COURSE DATA

Data Subject	
Code	36515
Name	Data Mining in Business
Cycle	Grade
ECTS Credits	6.0
Academic year	2023 - 2024

Study (s)

Degree	Center	Acad. year	Period
1332 - Degree in Business Intelligence and Analytics	Faculty of Economics	2	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
MARTINEZ DE LEJARZA ESPARDUCER, IGNACIO MA	110 - Applied Economics

SUMMARY

"Data mining in business" is a basic training subject assigned to the area of Quantitative Methods for Economics and Business that is taught in the first semester of the second year of the Degree of INTELLIGENCE AND ANALYTICS OF BUSINESS with a total teaching load of 6 ECTS credits.
Within the framework of a degree clearly aimed at training business professionals with deep knowledge of the analysis and process of large volumes of information, it is necessary to provide students with tools and methods for extracting knowledge from the basics. of complex and extensive data.
The reality of the company is multidimensional and multi-individual and therefore generates large volumes of information that require adequate treatment, able to choose the fundamental in order to consider all the information that is relevant for decision making. business.
Knowing how to handle large masses of data, sort them, classify them, detect the most important factors behind their high number of variables or classify individuals into groups of homogeneous behavior are, among others, some of the the purposes pursued with data mining techniques.

Therefore, the cataloging of individuals (customers, suppliers, products, etc.) into compact groups of

similar behavior from the available information; the use of large masses of information for the classification of customers, suppliers, products, etc. and / or the prediction of their behavior and the establishment or discovery of patterns of joint occurrence are some of the problems to be solved with the development of matter.

Without losing sight of the eminently practical orientation that inspires the degree, the subject aims to make a rigorous tour of both the main problems of extracting knowledge from the information available and the appropriate techniques and theoretical models for solving. of the same.

In the development of these issues it will always be tried, and in accordance with the objectives of other matters, not to lose sight of the practical applicability in the exploitation of the available information, the relations with the analysis of the information and the automation of the methods of analysis resulting indispensable the continuous appeal to the computer simulation and to the exploitation of the information; trying at all times to apply it to practical situations in the business environment.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

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

Other requirements

No prerequisites. It is assumed that to successfully complete this subject the student has a basic level of mathematics and statistics (the knowledge corresponding to first and second year of high school in the branch of science or social science) and is familiar with the contents of the subject Exploratory Data Analysis; Random, Uncertainty and Inference and other subjects taken previously and have already acquired some of the competencies of those previously scheduled on information management and use o

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

1332 - Degree in Business Intelligence and Analytics

- Students must be able to apply their knowledge to their work or vocation in a professional manner and have acquired the competences required for the preparation and defence of arguments and for problem solving in their field of study.
- Students must have the ability to gather and interpret relevant data (usually in their field of study) to make judgements that take relevant social, scientific or ethical issues into consideration.
- Students must be able to communicate information, ideas, problems and solutions to both expert and lay audiences.
- Students must have developed the learning skills needed to undertake further study with a high degree of autonomy.

- Acquire basic training that can be used to learn new methods and technologies and to adapt to new situations in academic and professional areas.
- Be able to solve problems and to communicate and spread knowledge, skills and abilities, taking account of the ethical, egalitarian and professional responsibility of the activity of business intelligence and analytics.
- Be able to make autonomous decisions in digital environments characterised by the abundance and dynamism of data.
- Know and know how to properly use the appropriate quantitative and qualitative methods to reason analytically, evaluate results and predict economic and financial magnitudes.
- Be able to apply analytical and mathematical methods for the analysis of economic and business problems.
- Be able to plan, organise, monitor and evaluate the implementation of business strategies.
- Demonstrate skills for analysis and synthesis.
- 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.
- Apply methods and techniques of analysis, synthesis and graphical representation by means of software tools.
- Express situations of uncertainty and randomness using mathematical, synthetic and graphic languages.
- Reorganise and restructure variables and databases.
- Use data mining software.
- Manage and distinguish the concepts of universe, population, sample, parameters and estimators in real problems.
- Use software tools to solve problems under uncertainty.
- Make predictions using appropriate software tools to manage time series.
- Apply supervised machine learning techniques using software.
- Apply unsupervised and semi-supervised machine learning techniques using software.
- Apply probability and non-probability sampling.
- Use software to collect and analyse survey data.

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

The aim of the course is that students could be able to model and deal with the practical situations in which it is necessary to extract non-explicit knowledge from large amounts of information. The type of knowledge to obtain will have to do above all with obtaining classification and prediction criteria, cataloging and clustering and generating patterns of occurrence.

