COURSE DATA

DATA SUBJECT

Code: 36516
Name: Unstructured Data
Cycle: Undergraduate Studies
ECTS Credits: 6
Academic year: 2026-27

STUDY (S)

Degree	Center	Acad. year	Period
1332 - Degree in Business Intelligence and Analytics	Facultat d'Economia	4	First quarter, Sin determinar

SUBJECT-MATTER

Degree	Subject-matter	Character
1332 - Degree in Business Intelligence and Analytics	Herramientas y Técnicas de Análisis de Datos	COMPULSORY

COORDINATION

MONTALVA CANDEL IGNACIO

MONTALVA CANDEL IGNACIO

MONTALVA CANDEL IGNACIO

SUMMARY

With the development of Web 2.0 and social networks came massive data. Most of this data has no predefined initial structure. We are referring to documents, photos, videos, e-mails, tweets, etc. Most of this data is generated by the users themselves. Every day, emails are sent, tweets are posted, photos or videos are uploaded to different social networks. All this data provides a lot of information about the people who are generating it. However, being unstructured data, the information they generate must be transformed in order to analyse it and turn it into knowledge. An example is a company's information on the social network Twitter. You can collect all the tweets that mention a company, and by transforming them into a structured database you can analyse sentiment and find out what people think about the company. You can also download competitors' tweets to find out what people think about those companies and modify the company's strategy.Unstructured data will require non-relational databases, such as MongoDB (document-based), Neo4j (graph-based), and so on. The new databases will allow a different analysis to be carried out to that performed with relational databases. The objective of the course will be that the students know and can work with these unstructured data and carry out analysis of these data in order to have a better knowledge of the company and the competition.

PREVIOUS KNOWLEDGE

RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

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

OTHER REQUIREMENTS

It is convenient to have knowledge and practice with R-software and RStudio.
In addition, it is convenient to have some skills on Basic Statistical and Statistical Inference. Machine learning methods can be appropriate.

COMPETENCES / LEARNING OUTCOMES

1332 - Degree in Business Intelligence and Analytics

Apply methods and techniques of analysis, synthesis and graphical representation by means of software tools.

Apply supervised machine learning techniques using software.

Apply unsupervised and semi-supervised machine learning techniques using software.

Be able to analyse and search for information from diverse sources.

Be able to define, solve and present complex problems systemically.

Be able to learn autonomously.

Be able to use ICT, both in academia and in professional practice.

Be able to work in a team demonstrating commitment to quality, ethics, equality and social responsibility.

Manage download APIs and capture and manipulate unstructured data values.

Reorganise and restructure variables and databases.

Students must be able to communicate information, ideas, problems and solutions to both expert and lay audiences.

Use data mining software.

DESCRIPTION OF CONTENTS

a. Data, Information and Knowledge

1. Introduction

a. Data, Information and Knowledgeb. Structured vs. Unstructured Data
c. The illusion of unstructured data analysis
d. Markup standards: XML, JSON,...

2. Web scraping I

a. Web scraping - direct scrapint
b. Applying web scraping to obtain an index (CPI)
C. Web scraping using API's
d. Good practice

3. Web Scraping II

a. Webpage structure: HTML
b. Dynamic HTML: JavaScript
c. Style Sheets: CSS, W3-CSS, Bootstrap(CSS)
d. Scraping an entire website
e. Web-specific scraping
F. Scripts that mimic human intervention: Bots

4. Text Mining

a. Types and uses of text mining techniques
b. Approaches: bag of words (BoW) and natural language processing (NLP)
c. The basic workflow
d. Data preparation and structures
e. Visualizations and descriptive analysis

5. Sentiment Analysis

a. Subjective content in texts, polarity and emotions
b. Dictionary-based approach vs machine learning
c. DSentiment dictionaries in Spanish
d. Sentiment analysis applications

a. Supervised and unsupervised learning techniques
b. Probabilistic topic modeling
c. Text classification
d. Predictive modeling

6. Machine Learning and Text Analytics

a. Supervised and unsupervised learning techniques
b. Probabilistic topic modeling
e. Advanced techniques for text analysis

7. Recommender Systems

a. Types of recommender systems
b. Implicit and explicit preferences
c. Construction of a recommender system
d. Evaluation of a recommender system
e. Applications

8. Networks and Social Networks

a. Definition and typologies
b. Statistical models of social networks
c. Random graph models
d. Social networks and segmentation
e. Applied Neural Networks

WORKLOAD

PRESENCIAL ACTIVITIES

Activity	Hours
Theory	15,00
Computer classroom practice	45,00
Total hours	**60,00**

NON PRESENCIAL ACTIVITIES

Activity	Hours
Attendance at other activities	0,00
Individual or group project	20,00
Independent study and work	0,00
Preparation of lessons	50,00
Preparation for assessment activities	10,00
Resolution of case studies	10,00
Total hours	**90,00**

TEACHING METHODOLOGY

- Classroom lectures to present the essential theoretical content of the subject.

- On-site practical classes, relating to problem solving, case studies, with the application of techniques, oral presentations, debates, individually and/or in teams.

- Supervised autonomous work based on exercises, case studies and questions to be debated or online experiments, with tutorial support.

EVALUATION

Throughout the course, several practical assignments and assessment tests will be carried out, through which students will be able to demonstrate the concepts acquired, both theoretical and practical. The set of practical assignments and tests will account for 60% of the final grade.

As a final course project, students must complete a paper in which they will apply the knowledge acquired in the subject. This report must be submitted no later than the date set for the first-call examination (40% of the final grade).

Although the practical assignments cannot be retaken, students who do not pass the subject in the first call will have the opportunity to resubmit the final report, which must be delivered no later than the date set for the second-call examination.

In order to apply the percentages mentioned above, it will be necessary to obtain at least a grade of 5 in the final course project.

REFERENCES

- Anandarajan, M.; Hill, C. y Nolan, T. (2019) Practical Text Analytics. Maximixing the Value of Text Data. Spinger Nature, Switzerland.
- Bali, R. Sarkar, D. y Sharma, T. (2017) Learning Social Media Analytics with R. Pack Publishing, Birmingan, UK.
- Gorakala, S.K. y Usuelli, M. (2015) Building a Recommendation System with R. Pack Publishing, Birmingan, UK.
- Kwartler, T. (2017). Text mining in practice with R. John Wiley & Sons.
- Munzert, S.; Rubba, C.; Mei¿ner, P. y Nyhuis, D. (2015) Automated data collection with R: a practical guide to web scraping and text mining. John Wiley & Sons, UK.
- Rusell, Matthew A. (2013) Mining the Social Web. O¿Reilly Media, CA.