



## Teaching Guide

Identifying Data				2022/23
Subject (*)	Data Mining	Code	710G04030	
Study programme	Grao en Xestión Dixital de Información e Documentación			
Descriptors				
Cycle	Period	Year	Type	Credits
Graduate	2nd four-month period	Third	Optional	6
Language	SpanishGalician			
Teaching method	Face-to-face			
Prerequisites				
Department	Matemáticas			
Coordinador	Navarro Burgos, Miguel Ángel	E-mail	miguel.navarro.burgos@udc.es	
Lecturers	Navarro Burgos, Miguel Ángel	E-mail	miguel.navarro.burgos@udc.es	
Web				
General description	The main objective of this course is that the student knows the fundamental concepts and the main models of data mining, and its application in the field of information science and documentation.			

## Study programme competences

Code	Study programme competences
A1	CE1 - Know and understand the theoretical and methodological principles of information and documentation management to apply them in their professional activity
A4	CE4 - Master the foundations of the behavior of individuals in the search, recovery and use of information, taking as a point of reference the aspects of motivation, environment and context
A8	CE8 - Master the different methods of representation of data, information and knowledge that ensure efficient recovery
A13	CE13 - Know and master the techniques and regulations for the creation and authentication, meeting, selection, organization, representation, preservation, recovery, access, dissemination and exchange, and evaluation of resources and information services
A20	CE20 - Master the bases to develop research activities using multidisciplinary methods and principles
A21	CE21 - Possess knowledge of statistics and quantitative analysis of information
A22	CE22 - Acquire computational skills and management of new ICT
B1	CB1 - 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
B2	CB2 - Apply the knowledge acquired and their ability to solve problems in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their area of study
B3	CB3 - Be able to integrate knowledge and face the complexity of making judgments based on information that, being incomplete or limited, includes reflections on social and ethical responsibilities linked to the application of their knowledge and judgments
B4	CB4 - Know how to communicate their conclusions -and the knowledge and ultimate reasons that sustain them- to specialized and non-specialized audiences in a clear and unambiguous way
B5	CB5 - Possess the learning skills that allow them to continue studying in a way that will be largely self-directed or autonomous
B6	CG1 - Capacity for cooperation, teamwork and collaborative learning
B7	CG2 - Capacity for reflection and critical reasoning
B8	CG3 - Capacity for planning, organization and management of resources, information and operations
B9	CG4 - Capacity for analysis, diagnosis and decision making
B10	CG5 - Ability to work in an international and global context
B11	CG6 - Ability to understand the importance, value and function of the Digital Information and Documentation Management in the current ICT society
C1	CT1 - Express correctly, both orally and in writing, in the official languages ??of the autonomous community
C2	CT2 - Use the basic tools of information and communication technologies (ICT) necessary for the exercise of their profession and for learning throughout their lives
C3	CT3 - Develop oneself for the exercise of a citizenship that respects democratic culture, human rights and the gender perspective
C4	CT4 - Understand the importance of the entrepreneurial culture and know the means available to entrepreneurs



C5	CT5 - Acquire skills for life and habits, routines and healthy lifestyles
C6	CT6 - Develop the ability to work in interdisciplinary or transdisciplinary teams, to offer proposals that contribute to a sustainable environmental, economic, political and social development
C7	CT7 - Assess the importance of research, innovation and technological development in the socio-economic and cultural progress of society
C8	CT8 - Have the ability to manage time and resources: develop plans, prioritize activities, identify criticisms, establish deadlines and comply with them

Learning outcomes			
Learning outcomes	Study programme competences		
Acquisition of skills for the selection, treatment, management and analysis of information through data mining techniques.	A8 A13 A20 A21 A22	B2 B4 B5 B6 B7 B8 B10 B11	C2 C4 C5 C6 C8
Choosing the quantitative techniques appropriate to the objectives for research, administration and management tasks.	A1 A8 A13 A20 A21 A22	B2 B4 B5 B6 B7 B8 B10 B11	C2 C4 C6 C8
Acquire knowledge of computational data analysis, including programs such as R statistical software.	A1 A4 A8 A13 A20 A21 A22	B1 B2 B3 B9 B11	C1 C3 C7
Knowledge and skills for the application of the main classification techniques.	A1 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C2 C4 C6 C7 C8



