



COURSE: DATA MINING		
DEGREE: DEGREE IN INFORMATICS	YEAR: 4º	SEMESTER: 1

WEEKLY PLANNING									
WEEK	SESSION	DESCRIPTION	GROUPS		SPECIAL ROOM FOR SESSION (Computer class room, audio-visual class room)	Indicate YES/NO If the session needs 2 teachers: Maximum 4 sessions	WEEKLY PROGRAMMING FOR STUDENT		
			LECTURE	SEMINAR			DESCRIPTION	CLASS HOURS	HOMEWORK HOURS Maximum 7 H
1	1	Introduction to Data Analysis and Data Mining. Concepts, applications and methodology. Phases in a Data Mining Project.	X			NO		1,5	7
1	2	Methodology and process of data analysis. Types of data and learning tasks	X			NO		1,5	
2	3	Data preparation. Cleaning, transformation, aggregation, discretization						1,5	7
2	4	Practical exercises. Data transformation and basic statistical data analysis						1,5	
3	5	Classical techniques. Statistical inference of parameters, variable comparison and linear regression	X			NO		1,5	7
3	6	Practical exercises. Statistical analysis and data visualization. Some well-known tests		X		SI		1,5	
4	7	Supervised learning. Bayesian classifiers	X			NO		1,5	7
4	8	Practical exercises: Building Bayesian classifiers		X		SI		1,5	
5	9	Symbolic classifier models. Decision trees and rule induction	X			NO		1,5	7

5	10	Practical exercises: Building symbolic classifiers		X		SI		1,5	
6	11	Machine learning for regression. Prediction trees and neural networks	X			NO		1,5	7
6	12	Practical exercises: learning of prediction functions		X		SI		1,5	
7	13	Evaluation of data mining models. Classification and prediction errors.	X			NO		1,5	7
7	14	Practical exercises: computing performance of a machine learning method		X		SI		1,5	
8	15	Validation techniques and comparative analysis. Cross validation and ROC curves	X						8
8	16	Laboratory session: Use of machine learning algorithms and evaluation analysis tools over real data samples		X	Computer Room	SI		1,5	
9	17	Attribute selection. Wrapper techniques for supervised selection	X			NO		1,5	7
9	18	Practical exercises: evaluation of attribute selection strategies		X		SI		1,5	
10	19	Nonsupervised learning techniques. Clustering numeric techniques (K-means, EM).	X			NO		1,5	7
10	20	Laboratory session: Evaluation tolos and model analysis		X	Computer Room	SI		1,5	
11	21	Symbolic hierarchical clustering. COBWEB. Evaluation	X			NO		1,5	7
11	22	Laboratory session: Implementing classification and prediction model building		X	Computer Room	SI		1,5	
12	23	Data association nonsupervised techniques. Evaluation	X			NO		1,5	7
12	24	Practical exercises: Building hypotheses of association rules		X		SI		1,5	
13	25	Other computing paradigms for data mining. Neural networks, SVMs, Genetic Algorithms, Fuzzy Systems,...	X			NO		1,5	7
13	26	Laboratory session: Evaluation of nonsupervised techniques (clustering, association) over real data sets		X	Computer Room	SI		1,5	
14	27	Techniques for aggregating classification models. Stacking, bagging, boosting	X			NO		1,5	7
14	28	Laboratory session: Performance analysis of advanced data		X	Computer	SI		1,5	

		mining techniques			Room				
SUBTOTAL								42	+ 98 = 140
15		Tutorial and extra class, assignments presentations, etc						5	
16-18		Preparation for exams and assessment						5	
TOTAL								150	

LABORATORIES CLASSES PROGRAMMING*									
SESSION	WEEK	DESCRIPTION	LABORATORY		SPECIAL ROOM FOR SESSION (Computer class room, audio-visual class room)	WEEKLY PROGRAMMING FOR STUDENT			
						DESCRIPTION	CLASS HOURS	HOMEWORK HOURS Maximum 7 H	
3	6	Complementary session: Excel tools for data analysis. Basic statistical and graphical analysis. Common tests for variable analysis.		YES	Computer Room		1,5	2	
7	14	Complementary session: Data preparation for analysis, introduction to WEKA analysis tool. Format, functions, options, etc. Tutorial with examples and introduction to data repositories		YES	Computer Room		1,5	2	
8	16	Complementary session: Introduction to model building tools for classification, prediction and visualization with WEKA.		YES	Computer Room		1,5	2	
12	24	Complementary session: Evaluation and comparison of alternative models with WEKA. Introduction to "Experimenter" tools for analysis and comparison of machine learning algorithms		YES	Computer Room		1,5	2	
TOTAL							14		

* 6 hours of complementary laboratories classes in EPS