# COURSE DATA

## DATA SUBJECT

**Code:** 44833
**Name:** Analysis of Web and Social Data
**Cycle:** Master's Degree
**ECTS Credits:** 4
**Academic year:** 2025-26

## STUDY (S)

| Degree | Center | Acad. year | Period |
|---|---|---|---|
| 2234 - Master's Degree in Web Technology, Cloud Computing and Mobile Applications | Escola Tècnica Superior d'Enginyeria | 1 | Second quarter |

## SUBJECT-MATTER

| Degree | Subject-matter | Character |
|---|---|---|
| 2234 - Master's Degree in Web Technology, Cloud Computing and Mobile Applications | Information and Content Management and Processing | COMPULSORY |

## COORDINATION

ROMERO GOMEZ VERONICA

# SUMMARY

In this course are presented the computational models aimed at the extraction of knowledge, potentially useful and previously unknown, from the World Wide Web. Within this field, known as Web Mining, the student is expected to know the computational techniques used for web data analysis, organized mainly in four groups: structure, content, use and social data/relations analysis. Each of these groups will be shown, as well as the characteristic problems of each one of them. The goal is to provide a broad view of the data analysis techniques that are applied to the web environment so that the student is able to handle them and solve analysis problems in this context. Additionally, the subject addresses the field of Web Analytics so that the students can develop measuring plans and analyse data in websites, mobile apps, social networks, advertising campaigns, newsletters and other platforms. Finally, students are expected to achieve an advanced command of Google Analytics, which will allow them to get the Google Analytics Individual Qualification (GAIQ).

IQ).

# PREVIOUS KNOWLEDGE

## RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

There are no specified enrollment restrictions with other subjects of the curriculum.

## OTHER REQUIREMENTS

Recommendations: Basic statistics, Basic programming

# COMPETENCES / LEARNING OUTCOMES

**-**

Ability to apply acquired knowledge and solve problems in new or little-known environments within broader and multidisciplinary contexts, being able to integrate this knowledge.

Ability to collect data and apply machine learning methods to model, design and develop applications and services.

Ability to design and evaluate servers, applications and systems based on distributed computing.

Ability to model, design, define the architecture, implement, manage, operate, and maintain applications, systems, services, networks and content in the field of Web technologies, cloud computing and mobile applications.

Ability to understand and apply the operation and organization of component models, intermediary software and services.

Students should apply acquired knowledge to solve problems in unfamiliar contexts within their field of study, including multidisciplinary scenarios.

Students should be able to integrate knowledge and address the complexity of making informed judgments based on incomplete or limited information, including reflections on the social and ethical responsibilities associated with the application of their knowledge and judgments.

Students should communicate conclusions and underlying knowledge clearly and unambiguously to both specialized and non-specialized audiences.

Students should demonstrate self-directed learning skills for continued academic growth.

Students should possess and understand foundational knowledge that enables original thinking and research in the field.

To foster, in academic and professional contexts, technological, social or cultural advancement within a society based on In knowledge and respect for: a) fundamental rights and equal opportunities between men and women; b) principles of equal opportunities and universal accessibility of persons with disabilities; and, c) the values of a culture of peace and democratic values.

## DESCRIPTION OF CONTENTS

### 1. Web mining

-Introduction: The challenges of the web, web mining, crawling the web
-Information retrieval and search: Vector Space Model, Word Embeddings, PageRank, Evaluation
-Document Classification: k-NN, Naïve Bayes, Deep Neural Networks (BERT), Feature Selection
-Document clustering: Hierarchical clustering, k-means, probabilistic clustering

### 2. Analysis of social data

- Introduction
- Definition of opinion
- Summary of opinions
- Sentiment analysis
- Polarity classification

### 3. Web Analytics

- Introduction to web analytics: Measurement methods, Measuring tools and Business objectives.
- Macroconversions and microconversions.
- Dimensions, Metrics and KPI.
- Measurement plan: practical cases.

### 4. Google Analytics, SEO strategies measurement, Google Search Console

- Google Analytics GA4: interface, reports, Analytics Intelligence search, Custom reports and explorations, performance and attribution models.
- Analytics SEO strategies, Technical SEO and Google Search Console

## WORKLOAD

### PRESENCIAL ACTIVITIES

| Activity | Hours |
|---|---|
| Theoretical and practical classes | 28,90 |
| Laboratory | 11,10 |
| **Total hours** | **40,00** |

### NON PRESENCIAL ACTIVITIES

| Activity | Hours |
|---|---|

| | |
|---|---|
| Attendance at other activities | 0,00 |
| Individual or group project | 6,00 |
| Independent study and work | 35,00 |
| Preparation of lessons | 16,00 |
| Preparation for assessment activities | 3,00 |
| Resolution of case studies | 0,00 |
| **Total hours** | **60,00** |

## TEACHING METHODOLOGY

- Lectures
- Problem resolution
- Project-oriented learning

## EVALUATION

The assesment modalities used in this subject are:

SE1: Online assessment and/or degree of participation

SE2: Assessment of problems, works, reports and/or memories

SE6: Laboratory assesment

SE4: Exam or face-to-face assessment

- First call:

In the first call the note will be obtained as follows:

SE1*0.1+SE2*0.3+SE6*0.4+SE4*0.2

- Second call:

The requested and undelivered works / memories / reports / code, etc. must be delivered throughout the course. The weights will be the same as those of the first call. SE1 is not recoverable.

The grading system is specified at the following link:

http://www.uv.es/uvweb/universidad/es/estudios-postgrado/informacion-administrativa-postgrado/permanencia-calificaciones/calificaciones-1285897761928.html

The applicable regulations can be found at the following link:

http://www.uv.es/uvweb/universidad/es/estudios-grado/informacion-academica-administrativa/normativas/normativas-universidad-valencia-1285850677111.html

## REFERENCES


- Natural Language Processing with Python. Steven Bird, Ewan Klein, and Edward Loper . http: //www.nltk.org/book/
- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython. Wes McKinney, W (2012). Ed. O'Reilly Media
- Building an Intelligent Web: Theory and Practice . Pawan Lingras , Rajendra Akerkar. Ed. Jonnes & Bartlett Learning
- Social big data mining. Ishikawa, H. CRC Press.
- Semantic mining of social networks. Synthesis Lectures on the Semantic Web: Theory and Technology, 5(2), 1-205. Tang, J., & Li, J.
- Networks, crowds, and markets: Reasoning about a highly connected world. Easley, D., & Kleinberg, J. Cambridge University Press.
- http://scikit-learn.org/stable/
- Analítica Web 2.0. Avinash KaushiK. Ed. Gestión 2000
- Google Analytics integrations. Daniel Waisberg. Indianapolis, Indiana: Wiley, 2015
- https://support.google.com/analytics/answer/10089681

- https://support.google.com/analytics/answer/9322688