Course Name : Data Mining II

Course Code	Course Type	Regular Semester	Lecture (hours/week)	Seminar (hours/week)	Lab. (hours/week)	Credits	ECTS
CMP 509	C	Fall	3.00	1.00	0.00	3.50	6.00

Lecturer	Migena Ceyhan, PhD
Assistant	
Course language	Albanian
Course level	Master
Description	The course "Data Mining I" provides general knowledge on the analysis and finding the desired data from a very large amount of data. Data Mining plays an essential role in understanding, preprocessing, categorizing and classifying data in multi-dimensional spaces. This course aims to equip students with the basic theoretical and practical knowledge for processing large amounts of data and obtaining desired information.
Objectives	• Familiarize students with data types. • To equip students with the different techniques and ways of analyzing large amounts of data. • To provide students with proper skills to implement data preprocessing methods. • To explain the importance and influence of Data Mining in in field of computer science for finding valuable information in Big Data. • To develop students' critical thinking in analyzing and finding patterns in multi-dimensional data.
Core Concepts	Data Mining Types of Data Kernel methods Multidimensional Data Finding Frequent Patterns Clustering Reduction of Dimensions Rules of Association K-means, Decision Tree

Course Outline

Week	Topic
1	Introduction to Data Mining This topic provides an overview of the course, and it will cover topics such as what is Data Mining, the origin and reason for the development of Data Mining as well as the main tasks that can be performed by using Data Mining techniques and methods. (Main Lit., pg. 21-42)
2	Types of Data This lecture covers topics such as attribute types, measurement and values, data categorization and transformation, data sets and their types, as well as data quality analysis due to data measurement and data collection problems. (Main Lit., pg. 43-69)
3	Data Preprocessing - 1 In this lecture, different similarity and distance measures will be covered. The main types of distances that will be treated are Euclidean Distance, Minkoski Distance, Mahalanobis Distance, the main types of Similarities that will be covered are Similarity between binary vectors, Cosinusoidal Similarity and Pearson Correlation. (Main Lit., pg. 91-110)
4	Data Preprocessing - 2 This lecture will cover data preprocessing techniques such as Aggregation, Sampling, Dimensionality Reduction, Feature Subset Selection, New Feature Creation, Discretization, Binarization, Variable Transformation as well as measurement units based on information. (Main Lit., pg. 70-90)
5	Data Exploration - 1 This lecture will cover the basic elements of data summary statistics, such as types of means, types of distributions, various measures of similarity and dissimilarity between different data objects, types of proximity units, mutual information and the use of proper techniques in selecting the appropriate measurement unit. (Recommended Lit., pg. 44 – 55)

6	6 Data Exploration - 2 This lecture will cover the basic elements of data summary statistics, such as types of means, types of distributions, various measures of similarity and dissimilarity between different data objects, types of proximity units, mutual information and the use of proper techniques in selecting the appropriate measurement unit. (Recommended Lit., pg. 56 – 64 and Main Lit., pg. 110-132)
7	Classification: Basic Concepts and Techniques - 1 This lecture will cover the basic concepts of classification, multiclass and binary classification, general approaches for building a classification model, basic methods for presenting test conditions, the calculation of different impurity measures for different data types, and basic classification algorithms. (Main Lit., pg. 133-167)
8	Midterm Exam
9	Classification: Basic Concepts and Techniques - 2 This lecture will cover the topics of overfitting in the selected classification model, evaluation and selection of different classification models, hyper parameters and limitations of the basic classification algorithms. (Main Lit., pg. 167-212)
10	Association Rules: Basic Concepts and Algorithms This lecture will cover the basic concepts and algorithms of association rules such as the Apriori principle, frequent itemset generation, candidate generation and pruning, techniques and methods for generation of association rules as well as the computational complexity of the basic association rules algorithms. (Main Lit., pg. 213-239)
11	Rules of Association: Issues in model selection and evaluation This lecture will cover topics such as compact representation of frequent itemsets, alternative methods for generating frequent itemsets, FP-growth algorithm, evaluation of association patterns and the effect of skewed support distribution. (Main Lit., pg. 240-306)
12	12 Cluster Analysis: Basic Concepts and Algorithms This lecture will cover the basic concepts and algorithms of cluster analysis such as what is Cluster Analysis, the different types of clustering methods, the different types of clusters as well as a detailed analysis of the K-means algorithm. (Main Lit., pg. 307-335)
13	Cluster Analysis: Issues in Model Selection and Evaluation This lecture will cover the basic concepts and algorithms of cluster analysis such as hierarchical agglomerative clustering, detailed analysis of the DBSCAN algorithm, as well as different techniques and methods for cluster evaluation. (Main Lit., pg. 336-394)
14	Classification: Alternative Techniques This lecture will cover topics such as types of classifiers, rule based classifiers, nearest neighbor classifiers, Naïve Bayes classifiers, Logistic Regression, Artificial Neural Network and Support Vector Machine. (Main Lit., pg. 395-463, 478 - 498)
15	General Review and Determining the topics for the Project Presentation
16	Final Exam

Prerequisites	The student must attend the course at a minimum rate of 75%.
Literature	• Introduction to Data Mining, Second Edition Global Edition, Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, Pearson Education Limited, 2019, ISBN 10: 0-273-76922-7, ISBN 13: 978-0-273-76922-4, eBook ISBN 13: 978-0-273-77532-4
References	• DATA MINING AND ANALYSIS, Fundamental Concepts and Algorithms, Mohammed J. Zaki, Wagner Meira Jr., Cambrige University Press, 2014, ISBN 978-0-521-76633-3

Course Outcome

1	Students will be able to understand different types of data.
2	Students will have knowledge of statistical data features and units of measurement.
3	Students will be able to understand the importance of Data Mining in finding valuable information.
4	Students will acquire the most important concepts related to the basic models and algorithms and will have the skills for finding valuable information.
5	Students will be ready to implement in practice given knowledge.
6	Students will be equipped with sufficient theoretical and practical knowledge to continue with other subsequent subjects related to Data Mining.

Course Evaluation

In-term Studies	Quantity	Percentage
Midterms	0	0
Quizzes	0	0
Projects	1	45
Term Projects	0	0
Laboratory	0	0
Class Participation	0	0
Total in-term evaluation percent		**45**
Final exam percent		**55**
Total		**100**

ECTS Workload (Based on Student Workload)

Activities	Quantity	Duration (hours)	Total (hours)
Course duration (Including the exam week: 16x Total hours of the course)	16	4	64
Study hours outside the classroom (Preparation, Practice, etc.)	14	4	56
Duties	1	15	15
Midterms	0	0	0
Final Exam	1	15	15
Other	0	0	0
Total Work Load			**150**
Total Work Load / 25 (hours)			**6.00**
ECTS			**6.00**