COURSE OUTLINE

(1) GENERAL

SCHOOLS	NATURAL SCIENCES		
ACADEMIC UNIT/UNITS	MATHEMATICS		
TITLE OF MASTER'S DEGREE	COMPUTATIONAL AND STATISTICAL DATA ANALYTICS (MCDA)		
LEVEL OF STUDIES	POSTGRADUATE		
COURSE CODE	MCDA203	**SEMESTER**	B
COURSE TITLE	DATABASES AND DATA MINING		

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	2	7.5
Laboratory exercises	1	
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*	General background
PREREQUISITE COURSES:	None
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	Greek
IS THE COURSE OFFERED TO ERASMUS STUDENTS	Yes
COURSE WEBSITE (URL)	https://eclass.upatras.gr/courses/MATH1085/

(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*

After successful completion of the course, the students will:

- know the principles of Databases (DB) and Database Management Systems (DBMSs),
- be able to clean and pre-process data,
- use an appropriate data mining algorithm to address a given problem,
- handle large amounts of data from business applications and social networks,
- use R language and its libraries related to data science.

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.
- Decision making.
- Working independently.
- Working in an interdisciplinary environment
- Production of free, creative and inductive thinking.
- Project planning and management.

(3) SYLLABUS

PART A: Theory

(**i**) Introduction to Databases. SQL. (**ii**) Preparing Data. The importance of data pre-processing and clearing. (**iii**) Missing data imputation. (**iv**) Introduction to supervised learning: decision trees, lazy learners, Bayesian classifiers, Ensembles of classifiers. (**v**) Introduction to regression: Multiple linear regression, Model Trees, Neural Networks. (**vi**) Dimensionality Reduction. Feature selection process. Principal Component Analysis with SVD. (**vii**) Un-supervised learning, Clustering. k-means algorithm. Hierarchical Clustering models, Density clustering. (**viii**) Association rules, Sparse matrices. (**ix**) Introduction to Big Data. Computational Methods for Large Data (Hadoop and MapReduce).

PART B: Laboratory

(**i**) Introduction to the R language for Data Science. (**ii**) Data Frames. Select data from a Data Frame and convert them to a Table. (**iii**) Introduction to SQL. Queries. Queries on multiple tables with the JOIN. (**iv**) Connection with R (SQLite). (**v**) Usage of R packages: sqldf, lattice, ggplot2, dplyr, party, C50, Rattle, mlr, randomForest, rpart, caret, factoextra, cluster, fpc, arules, arulesViz, RHadoop

(4) TEACHING and LEARNING METHODS - EVALUATION

<table>
<tr><td>DELIVERY
Face-to-face, Distance learning, etc.</td><td>Lectures (face to face)</td></tr>
<tr><td>USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY
Use of ICT in teaching, laboratory education, communication with students</td><td>• PowerPoint slides
• Support Learning through the eClass platform.</td></tr>
<tr><td>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</td><td>
<table>
<tr><th>Activity</th><th>Semester workload</th></tr>
<tr><td>Lectures</td><td>26</td></tr>
<tr><td>Laboratory</td><td>13</td></tr>
<tr><td>Study (no driven)</td><td>100</td></tr>
<tr><td>Solving suggested exercises</td><td>45</td></tr>
<tr><td></td><td></td></tr>
<tr><td>Final examination</td><td>2.5</td></tr>
<tr><td>Laboratory examination</td><td>1</td></tr>
<tr><td></td><td></td></tr>
<tr><td>Total number of hours for the Course (25 hours of work-load per ECTS credit)</td><td>187.5</td></tr>
</table>
</td></tr>
<tr><td>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.</td><td>Assessment Language: Greek
Assessment Language for Erasmus students: English

Assessment methods:
• Written examination (50%)
• Laboratory examination (25%)
• Exercises (25%)

Minimum passing grade: 5
Maximum passing grade: 10</td></tr>
</table>

(5) ATTACHED BIBLIOGRAPHY

- Beard, B. (2016). *Beginning SQL Server R Services: Analytics for Data Scientists*. Apress.
- Torgo, L. (2016). *Data Mining With R: Learning With Case Studies*. CRC press.
- Wickham, H., and Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, Inc.

(in Greek)

- Zaki, M.J. and Wagber, M. Jr. (2017). *Εξόρυξη και Ανάλυση Δεδομένων: Βασικές Έννοιες και Αλγόριθμοι*. Εκδόσεις Κλειδάριθμος ΕΠΕ.
- Βερύκιος, Β., Καγκλής, Β., και Σταυρόπουλος, Η. (2015). *Η Επιστήμη των Δεδομένων Μέσα από τη Γλώσσα R*. [ηλεκτρ. βιβλ.] Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο http://hdl.handle.net/11419/2965