

**Introduction to Big Data**

Code: 104748  
ECTS Credits: 6

| Degree                            | Type | Year | Semester |
|-----------------------------------|------|------|----------|
| 2503873 Interactive Communication | OB   | 3    | 2        |

## Contact

Name: Alessandro Bernardi

Email: [alessandro.bernardi@uab.cat](mailto:alessandro.bernardi@uab.cat)

## Teaching groups languages

You can check it through this [link](#). To consult the language you will need to enter the CODE of the subject. Please note that this information is provisional until 30 November 2023.

## Teachers

Alessandro Bernardi

## Prerequisites

This course does not have any compulsory requirements, but it is recommended that students have previously passed the following courses:

Information Systems

Information Storage and Retrieval

Advanced Web Services

## Objectives and Contextualisation

The main objective of the course is to introduce students to the basic concepts and main practices of Big Data.

The course also has the following specific objectives:

1. To develop a propaedeutic knowledge for the further development of Business Intelligence applications: the development of big data solutions for business intelligence and its influence on decision making.
2. To introduce the concepts of data sources and types of data (structure, classification, integration and quality).
3. To promote the exploration of requests and work with open data sources.
4. To make the first approaches to database analysis in a spreadsheet environment and other practical tools.

## Competences

- Act with ethical responsibility and respect for fundamental rights and duties, diversity and democratic values.
- Act within one's own area of knowledge, evaluating sex/gender-based inequalities.
- Determine and plan the technological infrastructure necessary for the creation, storage, analysis and distribution of interactive multimedia and social-networking products.
- Introduce changes in the methods and processes of the field of knowledge to provide innovative responses to the needs and demands of society.
- Manage time efficiently and plan for short-, medium- and long-term tasks.
- Promote and launch new products and services based on massive-scale mining and analysis of data from the Media.
- Search for, select and rank any type of source and document that is useful for creating messages, academic papers, presentations, etc.
- Students must be capable of applying their knowledge to their work or vocation in a professional way and they should have building arguments and problem resolution skills within their area of study.
- Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
- Students must develop the necessary learning skills to undertake further training with a high degree of autonomy.
- Take account of social, economic and environmental impacts when operating within one's own area of knowledge.

## Learning Outcomes

1. Analyse a situation and identify its points for improvement.
2. Communicate using language that is not sexist or discriminatory.
3. Critically analyse the principles, values and procedures that govern the exercise of the profession.
4. Cross-check information to establish its veracity, using evaluation criteria.
5. Describe the infrastructure needed to store big data.
6. Differentiate between the various types of existing architectures for working with big data.
7. Distinguish the salient features in all types of documents within the subject.
8. Evaluate the impact of problems, prejudices and discrimination that could be included in actions and projects in the short or medium term in relation to certain people or groups.
9. Explain the characteristics of the infrastructure needed to recover big data.
10. Explain the explicit or implicit deontological code in your area of knowledge.
11. Explain the infrastructure needed to process big data.
12. Extract large volumes of data from social networks and the new digital media in particular.
13. Identify situations in which a change or improvement is needed.
14. Identify the social, economic and environmental implications of academic and professional activities within one's own area of knowledge.
15. Plan and execute academic projects in the field of big data.
16. Propose new methods or well-founded alternative solutions.
17. Propose projects and actions that are in accordance with the principles of ethical responsibility and respect for fundamental rights and obligations, diversity and democratic values.
18. Propose projects and actions that incorporate the gender perspective.
19. Propose viable projects and actions to boost social, economic and environmental benefits.
20. Share experiences with the group as a path to learning, in order to work subsequently in multidisciplinary groups.
21. Solve basic problems in big data.
22. Submit course assignments on time, showing the individual and/or group planning involved.
23. Weigh up the risks and opportunities of both one's own and other people's proposals for improvement.

## Content

Unit 1. Big Data: Introduction to the subject: concept of Big Data, its processes and characteristics.

Unit 2. Sources, capture and storage of data: Presentation of data sources (mainly open sources. Processes of access and requests for public information and transparency laws. Processes for searching, downloading and storing different types of data (formats).

Unit 3. Data processing and analysis: Handling of data cleaning and analysis tools and functions for decision making.

Unit 4. Social Media data analysis and monitoring: Introduction to Social Media as a source of big data: presentation of techniques and tools to extract insights from the social networks.

Unit 5. Data visualization and data mapping: Presentation of different tools and possibilities of data visualization and cartographic representation of information for decision-oriented reporting.

(\*) The detailed calendar with the content of the different sessions will be displayed on the day of the presentation of the course. It will also be posted on the Virtual Campus where students will be able to find a detailed description of the exercises and practices, the various teaching materials and any information necessary for the proper monitoring of the course. In the event of a change of teaching modality for health reasons, the teaching staff will inform of the changes that will take place in the course programme and in the teaching methodologies.

