UNIVERSITY of NICOSIA

Course Syllabus

Course Code	Course Title	ECTS Credits
COMP-544DL	Machine Learning	10
Prerequisites	**Department**	**Semester**
COMP-542DL COMP-540DL	Computer Science	Spring
Type of Course	**Field**	**Language of Instruction**
Required	Data Science	English
Level of Course	**Lecturer(s)**	**Year of Study**
2nd Cycle	Ioannis Katakis	1st
Mode of Delivery	**Work Placement**	**Corequisites**
Distance Learning	N/A	None

Course Objectives:

The main objectives of the course are to:

- Provide understanding of what is Machine Learning
- Determine when and how we can use Machine Learning tools.
- Introduce the concepts and techniques of data classification (decision trees, Bayesian, support vector machines, lazy classifiers, neural networks).
- Explain the basic concepts of model evaluation and comparison.
- Provide practical experience on how ensemble methods can be of value for specific problems.
- Introduce the challenges of stream data classification.
- Explain the basic concepts of data clustering and its applications.
- Present the major algorithms of clustering.
- Provide the characteristics of the basic types of clustering (hierarchical, partitioning, density)
- Explain the principles and how association rules work.
- Define the different types of anomaly detection methods.
- Demonstrate a number of real-world applications about data clustering, association rule mining, and anomaly detection.
- Demonstrate a set of tools that a practitioner can use in order to apply the algorithms presented in the course.

UNIC

Learning Outcomes:

After completion of the course students are expected to be able to:

1. Breaking down a data mining task and devise a step-by-step solution.
2. Apply the basic machine learning algorithms on a diverse set of data sets.
3. Execute the pre-processing steps of data preparation and cleaning.
4. Apply Decision Tree algorithms to a diverse set of data sources.
5. Explain the problem of overfitting and provide solutions
6. Apply a different set of classification algorithms (decision tree, naïve bayes classifier, support vector machine, neural network, kNN) and be able to compare their performances in multiple aspects (training time, testing time, predictive accuracy, etc).
7. Explain the advantages and disadvantages of any machine learning classifier
8. Identify when an ensemble technique can increase predictive performance.
9. Indicate the challenges of data stream classification.
10. Apply Clustering Methods to analyze data
11. Extract Association Rules from data
12. Apply anomaly detection algorithms on real-time or offline data

Course Content:

1. Introduction to Machine Learning
2. Classification - Basic Concepts, Training, Testing, Models
3. Decision Trees and the ID3 Classifier
    a. Splitting Criteria – Information Gain, Entropy
4. Bayesian Classifiers
    a. The Bayes theorem
    b. The Naïve Bayes Classifier
5. Support Vector Machines
    a. Solving the optimization problem
    b. Special cases (data that are not linearly separable, slack variables)
6. Lazy Learners
    a. The k-Nearest Neighbor Classifier
7. Artificial Neural Networks
    a. General Principles and the relation with Biological Neural Networks
    b. The back-propagation algorithm
8. Model Evaluation, and Model Comparison
    a. Evaluation Metrics, Area Under the ROC Curve, Cross Validation
    b. Model Comparison and Tests of Significance
9. Ensemble Methods – Multiple Classifier Systems
    a. Boosting
    b. Bagging and Random Forests

10. Stream Data Classification
    a. Incremental and Batch Learning
    b. Concept drift
11. Prediction Methods
    a. Regression & Forecasting
    b. Time series classification
12. Introduction to Data Clustering
    a. Applications
    b. Basic Concepts, Similarity Metrics, Distance Metrics
13. Partitional Clustering
    a. The K-means algorithm
    b. K-means as an optimization problem
14. Hierarchical Clustering
    a. Basic Agglomerative Hierarchical Clustering
    b. Strengths and Weaknesses
15. Density Based Clustering
    a. The DBSCAN Algorithm
    b. Subspace Clustering
16. Cluster Evaluation
    a. Unsupervised Cluster Evaluation: Cohesion and Separation
    b. Unsupervised Cluster Evaluation: Proximity Matrix
    c. Supervised Measures of Cluster Validity
    d. Assessing the Significance of Cluster Validity Measures
17. Prototype-Based Clustering
    a. Fuzzy Clustering
    b. Mixture Models
    c. Self-Organizing Maps
18. Graph Based Clustering
    a. Minimum Spanning Tree
19. Association Analysis
    a. Frequent Item Generation (the Apriori principle) & Rule Generation
    b. The FP-Growth Algorithm
    c. Evaluation of Association Patterns
20. Anomaly Detection
    a. Statistical Approaches & Proximity-Based
    b. Density –Based & Clustering -Based

Learning Activities and Teaching Methods:

Lectures, Exercises, Guest Presentations, Projects, Discussions.

Assessment Methods:

Homework, Projects, Final Assessment* * The Final Assessment can be either a Final Exam or Final Assignment(s) with Viva

Required Textbooks / Readings:

Title	Author(s)	Publisher	Year	ISBN
Introduction to Data Mining	Tan, Steinbach, Karpatne, Kumar	Pearson	2018	0321321367

Recommended Textbooks / Readings:

Title	Author(s)	Publisher	Year	ISBN
Data Mining: Concepts and Techniques, Third Edition	Han, Kamber, Pei	Morgan Kaufmann	2011	9380931913
Data Mining: Practical Machine Learning Tools and Techniques	Witten, Frank, Hall	Morgan Kaufmann	2011	0123748569