DAMA510 Module Outline

1. GENERAL

SCHOOL	School of Science and Technology		
PROGRAMME	Data Science and Machine Learning		
LEVEL OF STUDIES	Level 7 of the Hellenic and European Qualifications Framework		
MODULE CODE	DAMA510	**SEMESTER**	2
MODULE TITLE	Machine Learning		
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*		**HOURS**	**CREDITS**
Weekly workload: 32-33 hours x 13 weeks		420	15 ECTS
MODULE TYPE *Compulsory/Elective/ Mandatory Optional*	Compulsory		
PREREQUISITE MODULES	None		
LANGUAGE OF INSTRUCTION and EXAMINATIONS	English		
IS THE MODULE OFFERED TO ERASMUS STUDENTS	Yes		
MODULE WEBSITE (URL)	The Module has a dedicated space in HOU's digital learning platform (http://courses.eap.gr, http://study.eap.gr), which students and tutors can access using their credentials.		

2. LEARNING OUTCOMES

Learning outcomes

The learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the Module 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*

<u>Knowledge:</u>

Upon successful completion of the Module, students will be able to:

- Explain the key phases of the data science process and the role of the data scientist.

<u>Skills:</u>

Upon successful completion of the Module, students will be able to:

- Assess the quality and characteristics of input data, such as data types, missing values and outliers.
- Apply data preprocessing techniques, such as data cleaning, transformation, and feature scaling, using appropriate tools and languages.
- Perform dimensionality reduction to reduce complexity in high-dimensional data.
- Compute similarity and distance measures for numerical and categorical attributes.
- Apply clustering algorithms to discover groupings in unlabeled data.
- Apply frequent itemset mining and association rule learning to extract patterns from transactional data.
- Apply regression and classification models to labeled datasets.
- Analyze dataset characteristics and prepare data for supervised learning by handling imbalance, selecting attributes, and encoding variables.

- Select and engineer relevant features using dimensionality reduction and feature selection techniques to improve predictive accuracy. - Apply learning based on Support Vector Machines and neural networks to classification tasks, and analyze their performance and behavior. **Competences:** Upon successful completion of the Module, students will be able to: - Communicate insights and results effectively, using appropriate visualization tools. - Evaluate the suitability of clustering paradigms for different problems. - Evaluate association rule models based on validation metrics and domain relevance. - Evaluate model performance using a variety of metrics.
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 Module aim?* *Search for, analysis and synthesis of data and information, with the use of the necessary technology* *Adapting to new situations* *Decision-making* *Working independently* *Team work* *Working in an international environment* *Working in an interdisciplinary environment* *Production of new research ideas* *Project planning and management* *Respect for difference and multiculturalism* *Respect for the natural environment* *Showing social, professional and ethical responsibility and sensitivity to gender issues* *Criticism and self-criticism* *Production of free, creative and inductive thinking*
The general skills that the students will acquire are: - Search for, analysis and synthesis of data and information, with the use of the necessary technology - Adapting to new situations - Decision-making - Working independently - Team work - Project planning and management - Showing social, professional and ethical responsibility and sensitivity to gender issues - Production of free, creative and inductive thinking

3. SYLLABUS

Purpose of Module The students will acquire a background on the algorithmic aspects and the computational requirements of key data science and machine learning approaches. They will learn fundamental concepts and principles that underlie the techniques for extracting knowledge from data, they will become acquainted with a number of practical considerations regarding the analysis and the interpretation of the data, the assessment of the quality of the input data and the derivation of insights from the results of mining the data. After completing this module, they will be able to apply theory, and use languages, algorithms and tools to solve real world problems and to interpret and communicate findings to any kind of audience. **Subjects covered** Data preprocessing, Feature engineering, Outlier detection, Dimensionality reduction, Clustering, Frequent itemsets, Association rules, Decision Trees, Regression, Support vector machines, Neural networks

4. TEACHING and LEARNING METHODS - EVALUATION

DELIVERY *Face-to-face, Distance learning, etc.*	- Distance teaching and learning with three (3) Group Counseling Meetings (GCMs) of 4-hour duration during the academic semester on weekends. - Personal communication and feedback (advisory role of Adjunct Faculty).
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY	During GCMs and teaching the following are used: - Remote meetings tools (webex, Teams),

Use of ICT in teaching, laboratory education, communication with students	- Presentation software (PowerPoint, educational video - animations etc.), - Specialized software/databases for the subjects under study. In addition, students use office automation tools, web browsers and e-reader for digital books. Communication with the students is supported by: - The digital platform of HOU ((https://courses.eap.gr/login/index.php / https://study.eap.gr/login/index.php) (course information, educational material posts, announcements, messages, examination results, user groups, discussion forums etc.). - e-mail and messaging.
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*	(see table below)

Activity	***Semester Workload***
3 GCMs (x 4 hours)	12
5 Educational Activities (x 4 hours)	20
2 Semester Assignments (x 30 hours)	60
Individual Study time (25 hours x 13 weeks)	325
Final examination	3
Total Workload	420

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* *Specifically-defined evaluation criteria are given, and if and where they are accessible to students.*	**Students' evaluation – Grade assessment of a Module:** a. Five (5) multiple-choice (quiz) Educational Activities (Q), which contribute equally to the final grade with a value of 2% each. b. Two (2) Semester Assignments (A) which contribute equally to the final grade with a value of 10% each. The scoring of educational activities and assignments is activated only if the student succeeds an overall score equal to or above the base (≥50%) in the final or repeat exams. c. Final or repeat exams (E) contribute to the final grade of the module by 70%. The Final Grade of the module is calculated as follows (with 10 being the maximum Grade): Final Grade = (Q1 x 2%) + (Q2 x 2%) + (Q3 x 2%) + (Q4 x 2%) + (Q5 x 2%) + (A1 x 10%) + (A2 x 10%) + (E x 70%) **Language of evaluation**: English

5. INDICATIVE BIBLIOGRAPHY

- Recommended bibliography:

- Michael R. Berthold, Christian Borgelt, Frank Höppner and Frank Klawonn. (2010). Guide to Intelligent Data Analysis. How to Intelligently Make Sense of Real Data. Springer.
- EMC, E. S., & EMC, E. S. (Eds.). (2015). Data science and big data analytics: Discovering, analyzing, visualizing and presenting data. Wiley.
- W. N. Venables, D. M. Smith and the R Core Team (2022). An Introduction to R.

* H. Wickham, M. Cetinkaya-Rundel, G. Grolemund (2023). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (2nd edition). O' Reilly Media.

Additional digital (and multimedia) material will be made available online.

Related scientific Journals:

* Journal of Machine Learning Research (http://www.jmlr.org)
* Machine Learning (https://www.springer.com/journal/10994*)*