

## 300271 - BIGDATA - Big Data & Data Mining

Coordinating unit: 300 - EETAC - Castelldefels School of Telecommunications and Aerospace Engineering  
 Teaching unit: 701 - AC - Department of Computer Architecture  
 Academic year: 2019  
 Degree: MASTER'S DEGREE IN APPLIED TELECOMMUNICATIONS AND ENGINEERING MANAGEMENT (MASTEAM) (Syllabus 2015). (Teaching unit Optional)  
 ECTS credits: 6 Teaching languages: English

### Prior skills

English, Programming, Probability.

### Requirements

English, Programming, Probability.

### Degree competences to which the subject contributes

Basic:

CB6. (ENG) CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

General:

03 DIS. (ENG) Diseñar aplicaciones de alto valor añadido basadas en las Tecnologías de la Información y las Comunicaciones (TIC), aplicadas a cualquier ámbito de la sociedad.

06 RES. (ENG) Resolver problemas y mejorar procesos en cualquier ámbito social a partir de la aplicación de las TIC, integrando conocimientos de diversos ámbitos y aplicando ingeniería de alto nivel tecnológico.

Transversal:

05 TEQ. TEAMWORK. Being able to work as a team player, either as a member or as a leader. Contributing to projects pragmatically and responsibly, by reaching commitments in accordance to the resources that are available.

03 TLG. THIRD LANGUAGE. Learning a third language, preferably English, to a degree of oral and written fluency that fits in with the future needs of the graduates of each course.

### Teaching methodology

The course is organized as a hands-on subject in which students work on projects related to the Big Data analytics. The main methodology is project based learning.

### Learning objectives of the subject

At the end of the course the student should be able to apply a number of data mining technologies over large data sets, extract useful information out of big data, program using the map-reduce paradigm and execute at large scale using cluster/cloud computers.

### Study load

Total learning time: 150h	Hours small group:	54h	36.00%
	Self study:	96h	64.00%

## 300271 - BIGDATA - Big Data & Data Mining

### Content

T1	Learning time: 10h Laboratory classes: 4h Self study : 6h
Description: Introduction to Big Data: Presentation of the course, examples of usage of big data technologies, available resources and developing environments.  Related activities: A1	
T2	Learning time: 25h Laboratory classes: 15h Self study : 10h
Description: Data sources, distributed file systems and databases, and data streaming: Technologies on Indexing, Memory, Streams, databases and evolution to big data. First examples on input sets.  Related activities: A1+A2	
T3	Learning time: 115h Laboratory classes: 35h Self study : 80h
Description: Processing and Data mining: Basic foundations and applications of map-reduce programming, learning models (search, classification, regression, clustering, information extraction), Bayesian inference, logic of reasoning, uncertainties and forecasting.  Related activities: A2	

## 300271 - BIGDATA - Big Data & Data Mining

### Planning of activities

A1	Hours: 26h Laboratory classes: 10h Self study: 16h
<p>Description: Guided exercises: Install of the programming environment and big data tools (i.e. Apache tools), basic examples and programs: hello world, lists, dictionaries, etc. Set up of data and machine learning libraries.</p> <p>Support materials: Atenea</p> <p>Descriptions of the assignments due and their relation to the assessment: A1 in Atenea (30%)</p>	
A2	Hours: 124h Laboratory classes: 44h Self study: 80h
<p>Description: Project: Classify objects based on features, using a variety of methods. Use Decision Trees and Bayesian Networks to explain phenomenon. Predict indicators using regression techniques. Display and analyze groups in your data using dimensionality reduction. Pre-process, extract, and select the learning features. Select the best parameters for your models using model selection.</p> <p>Support materials: Atenea</p> <p>Descriptions of the assignments due and their relation to the assessment: A2 in Atenea (70%)</p>	

### Qualification system

A1=30% + A2=70%

### Regulations for carrying out activities

Students should attend with their own personal laptop. Assistance is mandatory for at least 80% of class time. Activities are done in group.

## 300271 - BIGDATA - Big Data & Data Mining

### Bibliography

#### Basic:

Géron, Aurélien. Hands-on machine learning with scikit-learn & tensorflow : concepts, tools, and techniques to build intelligent systems [on line]. Sebastopol, CA: O'Reilly Media, Inc, [2017]Available on: <<https://proquest.safaribooksonline.com/9781491962282?uicode=politicat>>. ISBN 9781491962299.

#### Complementary:

Macias Lloret, Mario; Gómez Mauro; Tous Liesa, Rubén; Torres, Jordi. Introducción a Apache Spark : para empezar a programar el big data. Barcelona: UOC, 2015. ISBN 9788491160373.

Mohanty, Hrushikesh; Bhuyan, Prachet; Chenthati, Deepak. Big Data : A Primer [on line]. New Delhi: Springer India, 2015Available on: <<http://dx.doi.org/10.1007/978-81-322-2494-5>>. ISBN 9788132224945.

Leskovec, Jure; Rajaraman, Anand; Ullman, Jeffrey D. Mining of massive datasets [on line]. 2nd ed. New York, N.Y. ; Cambridge: Cambridge University Press, 2014Available on: <<https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=807230>>. ISBN 9781107077232.

Garreta, Raúl; Moncecchi, Guillermo. Learning scikit-learn : machine learning in Python. Birmingham: Packt Publishing, 2013. ISBN 9781783281930.

#### Others resources:

Atenea