COURSE DATA

DATA SUBJECT

Code: 36404
Name: Data mining and machine learning
Cycle: Undergraduate Studies
ECTS Credits: 6
Academic year: 2026-27

STUDY (S)

Degree	Center	Acad. year	Period
1400 - Degree in Computer Engineering	Escola Tècnica Superior d'Enginyeria	4	Second quarter
1403 - Degree in Telematics Engineering	Escola Tècnica Superior d'Enginyeria	4	Second quarter
1407 - Degree in Multimedia Engineering	Escola Tècnica Superior d'Enginyeria	4	Second quarter

SUBJECT-MATTER

Degree	Subject-matter	Character
1400 - Degree in Computer Engineering	Optional subject	ELECTIVES
1403 - Degree in Telematics Engineering	Optional subjects	ELECTIVES
1407 - Degree in Multimedia Engineering	Optatividad	ELECTIVES

COORDINATION

FERRI RABASA FRANCESC JOSEP

SUMMARY

Data mining and machine learning fundamentals are introduced from a computer science point of view. In particular, multimodal data processing is introduced along with associated learning algorithms including, parametric and nonparametric statistical, neural and metaheuristic methods. Several applications are considered to illustrate the course contents, e.g. content-based image retrieval, emotion detection from writing or web-browsing patterns, identity recognition, etc.

PREVIOUS KNOWLEDGE

RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

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

OTHER REQUIREMENTS

None

COMPETENCES / LEARNING OUTCOMES

1400 - Degree in Computer Engineering

C1 - Ability to know the fundamentals, paradigms and techniques in the field of intelligent systems, and to analyse, design and build computer systems, services and applications that use these techniques in any field of application.

C2 - Ability to acquire, obtain, formalise and represent human knowledge in a computable form for solving problems through a computer system in any field, particularly in those related to aspects of computing, perception and action in intelligent environments.

C3 - Ability to recognise and develop computational learning techniques and to design and implement applications and systems that use them, including those for the automatic retrieval of information and knowledge from large volumes of data.

1405 -

G1 - Be able to relate and structure information from different sources and to integrate ideas and knowledge. (RD1393/2007)

MM28 - Be able to solve problems with initiative, decision-making and creativity and to communicate and transmit the knowledge, abilities and skills of a multimedia engineer.

DESCRIPTION OF CONTENTS

1. Machine learning foundations and statistical background

Introduction to machine learning. Data representation, preprocessing and visualization

2. Parametric/nonparametric learners and distance-based methods

Bayes rule. Errors. Discriminant functions. Distance-based and neighbor-based techniques

3. Linear machines and extensions: kernels, layers and depth

Perceptrons. Adaline and extensions. Support vector machines. Introduction to kernels.

4. Unsupervised methods, estimates and clustering.

Clusters and quantization. Hierarchical methods. Prototype-based methods. Parametric and non parametric estimation. Semisupervised methods.

5. Extension and applications

Content-based multimedia information retrieval. Behavior pattern discovery. Optimal representations. Automatic identity recognition.

WORKLOAD

PRESENCIAL ACTIVITIES

Activity	Hours
Theory	30,00
Laboratory	20,00
Classroom practices	10,00
Total hours	**60,00**

NON PRESENCIAL ACTIVITIES

Activity	Hours
Attendance at other activities	0,00
Individual or group project	15,00
Independent study and work	30,00
Preparation of lessons	30,00
Preparation for assessment activities	10,00
Resolution of case studies	5,00
Total hours	**90,00**

TEACHING METHODOLOGY

-Theory and problem teaching with student participation

-Discussion sessions and problem solving

-Lab sessions

-Quiz solving both in class and remotely

-Monograph writing and bibliographic search, both individually and group wise

EVALUATION

Weighted average of the following items

(Weights for the 2nd round in brackets):

Assistance and participation: 10% (5%, non recoverable)

Partial tests: 15% (7.5%, non recoverable)

Labs: 25% (12.5%, non recoverable)

Final test: 50% (75%)

All individual marks must be superior to 4 out of 10 in order to compute the final mark.

In any case, the evaluation of this subject will be done in compliance with the University Regulations in this regard, approved by the Governing Council on 30th May 2017 (ACGUV 108/2017)

Copying or plagiarism of any activity that is part of the evaluation will result in the impossibility of passing the course, and the student will then be subject to the appropriate disciplinary procedures indicated in the ACTION PROTOCOL FOR FRAUDULENT PRACTICES AT THE UNIVERSITY OF VALENCIA (ACGUV 123/2020).

REFERENCES

- C.M. Bishop. Pattern recognition and machine learning, 2006
- D.J. Hand, H. Mannila, P. Smith. Principles of data mining, 2001

- R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2n ed, 2001
- R. Garreta, G. Moncecchi. Scikit-learn. Machine learning in Python, 2013
- E. Alpaydin, Introduction to machine learning, 2010
- W. McKinney. Python for data analysis, 2013
- D.G. Stork, E. Yom-Tov, Pattern Classification. Computer manual in Matlab, 2004
- S. Theodoridis, K. Koutroumbas, Pattern Recognition, 3r ed, 2006