

**COURSE DATA****Data Subject**

Code	36516
Name	Unstructured Data
Cycle	Grade
ECTS Credits	6.0
Academic year	2023 - 2024

Study (s)

Degree	Center	Acad. Period	year
1332 - Degree in Business Intelligence and Analytics	Faculty of Economics	4	First term

Subject-matter

Degree	Subject-matter	Character
1332 - Degree in Business Intelligence and Analytics	24 - Herramientas y Técnicas de Análisis de Datos	Obligatory

Coordination

Name	Department
MONTORO PONS, JUAN DE DIOS	110 - Applied Economics
MORILLAS JURADO, FRANCISCO GABRIEL	110 - Applied Economics

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.
Translated with www.DeepL.com/Translator (free version)

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

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

Other requirements

Es recomendable tener conocimientos del software R y RStudio.
Estadística Básica e Inferencia, así como Aprendizaje Automático (machine learning).

COMPETENCES (RD 1393/2007) // LEARNING OUTCOMES (RD 822/2021)

1332 - Degree in Business Intelligence and Analytics

- Students must be able to communicate information, ideas, problems and solutions to both expert and lay audiences.
- Be able to analyse and search for information from diverse sources.
- Be able to learn autonomously.
- Be able to use ICT, both in academia and in professional practice.
- Be able to define, solve and present complex problems systemically.
- Be able to work in a team demonstrating commitment to quality, ethics, equality and social responsibility.
- Apply methods and techniques of analysis, synthesis and graphical representation by means of software tools.
- Reorganise and restructure variables and databases.
- Use data mining software.
- Apply supervised machine learning techniques using software.
- Apply unsupervised and semi-supervised machine learning techniques using software.
- Manage download APIs and capture and manipulate unstructured data values.

LEARNING OUTCOMES (RD 1393/2007) // NO CONTENT (RD 822/2021)

- Synthesise information.
- Knowing how to design and implement research and its application to the economic environment.



- Apply automatic learning techniques to business problems.
- Know the different types of existing databases and the types of data they use.
- To learn the concepts of Graph Theory.
- To learn how a recommendation system works.- To learn how to process non-numerical data by means of computer programmes.
- To learn how to create dictionaries with which to carry out automatic processing of natural language for business exploitation.
- Know how to implement research and its application to the economic environment with textual data.
- Handle advanced quantitative tools and their application to the business environment.
- Learn to handle data download APIs in social networks.
- Learn to recognise patterns in data in text format.
- Learn processes for generating purchase recommendations.

DESCRIPTION OF CONTENTS

1. Introduction

- a. Data, Information and Knowledge
- b. 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

6. Machine Learning and Text Analytics

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

ACTIVITY	Hours	% To be attended
Computer classroom practice	45,00	100
Theory classes	15,00	100
Development of group work	10,00	0
Development of individual work	10,00	0
Preparation of evaluation activities	10,00	0
Preparing lectures	20,00	0
Preparation of practical classes and problem	30,00	0
Resolution of case studies	10,00	0
TOTAL	150,00	

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

Assessment consisting of one or more tests that will consider both theoretical and practical issues (50%).

Evaluation of the practical activities carried out by the student during the course, based on the elaboration of work/memories and/or oral presentations, with defence of the positions developed by the student. (30%)

Continuous assessment of each student, based on the participation and degree of involvement of the student in the teaching-learning process, taking into account regular attendance at the scheduled classroom activities and the resolution of questions and problems proposed periodically (20%).

REFERENCES



Basic

- Anandarajan, M.; Hill, C. y Nolan, T. (2019) Practical Text Analytics. Maximizing the Value of Text Data. Springer Nature, Switzerland.
- Bali, R. Sarkar, D. y Sharma, T. (2017) Learning Social Media Analytics with R. Pack Publishing, Birmingham, UK.
- Gorakala, S.K. y Usuelli, M. (2015) Building a Recommendation System with R. Pack Publishing, Birmingham, UK.
- Kwartler, T. (2017). Text mining in practice with R. John Wiley & Sons.
- Munzert, S.; Rubba, C.; Meiner, 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. OReilly Media, CA.