Teaching Guide				
Identifying Data				2020/21
Subject (*)	Computational intelligence for high dimensional data		Code	614522024
Study programme	Mestrado Universitario en Bioinformática para Ciencias da Saúde			
Descriptors				
Cycle	Period	Year	Type	Credits
Official Master's Degree	1st four-month period	Second	Obligatory	3
Language	SpanishEnglish			
Teaching method	Face-to-face			
Prerequisites				
Department	Ciencias da Computación e Tecnoloxías da InformaciónComputación			
Coordinador	Eiras Franco, Carlos	E-mail	carlos.eiras.franco@udc.es	
Lecturers	Eiras Franco, Carlos	E-mail	carlos.eiras.franco@udc.es	
Web	moodle.udc.es			
General description	We will work on the concepts and practical application of high-dimensional databases and on the application of data mining techniques in the bioinformatics.			
Contingency plan	1. Modifications to the contents 2. Methodologies *Teaching methodologies that are maintained *Teaching methodologies that are modified 3. Mechanisms for personalized attention to students 4. Modifications in the evaluation *Evaluation observations: 5. Modifications to the bibliography or webgraphy			

Study programme competences	
Code	Study programme competences
A2	CE2 ? To define, evaluate and select the architecture and the most suitable software for solving a problem in the field of bioinformatics
A3	CE3 ? To analyze, design, develop, implement, verify and document efficient software solutions based on an adequate knowledge of the theories, models and techniques in the field of Bioinformatics
A4	CE4 - Ability to acquire, obtain, formalize and represent human knowledge in a computable form for the resolution of problems through a computer system in any field of application, particularly those related to aspects of computing, perception and action in bioinformatics applications
A6	CE6 - Ability to identify software tools and most relevant bioinformatics data sources, and acquire skill in their use
B1	CB6 - Own and understand knowledge that can provide a base or opportunity to be original in the development and/or application of ideas, often in a context of research
B2	CB7 - Students should know how to apply the acquired knowledge and ability to problem solving in new environments or little known within broad (or multidisciplinary) contexts related to their field of study
B3	CB8 - Students to be able to integrate knowledge and deal with the complexity of making judgements from information that could be incomplete or limited, including reflections on the social and ethical responsibilities linked to the application of their skills and judgments
B6	CG1 -Search for and select the useful information needed to solve complex problems, driving fluently bibliographical sources for the field

B7	CG2 - Maintain and extend well-founded theoretical approaches to enable the introduction and exploitation of new and advanced technologies
C1	CT1 - Express oneself correctly, both orally writing, in the official languages of the autonomous community
C3	CT3 - Use the basic tools of the information technology and communications (ICT) necessary for the exercise of their profession and lifelong learning
C6	CT6 - To assess critically the knowledge, technology and information available to solve the problems they face to.

Learning outcomes			
Learning outcomes	Study programme competences		
Coñecer e comprender os paradigmas e aspectos máis relevantes do tratamento de bases de datos de alta dimensión	AJ2 AJ3 AJ4 AJ6	BJ1 BJ2 BJ3 BJ6 BJ7	CJ1 CJ3 CJ6
Coñecer e saber aplicar os principais métodos de minería de datos; coñecer as plataformas e as paradigmas principais que se empregan no campo.	AJ2 AJ3 AJ4 AJ6	BJ1 BJ2 BJ3 BJ6 BJ7	CJ1 CJ3 CJ6

Contents	
Topic	Sub-topic
Introducción ao Big data.	Qué é Big Data Principais características do Big data Principais campos de aplicación
Minería de datos e alta dimensión	Analítica Big data Técnicas de preprocesado MapReduce
Modelos de programación Batch	Hadoop Resilient Distributed datasets Programación batch en Spark
Modelos de programación streaming	Conceptos básicos Kafka, Apache Storm, Spark streaming

Planning				
Methodologies / tests	Competencies	Ordinary class hours	Student?s personal work hours	Total hours
Guest lecture / keynote speech	A4 C1 C6	7	14	21
Problem solving	A25 A33 A41 B1 B6 C3	8	16	24
Supervised projects	A21 B3 B6 C1 C2 C3 C6	4	4	8
Seminar	A21 B1 B3 B6	4	4	8
Mixed objective/subjective test	A2 A3 A4 A6 B1 B2 B3 B6 B7 C1 C3 C6	4	10	14
Personalized attention		0	0	0
(*)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	Empregada durante as clases presenciais teóricas para expor o núcleo básico de coñecementos que logo os alumnos terán que saber utilizar e ampliar nas prácticas, seminarios e traballos do curso
Problem solving	Emprego de técnicas de minería de datos en alta dimensión. Uso de paradigmas Big data Realización dunha práctica nunha plataforma específica de Big data
Supervised projects	Entrega dun breve traballo que discutirase na clase sobre algún aspecto concreto da materia.
Seminar	Exposición dun traballo específico de investigación que involucre tecnoloxías de alta dimensionalidade
Mixed objective/subjective test	Realizarase ao final do cuadrimestre sobre os contidos tratados ao longo do curso.

Personalized attention	
Methodologies	**Description**
Seminar Problem solving Supervised projects Mixed objective/subjective test Guest lecture / keynote speech	No esquema de carácter práctico utilizado nesta materia, as tutorías resultan un recurso fundamental moi empregado polos alumnos, sobre todo debido á complexidade dalgúns conceptos da materia, en función das titulacións de entrada dos diferentes alumnos. Os alumnos poden realizar dous tipos de tutorías: virtuais e presenciais. As primeiras poden utilizalas para dúbidas moi concretas de resposta rápida. As máis comúns iranse depositando nun apartado de %"Preguntas Frecuentes" que deberán consultar antes de enviar unha nova pregunta. .

Assessment			
Methodologies	**Competencies**	**Description**	**Qualification**
Seminar	A21 B1 B3 B6	Seminarios de temas específicos	0
Supervised projects	A21 B3 B6 C1 C2 C3 C6	Nota correspondente á parte práctica da materia, que comprende tanto os desenvolvementos realizados sobre as plataformas, como os traballos entregados.	50
Mixed objective/subjective test	A2 A3 A4 A6 B1 B2 B3 B6 B7 C1 C3 C6	Realizarase unha proba con cuestións relativas ás partes teóricas da materia	50
Guest lecture / keynote speech	A4 C1 C6	Clases presenciais	0

Assessment comments

Sources of information	
Basic	- Vladimir Bacvanski. (2015). Introduction to Big Data An Overview of Fundamental Big Data Concepts, Tools, Techniques and Practices.. O'Reilly Media - Venkat Ankam (2016.). Big Data Analytics. Packt Publishing - Tom White (2015). Hadoop: The Definitive Guide. O'Reilly Media - Thilina Gunarathne (2015). Hadoop MapReduce v2 Cookbook. Packt Publishing - Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia (2015). Learning Spark. O'Reilly Media - Sean T. Allen, Matthew Jankowski, and Peter Pathirana (2015). Storm Applied. . O'Reilly Media
Complementary	

Recommendations

Subjects that it is recommended to have taken before
Computational intelligence for bioinformatics/614522012 Advanced statistical methods in bioinformatics/614522009 High performance computing in bioinformatics/614522011 Introduction to programming/614522001 Foundations of Artificial Intelligence/614522003
Subjects that are recommended to be taken simultaneously
Subjects that continue the syllabus
Other comments

(*)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.