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: 2018
Degree: MASTERS 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

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Hours small group: 54h</th>
<th>36.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self study:</td>
<td>96h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
### Content

<table>
<thead>
<tr>
<th>T1</th>
<th>Learning time: 10h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laboratory classes: 4h</td>
</tr>
<tr>
<td></td>
<td>Self study: 6h</td>
</tr>
</tbody>
</table>

**Description:**
Introduction to Big Data: Presentation of the course, examples of usage of big data technologies, available resources and developing environments.

**Related activities:**
A1

<table>
<thead>
<tr>
<th>T2</th>
<th>Learning time: 25h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laboratory classes: 15h</td>
</tr>
<tr>
<td></td>
<td>Self study: 10h</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>T3</th>
<th>Learning time: 115h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laboratory classes: 35h</td>
</tr>
<tr>
<td></td>
<td>Self study: 80h</td>
</tr>
</tbody>
</table>

**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
### Planning of activities

<table>
<thead>
<tr>
<th>A1</th>
<th>Hours: 26h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laboratory classes: 10h</td>
</tr>
<tr>
<td></td>
<td>Self study: 16h</td>
</tr>
</tbody>
</table>

**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.

**Support materials:**
Atenea

**Descriptions of the assignments due and their relation to the assessment:**
A1 in Atenea (30%)

<table>
<thead>
<tr>
<th>A2</th>
<th>Hours: 124h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laboratory classes: 44h</td>
</tr>
<tr>
<td></td>
<td>Self study: 80h</td>
</tr>
</tbody>
</table>

**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.

**Support materials:**
Atenea

**Descriptions of the assignments due and their relation to the assessment:**
A2 in Atenea (70%)

### 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:**


**Complementary:**

Mario Macías, Mauro Gómez, Ruben Tous, Jordi Torres. *Introducción a Apache Spark*. UOC.


**Others resources:**

Atenea