COURSE DATA

DATA SUBJECT

Code: 36515
Name: Data Mining in Business
Cycle: Undergraduate Studies
ECTS Credits: 6
Academic year: 2025-26

STUDY (S)

Degree	Center	Acad. year	Period
1332 - Degree in Business Intelligence and Analytics	Facultat d'Economia	2	First quarter

SUBJECT-MATTER

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

COORDINATION

PAVIA MIRALLES JOSE MANUEL

SUMMARY

"**Data Mining in Business**" is a core course in the area of Quantitative Methods for Economics and Business. It is taught in the first semester of the second year of the Bachelor's Degree in **Business Intelligence and Analytics**, with a total workload of 6 ECTS credits.

Within a degree program clearly oriented toward training business professionals with deep knowledge of the analysis and processing of large volumes of information, it is necessary to provide students with solid understanding of the main methods and tools for extracting knowledge from complex and/or extensive databases.

Business reality is both multidimensional and multi-individual, generating large volumes of information that require appropriate handling, sometimes of a prospective nature, capable of identifying valuable knowledge relevant to business decision-making.

Being able to work with large datasets, organize and classify them, detect the most important factors behind a high number of variables, or classify objects (customers, products, suppliers, companies, etc.) into groups with homogeneous behavior are among the objectives pursued through data mining techniques. Categorizing objects into compact groups with similar behavior based on available information, using large

datasets to classify clients, suppliers, and products, and identifying or discovering joint occurrence patterns are some of the problems addressed in the course.

Without losing sight of the strongly practical orientation of the degree program, this course aims to provide a rigorous overview of the main knowledge extraction problems based on available information, as well as the appropriate theoretical techniques and models to solve them.

In this regard, the course is designed to train students in the most advanced prospective knowledge extraction methods and their application in business contexts. Understanding the philosophy behind unsupervised learning methods and being able to apply available algorithms to various business-related problems are among the core skills to be acquired. A deeper level of understanding, such as the ability to adapt existing algorithms to new data sets or specific problems, can begin to be developed through this course.

PREVIOUS KNOWLEDGE

RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

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

OTHER REQUIREMENTS

Although no restrictions have been established, it is assumed that to successfully complete this course, the student has previously been exposed to the contents of the degree courses "Exploratory Data Analysis and Databases", and "Probability, Uncertainty, and Inference"

COMPETENCES / LEARNING OUTCOMES

1332 - Degree in Business Intelligence and Analytics

Acquire basic training that can be used to learn new methods and technologies and to adapt to new situations in academic and professional areas.

Apply methods and techniques of analysis, synthesis and graphical representation by means of software tools.

Apply probability and non-probability sampling.

Apply supervised machine learning techniques using software.

Apply unsupervised and semi-supervised machine learning techniques using software.

Be able to analyse and search for information from diverse sources.

Be able to apply analytical and mathematical methods for the analysis of economic and business problems.

Be able to define, solve and present complex problems systemically.

Be able to learn autonomously.

Be able to make autonomous decisions in digital environments characterised by the abundance and dynamism of data.

Be able to plan, organise, monitor and evaluate the implementation of business strategies.

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 use ICT, both in academia and in professional practice.

Demonstrate skills for analysis and synthesis.

Express situations of uncertainty and randomness using mathematical, synthetic and graphic languages.

Know and know how to properly use the appropriate quantitative and qualitative methods to reason analytically, evaluate results and predict economic and financial magnitudes.

Make predictions using appropriate software tools to manage time series.

Manage and distinguish the concepts of universe, population, sample, parameters and estimators in real problems.

Reorganise and restructure variables and databases.

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 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.

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.

Use data mining software.

Use software to collect and analyse survey data.

Use software tools to solve problems under uncertainty.

DESCRIPTION OF CONTENTS

1. Introduction. Data mining and unsupervised learning.

2. Clustering and market segmentation.

3. Factor analysis and dimensionality reduction. Variable selection.

4. Correspondence analysis and positioning.

5. Anomaly detection.

6. Association rules. Decision trees.

7. Multidimensional scaling. Recommendations.

8. Other topics: Missing data. Classifiers. The Naive Bayes classifier. Conjoint analysis...

WORKLOAD

PRESENCIAL ACTIVITIES

Activity	Hours
Theory	15,00
Computer classroom practice	45,00
Total hours	**60,00**

NON PRESENCIAL ACTIVITIES

Activity	Hours
Attendance at other activities	0,00
Individual or group project	20,00
Independent study and work	20,00
Preparation of lessons	20,00
Preparation for assessment activities	10,00
Resolution of case studies	20,00
Total hours	**90,00**

TEACHING METHODOLOGY

The course is mainly structured around practical sessions, where the resolution of applied examples is used to introduce, consolidate, and reinforce the theoretical content delivered in lectures.

In the theoretical sessions (1 hour per week), the main topics of the course will be presented, introducing key elements and concepts and placing them within the context of prediction problems in a business environment. The primary teaching method in the lectures will be participatory lectures.

In the practical sessions (3 hours each), the instructor will present students with real or simulated problem situations or case studies that they will need to solve using appropriate techniques and software tools. Students will deliver oral presentations, participate in debates, and work individually and/or in teams. Projects and scenarios will be proposed in class, and students are expected to deliver outputs on time and in the required format.

EVALUATION

The evaluation of the course is based on a dual approach:

1. **Theoretical/practical exam**, which may involve the use of computational tools. Students will be required to solve a given problem or answer a set of questions applying the analytical methods and concepts developed throughout the course. The exam will account for between **20% and 40%** of the final grade, and a **minimum score of 4 out of 10** will be required to be eligible for grade averaging.
2. **Assessment of practical activities** carried out during the course, based on written reports, presentations, and the defense of the student's own work. Continuous assessment will account for between **60% and 80%** of the final grade.

Final percentages will be specified in a document posted on the virtual classroom after consultation with

students.

REFERENCES

Basics

- Berthold, M. R., Borgelt, C., Höppner, F. Klawonn, F. y Silipo, R. (2020). Guide to Intelligent Data Science. How to Intelligently Make Use of Real Data. Springer.
- Hair, J.E., Andersson, R.E., Tatham, R.L. y Black, W.C. (1998). Multivariate Data Analysis. 5th Edition. Prentice Hall.
- Hastie, T., Tibshirani, R. y Friedman, J. (2008) The Elements of Statistical Learning. Data Mining, Inference and Prediction. 2nd edition. Springer.
- Hernández, J., Ramirez,M.J. y Ferri,C. (2010). Introducción a la minería de datos. Pearson Prentice Hall.
- Kuhn, M. (2019) The caret Package. https://topepo.github.io/caret/index.html
- Greenacre, M. J. (2008). *La práctica del análisis de correspondencias*. Fundación BBVA.

Complementary

- Aggarwal, C.C. (2016) Recommender Systems: The Textbook. Springer.
- James, G., Witten, D., Hastie, T. y Tibshirani, R. (2021) An Introduction to Statistical Learning with Applications in R. 2nd edition. Springer.
- Gorakala, S. K. y Usuelli, M. (2015) Building a Recommendation System with R. Packt Publishing.
- Theobald, O. (2018) Machine Learning for Beginners: Make Your Own Recommender System. Scatterplot Press.
- Wiebold, T. (2019). Multivariate Statistical Analysis using R.
- Wickham, H. y Grolemund, G. (2017) R for Data Science. O'Reilly Media, Inc.