

## 29525 - Data Lifecycle

### Syllabus Information

**Academic Year:** 2022/23

**Subject:** 29525 - Data Lifecycle

**Faculty / School:** 175 - Escuela Universitaria Politécnica de La Almunia

**Degree:** 625 - Bachelor's Degree in Industrial Processes' Data Engineering

**ECTS:** 6.0

**Year:** 1

**Semester:** First semester

**Subject Type:** Compulsory

**Module:**

## 1. General information

### 1.1. Aims of the course

The subject data lifecycle pretends to give students a global vision of the natural cycle of the data in the industrial processes. The ultimate goal is that students could learn the data lifecycle stages, which constitutes a first approach to the data processing and its aim is to provide basics for being used as fundamentals for the subjects as data mining, neuronal networks or data visualisations among others. It is pretended that students acquire the required knowledge related with management plan and statistics analysis development.

These approaches and objectives are aligned with the following Sustainable Development Goals (SDGs) of the United Nations Agenda 2030 (<https://www.un.org/sustainabledevelopment/es/>) so that the acquisition of the learning outcomes of the subject provides training and competence to contribute to some extent to its achievement: Goal 4 Quality education.

### 1.2. Context and importance of this course in the degree

The subject is mandatory and it is part of the basic training for the Degree of Data in Industrial Processes Engineering students. (?) It is part of the first half in the first course of the curriculum of the Degree of Data in Industrial Processes Engineering and included in the data science module, being the first subject and having introduction character for this module.

### 1.3. Recommendations to take this course

There are not any requirements demanded beforehand for this subject. It is recommended a regular attendance to the classes, theory and practical classes and the involvement in both for a further utilisation.

## 2. Learning goals

### 2.1. Competences

#### Basic competences:

CB1 - That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects involving knowledge from the cutting edge of your field of study.

CB3 - That students have the ability to gather and interpret relevant data (usually within their study area) to make judgments that include a reflection on relevant issues of a social, scientific or ethical nature

CB5 - That students have developed those learning skills necessary to undertake further studies with a high degree of autonomy.

#### General competences:

CG1- Ability to design data management systems within the engineering field that include all stages of the data life cycle.

CG2 - Adequate knowledge and skills to efficiently analyze, synthesize, solve and communicate basic problems related to production processes in the industry.

CG3 - Knowledge of basic scientific and technical subjects that enable the acquisition, management and processing of data

in Engineering.

CG4 - Ability to work on problems understanding the ethical, social and professional responsibility of data processing activity.

CG5 - Ability to solve technological problems that may arise in data engineering in industrial processes.

CG6 - Ability to implement solutions derived from data analysis that optimize production processes in the industry.

#### **Transversal competences:**

CT01 - Work cooperatively, assuming and respecting the role of the different team members.

CT02 - Work in multidisciplinary environments.

CT03 - Search, select and manage information and knowledge responsibly.

CT04 - Develop critical thinking and reasoning

CT05 - Communication of results effectively.

CT06 - Generate innovative and competitive proposals in professional activity and research.

CT07 - Analyze and solve problems autonomously, adapt to unforeseen situations and make decisions.

CT08 - Demonstrate sensitivity towards ethical, social and environmental issues.

CT11 - Ability to plan and manage temporary, material and human resources.

## **2.2. Learning goals**

The student, to pass this subject, must demonstrate having acquired the following knowledge:

- Learn the necessary bases to extract information and visualize a data set.
- Identify the steps required for a data analysis.
- Determine non-relational structures in the data.
- Model the associations between variables with linear and nonlinear techniques.
- Validate and compare models using data segmentation or resampling techniques.
- Establish controls in a productive system.

## **2.3. Importance of learning goals**

Coursing this subject students develop competences for dealing with real problems related with extraction, treatment and analyse of data. The students have to know the stages for working with real data and learning how to recognize and use the models, which is used as a reflected of the basis points of the situations related with the engineering sphere.

It is pretended that the student is able to operate information, which inception is a database summarizing the information there included and having the possibility to make choices based on those results, using for it different statistical analysis. The whole results of the learning would be importants for the students in their professional development and the rest subjects of the data module.

# **3. Assessment (1st and 2nd call)**

## **3.1. Assessment tasks (description of tasks, marking system and assessment criteria)**

Students must show the foreseen learning outcomes through the following assessment activities:

Continuous assessment system:

- Exams: During the course two main exams will be conducted. They will focus on theoretical and / or practical aspects of the subject. Its valuation is 60% of the final grade.
- Applied work: Throughout the course, students will perform two works applied to matters of the subject, its valuation is 40% of the final grade.

Overall Assessment: Students who have not passed the subject with the system of continuous assessment, have to pass a global exam.

Evaluation criteria

In the written tests, controls and work participation will be evaluated:

Practical exercises must be properly raised. If a computer program is used in solving exercises, the code used and in any case the results are clearly explained be detailed. The probability distribution assigned to each random variable must be duly justified, identifying the value or values ??of the model parameters. Hypothesis testing will arise clear and defined manner.

