



POLITÉCNICA

INTERNATIONAL
CAMPUS OF
EXCELLENCE

COORDINATION PROCESS OF
LEARNING ACTIVITIES
PR/CL/001



E.T.S. de Ingenieros
Informáticos

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000541 - Data engineering

DEGREE PROGRAMME

10AM - Master Universitario en Ingeniería del Software

ACADEMIC YEAR & SEMESTER

2017/18 - Semester 1

Index

Learning guide

1. Description.....	1
2. Faculty.....	1
3. Prior knowledge recommended to take the subject.....	2
4. Skills and learning outcomes	2
5. Brief description of the subject and syllabus.....	4
6. Schedule.....	5
7. Activities and assessment criteria.....	7
8. Teaching resources.....	10

1. Description

1.1. Subject details

Name of the subject	103000541 - Data engineering
No of credits	4 ECTS
Type	Optional
Academic year of the programme	First year
Semester of tuition	Semester 1
Tuition period	September-January
Tuition languages	English
Degree programme	10AM - Master Universitario en Ingenieria del Software
Centre	Escuela Tecnica Superior de Ingenieros Informaticos
Academic year	2017-18

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Fco.javier Segovia Perez (Subject coordinator)	2305	javier.segovia@upm.es	M - 10:00 - 11:00 Hablar con el profesor
Ernestina Menasalvas Ruiz	4303	ernestina.menasalvas@upm. es	M - 10:00 - 11:00 hablar con la profesora

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

El plan de estudios Master Universitario en Ingeniería del Software no tiene definidas asignaturas previas recomendadas para esta asignatura.

3.2. Other recommended learning outcomes

- ? Conocimiento y aplicación de los principios fundamentales y técnicas básicas de los sistemas inteligentes y su aplicación práctica.
- ? Aptitud para aplicar los conocimientos sobre estadística y optimización.
- ? Design and implementation of relational databases
- SQL

4. Skills and learning outcomes *

4.1. Skills to be learned

CE3 - Elaborar una estimación de los parámetros del proyecto software.

CG1 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio (RD)

CG11 - Integración del conocimiento a partir de disciplinas diferentes, así como el manejo de la complejidad

CG2 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios (RD)

CG3 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades (RD)

CG7 E - Especificación y realización de tareas informáticas complejas, poco definidas o no familiares

CG8 - Planteamiento y resolución de problemas también en áreas nuevas y emergentes de su disciplina

CG9 - Aplicación de los métodos de resolución de problemas más recientes o innovadores y que puedan implicar el uso de otras disciplinas

CG125 - Comprender el procedimiento, valor y límites del método científico en el campo de la Informática, siendo capaz de identificar, localizar y obtener datos requeridos en un trabajo de investigación, de diseñar y guiar investigaciones analíticas, de modelado y experimentales, así como de evaluar datos de una manera crítica y extraer conclusiones

CGP 23 - Capacidad para desarrollar e implantar una solución informática en un entorno empresarial

CGP20 - Habilidad para hacer conexiones entre los deseos y necesidades del consumidor o cliente y lo que la tecnología puede ofrecer

CGP22 - Capacidad para comprender el mercado, sus hábitos y necesidades de productos o servicios tecnológicos

4.2. Learning outcomes

RA1 - Within an application field of Software Engineering, uses and designs the appropriate solution to solve some of its problems, describing the technical difficulties and the application limits

RA21 - Listening capability SC13, SC14, CG10 A

RA30 - Ante un problema real, elegir la solución de ingeniería del software más apropiada, analizando la viabilidad de su solución, lo que se puede y no se puede conseguir a través del estado actual de desarrollo de la solución seleccionada, lo que se espera que avance en el futuro

RA22 - Observing capability SC13, SC14, CG10 C

RA23 - Time organization capability SC13, SC14 K

RA26 - Group work skill SC13, SC14, CG17 A

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

5. Brief description of the subject and syllabus

5.1. Brief description of the subject

The course is mainly dedicated to the improvement of the development of software engineering projects by means of Data Mining.

The course is very interactive, with the development of many short projects and exposition at class. Learning by doing.

5.2. Syllabus

1. Introducción. El valor de los datos: variedad, velocidad, volumen
 - 1.1. Big Data
 - 1.2. Data Mining
2. Procesos de Data Mining
 - 2.1. Proceso de KDD
 - 2.2. CRISP-DM: fases
 - 2.3. El preproceso
3. Tipos de problemas de data mining
 - 3.1. Clasificación
 - 3.2. Clustering
 - 3.3. Asociación
4. . Evaluación de los resultados
 - 4.1. Técnicas de evaluación

6. Schedule

6.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Other face-to-face activities	Assessment activities
1	Tema 1 Duration: 03:00 Lecture			
2	Tema 2 Duration: 02:00 Lecture		Tema 2 Duration: 01:00 Problem-solving class	
3	Tema 2 Duration: 01:00 Lecture		Tema 2 Duration: 02:00 Problem-solving class	
4	Tema 2 Duration: 02:00 Lecture		Tema 2 Duration: 02:00 Problem-solving class	proyecto 1 Group work Continuous assessment Duration: 02:00
5	Tema 2 Duration: 01:00 Lecture		Tema 2 Duration: 02:00 Problem-solving class	
6	Tema 2 Duration: 01:00 Lecture		Tema 2 Duration: 02:00 Problem-solving class	
7	Tema 2 Duration: 01:00 Lecture		Tema 2 Duration: 02:00 Problem-solving class	
8	Tema 3 Duration: 02:00 Lecture		Tema 3 Duration: 02:00 Problem-solving class	proyecto 1 Group work Continuous assessment Duration: 02:00
9	Tema 3 Duration: 02:00 Lecture		Tema 3 Duration: 02:00 Problem-solving class	
10	Tema 3 Duration: 02:00 Lecture		Tema 3 Duration: 02:00 Problem-solving class	
11	Tema 3 Duration: 02:00 Lecture		Tema 3 Duration: 02:00 Problem-solving class	
12	Tema 3 Duration: 02:00 Lecture		Tema 3 Duration: 02:00 Problem-solving class	
13	Tema 3 Duration: 02:00 Lecture		Tema 3 Duration: 02:00 Problem-solving class	