The content of this course will be sensitive to aspects related to the gender perspective.

## Methodology

The structure of the course, in which different practical activities are carried out, seeks to internalise skills related to the management of Big Data (search, extraction, analysis and publication of data for decision-making). Its methodology is completely practical. Through laboratory activities, workshops and the final project, both the theoretical component of the subject and the practical application of the contents studied are evaluated.

The continuous assessment of the course, in which specific and continuous short-term practical activities are carried out, allows for a very precise monitoring of the student's learning and progression. In addition, activities are done progressively on the acquisition of knowledge that, step by step, is involved in the next activities.

The Introduction to Big Data course includes three types or categories of assessable training activities:

Laboratory exercises: individual or team work in which practical activities are carried out with a punctual deliverable with a time limit. Students must apply the knowledge, distribute the time and prepare the deliverables within the classroom and in the hours set aside for practice under the guidance of the professor.

Seminars: individual or team work involving more extensive practical activities with deliverables open to students' creativity. There are no time limits in the classroom, but there are deadlines. Students must apply knowledge, allocate time and prepare deliverables by starting their work in the classroom, but continuing it in the form of activities supervised by the teaching team.

Development of the final course work: practical group evaluation exercise in which students must solve, during the course, a practical application problem related to Big Data. Students must state the problem and carry out the four processes to provide a proposed solution based on large amounts of data: search, extraction, analysis and publication of a data report that includes a proposed decision based on the information collected and analysed.

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

## Activities

| Title  | Hours | ECTS | Learning Outcomes   |
|--|-------|------|---|
| Type: Directed   |       |      |   |
| Laboratory   | 33    | 1.32 | 3, 1, 20, 2, 5, 6, 11, 9, 12, 14, 13, 15, 23, 16, 17, 18, 19, 21, 8               |
| Theoretical sessions   | 15    | 0.6  | 3, 2, 6, 10, 11, 9, 18, 21  |
| Type: Supervised   |       |      |   |
| Mentoring  | 10    | 0.4  | 20, 7, 15, 22, 16, 19, 21   |
| Seminars   | 10    | 0.4  | 3, 1, 20, 2, 4, 5, 7, 6, 10, 11, 9, 12, 14, 13, 15, 23, 22, 16, 17, 18, 19, 21, 8 |
| Type: Autonomous   |       |      |   |
| Autonomous work: reading and coursework preparation and personal study | 60    | 2.4  | 3, 20, 4, 5, 7, 6, 11, 9, 12, 15, 22, 16, 18, 19, 21                              |

## Assessment

Continuous evaluation

The assessment activities are as follows:

Activity A: Laboratory exercises, which have a weight of 40% of the final qualification.

Activity B: Classroom exercises, with a weight of 30% of the final qualification.

Activity C: Courseworks, which weighs 30% of the final grade.

In order to pass the course, a minimum pass mark (5.0) must be obtained in each of the activities.

REVALUATION:

In the last three weeks of the course, students who have not passed the course may sit a revaluation test consisting of a theoretical test and a practical exercise. The compulsory condition to be eligible for the revaluation of the course is to have done at least 2/3 of the total number of practical exercises of the course (activities A, B and C) and to have obtained an average mark equal to or higher than 3.5 (and lower than 5) in all the assessment activities.

In accordance with the above criteria, if a student does not complete at least 66% of the practical activities of the assessment activities, he/she will be considered as not assessable in this course.

SECOND ENROLLMENT:

In the case of second enrollment, students may take a single synthesis test that will consist of a theoretical test and a practical exercise. The qualification of the subject will correspond to the qualification of the synthesis test. The student who wants to take this synthesis test must notify the coordinator of the subject.

#### PLAGIARISM:

In the event that the student performs any irregularity that may lead to a significant variation of an evaluation act, this evaluation act will be graded with 0, regardless of the disciplinary process that could be instructed. In the event, that several irregularities occur in the evaluation acts of the same subject, the final grade for this subject will be 0.

This subject doesn't provide for the single assessment system.

### Assessment Activities

| Title               | Weighting | Hours | ECTS | Learning Outcomes   |
|---------------------|-----------|-------|------|---|
| Classroom exercises | 30%       | 8     | 0.32 | 3, 1, 20, 2, 4, 5, 7, 6, 10, 11, 9, 12, 14, 13, 15, 23, 22, 16, 17, 18, 19, 21, 8 |
| Courseworks         | 30%       | 6     | 0.24 | 3, 1, 20, 2, 4, 7, 10, 12, 14, 13, 15, 23, 22, 16, 17, 18, 19, 21, 8              |
| Laboratory          | 40%       | 8     | 0.32 | 3, 1, 20, 2, 4, 5, 7, 6, 10, 11, 9, 12, 14, 13, 15, 23, 22, 16, 17, 18, 19, 21, 8 |

### Bibliography

Alcalde, Ignasi. (2015). *Visualización de la información. De los datos al conocimiento*. Editorial UOC.

