Machine Learning

(1) GENERAL

SCHOOL	SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY		
ACADEMIC UNIT	DEPARTMENT OF STATISTICS		
LEVEL OF STUDIES	POSTGRADUATE		
COURSE CODE		**TRIMESTER**	**2 OR 5**
COURSE TITLE	Machine Learning		
INDEPENDENT TEACHING ACTIVITIES		**WEEKLY TEACHING HOURS**	**CREDITS**
		3	5
		30	5
COURSE TYPE	Specialization Course in Sports Data Science and Elective Specialization in Applied Sports Analytics		
PREREQUISITE COURSES:			
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	ENGLISH		
COURSE DELIVERY METHOD	Distance		
IS THE COURSE OFFERED TO ERASMUS STUDENTS	NO		
COURSE WEBSITE (URL)	https://cloud.aueb.gr/index.php/s/RgDzPEQrYYwae2w		

(2) LEARNING OUTCOMES

Learning outcomes
• **Understand foundational concepts of Machine Learning** and their applications in data analysis. • **Recognize overfitting** and the bias–variance trade-off. • **Differentiate between prediction and inference** in statistical learning. • **Apply cross-validation techniques**, including train/test/validation splits and Bootstrap methods. • **Gain knowledge of supervised and unsupervised learning approaches**. • **Use classification methods** such as LDA, k-NN, neural networks, decision trees, random forests, SVM, and Naive Bayes. • **Evaluate model performance** using metrics including accuracy, sensitivity, specificity, ROC/AUC, Lift, Brier score, and F1. • **Apply clustering techniques**, including hierarchical, k-means, density-based, and stochastic methods. • **Use dimensionality reduction methods** such as PCA and SVD for data compression and representation. • **Familiarize with shrinkage and variable selection techniques**, including LASSO.
General Competences
• Analysis and synthesis of data and information, including the use of appropriate

data technologies.
- Decision-making skills.
- Ability to work in an interdisciplinary environment.

(3) SYLLABUS

The course examines fundamental concepts related to Machine Learning. Topics include overfitting and the bias–variance trade-off, as well as the distinction between prediction and inference. Cross-validation techniques are discussed, including train/test/validation splits and Bootstrap methods. Various learning approaches are also covered, such as supervised and unsupervised learning, along with classification methods including LDA, k-NN, neural networks, decision trees, random forests, support vector machines (SVM), and Naive Bayes.

The course also addresses model performance evaluation using metrics such as accuracy, misclassification rate, sensitivity, specificity, ROC curves, AUC, Lift, Brier score, and F1. Clustering techniques are introduced, including distance-based methods, hierarchical clustering, k-means, stochastic model-based clustering, and density-based clustering. Finally, dimensionality reduction methods, such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD), as well as shrinkage and variable selection techniques like LASSO, are analyzed.

(4) TEACHING and LEARNING METHODS - EVALUATION

DELIVERY *Face-to-face, Distance learning, etc.*	Distance
METHOD AND FREQUENCY OF COMMUNICATING WITH THE STUDENTS	Remotely via email and through weekly office hours (and whenever necessary) via teleconferencing or in-person office visits.
ENSURING COMMUNICATION AMONG STUDENTS	Teleconference, Chat via eclass and/ or TEAMS, and QA sessions
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY	MS Teams, e-class, MS Outlook, R, Python, WayGround or Kahoot educational games and quizzes
REQUIRED EQUIPMENT AND TECHNOLOGY KNOWLEDGE	Camera, mic, PC, MS Office, and TEAMS
COURSE POLICY ON PLAGIARISM AND PLAGIARISM DETECTION TOOLS	Turnitin
COURSE POLICY ON USE OF AI TOOLS	The use of Artificial Intelligence is permitted with explicit reference to the bibliography (2) for the verbal correction of assignments and as long as students have understood the basic principles and methods of the course.
TEACHING METHODS	
STUDENT PERFORMANCE EVALUATION	Evaluation language is English. Evaluation methods include lab exercises, written assignments, and/or written examinations. Oral examinations of the assignments may be conducted if clarifications are needed or if there are suspicions of plagiarism or unauthorized use of Artificial Intelligence tools. Assessment criteria are provided on the course eClass platform and in the course materials.

Activity	***Semester workload***
Lectures	30
Assignment writing	40
Lab Exercise	20
Self-Study hours	40
Course total	***130***

(5) RECOMMENDED BIBLIOGRAPHY

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: With applications in R. Springer.
- Murphy, K. P. (2012). Probabilistic machine learning: An introduction. MIT Press.