Teaching Guide				
Identifying Data				2023/24
Subject (*)	Machine Learning I		Code	614G02019
Study programme	Grao en Ciencia e Enxeñaría de Datos			
Descriptors				
Cycle	Period	Year	Type	Credits
Graduate	2nd four-month period	Second	Obligatory	6
Language	Spanish			
Teaching method	Face-to-face			
Prerequisites				
Department	Ciencias da Computación e Tecnoloxías da Información			
Coordinador	Rivero Cebrián, Daniel	E-mail	daniel.rivero@udc.es	
Lecturers	Molares Ulloa, Andrés Rabuñal Dopico, Juan Ramon Rivero Cebrián, Daniel Rodríguez Tajes, Álvaro	E-mail	andres.molares@udc.es juan.rabunal@udc.es daniel.rivero@udc.es a.tajes@udc.es	
Web				
General description	This course presents an overview of Machine Learning. The syllabus explains the different techniques and methods, including supervised and unsupervised learning. In the practical part, real cases will be solved.			

Study programme competences	
Code	Study programme competences
A24	CE24 - Comprensión e dominio dás principais técnicas básicas e avanzadas de aprendizaxe automática, incluíndo as dedicadas ao tratamento de grandes volumes de datos.
A25	CE25 - Capacidade para identificar a adecuación de cada unha das técnicas de aprendizaxe automática á resolución dun problema, incluíndo os aspectos relacionados coa súa complexidade computacional ou a súa capacidade explicativa, de acordo aos requisitos establecidos.
A26	CE26 - Coñecemento das ferramentas informáticas actuais no campo da aprendizaxe automática, e capacidade para seleccionar a máis adecuada para a resolución dun problema.
B2	CB2 - Que os estudantes saiban aplicar os seus coñecementos ao seu traballo ou vocación dunha forma profesional e posúan as competencias que adoitan demostrarse por medio da elaboración e defensa de argumentos e a resolución de problemas dentro da súa área de estudo
B3	CB3 - Que os estudantes teñan a capacidade de reunir e interpretar datos relevantes (normalmente dentro da súa área de estudo) para emitir xuízos que inclúan unha reflexión sobre temas relevantes de índole social, científica ou ética
B7	CG2 - Elaborar adecuadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables.
B8	CG3 - Ser capaz de manter e estender formulacións teóricas fundadas para permitir a introdución e explotación de tecnoloxías novas e avanzadas no campo.
B9	CG4 - Capacidade para abordar con éxito todas as etapas dun proxecto de datos: exploración previa dos datos, preprocesado, análise, visualización e comunicación de resultados.
B10	CG5 - Ser capaz de traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións.
C1	CT1 - Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida.

Learning outcomes			
Learning outcomes	Study programme competences		
Understand the relationship between the complexity of learning models, training data features and overfitting, and know the mechanisms to avoid it.	A24 A25		

Develop skills to design the stages of a complete data analysis process based on automatic learning techniques.		B2 B7 B9 B10	C1
Know how to correctly apply automatic learning techniques to obtain reliable and significant results.	A24	B3 B8	
Know the most representative and current techniques of unsupervised, semi-supervised and supervised learning.	A24	B8	
Know the most representative learning techniques for the classic problems of classification, regression and clustering, and other less classic ones such as sorting problems, one class problems or multitasking.	A24	B8	
Identify appropriate data analysis techniques according to the problem.	A25	B3 B8	
Manage the most current tools and work environments in the field of machine learning.	A26	B2 B10	

Contents	
Topic	**Sub-topic**
1. Introduction	1.1. Introduction to Machine Learning 1.2. Learning Paradigms 1.3. Inductive Learning 1.4. No free Lunch Theorems
2. Supervised learning	2.1. Introduction 2.2. Artificial Neural Networks 2.3. Support Vector Machines 2.4. Decision trees 2.5. Regression trees and regression model trees 2.6. Instance-based learning
3. Evolutionary Computation	3.1. Genetic Algorithms 3.2. Genetic Programming 3.3. Swarms and other Evolutionary Computation techniques
4. Methodologies in data analysis	4.1. Training, evaluation and model selection methodologies 4.2. Methodologies of a data analysis project
5. Unsupervised learning	5.1. Clustering methods 5.2. Self-organised networks

Planning				
Methodologies / tests	**Competencies**	**Ordinary class hours**	**Student?s personal work hours**	**Total hours**
Guest lecture / keynote speech	A24 A25 B3 B8 B9	30	38	68
Laboratory practice	A26 B2 B3 B10 C1	15	24	39
Supervised projects	B2 B3 B7 B9 B10	15	24	39
Objective test	A24 A25 B8 B9	2	0	2
Personalized attention		2	0	2
(*)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	Theoretical teaching of the subject matter of the course
Laboratory practice	Solve practical problems by using the different techniques that will be explained in the theory classes.

