DATA STRUCTURES

ACADEMIC YEAR 2024-2025

Degree	**B**achelor of **S**cience in **C**omputer **S**cience				
Qualification	**C**omputer **S**cience				
Professor	Varazdat Avetisyan				
Distribution of hours	CM 15h.	TD 0h.	TP 27h.	TI 0h.	ECTS 6

EXPECTED LEARNING OUTCOMES OF THE COURSE		
A- Knowledge		A1.1: Differentiate between various types of machine learning (supervised, unsupervised, reinforcement learning). A1.2: Recognize the historical context and key developments in the field of machine learning. A1.3: Identify and describe common supervised learning algorithms (e.g., linear regression, decision trees, support vector machines). A1.4: Identify and describe common unsupervised learning algorithms (e.g., k-means clustering, hierarchical clustering, principal component analysis). A1.5: Explain the concepts of overfitting and underfitting and their impact on model performance. A1.6: Understand the importance of data preprocessing in the machine learning pipeline. A1.7: Describe various data preprocessing techniques (e.g., handling missing values, encoding categorical variables, feature scaling). A1.8: Explain the role of feature engineering in improving model performance. A1.9: Understand the process of training machine learning models.
B-Skills	**B1 -** Skills to apply professional knowledge	B 1.1 Describe the differences among the three main styles of learning: supervised, reinforcement, and unsupervised. B 1.2 Determine which of the three learning styles is appropriate to a particular problem domain. B 1.3 Frame an application as a classification problem, including the available input features and output to be predicted (e.g., identifying alphabetic characters from pixel grid input).

		B 1.4 Apply two or more simple statistical learning algorithms (such as k-nearest-neighbors and logistic regression) to a classification or regression task and measure the models' accuracy.
		B 1.5 implement a statistical learning algorithm and the corresponding optimization process to train the model and obtain a prediction on new data.
		B 1.6 Identify overfitting in the context of a problem and learning curves and describe solutions to overfitting.
		B 1.7 Explain proper ML evaluation procedures, including the differences between training and testing performance, and what can go wrong with the evaluation process leading to inaccurate reporting of ML performance.
		B1.8 Compare two machine learning algorithms on a dataset, implementing the data preprocessing and evaluation methodology (e.g., metrics and handling of train/test splits) from scratch.
		B1.9 Compare and contrast each of the following techniques, providing examples of when each strategy is superior: decision trees, logistic regression, naive Bayes, neural networks, and belief networks.
	B2 - General (transversal) skills	B2.1 Analyze complex problems and develop innovative solutions using machine learning techniques.
		B2.2 Evaluate different approaches to problem-solving and select the most appropriate methodology.
		B2.3 Effectively communicate technical concepts and findings to both technical and non-technical audiences.
		B2.4 Develop clear and concise technical reports and presentations.
		B2.5 Adapt to changing environments and new challenges in the field of machine learning. "
		B2.6 Apply analytical skills to assess the validity and reliability of data and results.

KNOWLEDGE / SKILLS ASSESSMENT & EVALUATION		
Ongoing evaluation tasks *(max 1/3 of grade for the total course)*	**Midterm exam** *(max 1/3 of grade for the total course)*	**Final exam**
Assessment : **O**ral ☐ **W**ritten ☒	**A**ssessment : **O**ral ☐ **W**ritten ☒	**A**ssessment : **O**ral ☐ **W**ritten ☒
Duration : XXX h. **C**riteria :	**G**roup base: **Y**es ☒ **N**o ☐	**G**roup base: **Y**es ☒ **N**o ☐

Course project : Yes ☐ No ☐ Presentation : Yes ☐ No ☐	Duration : 1 h.	Duration : 2 h.
Tasks type & Weight : XXXXXX	Exam type : XXXXXX	Exam type : XXXXXX

TEACHING METHODS & TOOLS

This is an introductory course to machine learning that covers fundamental topics in ML, including ML types, supervised, unsupervised ML models and algorithms, data preprocessing, model and performance evaluation. The goal of this course is to provide students with a strong foundation in the theory, algorithms, and practical applications of machine learning. Through lectures, discussions, and homework assignments, students will learn basic ML concepts and principles and develop modeling and analytical skills to apply ML for problem-solving. The course will prepare students to further exploration of ML in both research and application domains. Students will be guided to develop problem-solving skills through interactive activities that are closely tailored to the lesson at hand. Students will be guided to discover activities that promote self-learning and help students develop critical thinking skills and retain knowledge that leads to self-actualization.

KNOWLEDGE & SKILLS PREREQUISITS

Basic notions in Coding in Python, Applied Statistics, Calculus, Probability Theory, Analytic Geometry and Algebra.

