COURSE OUTLINE

1. GENERAL

SCHOOL	NATURAL SCIENCES
DEPARTMENT	MATHEMATICS
LEVEL OF COURSE	UNDERGRADUATE

COURSE CODE	MAT_IC469	**SEMESTER OF STUDIES**	7th

COURSE TITLE	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*	**TEACHING HOURS PER WEEK**	**ECTS CREDITS**
Lectures and Laboratory exercises	4	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*	Elective course
PREREQUISITE COURSES:	Recommended prerequisite knowledge: PROGRAMMING WITH PYTHON, STATISTICAL INFERENCE I
TEACHING AND ASSESSMENT LANGUAGE:	Greek
THE COURSE IS OFFERED TO ERASMUS STUDENTS	Yes
COURSE WEBPAGE (URL)	https://eclass.math.upatras.gr/courses/MATHDEP235/

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 basic principles of Databases (DB) and Database Management Systems (DBMSs)
- be able to clean and pre-process data,
- use an appropriate machine learning 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.

- Search, analyze and synthesize data and information, using the necessary technologies.
- Decision making.
- Autonomous work.
- Working in an interdisciplinary environment
- Promote free, creative and inductive thinking.
- Project Design and management.

3. COURSE CONTENT

Data description with graphs and tables. Presentation of basic statistical measures for data description. Preparing Data. The importance of data pre-processing and clearing. Introduction to Databases. SQL. Introduction to supervised learning: decision trees, logistic regression. Introduction to regression: Multiple linear regression. Forecasts. Improving a model. The problems of over-fitting. Model Performance Evaluation. Dimensionality Reduction. Feature selection process. Principal Component Analysis with SVD. Un-supervised learning, Clustering. k-means algorithm. Application of Hierarchical Clustering models. Semi-supervised learning. Introduction to Metadata and Big Data. Computational Methods for Large Data (Hadoop and MapReduce).

Laboratory: (**i**) Introduction to the R language for Data Science. (**ii**) Create, select and compare categorical data using Factors. Save datasheets to Data Frames. Select data from a Data Frame and convert them to a Table. (**iii**) Basic graphics / visualization packages in R. (**iv**) Functions - Loops - Flow control. (**v**) Introduction to SQL. Queries. Queries on multiple tables with the JOIN. Subqueries. (**vi**) Rattle. (**vii**) R Hadoop.

ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΠΑΤΡΩΝ

4. TEACHING AND LEARNING METHODS - ASSESSMENT

TEACHING METHOD *Face-to-face, Distance learning, etc.*	Lectures (face to face)
USE OF INFORMATION AND COMMUNICATION TECHNOLOGIES *Use of ICT in teaching, laboratory education, communication with students*	✓ PowerPoint slides. ✓ Support Learning through the *eClass* platform.

TEACHING ORGANIZATION

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

Activity	*Semester workload*
Lectures	26
Laboratory exercises	26
Solving 4-5 exercises	30
Study for midterm examination	30
Midterm examination	2
Study for Final examination	33
Final examination	3
Total number of hours for the Course (25 hours of work-load per ECTS credit)	***150***

STUDENT ASSESSEMNT

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.

Assessment Language: Greek
Assessment Language for Erasmus students: English

Assessment methods:
✓ Written examination: 60%
✓ Laboratory examination: 15%
✓ Exercises: 10%
✓ Mid-term examination: 15%

Minimum passing grade: 5
Μαximum passing grade: 10

5. RECOMMENDED LITERATURE

(*in Greek*)
* Zaki Mohammed J. and Wagber Meira JR. *Εξόρυξη και Ανάλυση Δεδομένων: Βασικές Έννοιες και Αλγόριθμοι.* Εκδόσεις Κλειδάριθμος, 2017.
* Βερύκιος Β., Καγκλής Β. και Σταυρόπουλος Η. *Η επιστήμη των δεδομένων μέσα από τη γλώσσα R.* (e-book). Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Αποθετήριο "Κάλλιπος", 2015.

(in English)
* Wickham Hadley and Grolemund Garrett. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* O'Reilly, 2017.