MODULE OUTLINE

1. GENERAL INFORMATION

SCHOOL	SCHOOL OF SCIENCE AND TECHNOLOGY		
PROGRAM COURSE	DATA SCIENCE AND MACHINE LEARNING		
LEVEL OF STUDY	POSTGRADUATE		
MODULE CODE	DAMA-60	**YEAR OF STUDY**	2nd
MODULE TITLE	Algorithmic Techniques and Systems for Data Science and Machine Learning		
INDEPENDENT TEACHING ACTIVITIES *in case credits are awarded for separate components/parts of the course, e.g. in lectures, laboratory exercises, etc. If credits are awarded for the entire course, give the weekly teaching hours and the total credits*		**HOURS**	**CREDIS**
Weekly teaching hours: 21-25 *x 32 weeks		840	30 ECTS
COURSE TYPE Compulsory, Optional, Optional mandatory	Compulsory		
PREREQUISITE MODULES:	The selection of DAMA60 does not require the simultaneous selection or completion of any other DAMA module.		
LANGUAGE OF INSTRUCTION AND EXAMS	ENGLISH		
THE MODULE IS OFFERED TO ERASMUS STUDENTS	No (due to the annual duration of the module)		
MODULE WEBSITE (URL)	https://www.eap.gr/education/postgraduate/annual/data-science-and-machine-learning/topics/#dama60 Each module has its own space in the Learning Management System of EAP (http://study.eap.gr), with controlled access (use of code) for students and teaching staff.		

2. LEARNING OUTCOMES

Learning Outcomes
● *The course learning outcomes, specific knowledge, skills and competences of an appropriate (certain) level, which students will acquire upon successful completion of the course, are described in detail. It is necessary to consult:*
• Define the importance of Big Data in Machine Learning Applications • Identify key characteristics of Big Data • Compare serial and parallel processing techniques for mining data • Define scalability and fault tolerance for machine learning algorithms • Apply Big Data tools for solving real life problems • Identify the Big Data ecosystem • Describe the benefits of Cloud Computing for Big Data Applications

- Recognize the importance of Distributed File Systems and Map-Reduce
- Create parallel algorithms for mining large volumes of data
- Apply similarity search by using MinHashing and Locality Sensitive Hashing
- Describe Data Stream Processing
- Apply specialized algorithms for dealing with stream data
- Recognize the technology underlying the principles of search engine operation Use frequent itemset mining through Apriori and its improvements
- Describe algorithms for clustering Big Data
- Identify key problems for mining data from Web Applications
- Describe algorithms for analyzing social network graphs
- Apply techniques for obtaining the important properties of large datasets
- Use Machine Learning algorithms for mining large datasets

General Competences

Taking into consideration the general competences that students/graduates must acquire (as those are described in the Diploma Supplement and are mentioned below), at which of the following does the course attendance aim?

Search for, analysis and synthesis of data and information by the use of appropriate technologies,
Adapting to new situations
Decision-making
Individual/Independent work
Group/Team work
Working in an international environment
Working in an interdisciplinary environment
Introduction of innovative research
Project planning and management
Respect for diversity and multiculturalism
Environmental awareness
Social, professional and ethical responsibility and sensitivity to gender issues
Critical thinking
Development of free, creative and inductive thinking
……
(Other…….citizenship, spiritual freedom, social awareness, altruism etc.) …….

Search for, analysis and synthesis of data and information by the use of appropriate technologies,
Adapting to new situations
Decision-making
Individual/Independent work
Project planning and management
Critical thinking
Development of free, creative and inductive thinking

3. MODULE CONTENT

The students will acquire a strong background on the data structures, the algorithmic aspects and the computational requirements of data mining and machine learning approaches for analyzing very large volumes of data. Among other topics, the module will emphasize on tools for the parallelization of different machine learning algorithms such as Hadoop and Map Reduce, Recommender Systems, Dimensionality Reduction, Finding Nearest Neighbors and Similar Sets, Clustering, Link Analysis, Association Rules and Frequent Itemsets. The students are also expected to build on the basic programming skills that they acquired in DAMA50 and DAMA51 and enhance their understanding of how to apply these skills on a project where they will be asked to work on real data sets and computational infrastructure through R and/or Python and Azure and/or KNIME.