Supervised projects	Writing, under the supervision of the teacher, of the reports explaining the resolution of the problems carried out in the laboratory practices and the results obtained.
Objective test	This is a written assessment test in which the student must demonstrate the knowledge acquired from the subject.

Personalized attention	
Methodologies	**Description**
Laboratory practice Supervised projects	Practical work carried out with the advice of the teacher. Writing of the explanatory report under the teacher's supervision.

Assessment			
Methodologies	**Competencies**	**Description**	**Qualification**
Laboratory practice	A26 B2 B3 B10 C1	Development of a Machine Learning system based on explanations made in theory.	25
Supervised projects	B2 B3 B7 B9 B10	Writing of the report on the resolution of a real problem carried out in the laboratory practices. The writing of the report will include a bibliographic review of the most important works related, written in English for the most part, documentation on the problem to be solved, methodology used, and comparison of the results found in the application of the different techniques, as well as a critical evaluation of both the results obtained and the information used.	25
Objective test	A24 A25 B8 B9	Test questions about the contents of the course, based on the different machine learning techniques and their applications.	50

Assessment comments
In order to pass the subject, the student must obtain a minimum score of 5 out of 10 in the result of combining the grades of the objective test, the laboratory practices and the supervised works. In addition, the student must obtain a minimum score of 2 out of 5 points in the objective test. If the student does not obtain this minimum grade, the grade of the subject will be that corresponding to the grade of the objective test. In the second opportunity, the grade obtained in the laboratory practices and supervised works will be maintained, not being able to obtain again a grade since it results from the continuous evaluation of the work during the credits of practice of the subject. The student can retake the examination of the objective test, the criteria for obtaining the total score being those indicated at the beginning of this section. Part-time students must deliver in their reports on the same dates as full-time students, and attend the RGTs in which they will be corrected. Similarly, it is recommended that they attend the practice classes. No-show qualification: The student will receive the qualification of "no-show" when he/she does not take the final exam. Fraudulent performance of exercises or tests: For cases of fraudulent performance of exercises or tests, the provisions of the current regulation of the UDC about this topic will apply.

Sources of information

Basic	- D. Borrajo, J. González, P. Isasi (2006). Aprendizaje automático. Sanz y Torres - T.M. Mitchell (1997). Machine Learning. McGraw Hill - Basilio Sierra Araujo (2006). Aprendizaje automático: conceptos básicos y avanzados. Aspectos prácticos utilizando el software WEKA. Pearson Education - Saso Dzeroski, Nada Lavrac (). Relational Data Mining. Springer - David Aha (). Lazy Learning. Kluwer Academics Publishers - Andrew Webb (2002). Statistical Pattern Recognition. Wiley - Ethem Alpaydin (2004). Introduction to Machine Learning. MIT Press - Marcos Gestal, Daniel Rivero, Juan Ramón Rabuñal, Julián Dorado, Alejandro Pazos (2010). Introducción a los Algoritmos Genéticos y a la Programación Genética. Servicio de Publicaciones de la Universidade da Coruña
Complementary	

Recommendations
Subjects that it is recommended to have taken before
Design and Analysis of Algorithms/614G02011 Regression Models/614G02012 Statistical Modeling of High Dimensional Data/614G02013 Signals and Systems/614G02014 Fundamentals of Programming II/614G02009 Fundamentals of Programming I/614G02004 Statistical Inference/614G02007
Subjects that are recommended to be taken simultaneously
Information Theory/614G02018 Mathematical Optimisation/614G02020
Subjects that continue the syllabus
Large Scale Machine Learning/614G02032 Numerical Methods for Data Science/614G02033 Machine Learning III/614G02026 Image, Video and Audio Processing/614G02028 Machine Learning II/614G02021
Other comments