14	Tema 3 Duration: 01:00 Lecture		Tema 3 Duration: 01:00 Problem-solving class	
15	Tema 4 Duration: 01:00 Lecture		Tema 4 Duration: 01:00 Problem-solving class	proyecto 1 Group work Continuous assessment Duration: 02:00
16	Tema 4 Duration: 01:00 Lecture		Tema 4 Duration: 01:00 Problem-solving class	
17				PROYECTO FINAL Individual presentation Continuous assessment Duration: 00:00 examen final Written test Final examination Duration: 00:00

The independent study hours are training activities during which students should spend time on individual study or individual assignments.

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

* The subject schedule is based on a previous theoretical planning of the subject plan and might go through experience some unexpected changes along throughout the academic year.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
4	proyecto 1	Group work	Face-to-face	02:00	25%	5 / 10	CG9 CG11 CGI25 CGP22 CG2 CG3 CG8 CE3 CG1 CG7 E CGP 23 CGP20
8	proyecto 1	Group work	Face-to-face	02:00	25%	5 / 10	CG11 CG9 CGI25 CGP22 CG2 CG3 CG8 CE3 CG1 CG7 E CGP 23 CGP20
15	proyecto 1	Group work	Face-to-face	02:00	25%	5 / 10	CGI25 CGP22 CG2 CG11 CG9 CG3 CG8 CE3 CG1 CG7 E CGP 23 CGP20

17	PROYECTO FINAL	Individual presentation	Face-to-face	00:00	25%	5 / 10	CG11 CG9 CGI25 CGP22 CG2 CG3 CG8 CE3 CG1 CG7 E CGP 23 CGP20
----	----------------	-------------------------	--------------	-------	-----	--------	--

7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	examen final	Written test	Face-to-face	00:00	100%	5 / 10	CGI25 CGP22 CG2 CG3 CG11 CG9 CG8 CE3 CG1 CG7 E CGP 23 CGP20

7.1.3. Referred (re-sit) examination

No se ha definido la evaluación extraordinaria.

7.2. Assessment criteria

La asignatura de Procesos de data mining se evaluara mediante 1 proyecto final, y la asistencia y participación en clase presentando resultados de miniproyectos..

Los proyectos se realizarán en grupos de **hasta 3 alumnos** de entre los matriculados de la asignatura al inicio del curso.

Para poder superar la asignatura, en la **convocatoria de junio**, se establecen los siguientes requisitos:

1. Obtener un mínimo de 50 puntos sobre los 100 disponible en el cómputo global de la evaluación sumativa
2. Es OBLIGATORIO realizar el proyecto final y completar todas las entregas de los miniproyectos
3. En los informes y presentaciones de proyectos y casos prácticos se debe obtener una nota mínima igual o superior al 50% de la valoración del mismo (ver tabla de valoración sumativa)
4. El aspecto Participación en clase forma un 10% de la nota que se conformará a partir de la asistencia a clase, la participación en la misma y la respuesta correcta a las preguntas planteadas por los profesores de la asignatura, pudiéndose realizar en cualquier momento a lo largo del curso

Para poder superar la asignatura en la **convocatoria extraordinaria de julio**, se establecen los siguientes requisitos:

1. Haber entregado los informes de proyectos y casos prácticos habiendo obtenido la nota mínima.
2. Realizar las presentaciones de los casos prácticos y proyectos
3. Obtener un mínimo de 50 puntos sobre los 100 disponibles en el cómputo global de los dos módulos

Para aquellos alumnos que de forma extraordinaria, no puedan realizar la evaluación continua, y previa petición por escrito durante los primeros 15 días del curso, la forma de evaluación de la asignatura será la siguiente, **siendo excluyente con la evaluación continua. Estos alumnos solo realizarán el examen final.**

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Principles of Data Mining (Adaptive Computation and Machine Learning), D Hand, MIT Press, 2001.	Bibliography	
Jiawei Han, Micheline Kamber, Data Mining : Concepts and Techniques, 2nd edition, Morgan Kaufmann, ISBN 1558609016, 2006.	Bibliography	
Data Mining Techniques: Marketing, Sales and Customer Support, Michael J. A. Berry, Gordon Linoff, John Wiley & Sons, 1997.	Bibliography	
Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Addison Wesley (May, 2005). Hardcover: 769 pages. ISBN: 0321321367	Bibliography	
Ian Witten, Eibe Frank, Mark Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Morgan Kaufmann, ISBN 978-0-12-374856-0, 2011.	Bibliography	
Página web de la asignatura en moodle	Web resource	
WEKA	Others	
Sala de trabajo en grupo con ordenadores	Equipment	

aula	Equipment	
------	-----------	--