Bounegru, Liliana; Gary, Jonathan (Eds.). (2020). *The Data Journalism Handbook II. Towards a Critical Data Practice*. European Journalism Centre and Google News Initiative.

<https://datajournalism.com/read/handbook/two>

Bradshaw, Paul. (2017). *Scraping for Journalists. How to grab information from hundreds of sources, put it in data you can interrogate - and still hit deadlines (2nd edition)*. Leanpub

Bradshaw, Paul. (2019). *Finding Stories in Spreadsheets. Recipes for interviewing data - and getting answers*. Leanpub

Bradshaw, Paul., Maseda, Bárbara. (2015). *Periodismo de datos: Un golpe rápido. Cómo entrar, obtener los datos, escabullirse con la noticia... ¡Y asegurarse de que nadie salga herido!* Leanpub.

Cairo, Alberto. (2016). *The Truthful Art: Data, charts, and maps for communication*. New Riders.

Cairo, Alberto. (2017). ¿Visualización de datos: una imagen puede valer más que mil números, pero no siempre más que mil palabras. *El profesional de la información*, 26(6), 1025-1028.

Carlberg, Conrad. (2011). *Análisis estadístico con Excel*. Anaya.

CARTO (2018). The Top Trends in Data Visualization for 2018. *Medium*.  
<https://medium.com/@carto/the-top-trends-in-data-visualization-for-2018-54911e875375>

Charte Ojeda, Francisco (2016). *Excel 2016*. Anaya.

Ferrer-Sapena, Antonia; Sánchez-Pérez, Enrique. (2013). Open data, big data: ¿Hacia dónde nos dirigimos? *Anuario ThinkEPI*, 7, 150-156.

Fernández-Rovira C., Giraldo-Luque S. (2021). *La felicidad privatizada. Monopolios de la información, control social y ficción democrática en el siglo XXI*. Editorial UOC.

Fernández-Rovira C., Giraldo-Luque S. (Eds.). *Predictive Technology in Social Media*. CRC Press. Taylor & Francis Group.

Fuchs, Christian. (2017). "Dallas Smythe Today - The Audience Commodity, the Digital Labour debate, Marxist Political Economy and Critical Theory. Prolegomena to a Digital Labour Theory of Value". En: Fuchs, C., Mosco, V. (Eds.). *Marx and the Political Economy of the Media*. Haymarket Books. pp. 522-599.

Giraldo-Luque Santiago; Fernández-Rovira Cristina (2021) Economy of Attention: Definition and Challenges for the Twenty-First Century. En: Park S.H., Gonzalez-Perez M.A., Floriani D.E. (Eds.). *The Palgrave Handbook of Corporate Sustainability in the Digital Era*. Palgrave Macmillan. pp. 283-305.

Greene, Derek. (2014). Practical Social Network Analysis With Gephi. [Practical Social Network Analysis With Gephi · Derek Greene](#)

Greene, Derek; Cunningham, Pàdraig (2013). Producing a Unified Graph Representation from Multiple Social Network Views. Proc. ACM Web Science'13

Kauffmann, Erick; Peral, Jesús; Gil, David; Ferrández, Antonio; Sellers, Ricardo; Mora Higinio (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making, *Industrial Marketing Management*, 90, 523-537.

Mayer-Schönberger, Viktor; Cukier, Kenneth (2013). *Bigdata. La revolución de los datos masivos*. Turner.

O'Neil, Cathy. (2017). *Armas de destrucción matemática. Cómo el Big Data aumenta la desigualdad y amenaza la democracia*. Capitan Swing.

Patino, Bruno (2020). *La civilización de la memoria de pez. Pequeño tratado sobre el mercado de la atención*. Alianza.

Tascón, Mario (2013). Introducción. Big Data. Pasado, presente, futuro. *Telos: Cuadernos de comunicación e innovación*, 95, 47-50.

Turing, Alan. (1974). *¿Puede pensar una máquina?* Universidad de Valencia.

## Software

As this is a completely practical course, the software required is the usual one for the tasks of capturing, processing and analysing information in different formats.

Specifically, the following tools are required:

Text editing software: Word or similar

Data analysis software: Excel or similar

Data visualisation software: Flourish - Datawrapper - Gephi