SYLLABUS

1. Information on the study programme

1.1. Higher education institution	West University of Timisoara
1.2. Faculty	Mathematics and Computer Science
1.3. Department	Computer Science
1.4. Study program field	Computer Science
1.5. Study cycle	Master
1.6. Study programme / Qualification	Artificial Intelligence and Distributed Computing

2. Information on the course

2.1. Course title			Data Mining				
2.2. Lecture instructor			Daniela Zaharie				
2.3. Seminar / laboratory instructor			Daniela Zaharie				
2.4. Study year	2	2.5. Semester	1	2.6. Examination type	E	2.7. Course type	DI

3. Estimated study time (number of hours per semester)

3.1. Attendance hours per week	3	out of which: 3.2 lecture	2	3.3. seminar / laboratory	1
3.4. Attendance hours per semester	42	out of which: 3.5 lecture	28	3.6. seminar / laboratory	14

Distribution of the allocated amount of time*	**hours**
Study of literature, course handbook and personal notes	20
Supplementary documentation at library or using electronic repositories	26
Preparing for laboratories, homework, reports etc.	50
Exams	6
Tutoring	6
Other activities…	0

3.7. Total number of hours of individual study	108
3.8. Total number of hours per semester	150
3.9. Number of credits (ECTS)	6

4. Prerequisites (if it is the case)

4.1. curriculum	Algorithms and Data Structures, Probability and Statistics, Databases, Programming, Numerical Methods, Artificial Intelligence
4.2. competences	Knowledge of data structures, algorithms, statistics, numerical methods at undergraduate level and programming abilities

5. Requirements (if it is the case)

5.1. for the lecture	Lecture room with whiteboard and projector
5.2. for the seminar / laboratory	Lab room with computers having Weka (Java), Python and R software installed

6. Specific acquired competences

Professional competences	• Ability to analyze data and extract knowledge from them • Ability to identify the algorithm/method appropriate to classify and cluster data and to make predictions starting from data • Ability to solve a real-world problem using data mining tools.
Transversal competences	• Ability to conduct research activity and to prepare reports on a given topic • Team work ability

7. Course objectives

7.1. General objective	Getting familiar with techniques and methods used in data mining and knowledge extraction from data.
7.2. Specific objectives	*Knowledge objectives (OC)*: (1) to present basic concepts in data mining; (2) to present the main data mining techniques: classification, cluster, regression, association rules, outlier analysis; (3) to describe methods for extracting models from data and evaluating them (4) to present examples of efficient data mining methods. *Abilitation objectives (OAb)*: (1) to identify the techniques appropriate to a given problem; (2) to use software tools which are specific for data mining; (3) to implement efficient data mining algorithms; *Attitude objectives (OAt)*: (1) to argue the utility of data mining in solving real-world problems.

8. Content

8.1. Lecture	**Teaching methods**	**Remarks, details**
L1. Introduction in knowledge discovery from data. Basic concepts and main data mining tasks. Data categories and types of attributes. *(OC1)*	Discourse, conversation, illustration by examples	2 hours ([1]- ch 1,[2]- ch 1, [3]-ch 2)
L2. Data pre-processing. Basic transformations on data (discretization, normalization, standardization). Data cleaning and dealing with missing values. Attribute selection and feature extraction. Filter-based and wrapper-based methods. Principal Component Analysis. (OC2)	Discourse, conversation, illustration by examples	2 hours ([1]-ch 2)
L3-5. Classification methods. Basic concepts and performance measures (accuracy, precision, recall, specificity, sensitivity, ROC). Training, testing and cross-validation. Instance based classifiers (k Nearest Neighbour). Decision tables and rule-based classifiers. Decision trees (ID3, C45). Probability-based classifiers (Bayesian networks). Neural networks. Support	Discourse, conversation, illustration by examples	6 hours ([1]-ch 10; [2] –ch 4; [3] – ch 4, sect 5.2,5.3, 5.5, 5.6)