Knowledge and skills for the application of regression techniques, anomaly detection and time series.	A1 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C2 C4 C6 C7 C8
Capacity for analysis and synthesis applied to the management and organization of information.	A1 A4 A8 A13 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C1 C2 C3 C4 C5 C6 C7 C8
Acquisition of decision-making skills based on statistical analysis of complex databases.	A1 A4 A8 A13 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C1 C2 C3 C4 C5 C6 C7 C8

Contents	
Topic	Sub-topic
Introduction to data mining.	Preliminary concepts. Types of data mining problems: description, classification, prediction, clustering, anomaly detection, etc. Types of learning: supervised and unsupervised.
Unsupervised classification or clustering methods.	Basic concepts. Hierarchical classification methods. Partitioning clustering methods. Case studies in information science and documentation.
Supervised classification methods.	Basic concepts. Main models of supervised classification or pattern recognition. Validation of classification models (how well do they predict?). Case studies in information science and documentation.



Advanced regression methods.	Introduction. Univariate and multivariate regression models. Selection of relevant variables. Validation of regression models (how well does it fit the data, how well does it make predictions). Case studies in information science and documentation.
Time series	Basic concepts. Descriptive time series analysis. Practical use of time series models. Case studies.
Statistical techniques for text mining and information retrieval.	Basic concepts. Practical cases of application of text mining.

Planning				
Methodologies / tests	Competencies	Ordinary class hours	Student's personal work hours	Total hours
Guest lecture / keynote speech	A1 A4 A8 A20 A21 B1 B3 B7 C4 C7	19	0	19
ICT practicals	A13 A22 B11 C2	13	0	13
Case study	A1 A8 A21 B2 B3 B4 B5 B6 B7 B8 B9 C1 C8	7	7	14
Supervised projects	B2 B4 B5 B6 B8 B9 B10 C1 C3 C5 C6 C8	1	101	102
Objective test	A21 B1 B2	1	0	1
Personalized attention		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	
Methodologies	Description
Guest lecture / keynote speech	They will be expository sessions in which the various topics of the subject will be introduced and described through presentations (using appropriate audiovisual media) that will include theory and examples.
ICT practicals	Practical classes will be developed using statistical software, in which its programming and application will be introduced based on real and simulated cases.
Case study	The statistical techniques taught in the subject will be applied to solve exercises and real and simulated case studies in the field of digital information management.
Supervised projects	Individual and/or group work will be carried out, supervised by the teachers of the subject, in which the resolution, by means of the application of statistical techniques and R software, of practical exercises or particular case studies related to the field of communication and information sciences will be approached. A review study of a specific topic of the subject or of the software used may also be carried out. The works may be proposed by the teachers or by the students themselves (the proposals will be taken into account or not always according to the teacher's criteria).
Objective test	It will consist of a multiple-choice test on the contents taught in the subject.

Personalized attention	
Methodologies	Description



ICT practicals Case study Guest lecture / keynote speech Supervised projects	In the master classes, discussion among the students and between the students and the teacher will be encouraged at all times. For the resolution of problems it will be important to attend personally to the students in case of possible doubts that may arise. This attention will also serve, on the one hand, to the teacher to detect possible problems in the methodology used to teach the subject and, on the other hand, to the students to consolidate theoretical knowledge and to express their concerns about the subject. It will also be essential to give personalized attention to the student during the ICT practical classes, especially until he/she becomes familiar with the statistical software to be used, as well as in the resolution of case studies.
--	--

