UNIVERSITY of NICOSIA

Course Syllabus

Course Code	Course Title	ECTS Credits
COMP-344	Machine Learning and Data Mining II	6
Prerequisites	**Department**	**Semester**
COMP-244	Computer Science	Fall
Type of Course	**Field**	**Language of Instruction**
Required	Data Science	English
Level of Course	**Lecturer**	**Year of Study**
1st Cycle	Dr Ioannis Katakis	3rd
Mode of Delivery	**Work Placement**	**Corequisites**
Face to Face	N/A	None

Course Objectives:

The main objectives of the course are to:

- 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.

Learning Outcomes:

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

1. analyze data mining problems and find abstract solutions
2. use the basic data mining concepts and problem solving techniques
3. prepare data to be analyzed
4. use Clustering Methods to analyze data
5. extract Association Rules from data

6. Apply anomaly detection algorithms on real-time or offline data

Course Content:

1. Introduction to Data Clustering
    a. Applications
    b. Basic Concepts, Similarity Metrics, Distance Metrics
2. Partitional Clustering
    a. The K-means algorithm
    b. K-means as an optimization problem
3. Hierarchical Clustering
    a. Basic Agglomerative Hierarchical Clustering
    b. Strengths and Weaknesses
4. Density Based Clustering
    a. General Principles
    b. The DBSCAN Algorithm
    c. Subspace Clustering
5. Cluster Evaluation
    a. Unsupervised Cluster Evaluation: Cohesion and Separation
    b. Unsupervised Cluster Evaluation: Proximity Matrix
    c. Unsupervised Evaluation of Hierarchical Clustering
    d. Supervised Measures of Cluster Validity
    e. Assessing the Significance of Cluster Validity Measures
6. Prototype-Based Clustering
    a. Fuzzy Clustering
    b. Mixture Models
    c. Self-Organizing Maps
7. Graph Based Clustering
    a. Minimum Spanning Tree
8. Association Analysis
    a. Frequent Item Generation (the Apriori principle)
    b. Rule Generation
    c. The FP-Growth Algorithm
    d. Evaluation of Association Patterns

9. Anomaly Detection a. Preliminaries b. Statistical Approaches c. Proximity-Based d. Density –Based e. Clustering -Based

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