SYLLABUS

1. Information regarding the programme

1.1 Higher education institution	**Babeş Bolyai University**
1.2 Faculty	**Faculty of Mathematics and Computer Science**
1.3 Department	**Department of Computer Science**
1.4 Field of study	**Computer Science**
1.5 Study cycle	**Master**
1.6 Study programme / Qualification	**Distributed Systems in Internet - romanian**

2. Information regarding the discipline

2.1 Name of the discipline				**Data Mining**			
2.2 Course coordinator				**Lect. Dr. Anca Andreica**			
2.3 Seminar coordinator				**Lect. Dr. Anca Andreica**			
2.4. Year of study	**2**	2.5 Semester	**3**	2.6. Type of evaluation	**E**	2.7 Type of discipline	**Optional**

3. Total estimated time (hours/semester of didactic activities)

3.1 Hours per week	4	Of which: 3.2 course	2	3.3 seminar/laboratory	1 lab+ 1 project
3.4 Total hours in the curriculum	56	Of which: 3.5 course	28	3.6 seminar/laboratory	28

Time allotment:	hours
Learning using manual, course support, bibliography, course notes	40
Additional documentation (in libraries, on electronic platforms, field documentation)	40
Preparation for seminars/labs, homework, papers, portfolios and essays	40
Tutorship	10
Evaluations	14
Other activities:	

3.7 Total individual study hours	144
3.8 Total hours per semester	200
3.9 Number of ECTS credits	8

4. Prerequisites (if necessary)

4.1. curriculum	●
4.2. competencies	● Average programming skills in a high level programming language

5. Conditions (if necessary)

5.1. for the course	●

5.2. for the seminar /lab activities	• Laboratory with computers

6. Specific competencies acquired

Professional competencies	C5.3 Use of databases methodologies and design environments for particular problems C5.4 Quality evaluation of different database management systems in terms of structure, functionality and extensibility C5.5 Implementation of database projects
Transversal competencies	CT1 Application of organized and efficient work rules, of responsible attitudes towards the didactic and scientific domain, for the creative exploitation of their own potential according to the principles and rules of professional ethics CT2 Efficient conduct of activities organized in an interdisciplinary group and development of empathic capacity of interpersonal communication, networking and collaboration with diverse groups CT3 Use of effective methods and techniques of learning, information, research and development of the capacity to exploit knowledge, to adapt to the requirements of a dynamic society and communication in Romanian language and in a foreign language

7. Objectives of the discipline (outcome of the acquired competencies)

7.1 General objective of the discipline	• To learn data mining concepts, methods and techniques, from a database perspective.
7.2 Specific objective of the discipline	• The students will learn various data analysis techniques, and will apply these techniques for solving data mining problems using special software systems and tools. • A perception of data mining as a strong application field, as well as a significant database research domain, will be formed.

8. Content

8.1 Course	Teaching methods	Remarks
1. Introduction; Types of data explored in data mining; Data mining functionalities; Patterns; Data mining from a database perspective; Data mining applications	Exposure, description, explanation, examples, discussion of case studies	
2-3. Concept description; Definitions; Data generalization and summarization-based characterization; Analytical characterization: attribute relevance analysis; Class comparison: discriminating	Exposure, description, explanation, examples, discussion of case studies	

between classes; Descriptive statistical measures in large databases		
4. Data preprocessing; Data cleaning; Data transformation and integration; Data reduction	Exposure, description, explanation, examples, discussion of case studies	
5-6. Association rules; Problem definition; Algorithms for mining single-dimensional boolean association rules; Algorithms for mining multi-level association rules, multi-dimensional association rules, association rules with constraints; Correlation analysis	Exposure, description, explanation, examples, discussion of case studies	
7-9. Classification and prediction; Problem definition; Classification using k-nearest neighbor method; Classification using decision tree induction; Classification and estimation using neural networks; Bayesian classification	Exposure, description, explanation, examples, discussion of case studies	
10-11. Clustering (cluster analysis); Problem definition; Types of data in cluster analysis; Clustering methods; Outliers detection	Exposure, description, explanation, examples, discussion of case studies	
12. Data mining standards and software	Exposure, description, explanation, examples, discussion of case studies	
13-14. Students theoretical reports presentations		

Bibliography

1. S. Chakrabarti et al, Data Mining. Know It All, Morgan Kaufmann, 2009.
2. K. Cios, W. Pedrycz, R. Swiniarski, L. Kurgan, Data Mining. A Knowledge Discovery Approach, Springer, 2007.
3. J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2nd Edition, Morgan Kaufmann, 2006.
4. P. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Addison Wesley, 2006.
5. D. Larose, Discovering Knowledge in Data. An Introduction to Data Mining, John Wiley & Sons, 2005.
6. Han, J., Kamber, M., Data Mining: Concepts and Techniques, 1st Edition, Morgan Kaufmann, 2000.
7. Weka system and documentation (http://www.cs.waikato.ac.nz/ml/weka/). Weka is a suite of machine learning / data mining software. It contains Java implementation for various mining algorithms, data preprocessing filters, and experimentation capabilities. Weka is free open-source software under the GNU General Public License (GPL).

8.2 Laboratory	Teaching methods	Remarks
1. Data preprocessing	Practical projects	
2. Data Mining software tools		
3. WEKA software		
4-6. Data mining applications		

7. Students project presentations		

Bibliography

1. S. Chakrabarti et al, Data Mining. Know It All, Morgan Kaufmann, 2009.
2. K. Cios, W. Pedrycz, R. Swiniarski, L. Kurgan, Data Mining. A Knowledge Discovery Approach, Springer, 2007.
3. J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2nd Edition, Morgan Kaufmann, 2006.
4. P. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Addison Wesley, 2006.
5. D. Larose, Discovering Knowledge in Data. An Introduction to Data Mining, John Wiley & Sons, 2005.
6. Han, J., Kamber, M., Data Mining: Concepts and Techniques, 1st Edition, Morgan Kaufmann, 2000.
7. Weka system and documentation (http://www.cs.waikato.ac.nz/ml/weka/). Weka is a suite of machine learning / data mining software. It contains Java implementation for various mining algorithms, data preprocessing filters, and experimentation capabilities. Weka is free open-source software under the GNU General Public License (GPL).

9. Corroborating the content of the discipline with the expectations of the epistemic community, professional associations and representative employers within the field of the program

- The course exists in the studying program of all major universities in Romania and abroad;
- The content of the course is considered the software companies as important for average programming skills

10. Evaluation

Type of activity	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Share in the grade (%)
10.4 Course	- know the basic principle of the domain;	Written exam	50%
10.5 Lab activities	- data mining problems solving	Laboratory work	50%
10.6 Minimum performance standards			

  ➢ At least grade 5 at both written exam and laboratory work.

Date	Signature of course coordinator	Signature of seminar coordinator
24.04.2014	Assoc. Prof. Dr. Anca Andreica	Assoc. Prof. Dr. Anca Andreica

....…...........................

Date of approval   Signature of the head of department

....…...........................