The key subjects of the module are: A. “Machine Learning and Data Mining” B. “Big Data Analytics” C. ”Distributed Learning”

4. TEACHING METHODS--ASSESSMENT

<table>
<tr><td>MODES OF DELIVERY
Face-to-face, in-class lecturing, distance teaching and distance learning etc.</td><td>Distance education with five Group Counseling Meetings (OSS) during the academic year on weekends.</td></tr>
<tr><td>USE OF INFORMATION AND COMMUNICATION TECHNOLOGY
Use of ICT in teaching, Laboratory Education, Communication with students</td><td>We use :
Remote meetings tools (cisco webex),
Presentation software (e.g. PowerPoint),
Specialized software in the subjects under study (R, Python, KNIME, etc.).

Additionally, the students use office automation tools, web browsers, and e-reader for digital books.</td></tr>
<tr><td>MODULE DESIGN
Description of teaching techniques, practices and methods: Lectures, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, Internship, Art Workshop, Interactive teaching, Educational visits, projects, Essay writing, Artistic creativity, etc

The study hours for each learning activity as well as the hours of selfdirected study are given following the principles of the ECTS.</td><td>
<table>
<tr><th>Activity</th><th>Annual Workload</th></tr>
<tr><td>5 Tutorial Meetings (x 4 hours)</td><td>20</td></tr>
<tr><td>Multiple Choice Exercises (5 x 2 hours)</td><td>10</td></tr>
<tr><td>Preparation of Assignments (5 assignments x 10 hours)</td><td>50</td></tr>
<tr><td>Examination</td><td>3</td></tr>
<tr><td>Individual study</td><td>672-800</td></tr>
<tr><td>Total module workload (hours)</td><td>755-883</td></tr>
</table>
</td></tr>
<tr><td>STUDENT PERFORMANCE EVALUATION/ASSESSMENT METHODS
Detailed description of the evaluation procedures.

Language of evaluation, assessment methods, formative or summative (conclusive), multiple</td><td>Elaboration of written assignments during the academic year, the average of the grades of which participates in the formation of the final grade of module by 30%, if there is a passable in the final or repetitive examinations. In the final written exams the grade of the written assignments participates in the formation of the final grade of module by 70%.</td></tr>
</table>

choice tests, short- answer questions, open-ended questions, problem solving, written work, essay/report, oral exam, presentation, laboratory work, other……etc. *Specifically defined evaluation criteria are stated, as well as if and where they are accessible by the students*	All the criteria are posted, both in each written assignment (in the LMS study.eap.gr), as well as in the general regulation of HOU at: https://www.eap.gr/education/study-regulations/

(5) SUGGESTED BIBLIOGRAPHY

- P. Wentworth, J. Elkner, A.B. Downey & C. Meyers (2021). *Learn Python the right way: How to think like a computer scientist.* Chapters: 1 – 16, 18, 20.

o https://learnpythontherightway.com/

- J. Leskovec, A. Rajaraman & J.D. Ullman (2020). *Mining of Massive Datasets* (3rd edition). Cambridge University Press.

o http://www.mmds.org/

- P.-N. Tan, M. Steinbach, A. Karpatne & V. Kumar (2021). *Introduction to Data Mining* (2nd Edition). Pearson. Chapters: 3, 5, 7.

o https://www-users.cse.umn.edu/~kumar001/dmbook/index.php

- R. Zafarani, M.A. Abbasi & H. Liu (2014). *Social media mining: an introduction*. Cambridge University Press. Chapters: 2, 3, 6, 9.

o http://www.socialmediamining.info/book/

- T. Benschop & M. Welch (2019). *Statistical Disclosure Control for Microdata: Practice*. The World Bank.

o https://sdcpractice.readthedocs.io/en/latest/

- KNIME Analytics Platform Installation Guide (από https://docs.knime.com/).

- KNIME Workbench Guide (από https://docs.knime.com/).

- KNIME Best Practices Guide (από https://docs.knime.com/).