterization with techniques such as statistical regression, clustering or outlier analysis, and knowledge generation with techniques such as intuitive visualization or automatic classification. The course is taught in Spanish. English Friendly subject: International students may request from the teachers: a) resources and bibliographic references in English, b) tutoring sessions in English, c) exams and assessments in English.

Training and Learning Results

Code	
A2	CB2 Students must apply their knowledge and ability to solve problems in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their field of study.
A3	CB3 Students must integrate knowledge and handle complexity of formulating judgments based on information that was incomplete or limited, including reflections on social and ethical responsibilities linked to the application of their knowledge and judgments.
B4	CG4 Capacity for mathematical modeling, calculation and simulation in technological centers and engineering companies, particularly in research, development and innovation tasks in all areas related to Telecommunication Engineering and associated multidisciplinary fields.
B8	CG8 Ability to apply acquired knowledge and to solve problems in new or unfamiliar environments within broader and multidiscipline contexts, being able to integrate knowledge.
C25	CE25/TE2 Ability to manage the acquisition, structuring, analysis and visualization of data, extracting information and underlying knowledge, critically assessing the results, and applying it to strategic decision-making and innovation in different areas.

Expected results from this subject

Expected results from this subject	Training and Learning Results
- Knowledge of the different stages of knowledge extraction and the areas of application of data mining.	A2 A3 B4 B8 C25
- Knowledge of the importance of the preparation of the data and how to apply the main pre-processing techniques.	A2 B4 B8 C25

- Knowledge of the main techniques of data mining as well as the necessary premises for its application to a particular stage.	A2 A3 B4 B8
- Knowldge of the different types of data mining results evaluation and how to apply them.	C25
- Knowledge of statistical software and how to apply it to on-line and off-line data mining.	B4 C25
-Ability to to schedule, develop and evaluate a data analysis process.	B4 B8 C25

Contents

Topic	
Statistical analysis of data	- Correlation and causation. - Regressions. - Intervals of confidence and error. Hypothesis tests.
Data mining	- Cleaning, integration, reduction and transformation of data. - Classification and clustering.
Computational analysis of data	- Large-scale data analysis. - Visualisation of data and results. - Application scenarios.

Planning

	Class hours	Hours outside the classroom	Total hours
Project based learning	2	36	38
Laboratory practical	5	19	24
Lecturing	20	40	60
Problem and/or exercise solving	2	0	2
Essay	1	0	1
Essay	1	0	1

*The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

Methodologies

	Description
Project based learning	The students will solve a practical case of data analysis in an application scenario. A2 A3 B4 B8 C25.
Laboratory practical	During the course, students will develop solutions in laboratory sessions to grasp the course content. Software to be used: R (https://www.r-project.org/). A2 A3 B4 B8 C25.
Lecturing	Lectures that will illustrate the course content with small exercises. These will be solved by the lecturer of the students themselves, alone or in groups. The goal is to foster discussion and knowledge of course competencies. A2 A3 B4 B8.

Personalized assistance

Methodologies	Description
Lecturing	Individual atention will take place during official tutoring times published at https://atlanttic.uvigo.es/es/equipo/staff/francisco-javier-gonzalez-castano/ or via e-mail at any time.
Project based learning	Individual atention will take place during official tutoring times published at https://atlanttic.uvigo.es/es/equipo/staff/francisco-javier-gonzalez-castano/ or via e-mail at any time.
Laboratory practical	Individual atention will take place during official tutoring times published at https://atlanttic.uvigo.es/es/equipo/staff/francisco-javier-gonzalez-castano/ or via e-mail at any time.

Assessment

	Description	Qualification	Training and Learning Results		
Problem and/or exercise solving	Short-answer written exam.	40			C25
Essay	Deliverable reporting work on a dataset that will be handed at the beginning on the course.	30	A2 A3	B4 B8	C25
Essay	Deliverable reporting work on a dataset that will be handed at the beginning on the course.	30	A2 A3	B4 B8	C25

Other comments on the Evaluation

ORDINARY OPPORTUNITY

At the beginning of the course the student will have to choose between continuous and global evaluation. No change of decision will be allowed.

Continuous evaluation will consist in the following::

1. Short answer test (4 points maximum).
2. Two deliverables of the work on a common dataset (6 points maximum, 3 points each)

To pass the course, the student must obtain 1,5/4 points at least in the short answer test and an overal score (across all possible activities) above 5 points. Failure to reach the minimum grade in the short answer test limits the maximum achievable grade to 4. The maximum score is 10 points.

The contents of the short answer test and the deliverables will be balanced for a reasonable preparation effort.

Global evaluation will consist on a single exam covering the whole theoretical and practical course content (the maximum score of this exam will be 5 points. A minimum score of 2 is necessary to pass the course) and a deliverable based on a dataset selected by the professor (maximum score of 5 points). The minimum score to pass the course is 5 points overall. Failure to reach the minimum grade in the exam limits the maximum achievable grade to 4. The maximum score is 10 points.

EXTRAORDINARY OPPORTUNITY

The only possibility will be global evaluation, as previously described.

Sources of information
Basic Bibliography
Complementary Bibliography
Zummel, N., Mount, J., **Practical Data Science with R**, ISBN 9781617291562, Manning Publications,
James, G., Witten, D., Hastie, T., Tibshirani, R., **An Introduction to Statistical Learning with Applications in R**, ISBN 9781461471387, Springer,

Recommendations