UNIVERSITY of NICOSIA

Course Syllabus

Course Code	Course Title	ECTS Credits
COMP-244	Machine Learning and Data Mining I	6
Prerequisites	**Department**	**Semester**
COMP-240 COMP-212	Computer Science	Spring
Type of Course	**Field**	**Language of Instruction**
Core	Data Science	English
Level of Course	**Lecturer**	**Year of Study**
1st Cycle	Prof Ioannis Katakis	2nd
Mode of Delivery	**Work Placement**	**Corequisites**
Face to Face	N/A	None

Course Objectives:

The main objectives of the course are to:

- Develop a solid understanding of the fundamental concepts and scope of Data Mining.
- Analyse when and how Data Mining tools can be effectively applied to solve practical problems.
- Apply appropriate data pre-processing techniques to prepare data for mining tasks.
- Examine and compare key classification techniques, including decision trees, Bayesian classifiers, support vector machines, lazy classifiers, and neural networks.
- Interpret and assess the performance of classification models using established evaluation metrics.
- Gain practical experience in implementing ensemble learning methods to address specific data analysis challenges.
- Identify and discuss the key challenges associated with stream data classification, and evaluate potential solutions.

Learning Outcomes:

After completing the course, students are expected to be able to:

1. Analyse data-related problems and formulate abstract, structured solutions.
2. Apply fundamental data mining concepts and problem-solving techniques in practical contexts.
3. Pre-process and transform data in preparation for analysis.
4. Apply appropriate statistical methods to extract insights from data.

5. Implement and interpret Decision Trees for data analysis tasks.
6. Identify the issue of overfitting in predictive models and propose suitable mitigation strategies.
7. Apply and evaluate a range of classification algorithms (including decision trees, naïve Bayes, support vector machines, neural networks, and k-nearest neighbours), comparing their performance across multiple dimensions such as training time, testing time, and predictive accuracy.
8. Critically assess the strengths and limitations of various machine learning classifiers.
9. Determine when ensemble learning methods can improve predictive performance and justify their use.
10. Analyse the specific challenges associated with stream data classification and evaluate potential approaches.

Course Content:

1. Introduction to Data Mining
    a. What is Data Mining?
    b. What tasks can Data Mining accomplish?
2. Data preprocessing
    a. Data cleaning
    b. Handling missing Data
    c. Data transformation
3. Classification - Basic Concepts, Training, Testing, Models
4. Decision Trees and the ID3 Classifier
    a. Basic Principles
    b. Splitting Criteria – Information Gain, Entropy
5. Bayesian Classifiers
    a. The Bayes theorem
    b. The Naïve Bayes Classifier
6. Support Vector Machines
    a. Support Vectors
    b. Solving the optimization problem
    c. Special cases (data that are not linearly separable, slack variables)
7. Lazy Learners
    a. The k-Nearest Neighbor Classifier
8. Artificial Neural Networks
    a. General Principles and the relation with Biological Neural Networks
    b. Neurons, Hidden Layers, Activation Functions
    c. The back-propagation algorithm
    d. Applications of Neural Networks
9. Model Evaluation, and Model Comparison
    a. Evaluation Metrics, Area Under the ROC Curve, Cross Validation

b. Model Comparison and Tests of Significance
c. Unbalanced datasets

10. Ensemble Methods – Multiple Classifier Systems
    a. Boosting
    b. Bagging and Random Forests
    c. Stacking
11. Stream Data Classification
    a. Incremental and Batch Learning
    b. Concept drift
    c. Algorithms for data stream classification
12. Prediction Methods
    a. Regression & Forecasting
    b. Time series classification

Learning Activities and Teaching Methods:

Lectures, Demonstration of Data Mining Tools, Assignments, Projects.

Assessment Methods:

Assignments
Midterm
Project (Machine Learning Challenge)
Final Examination
Participation

Required Textbooks / Readings:

Title	Authors	Publisher	Year	ISBN
Introduction to Data Mining, 2nd Edition	Tan, Steinbach, Kumar	Pearson	2019	
Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd Edition	Aurélien Géron	O'Reilly	2019	

Recommended Textbooks / Readings:

Title	Authors	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