

MODULE SPECIFICATION

Name of Module		Data Mining Techniques and Applications					
Parent School/Dept		Computer Science/Information Systems					
Programme(s) where module is offered		BSc Computer Science with Electrical Engineering; BSc Computer Science with Economics; BSc Computer Science with Management; BSc Computer Science with International Relations; BSc Computer Science with Political Science; BSc Information Systems with Electrical Engineering; BSc Information Systems with Economics; BSc Information Systems with Management; BSc Information Systems with International Relations; BSc Information Systems with Political Science;					
Status (core, option, free choice)		Core		Pre-Requisite Modules or Qualifications		CSIS240 Database Systems	
FHEQ Level	5	Unit Value	6 ECTS	Module Code	CS355	Module coordinator	Amer Hadžikadić
Term taught		Fall		Applicable From		2016	

Educational Aims of the Module

The overall aim of this module is to introduce students to modern data mining techniques and their use in business and other areas of applications. In particular, the module explores basic concepts, principles and techniques of data mining with emphasis on both the technical and the practical issues. The module provides students with an understanding in evaluating and comparing data mining solutions for effective use of the solutions in practice. The module also equips students with some hands-on experience and skills in conducting a data mining project using a data mining software tools.

Module Outline/Syllabus

- *Introduction:* Concept of data mining. Data mining and KDD. Data mining process. Major data mining tasks. Data mining approaches. Overview of data mining solutions. Importance of evaluation. Evolution of data mining. Promises and challenges.
- *Understanding Data:* data and data sets. Data types. Data quality, Data pre-processing. Data summarisation and visualisation.
- *Data Mining Techniques I:* Problem of cluster detection. Proximity measures. Basic clustering methods: K-means and Agglomeration. Validation and evaluation of clusters. Overview of other types of clustering methods. Clustering in practice.
- *Data Mining Techniques II:* Problem of classification. Decision tree induction approach: ID3 and other tree induction solutions. Nearest neighbour approach: kNN and PEBLS methods. Statistical approach: Naïve Bayes method. Overview of other classification approaches. Evaluation of classifiers. Problem of overfitting and solutions. Classification in practice.
- *Data Mining Techniques III:* Problem of association rule discovery. Apriori algorithms for Boolean, generalised and quantitative association rules. Evaluation of association rules. Other types of association rules. Association rule discovery in practice.
- *Data Mining Projects.* Data Mining project life cycle. The industry standard: CRISP-DM guideline for data mining. A case study on customer segmentation
- *Data Mining software tool WEKA:* Overview of WEKA functions. Data Reprocessing in WEKA. WEKA visualisation facilities for data and patterns, WEKA data mining functions. WEKA evaluation parameters. Choosing the best data mining solutions in WEKA. Overview of data mining software tools.

- *Putting Everything Into Perspective*: Database, data warehouse, query, OLAP, data mining and decision support. Data, information and Knowledge in enterprises. Application areas of data mining. Ethical and professional issues regarding data mining.

<u>Student Engagement Hours</u>			
Type	Number per Term	Duration	Total Time
Lectures	30	2 hours	60 hours
Laboratory sessions	15	2 hours	30 hours
Total Guided/Independent Learning Hours			60
Total Contact Hours			90
Total Engagement Hours			150

<u>Assessment Method Summary</u>				
Type	Number Required	Duration / Length	Weighting	Timing/Submission Deadline
Final exam	1	180 minutes	50%	End of semester
Mid-term exam	1	90 minutes	10%	Week 8
Project (group)	1	2,000 words	10%	Week 15
Test	3	60 minutes	30%	Weeks 5, 11 and 14

Module Outcomes

<p><u>Intended Learning Outcomes:</u></p> <ol style="list-style-type: none"> 1. Understand key concepts and principles of data mining 2. Understand state-of-art approaches and techniques in data mining and visualisation 3. Develop a working application using a commercial data mining/data warehousing software tool 	→	<p><u>Teaching and Learning Strategy:</u></p> <ol style="list-style-type: none"> 1. Lectures introduce theoretical and conceptual materials from the recommended textbooks (ILO: 1, 2) 2. Laboratory sessions provide exercises to apply the theory and use data mining software tools (ILO: 2, 3) 3. Group project enables students to develop team-work skills and apply what they have learnt in the module to a practical problem (ILO: 3)
	→	<p><u>Assessment Strategy</u></p> <ol style="list-style-type: none"> 1. Test (ILO: 1, 2) 2. Mid-term exam (ILO: 1, 2) 3. Final exam (ILO: 1-3) 4. Project (ILO: 3)
<p><u>Practical Skills</u></p> <ol style="list-style-type: none"> 1. Understand the particular uses of algorithms in applications 2. Use Weka to conduct data mining 3. Accurately interpret the results and argue the gained knowledge and module of action as a result of the data mining. 	→	<p><u>Teaching and Learning Strategy:</u></p> <ol style="list-style-type: none"> 1. Laboratory sessions (PS: 1, 2, 3) 2. Group project (PS: 1, 2, 3) 3. Lectures (PS: 1, 3)
	→	<p><u>Assessment Strategy</u></p> <ol style="list-style-type: none"> 1. Test (ILO: 1) 2. Mid-term exam (ILO: 1) 3. Final exam (ILO: 1,3) 4. Project (ILO: 1- 3)
<p><u>Transferable Skills</u></p> <ol style="list-style-type: none"> 1. Oral and written presentation: ability to express ideas clearly and precisely 2. Critical thinking 3. Team working skills 4. Data analysis 	→	<p><u>Teaching and Learning Strategy:</u></p> <ol style="list-style-type: none"> 1. Personal and team projects to read, evaluate and present knowledge (TS: 1, 3) 2. Practical classes and application on real life data (TS: 1, 2, 3, 4)
	→	<p><u>Assessment Strategy</u></p> <ol style="list-style-type: none"> 1. Project (TS: 1-4) 2. Test (TS: 2, 4) 3. Mid-term exam (TS:2, 4) 4. Final exam (TS:2, 4)

Key Texts and/or other learning materials

Set Text

- Du,H., (2010) Data Mining Techniques and Applications, Cengage Learning
-

Supplementary Materials

- Kimball, R., Ross, M., (2013), The Data Warehouse Toolkit, 3rd Edition, John Wiley & Sons
- Tan, P-N., Steinbach, M., Kumar, V., (2013) Introduction to Data Mining. Pearson

The Data Warehousing Institute

www.tdwi.org

Dr R. Meredith: Data Warehousing Podcasts

<https://itunes.apple.com/gb/podcast/fit5095-data-warehousing/id306569323?mt=2>

Please note: This specification provides a concise summary of the main features of the module and the learning outcomes that a typical student might reasonably be expected to achieve and demonstrate if he/she takes full advantage of the learning opportunities that are provided. More detailed information on the learning outcomes, content and teaching, learning and assessment methods of each module and programme can be found in the departmental or programme handbook. The accuracy of the information contained in this document is reviewed annually by the University of Buckingham and may be checked by the Quality Assurance Agency.

Date of Production	Autumn 2016
Date approved by School Learning and Teaching Committee	28 th September 2016
Date approved by School Board of Study	12 th October 2016
Date approved by University Learning and Teaching Committee	2 nd November 2016
Date of Annual Review	December 2017