SUBJECT TEACHING GUIDE

G686 - Machine Learning and Data Mining

Degree in Computer Systems Engineering

Academic year 2023-2024

1. IDENTIFYING DATA				
Degree	Degree in Computer Systems Engineering		Type and Year	Optional. Year 3
Faculty	Faculty of Sciences			
Discipline	Subject Area: Computing Mention in Computing			
Course unit title and code	G686 - Machine Learning and Data Mining			
Number of ECTS credits allocated	6	Term	Semester based (2)	
Web				
Language of instruction	Spanish	English Friendly	Yes	Mode of delivery	Face-to-face

Department	DPTO. MATEMATICAS, ESTADISTICA Y COMPUTACION
Name of lecturer	CRISTINA TIRNAUCA
E-mail	cristina.tirnauca@unican.es
Office	Facultad de Ciencias. Planta: + 1. DESPACHO CRISTINA TIRNAUCA (1046)
Other lecturers	INES GONZALEZ RODRIGUEZ

3.1 LEARNING OUTCOMES
- Understanding and being able to choose and apply data mining and machine learning techniques, being able to distinguish between supervised and unsupervised learning, descriptive and predictive models, and understanding the most used algorithms in data analysis.

4. OBJECTIVES
Understanding the concepts and terminology of data mining techniques.
Recognizing the benefits of a systematic use of knowledge extraction techniques for obtaining predictive or descriptive models and patterns.
Knowing different machine learning and statistical techniques used in data mining , their potential, their computational cost and their limitations.
Choose, for a particular problem, which data mining techniques are more appropriate.
Being able to perform data analysis using a particular data mining tool (Weka or Knime).
Evaluate the quality of a model using simple evaluation techniques (cross-validation).
Implement specific data mining algorithms.

6. COURSE ORGANIZATION	
CONTENTS	
1	Lecture classes.
1.1	Introduction. Predictive and descriptive models. Supervised and unsupervised models. Python tutorial.
1.2	Linear regression. The gradient descent algorithm and the analytical method. Regularization.
1.3	Classification. Logistic regression (the gradient descent algorithm) with regularization. SVM.
1.4	Clustering. Hierarchical vs. partitional, exclusive vs. diffuse methods. K-means algorithm. Expectation Maximization.
1.5	Association rules. Frequent sets. Closures. Redundancy in association rules.
1.6	Artificial neural networks.
1.7	Meta-predictors. Bagging. Boosting. AdaBoost.
1.8	Dimensionality reduction (Principal Components Analysis).
1.9	Model selection and evaluation.
1.10	Grammatical inference
1.11	Weka and Knime.
2	Review of the main concepts discussed in class.
3	Individual project.
4	Final exam.

7. ASSESSMENT METHODS AND CRITERIA

Description	Type	Final Eval.	Reassessn	%
Final exam	Written exam	Yes	Yes	50,00
Final project	Work	No	Yes	20,00
Team work on practical applications	Work	No	Yes	20,00
Problems	Activity evaluation with Virtual Media	No	Yes	10,00
TOTAL				100,00

Observations
The grade obtained during the course is taken into consideration only for the ordinary examination session. In the extraordinary examination session, the exam has a 100% weight.

Observations for part-time students
For those students that for a justified reason (part-time students) could not follow the continuous evaluation, the exam has a 100% weight (in both ordinary and extraordinary sessions).

8. BIBLIOGRAPHY AND TEACHING MATERIALS

BASIC
Principles of data mining (David Hand, Heikki Mannila, Padrhraic Smyth).