COURSE OUTLINE

GENERAL

SCHOOL	Sciences and Engineering		
ACADEMIC UNIT	Computer Science		
LEVEL OF STUDIES	1st Cycle		
COURSE CODE	COMP-344	**SEMESTER**	Fall
COURSE TITLE	Machine Learning and Data Mining II		

INDEPENDENT TEACHING ACTIVITIES *if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for the whole of the course, give the weekly teaching hours and the total credits*	**WEEKLY TEACHING HOURS**	**CREDITS**
	2.5	6
Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).		

COURSE TYPE *general background, special background, specialised general knowledge, skills development*	Special background
PREREQUISITE COURSES:	COMP-244
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	English
IS THE COURSE OFFERED TO ERASMUS STUDENTS	
COURSE WEBSITE (URL)	

LEARNING OUTCOMES

Learning outcomes

The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described.

Consult Appendix A

- *Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area*
- *Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B*
- *Guidelines for writing Learning Outcomes*

After completing the course, students are expected to be able to:

- Analyse complex data mining problems and formulate abstract, structured solutions.
- Apply core data mining concepts and techniques to solve practical problems in a variety of contexts.
- Pre-process and prepare datasets for analysis, selecting appropriate techniques for data transformation and cleaning.
- Implement and evaluate clustering methods to identify patterns and structure in data.

- Extract and interpret association rules from data to uncover meaningful relationships.
- Apply and assess anomaly detection algorithms on both real-time and offline data, selecting suitable approaches based on data characteristics and problem requirements

General Competences

Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?

Search for, analysis and synthesis of data and information, with the use of the necessary technology	*Project planning and management*
Adapting to new situations	*Respect for difference and multiculturalism*
Decision-making	*Respect for the natural environment*
Working independently	*Showing social, professional and ethical responsibility and sensitivity to gender issues*
Team work	*Criticism and self-criticism*
Working in an international environment	*Production of free, creative and inductive thinking*
Working in an interdisciplinary environment	*……*
Production of new research ideas	*Others…*
	…….

Search for, analysis and synthesis of data and information, with the use of the necessary technology
Adapting to new situations
Decision-making
Working independently
Project planning and management
Criticism and self-criticism
Production of free, creative and inductive thinking

SYLLABUS

1.Introduction to Data Clustering
   a. Applications
   b. Basic Concepts, Similarity Metrics, Distance Metrics
2.Partitional Clustering
   a. The K-means algorithm
   b. K-means as an optimization problem
3.Hierarchical Clustering
   a. Basic Agglomerative Hierarchical Clustering
   b. Strengths and Weaknesses
4.Density Based Clustering
   a. General Principles
   b. The DBSCAN Algorithm
   c. Subspace Clustering
5.Cluster Evaluation
   a. Unsupervised Cluster Evaluation: Cohesion and Separation
   b. Unsupervised Cluster Evaluation: Proximity Matrix
   c. Unsupervised Evaluation of Hierarchical Clustering
   d. Supervised Measures of Cluster Validity
   e. Assessing the Significance of Cluster Validity Measures

6. Prototype-Based Clustering
    a. Fuzzy Clustering
    b. Mixture Models
    c. Self-Organizing Maps
7.Graph Based Clustering
    a. Minimum Spanning Tree
8.Association Analysis
    b. Frequent Item Generation (the Apriori principle)
    c. Rule Generation
    d. The FP-Growth Algorithm
    e. Evaluation of Association Patterns
9.Anomaly Detection
    a. Preliminaries
    b. Statistical Approaches
    c. Proximity-Based
    d. Density –Based
    e. Clustering -Based

TEACHING and LEARNING METHODS - EVALUATION

DELIVERY *Face-to-face, Distance learning, etc.*	Face-to-face
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY *Use of ICT in teaching, laboratory education, communication with students*	*Use of ICT in teaching / Χρήση ΤΠΕ* *Communication with students / Επικοινωνία με Φοιτητές*
TEACHING METHODS *The manner and methods of teaching are described in detail.* *Lectures, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc.* *The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS*	
STUDENT PERFORMANCE EVALUATION *Description of the evaluation procedure* *Language of evaluation, methods of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open-ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art interpretation, other*	Mid-term exam, Projects, Final Examination, Participation/Homework Assignments/Quizzes

Activity	***Semester workload***
Lectures	35
Preparation	*33*
Coursework	*50*
Exams Preparation	*30*
Final Exam	2
Course total	***150***

Specifically-defined evaluation criteria are given, and if and where they are accessible to students.	

ATTACHED BIBLIOGRAPHY

Required Textbooks / Readings:

Authors	Title	Publisher	Year	ISBN
Tan, Steinbach, Kumar	*Introduction to Data Mining, 2nd Edition*	Pearson	2019	978-0133128901
Aurélien Géron	*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd Edition*	O'Reilly	2019	978-1492032649

Recommended Textbooks / Readings:

Authors	Title	Publisher	Year	ISBN
Han, Kamber, Pei	Data Mining: Concepts and Techniques, Third Edition	Morgan Kaufmann	2011	9380931913
Witten, Frank, Hall	Data Mining: Practical Machine Learning Tools and Techniques	Morgan Kaufmann	2011	0123748569