COURSE OUTLINE

(1) GENERAL

SCHOOLS	**ENGINEERING, NATURAL SCIENCES**		
ACADEMIC UNIT/UNITS	**DEPARTMENT OF COMPUTER ENGINEERING AND INFORMATICS, DEPARTMENT OF MATHEMATICS**		
TITLE OF MASTER'S DEGREE	***MSc in Data Driven Computing and Decision Making***		
LEVEL OF STUDIES	Postgraduate		
COURSE CODE	**DDCD 101**	SEMESTER	Spring
COURSE TITLE	**ALGORITHMS FOR DATA SCIENCE**		

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
Lectures, Tutorials, Laboratory	3(L), 2(T),2(Lab)	7.5
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*	Specialized general knowledge, skills development
PREREQUISITE COURSES:	Recommended prerequisite knowledge (undergraduate level): Algorithms, Data Structures, Linear Algebra, Probability and Statistics, Combinatorial Optimization, or equivalent.
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	Greek (English if there are foreign students)
IS THE COURSE OFFERED TO ERASMUS STUDENTS	Yes
COURSE WEBSITE (URL)	https://www.ceid.upatras.gr/webpages/faculty/zaro/teaching/alg-ds/index.html

(2) 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*

Upon conclusion of the course the students ought to be able to:

- Understand techniques and applications of fundamental and advanced algorithms for harnessing large data sets.
- Apply algorithmic techniques along with their efficient implementation for analyzing large data sets.
- Conduct extensive experimental methodologies for big data analytics and for empirically assessing algorithms on real-world big data sets.

Upon conclusion of the course the students are expected to have the following skills/competences:

- Ability to connect data to underlying real-world phenomena and think critically about conclusions drawn from big data analysis.
- Ability to conceive abstractions and processes allowing inferential procedures to be embodied in algorithms for large data sets, ensuring scalability, robustness and

understandability.

- Hands-on experimentation with real-world large data sets such as economics data, document collections, geographic data, social and collaboration networks, etc.

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
- Production of new research ideas
- Criticism and self-criticism
- Production of free, creative and inductive thinking

(3) SYLLABUS

- Introduction to advanced programming techniques for large data sets (e.g., MapReduce, Hadoop, Spark)
- Exploration of Frequent Itemsets
- Locality-Sensitive Hashing
- Clustering
- Dimensionality Reduction
- Algorithms for link analysis of large graphs (e.g., PageRank)
- Introduction to Social Networks
- Efficient Algorithms on Large Graphs
- Recommender Systems
- Large-Scale Machine Learning
- Algorithms for Large Data Streams
- Computational Advertising
- Complexity of Algorithms for Large Data Sets

(4) TEACHING and LEARNING METHODS - EVALUATION

DELIVERY *Face-to-face, Distance learning, etc.*	Face-to-face. Tutorials and laboratory sessions with exemplary solutions of exercises.
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY *Use of ICT in teaching, laboratory education, communication with students*	ICT methods are used in both teaching and communication with the students. Lecture slides and supplementary material are uploaded in the course's web site.
TEACHING METHODS *The manner and methods of teaching are described in detail.* *Lectures, seminars, laboratory practice,*	

Activity	*Semester workload*
Lectures	3*13=39
Tutorials (exercises)	2*13=26

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*	Programming exercises	2*13=26
	Individual study, preparation and problem solving	3*13=39
	Weekend study	3*13=39
	Study for exam preparation	6*3=18
	Course total (25 hours per ECTS unit)	**187**
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.*	The language of instruction and examination is Greek. Special provisions (lecture notes and examinations in English) can be made for foreign students. Evaluation (criteria can be found in the web site of the course): • Class participation (attendance is mandatory). • Programming exercises (15% of final mark), aiming at familiarizing students with the efficient management and analysis of real-world big data sets in practice. • Examination on the presentation of a state-of-the-art topic (35% of final mark), after thorough study and analysis of relevant bibliography, aiming at familiarizing students with state-of-the-art technology. • Final examination (50% of final mark): oral (and written, if necessary) examination of graded difficulty based on the presentation of the state-of-the-art topic, including also short-answer questions on all topics discussed in the class, algorithm design for problem solving, as well as proofs of algorithm correctness and complexity.	

(5) ATTACHED BIBLIOGRAPHY

- Suggested bibliography:

- Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman: Mining of Massive Datasets, 2nd edition, Cambridge University Press (2014).
- Steven Skienna. The Data Science Design Manual, Springer (2017).
- Avrim Blum, John Hopcroft and Ravindran Kannan: Foundations of Data Science (2017).
- Reza Zafarani, Mohammad Ali Abbasi, Huan Liu: Social Media Mining – An Introduction. Cambridge University Press (2014)
- Jimmy Lin and Chris Dyer: Data-Intensive Text Processing with MapReduce. Morgan & Claypool Publishers (2010).
- Lecture notes and slides uploaded in the web site of the course.

- Related academic journals:

- IEEE Transactions on Data Engineering.
- Algorithmica, Springer.
- VLDB Journal
- IEEE Trans. on Pattern Analysis and Machine Intelligence
- Machine Learning, Springer
- Data Mining and Knowledge Discovery
- Data Science Journal
- International Journal of Data Science and Analytics, Springer
- Artificial Intelligence, Elsevier
- Journal of Big Data, Springer
- Big Data Research, Elsevier

-Related academic conferences:

- IEEE Int'l Conference on Data Engineering
- IEEE Int'l Conference on Data Mining
- Intl. Conf. on Very Large Databases
- ACM SIGIR Conference on Research and Development in Information Retrieval
- ACM Symposium on Principles of Database Systems
- Int'l World Wide Web Conference
- ACM SIGMOD Int'l Conference on Management of Data
- ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining
- ACM Int'l Conference on Web Search and Data Mining
- SIAM Int'l Conference on Data Mining
- Int'l Conference on Computational Statistics