LONDON SOUTH BANK UNIVERSITY

Data Mining and Business Intelligence

CIT-6-DMB

http://blackboard.lsbu.ac.uk

Faculty of Business

2011/2012

Level 6

Table of Contents

1. MODULE DETAILS

Module Title:	Data Mining and Business Intelligence
Module Level:	6
Module Reference Number:	CIT-6-DMB
Credit Value:	1
Student Study Hours:	150
Contact Hours:	45
Private Study Hours:	105
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):	None
Course(s):	Various
Year and Semester	Year 3, Semester 2
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: N306, 3rd Floor, Faraday Wing
Teaching Team & Contact Details (If applicable):	
Subject Area:	Information Systems and IT, Computer Science
Summary of Assessment Method:	100% Coursework

2. SHORT DESCRIPTION

Data mining is a fast-growing, exciting and challenging area in IT and computer science. '***Making sense and making use of data***' is the main theme of the subject. Data mining is the core of business intelligence process, and it enables 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. It will cover the main topics in data mining, business intelligence, and data warehousing. In particular, SAS® Enterprise Miner® and SAS® Enterprise Guide®, a powerful commercial business intelligence software suite, will be introduced and used to complete a data mining project on real-world datasets.

3. AIMS OF THE MODULE

This module aims to:

- Provide a solid background in data mining and business intelligence concepts, tools, and methodologies.
- Examine the role of data mining in business intelligence.
- 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 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 and business intelligence.
- Explain the role of data mining in business intelligence.

- Explain how and why data mining and business intelligence can be used to create competitive advantage for the enterprise.
- Explain when and why data mining should be considered a possible problem-solving strategy.
- Gain familiarity with the techniques and algorithms that are being used in data mining.
- Gain hands-on experience for using SAS® Enterprise Miner® and SAS® Enterprise Guide® to real-world problems.

4.2 Intellectual Skills

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

- Identify different types of data mining tasks, including descriptive modelling and predictive modelling, classification, prediction, regression, cluster analysis, and association analysis.
- Appreciate the strengths and weaknesses of the different data mining techniques, algorithms, and tools.

4.3 Practical Skills

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

- Select and apply appropriate data mining algorithms for a given real-world problem.
- Evaluate various models built from a data mining process.

4.4 Transferable Skills

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

- Demonstrate time management skills.
- Demonstrate team work skills.
- Demonstrate presentation skills.

5. ASSESSMENT OF THE MODULE

The module will be assessed by coursework only, which is a combination of team-based project assignment, individual portfolio (logbook) assessment, and an in-class test. A brief description of each coursework component is given below.

5.1 Team-based Assignment

The team-based coursework is to be undertaken in groups of four. Each group is required to carry out a data mining project on a real-world dataset. Your module tutor will assign a dataset to each project group, which is downloadable from the module blackboard site (http://blockboard.lsbu.ac.uk). You should apply a number of techniques and algorithms covered in lectures to analyse the assigned dataset and identify structural patterns and models (descriptive and predictive models) of the data.

There are two project deliverables: a written report (worth **50%** of the total module marks) and a presentation (worth **10%** of the total module marks). Each group is required to submit a written report on the project and give an in-class oral presentation. The mark awarded for both the report and the presentation will be a group mark. This team-based assignment is due **in week 12 at tutorial**.

The module will use SAS® Enterprise Miner® and SAS® Enterprise Guide® for the coursework 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 the Section 10 of this module guide. In addition, supplementary material for SAS® suite is available on the module Blackboard site. You are encouraged to use other software packages for the coursework, e.g. Microsoft SQL Server 2005/8, Oracle Data Mining/Data Miner, SPSS, and Weka, or implement data mining algorithms using a high level computing language of your own choice, e.g. Matlab, C, C++, Java, and VB.

5.2 Portfolio Assessment

This assignment is an individual work and is worth **30%** of the total module marks.