COURSE DESCRIPTION /SYLLABUS / RESOURCES

Course objective : The course aims to provide students with a comprehensive understanding of Machine Learning, also referred to as Statistical Learning. It encompasses a diverse range of algorithms designed to discern patterns within data, enabling the creation of models that can be implemented and potentially turned into practical products. By the end of the course, students will possess the essential theoretical foundations, hands-on skills, and analytical mindset required to address real-world machine learning obstacles effectively. This will empower them to explore advanced research prospects and seek career opportunities in the thriving domain of machine learning.

TOPIC	HOURS	CORE RESOURCES [1]	ADDITIONAL RESOURCES
1. Definition and examples of a broad variety of machine	3hCM		http://aima.cs.berkeley.edu/

[1] *For each topic max 20 -25 page of reading*

learning tasks ▪ Supervised learning (Classification, Regression) ▪ Unsupervised learning (Clustering) ▪ Reinforcement learning		Chapter 19	
2. Working with Data ▪ Data preprocessing ▪ Handling missing values (imputing, flag-as-missing), ▪ Encoding categorical variables, encoding real-valued data ▪ Normalization/ standardization ▪ Emphasis on real data, not textbook examples	1,5hCM 6 hTP	Chapter 19	https://jakevdp.github.io/PythonDataScienceHandbook/
3. Simple statistical-based supervised learning ▪ Learning Decision Trees	1,5hCM 3hTP	Chapter 19	http://aima.cs.berkeley.edu/
4. Machine learning evaluation ▪ The overfitting problem and controlling solution complexity (regularization, pruning –	1,5hCM 3hTP	Chapter 19	http://aima.cs.berkeley.edu/

intuition only) ▪ Separation of train, validation, and test sets ▪ Performance metrics for classifiers ▪ Estimation of test performance on held-out data ▪ Tuning the parameters of a machine learning model with a validation set o Importance of understanding what your model is actually doing, where its pitfalls/shortcomings are, and the implications of its decisions			
5. Supervised learning ▪ Nearest-neighbor (KNN) classification and regression ▪ Logistic regression	4,5hCM 6hTP	Chapter 19	http://aima.cs.berkeley.edu/ https://scikit-learn.org/stable/user_guide.html

▪ Linear regression ▪ Support vector machines (SVMs) and kernels ▪ Gaussian Processes ▪ Naïve Bayes			
6. Unsupervised learning and clustering ▪ K-means ▪ Gaussian mixture models ▪ Expectation maximization (EM) ▪ Self-organizing maps	1,5hCM 6 hTP	Chapter 19	http://aima.cs.berkeley.edu/ https://www.naftaliharris.com/blog/visualizing-k-means-clustering/
7. Basic neural networks ▪ Fundamentals of understanding how neural networks work and their training process, without details of the calculations	1,5hCM 3 hTP	Chapter 22	http://aima.cs.berkeley.edu/ https://aegeorge42.github.io/

STRUCTURE OF THE COURSE & ADDITIONAL INFORMATION REGARDING THE COURSE PREPARATION

Information regarding the course resources available in the web, Resources that students must have for the course, Resources & tools needed for course, including software and hardware with appropriate technical requirements descriptions

CORE REFERENCES

1. Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach (4th ed.), University of California at Berkeley, 2020

2. Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach (3rd ed.). Prentice Hall Press, 2009

ADDITIONAL REFERENCES

1. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition. 2019.
2. Sebastian Raschka, Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow, Packt Publishing, 3 edition, 2019.
3. Jake VanderPlas, Python Data Science Handbook , O'Reilly Media, 2016.
4. Christopher M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics).2006
5. Trevor Hastie , Robert Tibshirani , Jerome Friedman. The Elements of Statistical Learning .Data Mining, Inference, and Prediction, Second Edition.2006
6. Andrew Ng, Machine Learning Yearning, 2018.
7. Kevin P. Murphy, Machine Learning: A Probabilistic Perspective, The MIT Press Cambridge, Massachusetts London, 2012.
8. Marc Peter Deisenroth, A Aldo Faisal, Cheng Soon Ong, Mathematics for Machine Learning, Cambridge University Press, 2020.

WEB RESOURCES

1. https://docs.google.com/document/d/1CzyezFicBBHlFNHlYIwbHiEbsxnAF-sL2TN5iO2_VK8/edit?tab=t.0

2. https://colab.research.google.com/drive/1Nc9wFrTZdD881rlNGIJKATiQyh5aiSRz

3. http://aima.cs.berkeley.edu/

4. https://mlu-explain.github.io/

2. https://scikit-learn.org/stable/user_guide.html

3. https://ml-playground.com/

4. https://www.w3schools.com/ai/default.asp

5. https://jakevdp.github.io/PythonDataScienceHandbook/

6. https://www.intel.com/content/www/us/en/developer/topic-technology/artificial-intelligence/training/course-machine-learning.html#gs.2x2rxd

7. https://sebastianraschka.com/notebooks/ml-notebooks/

8. https://www.simplilearn.com/tutorials/machine-learning-tutorial

9. https://python-course.eu/machine-learning/

10. https://www.datacamp.com/tutorial