

Module Guide

Principles of Data Mining

BIF-6-PDM

<http://vle.lsbu.ac.uk>

School of Engineering

2015/16

Level 6

Table of Contents

1.	Module Details	3
2.	Short Description.....	3
3.	Aims of the Module	3
4.	Learning Outcomes.....	3
4.1	Knowledge and Understanding	3
4.2	Intellectual Skills.....	4
4.3	Practical Skills	4
4.4	Transferable Skills.....	4
5.	Assessment of the Module.....	4
5.1	Data Mining Project.....	4
5.2	Phase Test	4
5.3	Examination	4
5.4	Workbook	5
6.	Feedback	5
7.	Introduction to Studying the Module	5
7.1	Overview of the Main Content.....	5
7.2	Overview of Types of Classes	6
7.3	Importance of Student Self-Managed Learning Time	6
7.4	Employability	6
8.	The Programme of Teaching, Learning and Assessment	6
9.	Student Evaluation	7
10.	Learning Resources	8
10.1	Core Materials	8
10.2	Optional Materials	8
10.3	Useful Resources	9
	NOTES.....	9

1. MODULE DETAILS

Module Title:	Principles of Data Mining
Module Level:	6
Module Reference Number:	BIF-6-PDM
Credit Value:	20
Student Study Hours:	200
Contact Hours:	52
Private Study Hours:	148
Pre-requisite Learning (If applicable):	Essential knowledge of statistics and database systems. Elementary computer programming skills (in an arbitrary programming language). In all other respects the module will be self-contained.
Co-requisite Modules (If applicable):	Business Database Systems, Introduction to Business Intelligence, or equivalent.
Course(s):	Core for Business Intelligence pathway. Elective for all other pathways.
Year and Semester	Semester 1, Year 3
Module Coordinator:	Dr David D Chen
MC Contact Details (Tel, Email, Room)	Tel: 020 7815 7492 (Voice messaging available) Email: chend@lsbu.ac.uk Room: FW105, 1 st Floor, Faraday Wing
Teaching Team & Contact Details	
Subject Area:	Information Systems and IT, Computer Science
Summary of Assessment Method:	Coursework 60% Examination 40%
External Examiner appointed for module:	TBA

2. SHORT DESCRIPTION

Data mining is a fast-growing, exciting and challenging area in IT and computer science, both theoretically and practically. '*Making sense and making use of data*' is the main theme of the subject. Data mining techniques are essential to business intelligence, and enable business managers to make profitable use of the massive data their enterprises collect. This module will provide a broad introduction to the basic theory, concepts, and techniques of data mining, as well as its main application areas and its role in the context of business intelligence. In addition, SAS® Enterprise Miner® and SAS® Enterprise Guide®, a leading commercial business intelligence software suite, will be taught and used throughout the module in order for you to develop practical skills in solving real-world data mining problems. This module is the second specialist module for the Business Intelligence pathway.

3. AIMS OF THE MODULE

This module aims to:

- Provide a solid background in data mining concepts, methodologies, tools, and algorithms.
- Develop thorough understandings about the need for data mining, the nature of data mining, and benefits resulting from data mining, as well as their applicability in solving real-world problems.
- Provide practical experience of using appropriate data mining algorithms and tools in real-world data mining tasks.

4. LEARNING OUTCOMES

4.1 Knowledge and Understanding

On successful completion of this module you will be able to:

- Describe and explain the concepts of data mining including the techniques and algorithms for problem solving and creating competitive advantage.

4.2 Intellectual Skills

On successful completion of this module you will be able to:

- Critically evaluate different types of data mining tasks in relation to various business concerns, including descriptive modelling and predictive modelling, classification, regression, prediction, cluster analysis, and association analysis.

4.3 Practical Skills

On successful completion of this module you will be able to:

- Transfer a business problem into an appropriate data mining problem.
- Creatively apply standard industry data mining tools such as SAS® Enterprise Miner®.

4.4 Transferable Skills

On successful completion of this module you will be able to:

- Analyse and develop solutions for a wide range of business problems.