Vector Machines. (OC2,OC3)		
L6-8. Clustering methods. Basic concepts (cluster, centroid). Similarity and dissimilarity measures. Cluster quality measures. Partitional algorithms (kMeans, Fuzzy CMeans). Hierarchical algorithms (agglomerative, divisive). Statistical-based clustering (EM algorithm). Spatial clustering (DBSCAN). Graph-based clustering. Spectral techniques. (OC2,OC3, OC4)	Discourse, conversation, illustration by examples	6 hours ([1] – ch 6, [2]-ch 5, [3] – sect 5.8)
L9. Association rules. Basic concepts (support, confidence, frequent itemsets). Measures for rules quality. Apriori algorithm. (OC2,OC3, OC4)	Discourse, conversation, illustration by examples	2 hours ([1]-ch. 4; [2]-ch. 6, [3]-sect. 5.4)
L10-11. Regression and time series processing. Nonlinear regression models. Regression trees. Radial Basis Networks. Time series analysis (trend analysis, pattern detection, prediction models, anomaly detection). (OC2,OC3)	Discourse, conversation, illustration by examples	4 hours ([1] – ch. 8 ch. 11.5, 14, [2] – ch. 9)
L12-13 Ensemble methods. Voting. Bagging. Boosting (AdaBoost). Random forests. Stacking. (OC2,OC3,OC4)	Discourse, conversation, illustration by examples	2 hours ([1] – sect 11.8, [5])
L14. Text mining. Text pre-processing (tokenization, stemming). Vector space model (TF-IDF). Document classification and clustering. *Web mining.* Web structure mining. Web usage mining. Network analysis. (OC2,OC3).	Discourse, conversation, illustration by examples	2 hours ([1] –ch. 13, 18, [4], [2] – ch. 7)

Recommended literature

1. Charu C. Aggarwal. *Data Mining – the textbook*, Springer, 2015
2. M. H. Dunham. *Data Mining. Introductory and Advanced Topics*, Pearson Education 2003
3. F. Gorunescu, *Data Mining. Concepts, Models and Techniques*, Springer, 2011
4. C. D. Manning, P. Raghavan and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008.
5. I.H. Witte, E. Frank, M.A. Hall. *Data Mining – Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, 2011
6. D. Zaharie: Data Mining course materials (http://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/EN)

8.2. Seminar / laboratory	**Teaching methods**	**Remarks, details**
L1. Data sets and repositories. Introduction to Weka and Rattle packages. (Oab2)	Problem-based approach, dialogue, learning through collaboration	2 hours
L2. Data visualization. Data pre-processing. (Oab1,Oab2)	Problem-based approach, dialogue, learning through collaboration	2 hours
L3. Data classification using rules and decision	Problem-based approach, dialogue, learning through collaboration	2 hours

trees. (Oab1,Oab2)		
L4. Data classification using probabilistic models, neural networks and SVM. (Oab1,Oab2,Oab3)	Problem-based approach, dialogue, learning through collaboration	2 hours
L5. Data clustering using partitional, hierarchical and density-based algorithms. (Oab1,Oab2,Oab3)	Problem-based approach, dialogue, learning through collaboration	2 hours
L6. Extracting association rules. Applications in market basket analysis. (Oab1,Oab2,Oab3)	Problem-based approach, dialogue, learning through collaboration	2 hours
L7. Time series analysis. Prediction. Ensemble methods. Text mining. (Oab1,Oab2,Oab3)	Problem-based approach, dialogue, learning through collaboration	2 hours

Recommended literature:

1. Lab materials: http://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/EN
2. Datasets: http://archive.ics.uci.edu/ml/datasets, https://www.kaggle.com/
3. G. Williams, Data Mining with Rattle and R. The Art of Excavating Data for Knowledge Discovery, Sringer 2011
4. J. Grus, Data Science from Scratch. First Principles with Python, O'Reilly, 2015

9. Correlations between the content of the course and the requirements of the professional field and relevant employers.

- The content is in accordance with similar courses provided at other universities and it covers the basic aspects of data mining techniques in solving problems arising in various domains.

10. Evaluation

Activity	10.1. Assessment criteria	10.2. Assessment methods	10.3. Weight in the final mark
10.4. Lecture	Knowledge of basic data mining techniques	Written test	20%
	Correct identification of the appropriate technique to solve a given problem	Presentation of a project	60%
10.5. Seminar / laboratory	Usage of software tools for data analysis	Lab applications and homework	20%
10.6. Minimum needed performance for passing			

- Knowledge of basic concepts in data mining
- Knowledge of main classification, clustering and prediction algorithms
- Ability to identify the appropriate data mining method in solving real-world problems

Ability in using software tools for data mining. The final mark is computed as weighted average of the marks corresponding to the components specified at 10.4 and10.5. The exam is considered passed if the average is at least 5 (it is not required that each mark is at least 5) . In each session of exams (including re-examinations) the mark is computed using the same rule. The student can be re-examined only for the components for which the current mark is smaller than 5, excepting the cases when the student asks to be re-examined .

Date of completion	Signature (lecture instructor)	Signature (seminar instructor)
10.02.2018		

Date of approval	Signature (director of the department)