# **4. Methodology, learning tasks, syllabus and resources**

## 4.1. Methodological overview

Different teaching methods will be used in the learning process of the data life cycle subject, depending on the objectives set and the skills developed:

**Lectures** will be used for the theoretical classes, in order to analyze and develop the fundamental

concepts of the subject. Didactic forms of participation will be used, involving the student, to develop the ability to organize, plan and make decisions.

**Computer lab session** and case resolution will be used to address the skills of using technological tools and instruments, problem solving, and the ability to analyze and search for information from other sources.

**Practice session** will be held that allow the student to develop the ability to adapt to new situations and to apply knowledge in professional practice.

As support, the Moodle platform will be used where the theoretical and practical materials of the subject will be published, as well as all the information necessary for its development, starting with the teaching guide itself.

Application videos will be incorporated, of the part seen in the theory sessions, with the software defined for the subject. These sessions should be seen by each student before attending the practical sessions.

The approach, methodology and assessment of this guide are intended to be the same for any teaching scenarios. They will be adapted to the social-health situation at any particular time, as well as to the instructions given by the authorities concerned.

## 4.2. Learning tasks

The course is organized with 4 hours of class a week for the 15 weeks of the semester. Some of these hours are taught in the computer room, and in them, the teacher explains the more practical aspects of the subject, which are reinforced with practical work by using statistical analysis programs.

Tutored self-employment: 2 hours per week for 15 weeks where the student works autonomously in the computer room in performing work.

Personal work: 60 hours

## 4.3. Syllabus

In the course, the important phases of the data life cycle will be seen:

**Acquisition stage.** In this first phase, what data is part of this project will be defined and the variables with their formats will be detailed. They will see the collection of data in the different sources, from where they are created or transferred to the system and their accessibility: the governance of the data. In this first phase, the legal aspects of data processing will be defined in all phases of the data life cycle. At this point, the profiles that will form those who will work on the project will be defined and they will be given the necessary data access permissions.

**Deployment and maintenance stage.** This phase of the data life cycle will see the collection of all the data that will be part of the data model, as well as long-term preservation and automation. It is about giving special importance to data management. In this phase there is an important point that is the validation of the data, which will have to be done in a continuous way to ensure the quality of the data and the efficient processing of the same.

**Exploitation stage.** In this stage of the data life cycle, it will be seen from the management of the quality of the data to its descriptive or predictive analysis depending on the scope of the data project that is being worked on.

**Destruction stage.** At this stage of the data life cycle, the steps to be taken to destroy them will be seen once their study has been completed.

## 4.4. Course planning and calendar

The activities that will be developed in the subject are the following:

- **Lectures**, in which the fundamental concepts that constitute the body of basic knowledge that must be learned to achieve the related learning outcomes are exposed later. Theoretical concepts are complemented by detailed examples that illustrate how they work within a specific context.
- **Practical classes**, in which problems are proposed to be solved using the methods and concepts previously considered. Discussion, participation, cooperation and reflection are encouraged in these classes.
- **Autonomous works**, two group assignments will be proposed in which it is proposed to solve a little more complex and extensive exercises that make use of the central concepts of the subject and can be solved exclusively within it, making use of the computer tools of numerical calculation and relevant symbolic.
- **Evaluation sessions**, in which students take written tests on certain well-specified parts of the agenda that is covered, or publicly expose the group work proposed in the previous activity.
- **Seminars**, in which the computer tools used in the rest of the activities are introduced, such as the statistical analysis language R.
- **Personal work**, in which students spend time outside of class to study the concepts taught in class, solve problems that are analogous and / or complementary to those considered in class, and prepare the proposed work.
- **Global assessment test**, which includes both a written test and the assignments that may have been done during the course, although at this point the proposed assignments that have not been satisfactorily resolved in advance

can be returned at this point. There are two global tests, one for each official call, and both take place after the end of the classes and when the rest of the activities have concluded and been evaluated.

**Presentation of the subject:**

In the first session of the course, the objectives and contents of the subject are explained in detail, the teaching methodology used in the classes is presented and the evaluation criteria are clearly exposed.

**Continuous assessment system:** Two intermediate problem solving tests will be carried out using computer tools.

The assessable practical tests will be fixed in the hours of practical sessions

These dates are set at the beginning of the course by the teacher, and may be modified with prior notice if the development of the calendar so requires.

**Official announcement:** According to the calendar established by the center, in the examination period, the student will take a global test that will consist of a written exam (PE) where the competencies will be evaluated, as well as the two assignments.

These dates are set at the beginning of the year from the management of the center.

The material that will be made available to students throughout the course, as well as the examination sessions and the results of these will be displayed in the ADD.

#### **4.5. Bibliography and recommended resources**

<http://psfunizar10.unizar.es/br13/egAsignaturas.php?codigo=29525>