5. ASSESSMENT OF THE MODULE

This module will be assessed by both examination and coursework. The weighting between examination and coursework marks in this module is **40:60**. The coursework is completely an individual piece of work and has two components: individual data mining project and 3 phase tests. The weighting between these two coursework component marks is **70:30**.

5.1 Data Mining Project

You are required to carry out a data mining project on a real-world dataset. Your module tutor will assign a dataset to you, which is downloadable from the module Moodle site (<http://vie.lsbu.ac.uk>). You should follow an appropriate data mining methodology and apply a number of techniques and algorithms covered in lectures to analyse the dataset in order to identify structural patterns and models (descriptive and predictive models) of the data in relation to certain business concerns.

You are expected to submit a written report (worth **70%** of the total coursework marks) for this data mining project **in week 11 at tutorial**.

The module will use SAS® Enterprise Miner® and SAS® Enterprise Guide® for this coursework component and weekly tutorial exercises. An introductory material to SAS® Enterprise Miner® can be found at <http://support.sas.com/documentation/onlinedoc/miner/> and in some textbooks listed in **Section 10** of this module guide. In addition, supplementary material for SAS® suite is available on the module Moodle site. You are encouraged to use other software packages for this coursework, including Microsoft SQL Server 2008/2012, Oracle Data Mining/Data Miner, IBM SPSS, Weka, and Matlab. You are also encouraged to implement data mining algorithms using a high level computing language of your own choice, e.g. C, C++, Java, VBA, Python, and R.

5.2 Phase Test

You are expected to undertake 3 phase tests in weeks 4, 7, and 10, respectively. Each phase test is worth **10%** of the total module coursework marks. These tests are designed to examine your understandings of the key concepts and the main principles of the subject. The tests will be of the 'multiple choice' type. **No books, notes or calculator of any kind are allowed in these tests.**

5.3 Examination

The examination is of two-hour duration and has two sections. Section A is compulsory and contains **one** question (40 marks). Section B has **four** questions (30 marks each) and you are required to choose any two questions to attempt in this section. **No books, notes or calculator of any kind are allowed in the examination.**

5.4 Workbook

You are strongly advised to have a workbook throughout the module in order to keep notes on weekly tutorial exercises, and ideas, research findings, progress and problems in relation to your data mining project. Furthermore the workbook can also be used as revision material for the exam. Please note that the workbook is **not** an assessed component of the module.

6. [FEEDBACK](#)

Feedback will be given in **week 13** at the tutorial time.

7. [INTRODUCTION TO STUDYING THE MODULE](#)

7.1 Overview of the Main Content

Data mining is an inter-disciplinary computer-based process for finding hidden patterns within large, heterogeneous, and complex databases. It integrates techniques from different fields including pattern recognition, machine learning, database systems and statistics. Data mining techniques are essential in such areas as business intelligence, decision-making, CRM, and business management. Enterprises are embracing data mining as a must-have tool to gain and sustain competitive advantages in today's modern and ever-changing economy.

The module syllabus covers the following core topics in the area:

Data Mining Basics

- Data mining tasks and methodologies, the role of data mining in business intelligence.
- Typical data mining problems in various business sectors.
- Data mining tools, platforms, and the main vendors.
- Data and data pre-processing for data mining, different types of data, data quality issues.
- Learning: supervised vs. unsupervised; eager vs. lazy.

Data Mining Modelling Approaches

- Descriptive and predictive modelling.
- Cluster analysis.
- Association rules analysis.
- Decision tree induction.
- Rule-based classifiers.
- Regression models.
- Artificial neural networks.
- Nearest-neighbour classifiers.
- Statistical modelling: Naïve Bayesian classifiers.

Advanced Data Mining Topics

- Web mining.
- Text mining.
- Temporal-spatial mining.
- Big data analytics.

Module Project

- SAS® Enterprise Miner® and SAS® Enterprise Guide®: Essential skills.
- Data mining project using SAS® Enterprise Miner® and SAS® Enterprise Guide®.

In addition, some popular commercial and open source data mining tools will be introduced and evaluated.

