UNIVERSITY of NICOSIA

Course Syllabus

Course Code	Course Title	ECTS Credits
COMP-244	Machine Learning and Data Mining I	6
Prerequisites	**Department**	**Semester**
Sophomore Standing	Computer Science	Spring
Type of Course	**Field**	**Language of Instruction**
Core	Data Science	English
Level of Course	**Lecturer**	**Year of Study**
1st Cycle	Dr 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:

- Provide understanding of what is Data Mining.
- Determine when and how we can use Data Mining tools.
- Introduce the concepts and techniques of pre-processing of the data to be analysed.
- 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.

Learning Outcomes:

After completion of the course students are expected to be able to:
1. analyze problems and find abstract solutions
2. use the basic data mining concepts and problem solving techniques
3. prepare data to be analyzed
4. apply statistical methods to analyze data
5. use Decision Trees to analyze data
6. understand the problem of overfitting and to provide solutions

UNIC

7. 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).
8. explain the advantages and disadvantages of any machine learning classifier
9. understand when an ensemble technique can increase predictive performance
10. understand the challenges of data stream classification.

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:

Mid-term exam
Projects
Final Examination
Participation/Homework Assignments/Quizzes

Required Textbooks / Readings:

Title	Authors	Publisher	Year	ISBN
Introduction to Data Mining	Tan, Steinbach, Kumar	Pearson	2005	0321321367

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