Therefore, the student is intended to master the main models of Classification, clustering, association as well as dimensionality reduction techniques such as factor analysis and correspondence.

Additional pursued objective is also so that the student manages these methods with powerful computer tools that can be integrated into the analysis of situations with large masses of data.

DESCRIPTION OF CONTENTS

1. Data mining and knowledge extraction

1.1 Data mining and other disciplines
1.2.The (from data) extracting knowledge process: Preparation, modeling, evaluation and interpretation.
1.3.Main problems / tasks and methods / techniques of data mining: pattern extraction, clustering, classification, prediction and association.
1.4 Data mining, machine learning and big data

2. Data mining methodology.

2.1. Inductive learning and modeling
2.2. Training and validation (Training and test)
2.3. Expressiveness: Fit, Overfit, and Underfit
2.4. Validation and interpretation. Validation procedures

3. Classification. Classifiers and predictors

3.1. Functional methods
3.2. Bayesian methods
3.3. Rule-based methods
3.4. Trees
3.5. Neural methods
3.6. Predictors
3.7. Business applications

4. Clustering.

4.1. K-means and other direct methods
4.2. Hierarchical methods
4.3. Self-organizing maps and mesh methods.
4.4. Other methods.
4.5. Business applications.

5. Extraction of association rules

5.1.Extraction of association and dependency rules. Support, confidence, interest and correlation.
5.2. Multilevel Association Rules
5.3. Sequential rules
5.4, Sequential rules and Bayesian networks.

6. Reduction of the dimension.

6.1. Principal component analysis and factor analysis.
6.2. Correspondence analysis

7. Advanced issues in data mining and complex data mining

7.1. Extraction, selection and model combination (bagging)
7.2. Spatio-temporal data mining
7.3. Sequential and streaming data mining
7.4. Applications to multimedia, web mining and text mining

WORKLOAD

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

TEACHING METHODOLOGY

The development of the course is structured fundamentally around differentiated theoretical and practical sessions. In the theoretical sessions, the teacher will explain the concepts and methods of applying them to the analysis of situations involving the handling of large volumes of information from which you want to extract knowledge without neglecting the applications and the essential aspects of use and implementation through the appropriate Software (R, and Weka, fundamentally, without ruling out the use of other tools, CAEST or SPSS, for example) complementing the applications with practical situations and examples. Instead, the practical sessions will involve the student's use of these tools,

concepts and methods to solve practical questions under the tutelage of the teacher.

The predominant teaching method in the theoretical classes will be the participatory master class. This methodology allows the large groups of students to be led in an organized way, offering the advantages of a master class without limiting the participation of the students and the teacher-student interaction.

An attempt will be made to encourage participation and discussion in class, in order to offer the student a direct involvement with the content.

In the practical sessions, the teacher will propose to the students situations (real or

fictitious) for the resolution of problems or case studies that they must solve with the application of techniques and the use of appropriate computer programs, carrying out, if appropriate, oral presentations or debates ..., individually and / or in teams. In the practical classes, projects and situations will be proposed that the students must solve by delivering the outputs that are determined in a timely manner.

EVALUATION

Evaluation will be carried on fundamentally from the following double procedure:

1.- Theoretical / practical exam with the help of computer systems, in which a situation will have to be solved in which the analysis methods and concepts developed during the course will have to be applied to solve the questions raised. The exam will account for 40% of the final mark for the course and a minimum of 3 out of 10 will be required to obtain an average mark.

2.- Evaluation of the practical activities developed by the student during the course, based on the preparation of papers / reports and / or oral presentations, with defense of the positions developed by the student. Continuous assessment will account for 60% of the course grade.

REFERENCES

Basic

- -HERNANDEZ, J., Ramirez,M.J. y Ferri,C.: Introducción a la minería de datos Peearson Prentince Hall, 2010.
  -LARA,J.A.: Business intelligence. Ediciones CEF, 2016
  -PEREZ, M: Minería de datos . RC libros 2014
  -WITTEN,I.; FRANK,E.;HALL,M.; PAL,C.: DATA MINING. Elsevier, 2017

Additional

- - CEACES, Proyecto (Contenedor Hipermedia de Estadística Aplicada a las Ciencias Económicas y Sociales). Universitat de València. ON LINE: http://www.uv.es/ceaces
  - CRAWLEY, M.J.: The R book https://www.cs.upc.edu/~robert/teaching/estadistica/TheRBook.pdf
  -KERNS, G.J. : Package:prob-The R-project for Statistical Computing https://cran.rproject.