This module is essentially about **being analytical for business decision making**. To be able to get a deep understanding of the contents of this module you should have some basic knowledge of **statistics and regression analysis**. To make the learning process a lot easier and more

enjoyable, it is essential that **you have a passionate interest in manipulating and analysing data using appropriate analytical tools.**

7.2 Overview of Types of Classes

The module will be delivered using a combination of lectures, tutorials and lab sessions. Teaching takes place over 13 weeks of the semester when there will be 4 hours of direct class contact. **Regular workshops are also organised outside the class** in order to support your studies (see **Notes** below at the end of this guide). You will also be expected to undertake appropriate follow-up private study.

7.3 Importance of Student Self-Managed Learning Time

The module covers a wide range of topics, both theoretical and practical, in the area of data mining. In order to complete the module successfully you should manage your time effectively and ensure to spend sufficient private study time doing homework, preparing for weekly tutorial discussions, and conducting assigned data mining project. In particular, you should be aware that, like any commercial software, SAS® Enterprise Mine is very complex, and requires significant amount of practice time.

7.4 Employability

Data mining and business intelligence are becoming very popular in recent years across almost all industries. According to Gartner EXP survey 2007-2015, analytics and business intelligence has been scored as a top 5 technology priority for businesses for the last nine years. Having sound analytical skills and a deeper fluency with commercial software suite such as SAS® Enterprise Miner® will help improve your profile and increase employability significantly.

8. THE PROGRAMME OF TEACHING, LEARNING AND ASSESSMENT

Week No	Week Commencing	Lecture Topic	Reading	Tutorial Work/ Assessment
1	28/09/2015	Module Overview & Introduction to Data Mining	Ch 1, Myatt Ch 1, Bramer	Introduction to the module Moodle site Coursework issued
2	05/10/2015	Data Mining Methodology Data and Data Types	Chs 2 & 3, Myatt Ch 2, Bramer	Introduction to SAS® Enterprise Miner® 14.1 and SAS® Enterprise Guide® 7.1
3	12/10/2015	Data Pre-processing & Exploration	Chs 3, 4, & 5, Myatt Ch 2, Bramer	Module coursework project planning
4	19/10/2015	Descriptive Modelling: Cluster Analysis	Secs 6.1 & 6.2, Myatt Ch 19, Bramer	Dataset selection, preparation, and initial exploration Data summarisation Phase Test 1
5	26/10/2015	Descriptive Modelling: Association Rules	Sec 6.3, Myatt Chs 16 & 17, Bramer	Cluster analysis

6	02/11/2015	Predictive Modelling: Decision Tree Induction	Sec 6.4, Myatt Chs 4 & 5, Bramer	Mining association rules
7	09/11/2015	Predictive Modelling: More on Decision Tree Induction Rule-based Classifiers	Secs 6.4, & 7.1, Myatt Chs 7, 9, 10 & 12, Bramer	Classification tasks Phase Test 2
8	16/11/2015	Predictive Modelling: Simple Regressions Artificial Neural Networks	Secs 7.2 & 7.5, Myatt Chs 7, 9, 10 & 12, Bramer	Classification tasks
9	23/11/2015	Predictive Modelling: Eager and Lazy Learning	Secs 7.3 & 7.4, Myatt Chs 3 & 14, Bramer	Model evaluation, interpretation, and comparison
10	30/11/2015	Predictive Modelling: Bayesian method Data Mining Modelling approaches: A Comparison	Chs 5 & 8, Myatt Chs 3 & 14, Bramer	Data mining project draft report check Phase Test 3
11	07/12/2015	Advanced Topics: Web Mining, Text Mining, Spatial-Temporal Mining, Big Data Analytics	Chs 8 & 9, Myatt Chs 13 & 20, Bramer Supplementary material	Data mining project report submission
12	14/12/2015	Advanced Topics: Web Mining, Text Mining, Spatial-Temporal Mining, Big Data Analytics	Chs 13 & 20, Bramer Supplementary material	Research paper review
	21/12/2015- 08/12/2016	Christmas Vacation		
13	11/01/2016	Module Revision		Coursework feedback
14	18/01/2016	Exam week 1		
15	25/01/2016	Exam week 2		

Coursework Assessment Schedule

Assignment	Hand-out Date	Hand-in Date	Description	Marks
1	Week 1	Week 11	Individual data mining project report	70%
2	Week 4	Week 4	Phase Test 1	10%
3	Week 7	Week 7	Phase Test 2	10%
4	Week 10	Week 10	Phase Test 3	10%

9. STUDENT EVALUATION

There were 28 students taking this module last year. The weekly attendance rate was high. The students found the subject quite challenging, very technical, and potentially very useful. In the returned MEQs, all of the areas were rated mainly as either “Acceptable” or “Vary acceptable” by the students.