You should keep a logbook (**A4-sized**) throughout the course to record individual work on all required weekly exercises. These exercises will be specified in the weekly lecture handouts. In addition, you can use your logbook to keep notes on ideas, research findings, progress and problems in relation to the coursework project. The logbook will be **checked and signed** by your module tutor on a weekly basis and should be submitted **in week 13** at the tutorial.

5.3 In-class Test

The in-class test is of **40-minute** duration and is worth **10%** of the total module marks. It is designed to examine your understandings of the key concepts and the main principles of the subject. The test will be of the 'multiple choice' type. **No books, notes or calculator of any kind are allowed in this test**. This test will take place during the normal **week 13** tutorial.

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, 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 tasks and methodologies, the concept of business intelligence, the role of data mining in business intelligence.
- Popular business intelligence tools including data warehousing, OLAP, dashboards, and scorecards.
- Descriptive and predictive modelling, cluster analysis, association rules analysis, decision tree induction, regression models, artificial neural networks, rule-based classifiers, eager and lazy learning, and simple statistical modelling etc.

In addition, SAS® Enterprise Miner® and SAS® Enterprise Guide® will be utilised for the module coursework project. Some popular commercial and open source data mining tools will be introduced and evaluated.

This module is **math intensive** and it is essentially about **being analytical** business decision marking. To be able to get a deep understanding of the content 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. The student 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 and business intelligence. 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 Miner® is very complex, and requires significant amount of practice time.

7.4 Employability

Business intelligence is becoming popular in recent years across various industries. 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	30/01/12	Module Overview *&* Introduction to Data Mining	Ch 1, Myatt Ch 1, Cerrito	Introduction to the unit Blackboard site ***Coursework issued***
2	06/02/12	Business Intelligence Basics Data Mining Methodology and Project Management	Ch 2, Myatt Chs 1 & 4, Turban Ch 1, Cerrito	Introduction to SAS® software suite
3	13/02/12	Business Intelligence Tools: OLAP, Dashboards and Scorecards	Chs 1 & 2, Myatt Ch 2, Turban Chs 1 & 2, Cerrito	Introduction to SAS® Enterprise Miner® 6 and SAS® Enterprise Guide® 4
4	20/02/12	Data Pre-processing *&* Exploration	Chs 3, 4, & 5, Myatt Chs 3 & 4, Turban Chs 2 & 4, Cerrito	Module coursework project planning
5	27/02/12	Descriptive Modelling: Cluster Analysis	Secs 6.1 & 6.2, Myatt Ch 8, Cerrito	Dataset selection, preparation, and initial exploration
6	05/03/12	Descriptive Modelling: Association Rules	Sec 6.3, Myatt Chs 5 & 6, Cerrito	Data summarisation
7	12/03/12	Predictive Modelling: Decision Tree Induction	Sec 6.4, Myatt Ch 7, Cerrito	Mining association rules
8	19/03/12	Predictive Modelling: Rule-based Classifiers and Simple Regressions	Secs 6.4, 7.1 & 7.2, Myatt Ch 7, Cerrito	Classification tasks
9	26/03/11	Predictive Modelling: Neural Networks	Sec 7.5, Myatt Ch6 Turban Ch 7, Cerrito	Classification tasks
	04/04/12 – 16/04/12	***Easter vacation***		
10	23/04/12	Predictive Modelling: Eager and Lazy Learning	Secs 7.3 & 7.4, Myatt Ch5 Turban Chs 7 & 9, Cerrito	Model evaluation, interpretation, and comparison

11	30/04/12	Predictive Modelling: Bayesian method	Chs 5 & 8, Myatt Chs 7 & 10, Cerrito	***Coursework draft report check***
12	07/05/12	Advanced Topics: Web Mining, Spatial and Temporal Mining	Chs 8 & 9, Myatt Chs 9 & 10, Cerrito	***Coursework report submission*** ***Group presentation***
13	14/05/12	Module Revision Coursework feedback	Chs 9, Myatt Chs 10 & 11, Cerrito	***Logbook submission*** ***In-class test***

