Universidad Zaragoza

Academic Year/course: 2023/24

30253 - Data Warehouses and Data Mining

Syllabus Information

Academic year: 2023/24
Subject: 30253 - Data Warehouses and Data Mining
Faculty / School: 110 - Escuela de Ingeniería y Arquitectura
326 - Escuela Universitaria Politécnica de Teruel
Degree: 439 - Bachelor's Degree in Informatics Engineering
443 - Bachelor's Degree in Informatics Engineering
ECTS: 6.0
Year: 4
Semester: First semester
Subject type:
Module:

1. General information

Students will become familiar with the design of data warehouses and their subsequent exploitation using data mining techniques . Theoretical concepts, the different existing design techniques, and the different approaches that can be adopted will be discussed in order to formulate questions about the data warehouse. Whenever possible, use real examples . Computer tools that support the solution of this type of problems will also be discussed.

Since data warehouses enable the extraction of information useful for decision making in organizations, can also be used in scenarios that address SDG challenges. Therefore, it is considered that the acquisition of the learning results of the subject provides training and competence to the student to contribute to some extent to the achievement of the Sustainable Development Goals in a cross-cutting manner.

2. Learning results

- Be able to describe the basic aspects of data warehouses and explain the processes required for their design.
- Be able to design a data warehouse effectively.
- Know the basics of data warehouses and is able to search for additional information on other advanced related topics.
- Be able to compare various data warehouse designs and choose an appropriate commercial system for implementation.
- To know the basic techniques of machine learning applied to data mining.
- Use basic techniques in real problems.

3. Syllabus

1. Introduction to data warehouses
- Basic concepts:
  - User requirements analysis.
  - Life cycle.
  - The problem of data source reconciliation.
  - OLTP versus OLAP transactions.
- Data warehouse architecture:
  - Conceptual, logical and physical design.
  - ETL procedures.
- Commercial systems.

2. Introduction to data mining:

- Knowledge and data discovery.

- Web Mining.

- Data analysis tools.

- Areas of application, e.g:

  - Decision-making (banking-finance-insurance, marketing, health/demographic policies, etc.).
  - Industrial Processes.
  - Reverse Engineering.

4. Academic activities

Lectures: sessions with the teachers in which the subject matter will be explained: 30 hours.

Problems and cases: sessions to solve practical cases presented by the teacher: 15 hours.

Laboratory practices: 15 hours.

Study of the subject; class preparation; practical activities: 84 hours.

Assessment tests. 6 hours.

Estimated hours for each type of activity are indicated.

5. Assessment system

Continuous assessment:

1. Practical computer work (internships): 40% of the final grade.
2. Practical work/projects: 20% of the final grade.
3. Written test with theoretical and practical assumptions: 40% of the final grade.

In the continuous assessment, the delivery of assignments and their presentations will be made throughout the development of the subject on the dates set by the teaching staff. The final grade will be obtained as the weighted average of the parts of which consists the assessment, taking into account that it is necessary to have at least 5.0 points out of 10.0 in each of them. In case of not reaching this minimum in any of the parts, the final grade will be the minimum between 4.0 and the result of weighting with the percentages of each part.

Final global test:

It will include the written test. In addition, students who have not submitted the practices and/or assignments regularly on the established partial due dates (or whose submissions are deficient) must submit said internships and/or assignments as part of the overall final exam, and must also pass an additional related test consisting of the resolution of a statement. The final grade is calculated as in the continuous assessment.