10. LEARNING RESOURCES

10.1 Core Materials

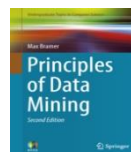
Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining

Glenn J. Myatt
John Wiley & Sons, Inc 2007
ISBN: 978-0-470-07471-8



Principles of Data Mining (2nd Edition)

Max Bramer
Springer 2013
ISBN-13: 978-1447148838



Introduction to Data Mining: Using SAS® Enterprise Miner

Patricia B. Cerrito
SAS Institute Inc., 2006
ISBN: 1-59047-829-0



Predictive Modelling with SAS® Enterprise Miner: Practical Solutions for Business Applications

SAS Institute Inc., 2007
ISBN: 978-1-59047-703-8



10.2 Optional Materials

Data Mining: Know It All

Soumen Chakrabarti, Earl Cox, and Eibe Frank, *et al.*
Morgan Kaufmann Publishers 2009
ISBN: 978-0123746290



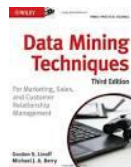
Learning Data Mining with Python

Robert Layton
Packt Publishing, 2015
ISBN-13: 978-1784396053



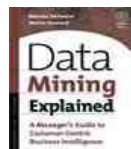
Data Mining techniques: For Marketing, Sales and Customer Relationship Management (3rd Edition)

Michael Berry and Godeon Lindoff
John Wiley & Sons Inc, 2011
ISBN: 978-0470650936



Data Mining Explained: A Manager's Guide to Customer-Centric Business Intelligence

Rhonda Delmater and Monte Hancock
Digital Press, 2001
ISBN-13: 978-1555582319



Doing Data Science: Straight Talk from the Frontline

Cathy O'Neil and Rachel Schutt
O'Reilly Media, 2013
ISBN-13: 978-1449358655



Data Mining: Introductory and Advanced Topics

Margaret H. Dunham
Prentice and Hall, 2003
ISBN: 0-13-088892-3



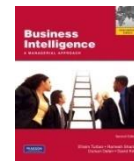
Introduction to Data Mining

Pang-Ning Tan, Michael Steinbach, and Vipin Kumar
Addison Wesley, 2006
ISBN: 0-321-42052-7



Business Intelligence: A Managerial Approach (2nd Edition)

Efraim Turban, Ramesh Sharda, Jay E. Aronson and David King
Pearson Education, Inc., 2011
ISBN: 978-0-13-247882-3



10.3 Useful Resources

- SAS® system: a leading business intelligence software suite, <http://support.sas.com/> and <http://www.sas.com>.
- SAS® Global Forum and Online Proceedings, <http://support.sas.com/events/sasglobalforum/previous/online.html>.
- KDNuggets: Knowledge Discovery Neggets, a leading information repository for data mining, <http://www.kdnuggets.com/>.
- Kaggle, the leading platform for predictive modelling competitions, <https://www.kaggle.com/competitions>.
- KDD Cup: the leading data mining and knowledge discovery competition in the world, organized by [ACM SIGKDD - Special Interest Group on Knowledge Discovery and Data Mining](#), the leading professional organization of data miners, <http://www.sigkdd.org/kddcup/index.php>

NOTES

Office hours: Tuesday 14.00-16.00; Friday 14.00-16.00.

Workshops: During all the scheduled teaching weeks, regular drop-in workshops will take place on every Friday afternoon from 14.00 onwards in **N112** for all students on the module to exchange ideas and experiences regarding the subject of the module and SAS® software suite.