

Course Name : Data Mining							
Course Code	Course Type	Regular Semester	Lecture (hours/week)	Seminar (hours/week)	Lab. (hours/week)	Credits	ECTS
EMS 331	C	Fall	3.00	1.00	0.00	3.50	5.00
Lecturer		Olsi Shehu, MSc					
Assistant							
Course language		English					
Course level		Bachelor					
Description		This course explores the concepts and techniques of knowledge discovery and data mining. As a multidisciplinary field, data mining draws on work from areas including statistics, machine learning, pattern recognition, database technology, information retrieval, network science, knowledge-based systems, artificial intelligence, high-performance computing, and data visualization. This course focuses on issues relating to the feasibility, use-fulness, effectiveness, and scalability of techniques for the discovery of patterns hidden in large data sets. As a result, this course is not intended as an introduction to statistics, machine learning, database systems, or other such areas, although it does provide some background knowledge to facilitate the reader's comprehension of their respective roles in data mining.					
Objectives		This course aims to: - Familiarize students with data types. - To acquaint students with the different techniques and ways of analyzing large amounts of data. - To acquaint students with data preprocessing methods. - To explain the importance, influence and close connection of Data Mining in the implementation in the field of computer science for finding valuable information. - To develop students' critical thinking in analyzing and finding patterns in multi-dimensional data.					
Core Concepts		Students make everyone able to understand: 1. Different types of data. 2. Features and statistical measurement units of data. 3. The most important concepts about me and images 4. Basic algorithms for finding valuable information. 5. The importance of Data Mining in finding information valuable. Students do all the rest for applying knowledge based on data. Students will argue with knowledge of theory and practices to follow with other subsequent subjects.					
Course Outline							
Week	Topic						
1	Introduction to Data Mining In this lecture, a general knowledge of the subject will be realized, as well as topics such as what Data Mining is, the origin and reason for the development of Data Mining, as well as the main tasks that can be performed by means of Data Mining. (Basic Lit., pp. 21-42)						
2	Data Types In this lecture, topics such as properties, attribute types and values, data categorization and transformation, data sets and their types, as well as analyzing data quality due to data problems will be covered. measurements and data collection. (Basic Lit., pp. 43-69)						
3	Data Preprocessing - 1 In this lecture, different measures of similarity and distances will be covered. The main types of distances that will be treated are Euclidean Distance, Minkoski Distance, Mahalanobis Distance, the main types of separations that will be treated are Similarity between binary vectors, Cosine Similarity and Pearson Correlation. (Lit. base, pp. 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 and information-based units of measure. (Basic Lit., pp. 70-90)						

5	Data Exploration - 1 This lecture will cover the basic elements of summary data statistics, such as types of means, types of distributions, various measures of similarity and differences between different data objects, types of units of measurement proximity, mutual information as well as techniques in selecting the appropriate unit of measurement. (Recommended Lit., pp. 44 – 55)
6	Data Exploration - 2 This lecture will cover the basic elements of summary data statistics, such as types of means, types of distributions, various measures of similarity and differences between different data objects, types of units of measurement proximity, mutual information as well as techniques in selecting the appropriate unit of measurement. (Recommended Lit., pp. 56-64 as well as Basic Lit., pp. 110-132)
7	Classification: Basic Concepts and Techniques - 1 In this lecture, the basic concepts of classification, multiclass and binary classification, general approaches for construction of the classification model, basic methods for presenting test conditions, calculation of different impurity measures will be covered. for different types of data as well as basic classification algorithms. (Lit. base, pp. 133-167)
8	Semi-final exam
9	Classification: Basic Concepts and Techniques - 2 In this lecture, the topics of overfitting of the selected classification model, evaluation and selection of different classification models, hyper parameters and limitations of basic classification algorithms will be addressed. (Lit. base, pp. 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, generating frequent sets of items, candidate selection, techniques and methods for generating association rules, and complexity computing of basic algorithms of association rules. (Lit. base, pp. 213-239)
11	Association Rules: Issues in model selection and evaluation In this lecture, more detailed topics of association rules will be addressed, such as the compact representation of frequent item groups, alternative methods for generating frequent item groups, the FP growth algorithm , estimation of association models as well as the effect of skewed distributions. (Lit. base, pp. 240-306)
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 and a detailed analysis of the algorithm. K-means. (Basic Lit., pp. 307-335)
13	Cluster Analysis: Issues in Model Selection and Evaluation In this lecture, the basic concepts and algorithms of cluster analysis such as hierarchical agglomerative clustering, the detailed treatment of the DBSCAN algorithm, as well as the various techniques and methods for cluster evaluation will be addressed. (Lit. base, pp. 336-394)
14	Classification: Alternative Techniques This lecture will cover the types of classifiers, rule based classifiers, nearest neighbor classifiers, Naive Bayes classifiers, Logistic Regression, Artificial Neural Network and Support Vector Machine. (Lit. base, pp. 395-463, 478 - 498)
15	Presentation of Projects and Recapitulation
16	Final Exam
Prerequisites	The student must attend the course at a minimum rate of 75%.
Literature	• Introduction to Data Mining 2nd Edition 2018
References	• The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han- Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition- Morgan-Kaufmann-2011

Course Evaluation			
In-term Studies	Quantity	Percentage	
Midterms	1	25	
Quizzes	0	0	
Projects	0	0	
Term Projects	0	0	
Laboratory	0	0	
Class Participation	1	10	
Total in-term evaluation percent			35
Final exam percent			65
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	0	0	0
Midterms	1	2	2
Final Exam	1	4	4
Other	0	0	0
Total Work Load			126
Total Work Load / 25 (hours)			5.04
ECTS			5.00