Coursework Assessment Schedule

Assignment	Hand-out Date	Hand-in Date	Description	Marks
1	Week 1	Week 12	Team-based data mining project report	50%
2	Week 1	Week 12	Team-based data mining project presentation	10%
3	Week 1	Week 13	Individual portfolio	30%
4	Week 13	Week 13	In-class test	10%

9. STUDENT EVALUATION

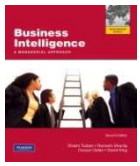
There were 33 students taking this module last year. In the returned Unit Evaluation Questionnaires, all of the areas were rated mainly as either “acceptable” or “vary acceptable” by the students. In general the students found the subject quite challenging, very technical, and potentially very useful. The weekly attendance rate was high.

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 Wiely & Sons, Inc 2006
ISBN: 978-0-470-07471-8

Business Intelligence: A Managerial Approach (Second edition)
Efraim Turban, Ramesh Sharda, Jay E. Aronson and David King
Pearson Education, Inc., 2011
ISBN: 978-0-13-247882-3



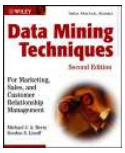
Introduction to Data Mining: Using SAS® Enterprise Miner
Patricia B. Cerrito
SAS Institute Inc., 2006
ISBN: 1-59047-829-0

Predictive Modelling with SAS® Enterprise Miner: Practical Solutions for Business Applications
SAS Institute Inc., 2007
ISBN: 978-1-59047-703-8

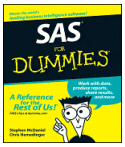
10.2 Optional Materials

Data Mining Techniques and Applications: An Introduction
Hongbo Du
CENGAGE Learning Business Press, 2010
ISBN: 978-1844808915

Data Mining techniques: For Marketing, Sales and Customer Relationship Management
Michael Berry and Godeon Lindoff
John Wiley & Sons Inc, 2004
ISBN: 9780471470649.



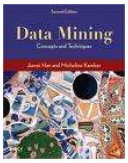
SAS® for Dummies
Stephen McDaniel and Chris Hemedinger
Wiley Publishing, Inc, 2007
ISBN: 978-0-471-78832-4



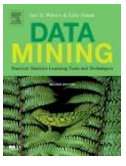
Business Intelligence: The Savvy Manager's Guide
David Loshin
Morgan Kaufmann Publishers, 2003.
ISBN: 1-55860-916-4.

Data Mining Explained: A Manager's Guide to Customer-Centric Business Intelligence
Rhonda Delmater and Monte Hancock
Digital Press, 2001
ISBN-13: 978-1555582319

Data Mining: Concepts and Techniques (Second Edition)
Jiawei Han and Micheline Kamber
Morgan Kaufmann Publishers, 2006.
ISBN: 1-55860-901-6.



Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)
Ian H. Witten and Eibe Frank
Elsevier Inc., 2005.
ISBN: 0-12-088407-0.



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.

10.3 Useful Resources

- SAS® system: a leading business intelligence software suite, http://support.sas.com/document and http://www.sas.com.
- SAS® Global Forum and Online Proceedings, http://support.sas.com/events/sasglobalforum/previous/online.html.
- KDD Cup 2011: Recommending Music Items based on the Yahoo! Music Dataset, http://www.kdd.org/kdd2011/kddcup.shtml.
- KDNuggets: Knowledge Discovery Neggets, a leading information repository for data mining, http://www.kdnuggets.com/.

NOTES

Office hours: Monday 16.00-18.00; Wednesday 12.00-14.00.
Workshops: during the scheduled teaching weeks, an informal drop-in workshop will be held on every Friday afternoon in **N306** for all students on the module to exchange ideas and experiences regarding the subject and SAS® software suite.