Assessment			
Methodologies	Competencies	Description	Qualification
ICT practicals	A13 A22 B11 C2	Attendance and/or student performance with the statistical software will be assessed.	20
Objective test	A21 B1 B2	A multiple-choice test consisting of a number of questions between 10 and 20 with 3 possible answers.	40
Supervised projects	B2 B4 B5 B6 B8 B9 B10 C1 C3 C5 C6 C8	Individual and/or group work will be carried out, supervised by the teachers of the subject, in which the resolution, by means of the application of statistical techniques and R software, of practical exercises or particular case studies related to the field of communication and information sciences will be approached. A review study of a specific topic of the subject or the software used may also be carried out. The works may be proposed by the teachers or by the students themselves (the proposals will be taken into account or not always according to the teacher's criteria).	40

Assessment comments
---------------------



## First opportunity

There will be a multiple-choice test of 10 to 20 questions that represents 40% of the grade. On the other hand, the continuous evaluation will consist of the attendance and/or delivery of practices related to the learning and application of the statistical software R for the resolution of problems in the field of digital information management (20% of the overall grade), in addition to the delivery of one or several works of application of statistics for the resolution of case studies in digital documentation (alternatively they can be works of revision or extension of the subject) that represents 40% of the total grade.

## Second opportunity

In the evaluation of the second opportunity, the same criteria corresponding to the first opportunity will be followed.

## Advanced call

All the previous observations are applicable to students who request the advanced call of the exam.

## No-show grade

In any of the two annual opportunities a NON PRESENTADO will appear in those cases in which the student does not attend the official exam of the subject.

Student with recognition of part-time dedication and academic dispensation of exemption from attendance.

In the case of students with recognition of part-time dedication and academic dispensation of exemption from attendance who decide not to attend classes, they will be evaluated in both opportunities as the rest of the students who are in a similar situation.

The fraudulent performance of the tests or evaluation activities will directly imply the grade of "0" in the subject in the corresponding call, thus invalidating any grade obtained in all evaluation activities for the extraordinary call.

## Sources of information

<b>Basic</b>	<ul style="list-style-type: none"> <li>- Williams, G. (2011). Data mining with Rattle and R: The art of excavating data for knowledge discovery. . Springer Science &amp; Business Media.</li> <li>- Cirillo, A. (2017). R Data Mining: Implement Data Mining Techniques Through Practical Use Cases and Real-world Datasets. . Packt Publishing.</li> <li>- Silge, J. y Robinson, D. (2017). Text Mining with R: A Tidy Approach. O'Reilly</li> <li>- Jockers, M.L. (2014). Text Analysis with R for Students of Literature. Springer</li> </ul>
<b>Complementary</b>	

## Recommendations

### Subjects that it is recommended to have taken before

Data Science/710G04026  
Fundamentals of Statistics/710G04040

### Subjects that are recommended to be taken simultaneously

### Subjects that continue the syllabus

### Other comments



To help to achieve a sustainable environment and meet the objective of action number 5: "Healthy and sustainable environmental and social teaching and research" of the "Green Campus Ferrol Action Plan":

- 1.- The delivery of the documentary works carried out in this subject:
  - 1.1. It will be requested in virtual format and/or computer support.
  - 1.2. It will be done through Moodle, in digital format without the need to print them.
  - 1.3. If done on paper:
    - Plastics will not be used.
    - Double-sided prints will be made.
    - Recycled paper will be used.
    - Draft printing will be avoided.
- 2.- A sustainable use of resources and the prevention of negative impacts on the natural environment must be made.
- 3.- The importance of ethical principles related to the values of sustainability in personal and professional behavior must be taken into account.
- 4.- As it is included in the different regulations of application for university teaching, the gender perspective must be incorporated in this subject (non-sexist language will be used, bibliography of authors of both sexes will be used, intervention in student class will be encouraged and students...).
- 5.- We will work to identify and modify prejudices and sexist attitudes, and the environment will be influenced to modify them and promote values of respect and equality.
6. Situations of discrimination based on gender must be detected and actions and measures will be proposed to correct them.
7. The full integration of students who, due to physical, sensorial, psychic or sociocultural reasons, experience difficulties in an ideal, egalitarian and profitable access to university life will be facilitated.

(\*The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.